



D2.3 Report on Methodology development

MAIL: Identifying Marginal Lands in Europe and strengthening their contribution potentialities in a CO₂ sequestration strategy



Project title	Identifying Marginal Lands in Europe and strengthening their contribution potentialities in a CO2 sequestration strategy
Call identifier	H2020 MSCA RISE 2018
Project acronym	MAIL
Starting date	01.01.2019
End date	31.12.2021
Funding scheme	Marie Skłodowska-Curie
Contract no.	823805
Deliverable no.	D2.3
Document name	MAIL_D2.3.pdf
Deliverable name	Report on Methodology development
Work Package	WP2
Nature ¹	R
Dissemination ²	PU
Editors	Charalampos Georgiadis and Jesús Torralba Pérez
Authors	Rodrigo Gómez Conejo Natalia Verde Charalampos Georgiadis Maria Tassopoulou Michał Krupinski Ewa Grommy Sebastian Aleksandrowicz Jesús Torralba Pérez
Contributors	All project partners
Date	30/01/2021

¹ **R** = Report, **P** = Prototype, **D** = Demonstrator, **O** = Other

² **PU** = Public, **PP** = Restricted to other programme participants (including the Commission Services), **RE** = Restricted to a group specified by the consortium (including the Commission Services), **CO** = Confidential, only for members of the consortium (including the Commission Services).



MAIL CONSORTIUM

 <p>Aristotle University of Thessaloniki (AUTH) Greece</p>	 <p>Industriebetriebe Betriebsgesellschaft MBH (IABG) Germany</p>
 <p>Gounaris N. – Kontos K. OE (HOMEOTECH) Greece</p>	 <p>Centrum Badan Kosmicznych Polskiej Akademii Nauk (CBK PAN) Poland</p>
 <p>UNIVERSITAT POLITÈCNICA DE VALÈNCIA</p> <p>Universitat Politècnica de València (UPV) Spain</p>	 <p>Fundación Centro de Servicios y Promoción FOrestal y de su Industria de Castilla y León (CESEFOR) Spain</p>



ABBREVIATIONS

Term	Explanation
ASTER	Advanced Spaceborne Thermal Emission and Reflection Radiometer
AWC	Available Water Capacity
BOA	Bottom of Atmosphere
C	Carbon
CEC	Cation Exchange Capacity
CLC	CORINE Land Cover
Degr	Degrees
DSM	Digital Surface Model
EEA	European Environment Agency
ESDAC	European Soil Data Centre
ETRS89	European Terrestrial Reference System 1989
EU-DEM	European Digital Elevation Model
EUE	Energy-Use Efficiency
EUROSTAT	European Statistical Office
ETM+	Enhanced Thematic Mapper Plus
GDEM	Global Digital Elevation Map
GDP	Gross Domestic Product
GDP	Gross Domestic Product
GEE	Google Earth Engine
GLASS	Geoscience Laser Altimeter System
HR	High-Resolution



HRL	High-Resolution Layers
HWSD	Harmonized World Soil Database
ILSWE	Index of Land Susceptibility to Wind Erosion
JRC	Joint Research Centre
LAEA	Lambert Azimuthal Equal-Area projection
LGP	Length of Growing Period
LST	Land Surface Temperature
LULC	Land Use and Land Cover
LULUCF	Land Use, Land Use Change and Forestry
MLs	Marginal Lands
MMU	Minimum Mapping Unit
MODIS	Moderate Resolution Imaging Spectroradiometer
MSI	Multispectral Instrument
NCCPI	National Commodity Crop Productivity Index
NDVI	Normalized Difference Vegetation Index
NDWI	Normalized Difference Water Index
NPP	Net Primary Production, or Net Primary Productivity
NUTS	Nomenclature of Territorial Units for Statistics
OLI	Operational Land Imager
OM	Organic Matter
PCM	Pairwise Comparison Matrix
RUSLE	Universal Soil Loss Erosion and the revised version
S2GLC	Sentinel-2 Global Land Cover
SEDEM	Sediment Delivery Model



SOC	Soil Organic Carbon
SOM	Soil Organic Matter
SOTER	Soil-Terrain Database
SQR	Soil Quality Rating
STRM	Shuttle Radar Topography Mission
STUs	Soil Typological Units
TCD	Tree Cover Density
TOA	Top of Atmosphere
TM	Thematic Mapper
USDA	United States Department of Agriculture
USGS	United States Geological Survey
VHR	Very High-Resolution
WaTEM	Water and Tillage Erosion Model,
WGS84	World Geodetic System 1984
WISE	World Inventory of Soil Emission Potentials



CONTENTS

MAIL Consortium.....	3
Abbreviations	4
Executive Summary	10
1. Introduction.....	12
2. Definition of minimum marginal land parcel area.....	14
3. Datasets used.....	14
3.1 CORINE Land Cover.....	14
3.2 Pan-European High-Resolution Layers (HRL).....	15
3.3 Digital European elevation model.....	16
3.4 Environmental dataset	17
3.5 European Soil Database Derived data	18
3.6 ISRIC World Soil Information	19
3.7 S2GLC dataset description	20
4. Definition of GIS specifications	22
4.1 Coordinate Projection.....	23
4.2 File format.....	23
4.3 Coverage	23
5. Review of existing research projects and literature for available GIS models & workflows	24
5.1 Marginality index	24
5.2 Land use change	24
5.3 Tree Cover Density	25
5.4 Digital Elevation Model of Europe	25
5.5 Protected areas. European network of protected sites (Natura 2000)	25
6. Development of the final ML detection, criteria and classification scheme.	26
6.1 ML definition refinement.....	26
6.2 Methodology overview	26
6.3 Indicator / criteria selection.....	31
6.4 Initial indicator classification categories and thresholds.....	33
6.4.1 Land Use/Land Cover indicators	34
6.4.2 Terrain and Soil indicators.....	34
6.4.3 Sustainability indicators.....	38
6.4.4 Productivity indicators	42
6.4.5 Climate indicators.....	44



6.5 ML classification schemes	46
6.6 Production of intermediate layers	51
7. Development and realization of a GIS using the datasets collected in T2.2	52
7.1 Datasets overview	52
7.2 Dataset transformation to common file formats, projection, etc.	56
7.3 GIS implementation	56
7.4 Layers	59
8. Evaluation of existing and development of indicator sets	60
8.1 Selection of test sites	60
8.2 Testing different sets of existing indicators and thresholds	63
8.2.1. Selection of appropriate Land Use/Land Cover datasets	63
8.2.2. Selection of appropriate Marginality indexes	78
8.3 Development of new indicators sets	83
8.4 Initial results according to the proposed methodology	84
8.4.1 Initial results after the implementation of “hard” thresholds	84
8.4.2 Results after the implementation of additional marginality indexes	87
9. Design and development of computation workflows for all Europe	91
9.1 Workflow design for indicator sets and algorithm development to produce new Hard layers	91
9.1.1 Workflow design to produce intermediate hard layer “Other Land Cover” .	91
9.1.2 Workflow design to produce intermediate hard layer “Forest”	92
9.1.3 Workflow design to produce intermediate hard layer “Croplands”	93
9.1.4 Workflow design to produce intermediate hard layer “Protected areas”	95
9.1.5 Workflow design to produce intermediate hard layer “Impervious”	95
9.1.6 Workflow design to produce intermediate hard layer “Changed Areas”	96
9.1.7 Workflow design to produce intermediate hard layer “ML_HARD”	97
9.2 Workflow design for indicator sets and algorithm development to produce new SOFT layers	100
9.2.1 Workflow design to produce intermediate soft layer related to “Terrain and Soil”	101
9.2.2 Workflow design to produce intermediate soft layer related to “Sustainability”	103
9.2.3 Workflow to produce intermediate soft layer related to “Productivity”	104
9.3 Workflow design to produce final layer “ML_FINAL”	106
10. Implementation of workflows to the gis	108
10.1 Automatization of procedures and tools description for hard layers	108



10.2	Automatization of procedures and tools description for soft layers.....	114
10.3	Automatization of procedures and tools description for Final MLs.....	122
11.	Development of guidelines for the usage of indicators set and GIS workflows	125
11.1	General guidelines.....	125
11.2	Specific guidelines for the software and GIS workflows	126
11.3	Specific guidelines on databases.....	128
11.4	Specific guidelines for nomenclature, projections, and spatial resolution ..	128
12.	Experimental results for MLs identification	129
12.1	Experimental results for Hard marginal land indicators	129
12.2	Experimental results for marginal land identification applying soft indicators	136
13	1st methodology refinement (indicator fine-tuning based on the experimental results)	143
14	Conclusions	145
	References.....	146
	Annex I: Marginality index calculation and CLC reclassification.....	151
	Annex II: Table of figures.....	154
	Annex III: List of Tables	160
	Annex IV: Database structure.....	163



EXECUTIVE SUMMARY

The current methods for identifying marginal lands are diverse and mostly reflect specific management purposes. In these methods, many of the parameters and variables used in the identification of marginal lands are linked to land use/land cover and the specific characteristics of the Earth's surface, focusing on biophysical criteria.

There are several ways to analyze the marginality of a territory, but they are based on three approaches, firstly, on soil health criteria: erosion, poorly drained, frequently flooded land, low crop productivity, and high slopes. Secondly, it is to analyze the use of the soil focusing on those crops with very low productivity and fallow lands. The third approach is based on analyzing areas with a degree of environmental degradation such as abandoned mining and industrial areas, contaminated zones, water resource contamination, and farmland where irrigation could lead to water resource depletion. For the objective identification of these parameters and consequently of marginal lands at different scales, remote sensing and modern interpretation techniques are increasingly used.

The objective of this task was the development of a methodology for the detection of M/SM MLs. This methodology integrated land characteristics based on available information sources so that the criteria for defining marginal lands were satisfied. In the framework of the MAIL project, this task used the knowledge acquired in tasks 2.1 and 2.2, mainly the bibliographic review, the set of indicators, the thematic maps, and the criteria for the definition of marginal lands. Throughout this task, indicators, and guidelines for the identification of marginal lands were evaluated, selected, and developed, as well as mapping and database of marginal lands at the European level was generated.

The methodology implemented consisted of combining land use data and soil quality data. The land-use constraints were divided into two, "soft" constraints which were factors with variable thresholds (e.g., slope), and "hard" constraints we considered as binary exclusion factors (e.g., protected areas). The majority of the factors that indicate marginality could be considered "soft". These included biophysical factors such as slope, elevation, soil quality/fertility, and erosion, which are inherent properties of the land or soil. Current land-use and policy can be classified as "hard" constraints.

The methodology was divided into five steps. The first step was the selection of the studies analyzed in Task 2.1 to establish the criteria and thresholds for identifying marginal lands. The second was the selection of the datasets proposed in task 2.2 and pre-processing of the data. The third step was the implementation of the "hard" thresholds and constrains methodology



which relies on the exclusion of areas that do not meet the requirements of the definition of Marginal Lands (i.e., agricultural lands, forest and impervious areas, permanent water, and snow areas, peat bogs, marshes, and more) and fourthly, the “additional indicators/thresholds” phase based on the development of specific sets of additional indicators (i.e., “soft” constraints). A weighted overlap analysis was then performed considering the physical characteristics of the possible areas of interest across Europe. The final step in the mapping of marginal lands, was the reclassification of the resulting product of the weighted overlay into 3 classes, depicting marginality: 1) *Marginal lands with high plantation suitability*, 2) *Marginal lands with low plantation suitability* and 3) *Potentially unsuitable lands*.

As a result, marginal land maps (hard constraints) and marginality maps (soft constraints) were obtained for Europe and the pilot sites proposed in the MAIL project. The results are currently being validated in task 2.4. The developed methodology will be adjusted and fine-tuned based on the results of the accuracy assessment evaluation.

1. INTRODUCTION

This document contains the methodology and key aspects to be taken into account for marginal land detection under the scope of **MAIL** project.

The starting point is the definition of marginal land as pointed out in Deliverable 2.1. As per D2.1 marginality is driven by three main forces: environmental factors, socioeconomic factors and cultural factors.

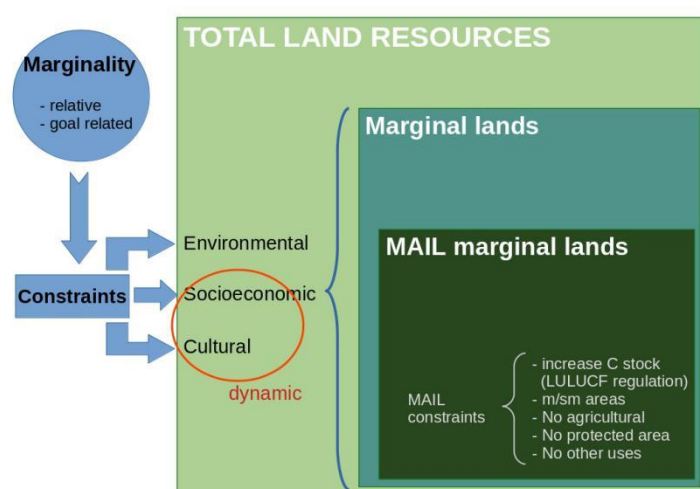


Figure 1: Transition between marginality and the definition on marginal lands in the framework of MAIL project.

These forces or factors directly influence land use, so that land cover data can be considered an aggregate of all these variables (Bertaglia, Joost, & Roosen, 2007). In addition, there are some constraints to relative marginality related with **MAIL** scope:

- Increase of C stock as per LULUCF regulation³
- Mountainous and semi-mountainous areas
- No agricultural zones
- No protected areas
- No other uses

³ The EU has revised its legislative framework in order to meet the requirements for climate change mitigation under the 2015 Paris Agreement. As part of this framework the Land Use, Land Use Change and Forestry (LULUCF) regulation (2018/841) was adopted in May 2018 (European Parliament, 2018).



Marginal land detection was performed using primary data (forces related to the general definition of marginal lands) and secondary data (related to **MAIL** definition of marginal land). In the next tables the variables chosen for marginal land identification are summarized:

Table 1: Factors for marginal land identification under *MAIL* scope. Source: Personal compilation.

Factor	Dataset	Variable
Environmental factors	CORINE Land Cover	Land cover
Cultural factors		

Table 2: Constrains for marginal land identification under *MAIL* scope Source: Personal compilation.

Constrain	Dataset	Variables
Increase of C stock as per LULUCF regulation	Pan-European High-Resolution Layers (HRL)	Forest variables (Tree Cover Density)
Mountainous and semi-mountainous areas	Digital European Elevation Model	Elevation
No agricultural zones	CORINE land cover	Land cover
No protected areas	Environmental dataset	Protected areas
No other uses	S2GLC	Land cover ⁴
Other factors ⁵	European Soil Database Derived data	Soil variables

⁴ Land use is proposed as variable to detect other uses, as extensive grazing, that could compete with the uses proposed under the scope of **MAIL** project. Land uses classes will be analyzed regionally by specialists in order to detect uses that will constrain afforestation or other actions proposed by **MAIL**.

⁵ Variables related with extra factors (mainly soil variables) will be utilized in order to improve marginal land detection under the scope of **MAIL** project.



2. DEFINITION OF MINIMUM MARGINAL LAND PARCEL AREA

Under **MAIL** project, the minimum area of marginal land⁶ was set (a priori) as 1 ha (10.000 m²). Marginal land with area less than this minimum parcel size will not be studied under the scope of **MAIL** project. Considering the European scale of the project, the aforementioned size was considered as an equilibrium point between available resources and marginality characteristics at local scale.

3. DATASETS USED

3.1 CORINE Land Cover

In this study the CORINE Land Cover (CLC) database will be used as primary source of data. It provides a pan-European inventory of biophysical land-cover. The CLC inventory was initiated in 1985 (reference year 1990) and was created from remotely sensed data. Updates have been produced in 2000, 2006, 2012, and 2018. CLC provides information about land-cover changes for a substantial part of Europe. CLC uses a Minimum Mapping Unit (MMU) of 25 hectares for areal phenomena and a minimum width of 100 m for linear phenomena. The time series are complemented by change layers, which highlight changes in land cover with an MMU of 5 ha.

The CORINE Land Cover database has been validated. The official classification accuracy of CORINE is $\geq 85\%$ except for the release of 2002 (G Büttner & Kosztra, 2017). CLC dataset will be utilized for general marginality detection and for the detection of land uses as agricultural and extensive grazing, that will compete with afforestation or other actions proposed under **MAIL**'s scope.

Table 3: Technical specifications of CLC dataset. Source: MAIL Deliverable D2.2.

Specification	Source data specification
File name: Corine Land Cover (CLC) 2018, Version 20	Sensor: -
Coordinate system: ETRS89 LAEA	Data type: -

⁶ Minimum parcel size of marginal land is subject to review according the first outputs of the detection methodology and the knowledge depicted from pilot cases.



Specification	Source data specification
Production date: 14 - 06 - 2019	Sensor resolution: -
Coverage (top L, BR coordinates): EEA 39	Acquisition date: 2012 - 2018
Grid size: 25 ha / 500 m	Grid size: -
Position accuracy: 100 m	Positional accuracy: -
Vertical accuracy: -	Vertical accuracy: -
Completeness: Complete	
File type, format: Vector, AutoCAD Slide (.sld) & ArcGIS Layer (.lyr)	

3.2 Pan-European High-Resolution Layers (HRL)

The Pan-European High-Resolution Layers (HRL) (European Environment Agency, 2015) provide information on specific land cover characteristics, and are complementary to land cover / land use mapping such as the CLC datasets. The HRLs are produced from satellite imagery through a combination of automated processing and interactive rule-based classification. Since the compilation of the 2015 reference year the production is increasingly based on analyzing time series of satellite images from several different sensors, including the combination of optical and radar data. The main sources are the Sentinel Satellites, in particular Sentinel-2 and Sentinel-1. In addition to High-Resolution (HR) data, since 2015, very High-Resolution (VHR) imagery was also used for some of the products.

Five themes have been identified so far, corresponding with the main themes from CLC, i.e., the level of sealed soil (imperviousness), tree cover density and forest type, grasslands, wetness and water, and small woody features.

Tree Cover Density (TCD) will be utilized as the forestry variable in order to select the marginal lands where activities such as afforestation and reforestation will have more impact on the C stock accounting system as defined by the LULUCF directive⁷. According to this regulation, not all of the forest related sinks will count toward the mitigation target (Grassi et al., 2019), therefore carbon stock contribution, for legal purposes, will depend if performed on already managed forest lands or afforested and forested lands. Distinction between those categories

⁷ Regulation 2018/841 adopted in May 2018. This normative establish a land-based approach for accounting the emissions and removals from the LULUCF sector in five land accounting categories: (1) afforested and forested land; (2) managed cropland, grassland and wetland; (3) managed forest land; (4) harvested wood products; and (5) natural disturbances (Romppainen, 2019).



will be performed through TCD thresholds in order to select the most suitable marginal lands for MAIL project.

Table 4: Technical specifications of Pan-European High-Resolution Layers (HRL) dataset.

Source: MAIL Deliverable D2.2.

Specification	Source data specification
File name: Tree Cover Density (TCD)	Sensor: <u>Sentinel-2</u> : Multispectral instrument (MSI). <u>Landsat-8</u> : Operational Land Imager (OLI)
Coordinate system: ETRS89 LAEA	Data type: Sentinel-2: TOA reflectances (Level 1), TOA radiances in sensor geometry (L1B) (Level 1) and BOA reflectances in cartographic geometry (L1C) (Level 2)
Production date: 22 - 03 - 2018	Sensor resolution: <u>Sentinel-2</u> : 10-60 m <u>Landsat-8</u> : 30 m (visible, NIR & SWIR), 100 m (thermal) and 15 m (panchromatic)
Coverage (top L, BR coordinates): *	Acquisition date: 2012 - 2015
Grid size: 20 m	Grid size: -
Position accuracy: Less than one pixel	Positional accuracy: -
Vertical accuracy: -	Vertical accuracy: -
Completeness: Complete	
File type, format: Raster, TIFF image	

3.3 Digital European elevation model

The Digital European elevation model (EU-DEM) is a digital surface model (DSM) of EEA member and cooperating countries representing the first surface as illuminated by the sensors. It is a hybrid product based on SRTM and ASTER GDEM data fused by a weighted averaging approach (Digital Elevation Model over Europe (EU-DEM), (European Environment Agency, 2017)). This is the v1.1 of EU-DEM, based on data acquired in 2011.

EU-DEM will be used as elevation source in order to identify mountainous and semi-mountainous marginal lands under **MAIL**'s scope.

Table 5: Technical specifications of Digital European elevation model dataset. Source: MAIL Deliverable D2.2.

Specification	Source data specification
File name: Digital Elevation Model of Europe v1.1	Sensor: (GLASS) Geoscience Laser Alimeter System
Coordinate system: ETRS89 LAEA	Data type: Level 1A, 1B, 2 and 3 data products
Production date: 20 - 04 - 2016	Sensor resolution: 60-70 m x 60-70 m
Coverage (top L, BR coordinates): **	Acquisition date: 2011
Grid size: 25 m	Grid size: -
Position accuracy: -	Positional accuracy: -
Vertical accuracy: +/- 7 m RMSE	Vertical accuracy: -
Completeness: Complete	
File type, format: Raster, Geotiff 32 bits	

3.4 Environmental dataset

In order to identify the legally protected areas, the European network of protected sites (Natura 2000) was used. Natura 2000 is the key instrument to protect biodiversity in the European Union. It is an ecological network of protected areas, set up to ensure the survival of Europe's most valuable species and habitats. Natura 2000 is based on the 1979 Birds Directive and the 1992 Habitats Directive. This version covers the reporting of 2018.

The European database on Natura 2000 sites consists of a compilation of the data submitted by Member States to the European Commission. This European database is generally updated once per year. (Natura 2000 data - the European network of protected sites (European Environment Agency, 2020)).

Table 6: Technical specifications of Environmental dataset. Source: MAIL Deliverable D2.2.

Specification	Source data specification
File name: Natura 2000 data - the European network of protected sites	Sensor: -
Coordinate system: ETRS89 LAEA	Data type: -



Specification	Source data specification
Production date: 12 - 04 - 2019	Sensor resolution: -
Coverage (top L, BR coordinates): Europe	Acquisition date: 2018
Grid size: 1: 100.000	Grid size: -
Position accuracy: -	Positional accuracy: -
Vertical accuracy: -	Vertical accuracy: -
Completeness: Complete	
File type, format: Vector, Shapefile	

3.5 European Soil Database Derived data

Regarding soil variables, the European Soil Database Derived data compiled by the European Soil Data Centre (ESDAC) was used. Several layers for soil properties have been created based on data from the European Soil Database in combination with data from the Harmonized World Soil Database (HWSD) and Soil-Terrain Database (SOTER). The available layers include:

- Total available water content
- Depth available to roots
- Clay content
- Silt content
- Sand content
- Organic carbon
- Coarse fragments
- Soil pH
- Soil erosion water (RUSLE 2015)
- Soil biomass productivity maps

The layers of soil properties of Soil Typological Units (STUs) are only intended to facilitate modelling purposes. The final result of the modelling activity should be aggregated to SMUs or another larger mapping unit.



Table 7: Technical specifications of the European Soil Database Derived data dataset. Source: MAIL Deliverable D2.2.

Specification	Source data specification
File names: Various ⁸	Sensor: -
Coordinate system: ETRS89 LAEA	Data type: -
Production date: 2013	Sensor resolution: -
Coverage (top L, BR coordinates): Europe	Acquisition date: 2013
Grid size: 1 km	Grid size: -
Position accuracy: -	Positional accuracy: -
Vertical accuracy: -	Vertical accuracy: -
Completeness: Complete	
File type, format: Raster, Idrisi raster format	

3.6 ISRIC World Soil Information

To complete the set of soil variables the ISRIC World Soil dataset was used. The ISRIC is a science-based organization than providing quality evaluated soil data and interpreted soil information. They maintain a thorough knowledge of soil evaluation, soil analysis and soil data management. The ISRIC provides among other data sets, the World Inventory of Soil Emission Potentials (WISE) derived soil properties on a 30 by 30 arc-seconds global grid and the World Reference Base for Soil Resources (WRB). Available layers that have been used in **MAIL** include:

- Soil classes and probabilities.
- Soil moisture
- Cation exchange capacity
- Gypsum
- Sodium
- Nitrogen

⁸ Area of STU allocation, Depth available to roots, Clay content (topsoil & subsoil), Sand content (topsoil & subsoil), Silt content (topsoil & subsoil), Organic carbon content (topsoil & subsoil), Bulk density (topsoil & subsoil), Coarse fragments (topsoil & subsoil), Total water content from PTR and PTF (topsoil & subsoil).

**Table 8: Technical specifications of the ISRIC World Soil Information dataset.**

Specification	Source data specification
File names: WRB & WISE	Sensor:
Coordinate system: WGS84	Data type: -
Production date: WRB: 2020 & WISE: 2016	Sensor resolution:
Coverage (top L, BR coordinates): World	Acquisition date: WRB: 1905 – 2016 & WISE: 1950 - 2015
Grid size: WRB: 250 m & WISE: 30 arc-seconds (\approx 900 m)	Grid size: -
Position accuracy: -	Positional accuracy: -
Vertical accuracy: -	Vertical accuracy: -
Completeness: Complete	
File type, format: Raster	

3.7 S2GLC dataset description

S2GLC land cover map is one the most detailed pan-European land cover products. It was produced using automatic classification approach and Sentinel-2 images from 2017. It contains 13 classes with MMU equal to Sentinel-2 pixel which is 10x10 m. The overall accuracy is 86%. The product is available in two forms: mosaic for the whole Europe, Sentinel-2 tiles.

Land cover/land use classes of S2GLC:

- Artificial surfaces
- Natural material surfaces (consolidated and un-consolidated)
- Broadleaf tree cover
- Coniferous tree cover
- Herbaceous vegetation
- Moors and heathland
- Sclerophyllous vegetation
- Cultivated areas
- Vineyards
- Marshes



- Peatbogs
- Water bodies
- Permanent snow cover

Table 9: Technical specifications of the S2CLG layer.

Specification	Source data specification
File names: 2GLC_Europe_2017_v1.2_grey.tif	Sensor: Sentinel-2
Coordinate system: ETRS89 LAEA	Data type: Thematic mapper
Production date: 2020	Sensor resolution: 10 m
Coverage (top L, BR coordinates): Europe	Acquisition date:
Grid size: 110 x 110 m	Grid size: -
Position accuracy: -	Positional accuracy: -
Vertical accuracy: -	Vertical accuracy: -
Completeness: Complete	
File type, format: Raster	



4. DEFINITION OF GIS SPECIFICATIONS

In the table below the original characteristic of each dataset used are summarized.

Table 10: Main specifications of the datasets used. Source: Personal compilation.

Dataset	Layer	File format	Projection	Coverage	Unit
CORINE Land Cover	Land cover	Vector	ETRS89 LAEA	EEA 39 ⁹	quantitative
Pan-European High-Resolution Layers	Tree Cover Density	Raster	ETRS89 LAEA	EEA 39 ⁹	%
Digital European elevation mode	EU-DEM	Raster	ETRS89 LAEA	EEA 39 ⁹	meters
Environmental dataset	Natura 2000 network	Vector	ETRS89 LAEA	Europe	quantitative
European Soil Database Derived data	10 layers ¹⁰	Raster	ETRS89 LAEA	Europe	various
Joint Research Centre	1 Layer	Raster	ETRS89 LAEA	World	%
ISRIC	6 Layers	Raster	various	World	various
TerraClimate	2 Layer	Raster	ETRS89 LAEA	World	mm
S2GLC	1 Layer	Raster	ETRS89 LAEA	World	meters

In order to homogenize and allow the required interoperability between each dataset, the following specifications should be followed.

⁹ European Economic Area (EEA) 39: 33 member countries and six cooperating countries. The 33 member countries are the 28 European Union Member States, together with Iceland, Liechtenstein, Norway, Switzerland and Turkey. The six cooperating countries are Albania, Bosnia and Herzegovina, Kosovo*, Montenegro, North Macedonia and Serbia.

¹⁰ Total available water content, Depth available to roots, Clay content, Silt content, Sand content, Organic carbon, Bulk Density, Coarse fragments



4.1 Coordinate Projection

For all layers the horizontal coordinate system shall be European Terrestrial Reference System 1989 (ETRS89) using Lambert Azimuthal Equal-Area projection (LAEA).

4.2 File format

File format will be raster data (discrete) due to the fact that those are very useful for analysis and for storing data that varies continuously.

4.3 Coverage

The extent used will be the 28 European Member States.



5. REVIEW OF EXISTING RESEARCH PROJECTS AND LITERATURE FOR AVAILABLE GIS MODELS & WORKFLOWS

As already presented on **MAIL**'s Deliverable 2.1, there is not a unique definition of marginal land. Therefore, diverse methodologies for marginal land definition have been developed. Considering the **MAIL** scope and the literature reviewed, the most suitable approach for marginal land identification is described as follows.

5.1 Marginality index

In compliance with (Bertaglia et al., 2007), a marginality index could be computed. The objective is to provide a quantitative operational tool, an integrated index based on land cover data, as a measure of integrated marginality. This variable aims at synthesizing information on the marginal characteristic of a region given its land use pattern.

Relative marginality concerning land use, is defined as those areas where there is relatively more non-productive and/or less productive land, compared both to land used for arable farming and to urban areas and transportation infrastructure (more productive classes)¹¹. The method chosen is to compute a ratio of land use type that can describe the relative abundance of marginal land use in a certain area. As such, an area could be less marginal because it has a high proportion of high-intensity arable-land-based farming systems and/or a high proportion of urban or transport infrastructure. The ratio obtained is thus a good indicator of integrated relative marginality resulting from a different combination of factors.

5.2 Land use change

In order to include the dynamic aspect of marginality definition (see Figure 1) in the identification methodology, the change of land use will be included. During the literature review two approaches to integrate dynamism were found:

1. CLC change layer. Under this approach CLC change layers of different periods could be used to calculate the probability of change of each use.
2. Change detection. Computation of Transition Potential matrix used for the prediction of land use change using CA-Markov model (Dzieszko, 2014; Hamad, Balzter, & Kolo, 2018).

¹¹ Detailed description concerning the calculation of the marginality index and the reclassification of the CLC classes required can be found in Annex I: Marginality index calculation and CLC reclassification.



5.3 Tree Cover Density

The Tree Cover Density (pan-European High-Resolution Layer) will be utilized as forestry variable in order to select the marginal lands where activities such as forestation and reforestation will have more impact on the C stock accounting system as defined by the LULUCF directive¹². According this regulation, not all of the forest related sinks will count toward the mitigation target (Grassi et al., 2019), therefore carbon stock contribution, for legal purposes, will depend if performed on already managed forest lands or afforested and forested lands. Distinction between those categories will be performed through TCD thresholds in order to select the most suitable marginal lands for the **MAIL** project.

5.4 Digital Elevation Model of Europe

The European Digital Model will be used for the selection of marginal lands located on mountainous or semi-mountainous areas. Thresholds for this variable should be set in regionally taking into consideration the climate regions of Europe.

5.5 Protected areas. European network of protected sites (Natura 2000)

This layer, where all protected areas under Natura 2000 network are included, will be used as restriction layer to exclude from the **MAIL** set of marginal lands those with any kind of protection status.

¹² Regulation 2018/841 adopted in May 2018. This normative establishes a land-based approach for accounting the emissions and removals from the LULUCF sector in five land accounting categories: (1) afforested and forested land; (2) managed cropland, grassland and wetland; (3) managed forest land; (4) harvested wood products; and (5) natural disturbances (Romppainen, 2019).



6. DEVELOPMENT OF THE FINAL ML DETECTION, CRITERIA AND CLASSIFICATION SCHEME.

6.1 ML definition refinement

In the previous work done in T2.1 “Literature review on Marginal Land definition”, in Deliverable 2.1, marginal lands for the **MAIL** project are defined as:

Lands with significant, either environmental (biophysical variables) or socioeconomic, constraints and with potential to impact national accounting for C stock, excluding agricultural lands and other valuable areas (protected areas, uses with local importance, etc.). Dynamic and variability are key concepts for marginal land identification.

Examples of these areas include, but are not limited to, degraded and/or abandoned lands, lands with naturally low productivity due to biophysical constraints, and other degraded lands that have not (yet) been converted to other uses, e.g., post-industrial and post-mining sites.

This definition is further clarified in this document by taking into consideration the final indicators and criteria used in **MAIL** methodology for defining the MLs across Europe. Therefore, based on the indicators and criteria that will be analyzed in paragraphs 6.3 and 6.4, up until now, the definition can be modified to:

Lands with significant, either environmental (biophysical variables) or socioeconomic, constraints and with potential to impact national accounting for C stock, excluding *agricultural lands, forest and impervious areas, permanent water and snow, peatbogs and marshes, as well as protected areas.* MLs are further defined by constraints in a) *terrain and soil physical variables such as slope, depth, texture, stoniness, drainage, water capacity, moisture, clay and sand,* b) *soil sustainability variables such as salinity, acidity, erosion, flood, sodicity, contamination, dryness and toxicity,* c) *soil productivity variables such as organic matter and cation exchange capacity.* Dynamic and variability are key concepts for marginal land identification.

6.2 Methodology overview

The literature review performed in T2.1 showed that most authors use a combination of land use/land cover and soil data to classify MLs. Additionally, climate, socio-economic and elevation data were regularly considered. The temporal dimension was investigated by evaluating multi-temporal datasets to assess land use and land productivity trends. Spatial classification was mostly implemented through geographic overlay of input data, using either



a combination of binary constraints or fuzzy logic for ML classification. Constraints were divided into “soft” or “hard” constraints, “soft” constraints being factors with variable thresholds (e.g., elevation) and “hard” constraints binary exclusion factors, e.g., protected areas. Most factors indicative of marginality can be considered as “soft”. These include biophysical factors such as slope, elevation, soil quality/fertility and erodibility, which are inherent properties of the land or soil. Current land use and policy can be classified as “hard” constraints. Land that is currently in active use for agriculture cannot be seen as marginal, even if it has all characteristics of ML. This includes lands temporarily fallow as part of crop rotation. Protected areas are also excluded from marginal land classification within the **MAIL** project.

For the proper selection of indicators/criteria, along with the thresholds that will be used in the **MAIL** project, it was necessary to develop a methodological outline that will act as a roadmap for the next project tasks. The scope of the project is the development of a common methodology that will be applied in large areas across Europe. However, an important parameter affecting the implementation of this common methodology is the physical characteristics (i.e., geological, hydrological, and more) of each selected test site. Namely, the implementation of a specific set of indicators and thresholds, returned adequate results when applied in semi-mountainous areas. Nevertheless, the same methodology was not acceptable for lowland or semi-urban areas.

Thus, within the **MAIL** project, the proposed methodology is divided into two main steps: The implementation of the “hard” thresholds and constrains methodology which relies on the exclusion of areas that do not meet the requirements of the definition of Marginal Lands (i.e. agricultural lands, forest and impervious areas, permanent water and snow areas, peatbogs, marshes, and more) and the “additional indicators/thresholds” phase based on the development of specific sets of additional indicators (i.e. “soft” constrains) that will be applied according to the physical characteristics of each selected test site. These indicators are considering additional variables including the terrain and soil characteristics of an area and soil productivity constraints.

In the first phase, a top-down stepwise approach is followed (Figure 2), in which areas that are not MLs are incrementally removed, based on thresholds of various marginality criteria and indicators. Moving on, in phase two, the resulting potential MLs are further downsized using additional indicators, related to Terrain, Soil, Sustainability, Climate and Productivity thresholds. It should be noted that the selection of the final sets of indicators is performed considering the physical characteristic of each area of interest. Finally, the resulting MLs are classified into 3 classes.



Figure 2: The *MAIL* methodology overview of the first phase of the top-down stepwise approach, Source: personal compilation of Maria Tassopoulou.

In phase one, starting with the whole European area, areas are excluded, based on land cover type (i.e., urban areas, protected areas, water and forest areas, areas covered with snow, and more). This phase is the “hard” thresholds and constraints methodological approach. In the second phase additional indicators (i.e., slope, moisture, salinity, productivity, etc.) are also implemented based on thresholds found in literature. The indicators/criteria used, are grouped by type, according to categorization made in D2.1, Table 7 (soil, climate, terrain, sustainability, productivity).

The aforementioned indicators of this second phase that constitutes a “soft” constraint and has non-thematic data (numerical data), is ranked according to its importance (times found in literature) and weights are assigned to each one based on a Pairwise Comparison Matrix (PCM) of ranks (Zolekar & Bhagat, 2015). The remaining values of each indicator are grouped into 3 classes according to their physical characteristics.

Taking into account the results of the weighted overlay analysis and the physical characteristics of possible areas of interest across Europe, finally, three different sets of indicators of “soft” constraints are developed as analyzed in the following chapters of this report (see chapter 8), (Figure 3).

The final step in the mapping of marginal lands, is the reclassification of the resulting product of the weighted overlay into 3 classes, depicting marginality: 1) *Marginal lands with high plantation suitability*, 2) *Marginal lands with low plantation suitability* and 3) *Potentially unsuitable lands*.

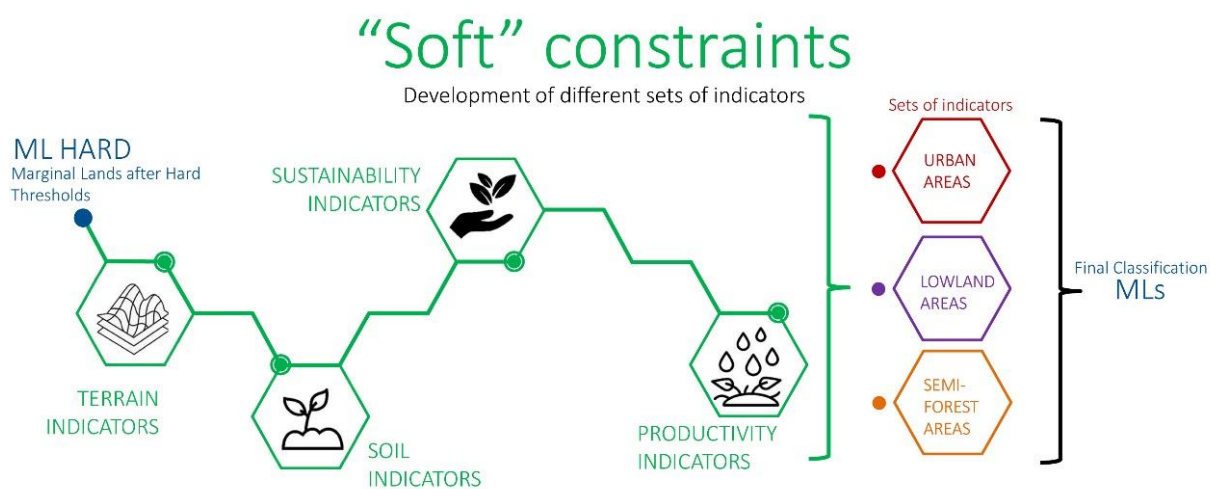


Figure 3: The *MAIL* methodology overview of the second phase of the proposed methodology regarding the “Soft” constraints and indicators, Source: personal compilation of Maria Tassopoulou.

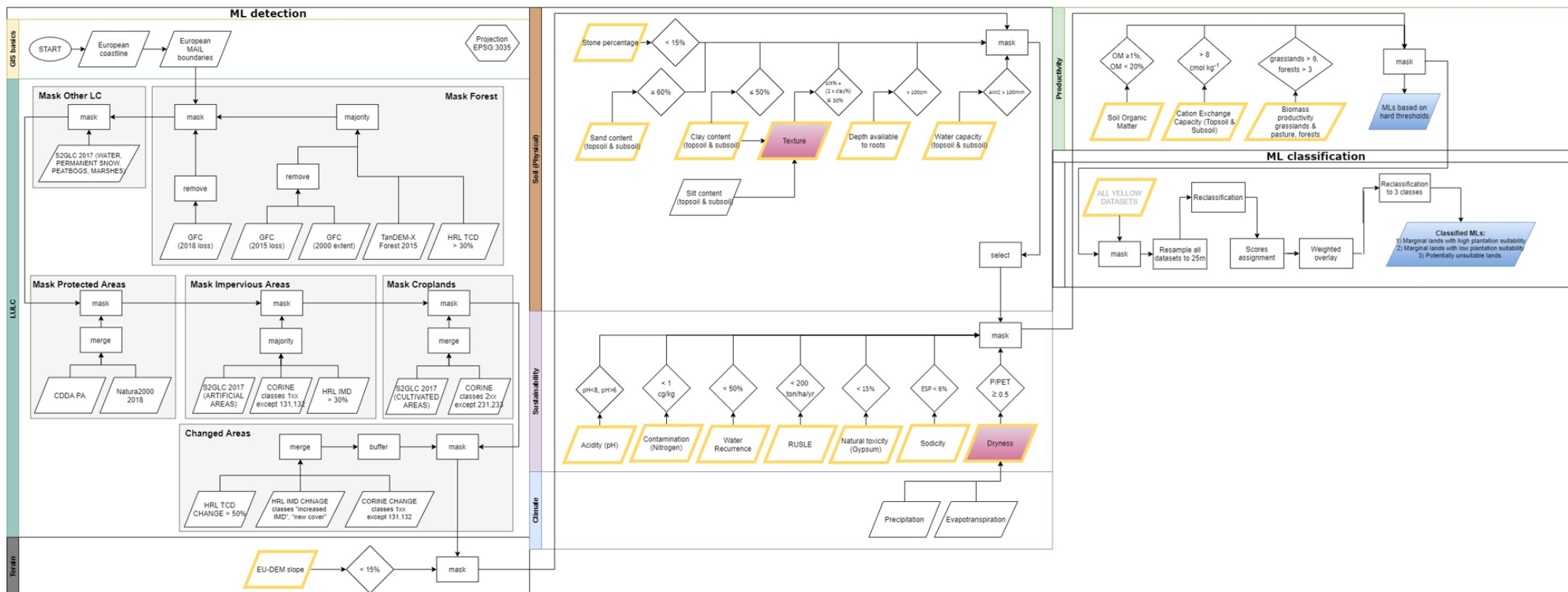


Figure 4: Workflow of the methodology for mapping and classifying MLs in *MAIL*. Yellow parallelograms depict datasets used in the weighted overlay, while pink parallelograms depict “synthetic” indicators/layers.

6.3 Indicator / criteria selection

In general, the indicators and criteria for defining the MLs in the **MAIL** project, were selected by juxtaposing the indicators and criteria used in selected ML studies from T2.1 “Literature review on Marginal Land definition” and the available datasets from T2.2 “Collection of appropriate existing European/Global datasets”, taking into consideration the indicator units in both datasets and literature (Figure 5).

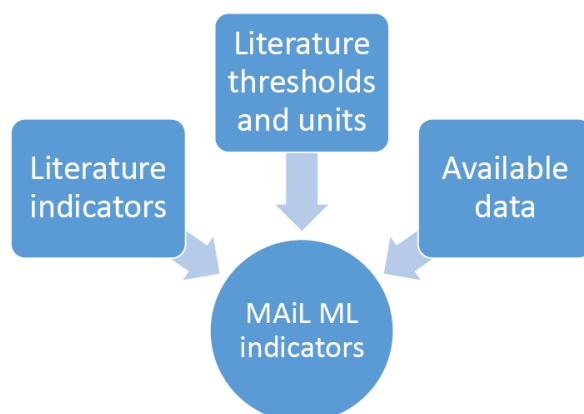


Figure 5: Overview of the indicator and criteria selection in the **MAIL project.**

A subset of 24 studies was made (Table 11), by selecting studies used in the literature review in T2.1 “Literature review on Marginal Land definition”. The subset contained studies only focusing on methodological aspects of mapping MLs. For each study, information regarding various methodological aspects were noted (study extent, MMU, technology used, datasets, creation of synthetic layers/indicators, ML classification scheme), as well as all the indicators used, grouped by type, according to categorization made in D2.1, Table 7 (soil, climate, terrain, sustainability, productivity, LULC, socioeconomic). Additionally, indicator thresholds were also noted, where a clear threshold was used.

Table 11: The subset of 24 studies on which the indicator selection was based on.

No.	Year	Authors	Study Title
1	2019	Ciria, Sanz, Carrasco, & Ciria	Identification of Arable Marginal Lands under Rainfed Conditions for Bioenergy Purposes in Spain
2	2018	Peter, Messina, & Snapp	Multiscalar approach to mapping marginal agricultural land: Smallholder agriculture in Malawi
3	2018	Vlachaki, Gounaris, Dimitriadis, & Galatsidas	Final guidelines for the sustainable exploitation of Marginal Lands for bioenergy (D6.8)
4	2018	Gerwin et al.,	Assessment and quantification of marginal lands for biomass production in Europe using soil-quality indicators



5	2018	Sallustio et al.,	Assessing the economic marginality of agricultural lands in Italy to support land use planning
6	2018	Elbersen et al.,	Mapping Marginal land potentially available for industrial crops in Europe
7	2017	Li, Messina, Peter, & Snapp	Mapping Land Suitability for Agriculture in Malawi
8	2016	Ivanina, Roik, & Hanzhenko	Report on MagL concepts, debate and indicators (D2.3)
9	2015	Zolekar & Bhagat	Multi-criteria land suitability analysis for agriculture in hilly zone: Remote sensing and GIS approach
10	2013	Bandaru et al.,	Soil Carbon Change and Net Energy Associated with Biofuel Production on Marginal Lands: A Regional Modeling Perspective
11	2013	Gelfand et al.,	Sustainable bioenergy production from marginal lands in the US Midwest
12	2013	Kang et al.,	Hierarchical marginal land assessment for land use planning
13	2012	Liu et al.,	Bioenergy production potential on marginal land in Canada
14	2011	Cai, Zhang, & Wang	Land Availability for Biofuel Production
15	2011	Gopalakrishnan, Cristina Negri, & Snyder	A Novel Framework to Classify Marginal Land for Sustainable Biomass Feedstock Production
16	2011	Schweers et al.,	Identification of potential areas for biomass production in China: Discussion of a recent approach and future challenges
17	2010	Eliasson et al.,	Common criteria for the redefinition of Intermediate Less Favored Areas in the European Union
18	2010	James	Theory and identification of marginal land and factors determining land use change
19	2009	Milbrandt & Overend	Assessment of Biomass Resources from Marginal Lands in APEC Economies
20	2008	Bai, Dent, Olsson, & Schaepman	Global Assessment of Land Degradation and Improvement 1. Identification by remote sensing
21	2007	Bertaglia et al.,	Identifying European marginal areas in the context of local sheep and goat breeds conservation: A geographic information system approach
22	2006	Niu & Duiker	Carbon sequestration potential by afforestation of marginal agricultural land in the Midwestern U.S.
23	2005	Roehrig & Menz	The Determination of Natural Agricultural Potential in Western Africa Using the Fuzzy Logic Based Marginality Index
24	2018	Elbersen et al.,	Deliverable 2.6 Methodological approaches to identify and map marginal land suitable for industrial crops in Europe

Emphasis was given to soil, climate, terrain, sustainability, productivity and LULC constraints. In the following paragraphs, indicators and thresholds are reviewed. Table 12, Table 13 and Table 14 summarize the final indicators chosen for soil, climate, terrain, sustainability, and productivity constraints which are analyzed in paragraph 6.4.

Table 12: Soil and terrain marginality indicators used in *MAIL*.

Indicator	Times Found in Literature
slope	18
depth available to roots	18
texture	9
stoniness	8
drainage	7
water capacity	6
moisture	5
clay	4
sand	4

Table 13: Sustainability marginality indicators used in *MAIL*.

Indicator	Times Found in Literature
salinity	10
acidity (pH)	9
erosion	8
flood	6
sodicity	5
contamination	4
dryness	2
natural toxicity	1

Table 14: Productivity marginality indicators used in *MAIL*.

Indicator	Times Found in Literature
soil organic matter	8
cation exchange capacity	4
productivity	3

6.4 Initial indicator classification categories and thresholds

In general, the indicators and criteria for defining the MLs in the *MAIL* project, were selected by combining the indicators and criteria resulting from the literature review with the available datasets from T2.2, taking into consideration the indicator units in both datasets and literature (Figure 5).

In all indicator cases, when more than one threshold was observed, the final threshold chosen for the *MAIL* project was the one producing larger MLs area. Following this, a match was tried to be achieved between each indicator to a corresponding dataset from T2.2. In cases where a match could not be found between an indicator and a dataset (conceptually and in units), the indicator was omitted. For example, no dataset existed for “Phosphorus level”. Evidently, the



study in which this indicator was used (Zolekar & Bhagat, 2015), used national data for the country studied which is not available in a European level.

6.4.1 Land Use/Land Cover indicators

Many authors have used Land Use/Land Cover (LULC) in detecting MLs (Bai et al., 2008; Bertaglia et al., 2007; Gelfand et al., 2013; Gerwin et al., 2018; Niu & Duiker, 2006). LULC is considered as a single indicator with multiple sub-variables that form the “hard” constraints to marginality. LULC datasets are thematic datasets and thus thresholding is implemented in the form of masking.

In the **MAIL** project, LULC datasets are used in phase one. Starting with the whole European area, areas are excluded, initially based on land cover type. More specifically, areas excluded from the analysis, based on LULC are forest areas, croplands, impervious areas, protected areas, water, permanent snow, marshes, and peatbogs.

Moreover, to incorporate the dynamic aspect of MLs, changed areas are also considered. In particular, Tree Cover Density (TCD) will be utilized as forestry variable in order to select the marginal lands where activities such as afforestation and reforestation will have more impact on the C stock accounting system. In addition, an increase in imperviousness over the past years will also be considered, by masking around a buffered zone of areas marked as such.

6.4.2 Terrain and Soil indicators

6.4.2.1 Slope

The slope is the angle the soil surface makes with the horizontal, expressed in degrees or as a percentage. It is considered as one of the main criteria for MLs. Not only it affects water drainage and erosion, but it is also associated with shallow soils and mechanical constraints in agriculture (Eliasson et al., 2010). Sallustio et al. (2018), used a >30% slope threshold for unsuitable agricultural lands in Italy. Elbersen et al. used a threshold of >17.5 degrees (30%) in pair wise combinations for mapping MLs suitable for growing industrial crops. Marginally suitable agricultural lands had slopes 6-16% and marginally unsuitable 16-30% in Li, Messina, Peter, & Snapp, (2017). Ivanina, Roik, & Hanzhenko, (2016) excluded areas with slopes $\geq 15\%$ for mapping agricultural ML suitable for bioenergy crops. Zolekar & Bhagat, (2015) considered slopes 12-20 degrees (21-36%) marginally suitable for agriculture, while slopes >20 degrees (36%), unsuitable. For Gelfand et al., (2013), MLs have slopes <20%. Marginal agricultural land for biomass feedstock production was mapped, by selecting areas with slope >15% in Gopalakrishnan, Cristina Negri, & Snyder, (2011). Schweers et al., (2011) marked slopes >25 degrees (47%), as unsuitable for agriculture. Unfavorable agriculture areas in the EU have



slopes >15%, according to Eliasson et al., (2010). Finally, lands with moderate (8-16%) or steep slope (16-30%), are considered as marginal for biomass production in Milbrandt & Overend, (2009).

In the **MAIL** project the threshold chosen was the smallest one, in other words MLs in **MAIL** have slopes $\geq 15\%$. It should be noted that slope is the marginality index with the most significant impact when identifying marginal lands. Thus, we suggest using the slope index in the final classification phase, after the implementation of the remaining soft thresholds (i.e., after the implementation of other indexes such as depth available to roots, texture, water storage capacity and more).

6.4.2.2 Rooting depth

Rooting depth is the depth from the soil surface to a hard rock, in which plant roots can grow. Apart from rooting depth being a constraint itself, shallow soil affects also water capacity (Zolekar, 2018) and is thus a major factor for characterizing a land as marginal. Ciria, Sanz, Carrasco, & Ciria, (2019) used a threshold of 100 cm to map arable MLs below this. Elbersen et al. used depths <30 cm in pair wise combinations for mapping MLs suitable for growing industrial crops. Li et al., (2017) classify soil depth in 5 classes, with MLs having depths <50 cm. Ivanina, Roik, & Hanzhenko, (2016) excluded areas with rooting depth <35 cm for mapping agricultural ML suitable for bioenergy crops. Marginally suitable agriculture soils were the ones with depths 30-50 cm in Zolekar & Bhagat, (2015). In a study by Kang et al., (2013), MLs had soil depths < 50 cm while in Eliasson et al., (2010), unfavorable agriculture areas in the EU have rooting depths <30 cm. Milbrandt & Overend, (2009) used depths of <50 cm to map marginal lands for biomass production.

In the **MAIL** project the threshold chosen was the largest one, in other words MLs in **MAIL** have rooting depths <100 cm.

6.4.2.3 Soil texture (soil structure/substrate)

Soil texture (also found as “soil structure” or “substrate”), refers to the relative proportions of different-sized soil particles, consisting of sand, silt, and clay. It controls the soil structure, fertility, and water availability (Eliasson et al., 2010; Zolekar, 2018) and is an important soil marginality indicator. Soils with high sand, clay or rock proportions are unfavorable for growing crops (Ivanina et al., 2016). For this indicator many authors use thresholds in percentages of soil particles while others use soil type as a proxy for texture. Ciria et al., (2019) use stoniness >15% to characterize lands as marginal. Elbersen et al., (2018) used a combination threshold of silt and clay ($\text{silt}\% + (2 \times \text{clay}\%) \leq 30\%$), sand <60% and stoniness >35%, for ML mapping.



Eliasson et al., (2010) used percentages of soil particles (>60% clay and >15%(v/v) of coarse fragments), but also soil type classes (clay, silty clay, or sandy clay with vertic properties) to detect MLs. Soil type was used as a proxy for texture in Li et al., (2017) to detect agriculture MLs, by selecting Sandy-Clay and Clay soil types in Malawi. Zolekar & Bhagat, (2015), used Loam Soil texture class for marginally suitable agriculture areas and last, Milbrandt & Overend, (2009), used coarse textured or sandy soils (Arenosols, Regosols, and Vitric Andosols with coarse texture, and all soils with petric and stony phase), to map marginal lands for biomass production.

In the **MAIL** project the percentage thresholds of different soil particles were chosen. Texture is considered as a “synthetic” indicator which depends on several other indicators (stoniness, silt, clay, sand). Thus, MLs in **MAIL** have stoniness% >10%, $\text{silt\%} + (2 \times \text{clay\%}) \geq 30\%$, clay $\geq 50\%$ and sand >60%.

6.4.2.4 Soil drainage

Soil drainage refers to the air (and thus oxygen) supply in soil pores by removal (or non-addition) of water (Eliasson et al., 2010; Li et al., 2017). It is usually assessed by soil type. Marginal lands in Elbersen et al. have Gleysols, Histosols, Stagnosols, Planosols, and soils with Histic, Gleyic and Stagnic primary qualifiers. Gleysols and Stagnosols are also found in Eliasson et al., (2010) for unfavorable agriculture areas in the EU. In some studies, drainage is found in local/state data as a variable per se (ex. MASDAP soil data record in Li et al., (2017), USDA-NRCS in Kang et al., (2013).

The only available datasets found in D2.2, associated with soil drainage throughout whole Europe, were soil type datasets. For the **MAIL** project, they were discarded as MLs indicator variables because they were considered dichotomous qualitative variables and involved a severe restriction of the MLs area.

6.4.2.5 Water storage capacity (maximum water holding capacity/profile available water)

Water storage capacity of soils determines soil depth, leaching process of nutrients and pesticides, the water availability of a soil profile available for vegetation, the cropping pattern, irrigation facilities, etc., and is a crucial parameter of soil quality, in particular in regions of rain deficit during the vegetation period (Mueller, Schindler, Behrendt, Eulenstein, & Dannowski, 2007; Zolekar & Bhagat, 2015). MLs in Ivanina et al. (2016) had a level of underground water in 0-80 cm, and a gleyic pattern in 0-40 cm, according to field surveys. Marginally suitable agriculture lands had Maximum Water Holding Capacity MWHC 100-200 and unsuitable lands



MWHC <100 in Zolekar & Bhagat, (2015). Kang et al. (2013) used the water table data by FAO as a marginality indicator with MLs having a water table <30 cm.

As none of the variables used in literature could match a European dataset from the [MAIL](#) task 2.2, it was decided to use the Available Water Capacity (AWC) from the ESDAC dataset “Topsoil physical properties for Europe (based on LUCAS topsoil data)”. AWC in this dataset was derived as the difference between the – 33 kPa and the – 1500 kPa water content (expressed as volume fraction). The threshold chosen was AWC <100 mm, based on Zolekar & Bhagat, (2015).

6.4.2.6 Soil moisture

Soil moisture is measured as the water availability during the plant’s growing period, which is crucial for plant normal growth and development and is thus an important ML indicator (Eliasson et al., 2010). Soil moisture is linked with numerous sub-factors such as length of growing period, temperature, precipitation, evapotranspiration, soil water content, and finally soil type. Eliasson et al. (2010), analyzed the number of days within the growing period (as defined by temperature >5°C), for which the amount of precipitation and water available in the soil profile exceeds half of potential evapotranspiration. Areas where the number of days is less than 90 are considered as unfavorable. Marginal lands had water content in the soil exceeding field capacity for at least 210 days in Elbersen et al. On the other hand, Zolekar & Bhagat (2015), classified NDWI to four classes (good, medium, less, very less), with class “less” being marginally suitable for agriculture. Last, in a more simple approach, Kang et al. (2013) defined soil moisture by soil type, with dry soils Aridic and Torric being an indicator for MLs.

In the [MAIL](#) project MLs have low soil moisture and are defined by soil types Aridic and Torric. For the [MAIL](#) project, they were discarded as MLs indicator variables because they were considered dichotomous qualitative variables and involved a severe restriction of the MLs area.

6.4.2.7 Soil type

Soil type is an indicator widely used as a marginality indicator, although it is mainly connected to other factors. Soil type is associated to *drainage* (Gleysols, Histosols, Stagnosols, Planosol, Soils with primary qualifiers Histic, Gleyic and Stagnic), in Elbersen et al. In Milbrandt & Overend, (2009), marginal land soil types are associated to numerous factors: *texture* (Arenosols, Regosols, and Vitric Andosols with coarse texture, and all soils with petric and stony phase), *salinity* (Solonchaks, Solonetz, and Solodic Planosols), *gypsum levels* (Gypsic Xerosols and Gypsic Yermosols), *calcium levels* (Calcisols) etc. In Kang et al., (2013), soil type is used as a proxy for *moisture* in MLs (Aridic, Torric). The soil type is used as an indicator

per se in another study, for mapping agricultural ML in Spain (Entisol and Aridisol soil types) (Ciria et al., 2019).

Because the soil type is used for defining other indicators in the **MAIL** project (drainage, moisture, salinity), it was not used as an indicator per se.

Table 15: Terrain and soil indicators and thresholds, sorted by importance (times used in the literature review).

Indicator	Data source	MAIL threshold	MAIL threshold based on	Literature Thresholds
slope	EU-DEM 2016 25m	$\geq 15\%$	(Gopalakrishnan et al., 2011)	$> 15\% - 47\%$
depth available to roots	ESDAC 2013 1km	$< 100\text{cm}$	(Ciria et al., 2019)	$< 30\text{cm} - 100\text{cm}$
texture	ESDAC 2013 1km	$\text{silt}\% + (2 \times \text{clay}\%) \geq 30\%$	(Elbersen et al., 2018)	$\text{silt}\% + (2 \times \text{clay}\%) \geq 30\%$
stoniness	ESDAC 2008	$> 10\%$	(Ciria et al., 2019)	$> 15\%$
drainage	WRB "Soil type" 2008	By soil type: Gleysols, Histosols, Stagnosols, Planosols	(Eliasson et al., 2010)	By soil type: Gleysols, Histosols, Stagnosols, Planosols
water capacity	ESDAC "Total available water content from PTF" 1km 2013	$\text{AWC} < 100 \text{ mm}$	(Zolekar & Bhagat, 2015)	$\text{AWC} < 100$, water table $< 0.3\text{m}$
moisture	WRB "Soil type" 2008	By soil type: Aridic, Torric	(Kang et al., 2013)	Various calculation methods and thresholds.
clay	ESDAC 2013 1km	$\geq 50\%$	(Eliasson et al., 2010)	$\geq 50\%$, $>60\%$
sand	ESDAC 2013 1km	$> 60\%$	(Elbersen et al., 2018)	$> 60\%$

6.4.3 Sustainability indicators

6.4.3.1 Soil salinity/alkalinity

Soil salinity is the presence of salts of alkalis (sodium, potassium, magnesium, and calcium) on the land surface, in soil. Soil salinity can be caused by environmental factors or human induced factors that disturb natural ecosystems. Soil salinity affects plant productivity and soil structure, leading to the creation of toxic substances and serious soil erosion (Eliasson et al., 2010), and is thus a major marginality indicator. Salinity levels above 6 dSm^{-1} , make winter cereal growth impossible for Spain in the study by (Ciria et al., 2019), whereas in Elbersen et al., (2018) Solonchaks and soils with a salic qualifier, soils with salt levels $>15 \text{ dSm}^{-1}$ in more than 50% of the mapping unit area were used in pair wise combinations for mapping MLs suitable for growing industrial crops. Soil type is again used for characterizing salinity, with saline soils being Solonchaks, Solonetz, and Solodic Planosols (Milbrandt & Overend, 2009).

Due to lack of European datasets associated with salinity in the units described in the literature, in the **MAIL** project it was not used as an indicator per se.

6.4.3.2 Soil acidity (potential of hydrogen/pH)

Acidification is a major soil degradation factor and affects soil fertility. Acidity is measured by determining the pH of a soil (Mueller et al., 2007), and is a major marginality indicator. In Kang et al. (2013), marginal areas are characterized by pH levels in soil > 9 or < 5.5 . Elbersen et al. used values of pH in topsoil water < 5 , for mapping MLs suitable for growing industrial crops. Similarly, Ivanina et al. (2016) used values of pH in topsoil water ≤ 5.5 for mapping agricultural ML suitable for bioenergy crops. Marginal lands for biomass production had pH values 5.5 - 4.5 or < 4.5 in (Milbrandt & Overend, 2009), and finally, Ciria et al. (2019) used pH values < 6 and > 8 to map arable agriculture MLs.

In the **MAIL** project, soft pH thresholds were chosen, in order to result in larger areas as potential MLs, i.e. $[pH > 8, pH < 6]$, $[pH > 8.5, pH < 5.25]$ and $[pH > 9, pH < 4.5]$.

6.4.3.3 Soil erosion

Being a major hazard, soil erosion is inseparably connected to slope and soil texture, but is also affected by climate, vegetation cover and agricultural practices (Mueller et al., 2007). Li et al. (2017) used erosion as a marginality indicator by calculating the Universal Soil Loss Erosion and the revised version (RUSLE), classified in 11 classes. Marginally suitable agricultural lands had moderate erosion risk (aprox. 200-1500 ton/yr/ha), and marginally unsuitable agricultural lands had higher erosion risk (aprox. 1500-2500 ton/yr/ha). Zolekar & Bhagat, (2015) used elevation, classified in 3 classes, as a proxy to depict soil erosion risk, while Roehrig & Menz (2005), in this same context, used slope. Both soil erosion tolerance rate of 2 ton/ha, and erosion K factor $K=8$ were used to assess marginal lands in (Kang et al., 2013). Lastly, Elbersen et al. used the classes 'High' and 'Very High' of the Index of Land Susceptibility to Wind Erosion (ILSWE), as well as the WaTEM/SEDEM spatial database for sensitivity for erosion by water (> 100 ton/ha/yr) as thresholds for MLs suitable for growing industrial crops. The detailed approach for the definition of the threshold value for soil erosion is presented in chapter 13.

6.4.3.4 Flooding (ponding)

Flooding is considered as a marginality index by some authors, since flooding implies a risk to most plant species. Elbersen et al. used a threshold of > 1 -2 m flood in a 2-year return time during the growing season, for mapping MLs suitable for growing industrial crops. Flood



duration and frequent ponding are also used by Kang et al. (2013) for assessing MLs. Gopalakrishnan et al., (2011) classified lands as marginal for biomass production when chance of flooding in a year exceeded 50%.

Based on the last study, in the **MAIL** project, the threshold for describing a land as marginal, decided to be a recurrence of water >50% from JRC Global Surface Water. Recurrence provides information concerning the inter-annual behavior of water surfaces and is defined as the frequency with which water returns from years to year, expressed as a percentage (JRC, 2018).

6.4.3.5 Sodicty

Soil sodicty is a characteristic of land for which the proportion of adsorbed sodium in the soil clay fraction is too high for plants to perform or survive. The effects of sodicty are often indirect as they affect vital soil properties rather than plant growth itself (Eliasson et al., 2010). Kang et al. (2013), classified lands as marginal when Na saturation exceeded 6–15%. Elbersen et al. and Eliasson et al. (2010) used Saturation with Exchangeable Sodium (ESP) to assess sodicty (ESP >15% in more than 50% of the mapping unit area and ESP >6, respectively).

In **MAIL**, ESP >6 was chosen as a threshold for sodicty in MLs.

6.4.3.6 Toxicity from contamination

Soil toxicity by pollutants, other chemicals is a serious obstacle for conventional agriculture because of increasing risks of phytotoxicity and food contamination. Areas with high toxicity from contamination are usually referred to as marginal (Ivanina et al., 2016). One of the most widespread contaminants is Nitrogen, with Nitrogen $\geq 10\text{mg L}^{-1}$ in groundwater being the lower limit of classifying a land as marginal in some studies (Gopalakrishnan et al., 2011; Ivanina et al., 2016). This threshold is also used in **MAIL**, by slightly changing the units to match the corresponding dataset (1cg/kg).

6.4.3.7 Dryness (Aridity Index)

Dryness is defined as an overall low soil water content, resulting from a natural imbalance in the water availability (through low annual precipitation and high annual evaporative demand) (Ivanina et al., 2016). Elbersen et al. and Ivanina et al. (2016) used dryness as a marginality indicator, by computing the Aridity Index (AI). The AI is the ratio of the total annual precipitation (P) to the total annual potential evapotranspiration (PET), and it expresses the relationship between severity of dry conditions and biomass production. Severe conditions correspond to AI values ≤ 0.5 . In (Bai et al., 2008), for calculating the AI, PET is defined as $PET = P / (0.9 +$



$(P/L)^2$), and $L = 300 + 25T + 0.05T^3$, where T is mean annual temperature and P is annual precipitation in mm. AI is also used by Cai, Zhang, & Wang (2011), but with a different formula.

In **MAIL**, the AI was chosen for assessing the dryness indicator with MLs having $AI \leq 0.5$.

6.4.3.8 Natural toxicity

Natural toxicity is an adverse chemical condition for soils, affecting soil fertility, caused by chemical elements that naturally exist such as Aluminum, Sulfur or Gypsum. Eliasson et al. (2010) marks lands as unfavorable when Gypsum presence $>15\%$. In **MAIL**, gypsum $> 15\%$ (for WISE GYPS > 150 g/Kg) was chosen as a threshold for natural toxicity in MLs.

Table 16: Sustainability indicators and thresholds, sorted by importance (times used in the literature review).

Indicator	Data source	MAIL threshold	MAIL threshold based on	Literature Thresholds
acidity (pH)	ESDAC 2009	[pH >8 , pH <6], [pH >8.5 , pH <5.25] [pH >9 , pH <4.5].	(Ciria et al., 2019)	pH >9 or <5.5 , pH <5 , 5.5 - 4.5 or <4.5 , pH <6 and pH >8 , \leq 5.5
erosion	RUSLE 2015 100m, ESDAC 2015	RUSLE > 200 ton/ha/yr	(Li et al., 2017)	By elevation, by slope, Soil erosion tolerance rate 2 ton/ha, Erosion by water > 100 ton/ha/yr
flood	JRC Global Surface Water- "recurrence" 1984-2018 30m	$> 50\%$	(Gopalakrishnan et al., 2011)	$>1\text{-}2$ m flood in 2yr, $>50\%$ chance of flooding per year
sodicity	WISE 2015 10km	$> 6\%$	(Eliasson et al., 2010)	$>6\text{--}15\%$, $>$ 15% ESP in more than 50% of the mapping unit area
contamination	WISE 2015 10km	Nitrogen $> 1\text{cg/kg}$	(Gopalakrishnan et al., 2011; Ivanina et al., 2016)	nitrate ≥ 10 mg L $^{-1}$
dryness	TerraClimate "Precipitation accumulation" and "evapotranspiration" 2018 4.5km	$P/PET \leq 0.5$	(Elbersen et al.; Ivanina et al., 2016)	$P/PET \leq 0.5$
natural toxicity	WISE 2015 10km	gypsum $> 15\%$ (for WISE GYPS > 150 g/Kg)	(Eliasson et al., 2010)	gypsum $> 15\%$



6.4.4 Productivity indicators

6.4.4.1 Soil Organic Matter (organic carbon)

According to the United States Department of Agriculture (USDA), “*Soil organic matter (SOM) is the organic component of soil, consisting of three primary parts including small (fresh) plant residues and small living soil organisms, decomposing (active) organic matter, and stable organic matter (humus).*” SOM is a reservoir of nutrients for plants, inextricably related to soil fertility, and is thus used in many studies as a marginality indicator. Although some studies use only low SOM values to depict MLs (<1% SOM in the study of Ciria et al., (2019) and <0.8% SOM in Li et al., (2017)), others also used the high values of SOM to take histic soils into consideration. Elbersen et al. used a threshold of SOM <0.75% (at depth range 0-30 cm) for low fertility MLs and SOM ≥20% to express Histosols (Peat soils). Eliasson et al. (2010) used high values (SOM >30%) to assess histic soil areas as less favorable in the EU.

In the **MAIL** project, soft SOM thresholds were chosen, in order to result in larger areas as potential MLs, i.e. SOM <1% - SOM ≥ 20% and SOM <0.75% - SOM ≥ 30%.

6.4.4.2 Cation Exchange Capacity (CEC)

According to Li et al. (2017) “*Cation exchange capacity (CEC) is the negatively charged content (usually clay and organic matter particles) in soils that can hold positively charged ions such as calcium, magnesium and potassium through electrostatic forces. It is used as a measure of fertility and nutrient retention capacity*”. In the same study, it was found that MLs had low CEC, namely CEC <8 cmol kg⁻¹. In (Kang et al., 2013) though, a lower threshold was used of CEC <4 meq/100g at pH=7. The detailed approach for the definition of the threshold value for CEC is presented in chapter 13.

6.4.4.3 Productivity

Crop productivity, as an indicator for defining ML, is used mostly by creating a threshold for a main crop type cultivated in the study area. Ciria et al. (2019), found that annual productivity of ≤1.5 Mgha⁻¹ of wheat and barley was insufficient to cover the calculated cultivation costs, and thus land with productivity below this is considered as marginal. Bandaru et al. (2013), used the National Commodity Crop Productivity Index to classify lands as marginal for NCCPI 0.42–0.14. Lands with productivity for the main grain crop (no irrigated yields of corn) <9 tons/ha (4 tons/acre), are defined as marginal in Gopalakrishnan et al. (2011). On the other hand, in (Cai et al., 2011), MLs are classified according to *land* productivity, and productivity is used as a synthetic indicator according to factors such as slope, soil type, soil temperature regime etc., based on land cover types.



It is evident that, crop productivity is tightly related to crop type and on national data, and thus cannot be used in the **MAIL** project due to unavailable data. Nevertheless, based on the ESDAC dataset “Soil Biomass Productivity maps of grasslands and pasture, of croplands and of forest areas in the European Union”, productivity as an indicator will be used for grasslands, pastures and forest areas, taking as a threshold the lowest value classes.

6.4.4.4 Net primary productivity (NPP)

Net primary production, or Net primary productivity (NPP), is the difference between total photosynthesis and total plant respiration in an ecosystem (Clark et al., 2001). Peter, Messina, & Snapp, (2018) used interannual productivity (from MODIS NPP data), for classifying agricultural areas as marginal or not. Roehrig & Menz (2005) used NPP in a fuzzy logic marginality index. Furthermore, NPP is used for measuring land degradation, in relation to other factors in the study by Bai et al. (2008).

Due to no clear threshold used in the examined literature, it was decided not to use NPP in the **MAIL** project as a marginality indicator.

6.4.4.5 Soil fertility

Soil fertility, being a complex variable, dependent on many other factors and other marginality indicators, is also used as an independent indicator in some studies. Soil fertility was assessed using the Soil Quality Rating (SQR) System in (Mueller et al., 2007) and (Ivanina et al., 2016), by taking field soil samples. More specifically, in the last, regions with $SQR \leq 25$ are considered as marginal. Fertility factor S_f was chosen from the Leemans and van den Born database of soil properties in (Roehrig & Menz, 2005). Finally, soil fertility is used as a combination of soil pH and SOC content for defining MLs in Elbersen et al.

In three out of four studies reported above, there is no clear threshold, or local datasets are used. In the study by Elbersen et al., (2018) the variables used (pH, SOC) will be used in **MAIL** as independent indicators. Therefore, it was decided to omit the soil fertility indicator from the analysis.

Table 17: Productivity indicators and thresholds, sorted by importance (times used in literature review).

Indicator	Data source	MAIL threshold	MAIL threshold based on	Literature Thresholds
soil organic matter	LUCAS “Topsoil Soil Organic	OM <1%, OM ≥ 20% OM<0.75% – OM ≥ 30%	(Ciria et al., 2019;	OM <1%, OM < 0.75%, OM ≥ 20%, OM ≥ 30%, Histosols



	Carbon ² 2015 1km		Elbersen et al.)	
cation exchange capacity	WISE 2015 10km	CEC < 8 cmol kg ⁻¹	(Li et al., 2017)	CEC<8 cmol kg ⁻¹ , CEC<4 meq/100g at pH=7
productivity	“Soil biomass productivity of grasslands and pastures” “Soil biomass productivity of forest areas” 2016 1km	grasslands < 6, forests < 3	-	No clear threshold.

6.4.5 Climate indicators

6.4.5.1 Precipitation

Mostly used for calculating other indicators such as dryness and aridity, precipitation is used in many studies as a marginality indicator per se. Rain fed arable land with annual precipitation less than 400 mm is marginal according to Ciria et al. (2019). Peter et al. (2018) used accumulated growing season precipitation <750 mm and >1217 mm to define areas as marginal for maize growth. In (Li et al., 2017), annual precipitation data for Malawi were obtained by accumulating mean monthly rainfall data. Average annual precipitation is used, not as a marginality indicator, but as a suitability indicator for bioenergy crops (300-1000 mm) in the study of Ivanina et al. (2016). In another approach, internal variability of seasonal precipitation, based on negative anomalies of monthly growing season precipitation was used, along with a fuzzy logic marginality indicating system, in (Roehrig & Menz, 2005). Finally, similar to the method followed in [MAIL](#), precipitation is used for calculating the “dryness” marginality indicator Elbersen et al., (2018) and to calculate the Global Aridity Index (Bai et al., 2008; Cai et al., 2011), as already described in 6.4.3.

In [MAIL](#), precipitation is used only as a variable for calculating the synthetic dryness/aridity indicator.

6.4.5.2 Evapotranspiration

The United States Geological Survey (USGS) defines evapotranspiration as “*the sum of evaporation from the land surface plus transpiration from plants*”. In all the reviewed studies of this document, evapotranspiration is used for calculating other marginality indicators such as dryness/aridity (Bai et al., 2008; Cai et al., 2011; Elbersen et al.; Ivanina et al., 2016). Based on these, the same approach is used in in the [MAIL](#) project, thus evapotranspiration was used for calculating the dryness/aridity indicator and applying the threshold accordingly.

6.4.5.3 Air temperature

Like precipitation, air temperature is also used for calculating other indicators such as length of growing period, aridity index, energy-use efficiency etc. In Elbersen et al., (2018) temperature is used to calculate the Length of Growing Period (LGP), as the days with average temperature >5 . If $LGP \leq 180$ days, then the area is considered as ML. In (Cai et al., 2011), as mentioned above, mean monthly temperature is used to calculate the global aridity index. Moreover, it was found that air temperature correlates highly with land surface temperature (LST) so in this study, air temperature was used instead of LST for the fuzzy logic modeling for ML detection. In (Bai et al., 2008), annual accumulated temperature is used to calculate energy-use efficiency (EUE) (ratio of annual sum NDVI to annual accumulated temperature). Areas with a negative EUE indicator are marked as degraded.

Due to no clear threshold used in the examined literature, it was decided not to use air temperature in the **MAIL** project as a marginality indicator.

Table 18: Climate indicators and thresholds, sorted by importance (times used in the literature review).

Indicator	Data source	MAIL threshold	MAIL threshold based on	Literature Thresholds
precipitation	TerraClimate "precipitation accumulation" 2018 4.5km	Calculation of other indicators	-	$P_{\text{annual, sum}} < 400 \text{ mm}$, $P_{\text{GS, sum}} < 750 \text{ mm}$ and $P_{\text{GS, sum}} > 1217 \text{ mm}$
evapotranspiration	TerraClimate "evapotranspiration" 2018 4.5km	Calculation of other indicators	-	Calculation of other indicators.

6.5 ML classification schemes

The term marginal land does not have a unique definition. Categorization of MLs in literature is based on the definition of MLs used by each approach. Therefore, there are different classifications depending on each study's goals. The **MAIL** project aims to map MLs which can potentially become carbon sinks, thus MLs must be classified in Carbon sequestration capacity groups based on the indicators. Therefore, MLs are reclassified into 3 classes, depicting marginality: 1) *Marginal lands with high plantation suitability*, 2) *Marginal lands with low plantation suitability* and 3) *Potentially unsuitable lands*.

The indicators that used to derive the final MLs classes are the non-thematic data ones (numerical data indicators). A rank is given to each indicator (Table 19), according to its importance (times found in literature). Alternatively, the rank given to each indicator can be based on expert opinions in Carbon sequestration. Subsequently, the Pairwise Comparison Matrix (PCM) of ranks (Zolekar & Bhagat, 2015) is calculated (Table 20), along with the normalized PCM, and the weight for each indicator. The cell values of PCM were divided by sum of the column to obtain the cell values in the normalized PCM and averaged in row to calculate the weights of each indicator (Table 21). The calculated weights are scaled from 0 to 1 in ascending order, and have a sum equal to 1, to maintain hierarchy according to their importance in marginality. The remaining values of each indicator are grouped into 3 classes and are given scores, based on their contribution to marginality (Table 22, Table 23, Table 24). This is followed by a weighted overlay in a GIS, based on the scores and weights of each indicator. The final step in the mapping of marginal lands, is the reclassification of the resulting product of the weighted overlay into the 3 aforementioned ML classes.

Table 19: The ranks given to each indicator, in order to calculate the weights.

Indicator	Times Found in Literature	Rank
slope	18	1
depth available to roots	18	1
acidity (pH)	9	2
texture	9	2
erosion	8	3
stoniness	8	3
soil organic matter	8	3
water capacity	6	4
flood	6	4
sodicity	5	5
clay	4	6

Indicator	Times Found in Literature	Rank
sand	4	6
contamination	4	6
cation exchange capacity	4	6
productivity	3	7
dryness	2	8
natural toxicity	1	9

**Table 20: The pairwise comparison matrix of the marginality indicators, based on the ranks.**

	slope	depth available to roots	acidity (pH)	texture	erosion	stoniness	soil organic matter	water capacity	flood	sodicity	clay	sand	contamination	cation exchange capacity	productivity	dryness	natural toxicity
slope	1.00	1.00	2.00	2.00	3.00	3.00	3.00	4.00	4.00	5.00	6.00	6.00	6.00	6.00	7.00	8.00	9.00
depth available to roots	1.00	1.00	2.00	2.00	3.00	3.00	3.00	4.00	4.00	5.00	6.00	6.00	6.00	6.00	7.00	8.00	9.00
acidity (pH)	0.50	0.50	1.00	1.00	1.50	1.50	1.50	2.00	2.00	2.50	3.00	3.00	3.00	3.00	3.50	4.00	4.50
texture	0.50	0.50	1.00	1.00	1.50	1.50	1.50	2.00	2.00	2.50	3.00	3.00	3.00	3.00	3.50	4.00	4.50
erosion	0.33	0.33	0.67	0.67	1.00	1.00	1.00	1.33	1.33	1.67	2.00	2.00	2.00	2.00	2.33	2.67	3.00
stoniness	0.33	0.33	0.67	0.67	1.00	1.00	1.00	1.33	1.33	1.67	2.00	2.00	2.00	2.00	2.33	2.67	3.00
soil organic matter	0.33	0.33	0.67	0.67	1.00	1.00	1.00	1.33	1.33	1.67	2.00	2.00	2.00	2.00	2.33	2.67	3.00
water capacity	0.25	0.25	0.50	0.50	0.75	0.75	0.75	1.00	1.00	1.25	1.50	1.50	1.50	1.50	1.75	2.00	2.25
flood	0.25	0.25	0.50	0.50	0.75	0.75	0.75	1.00	1.00	1.25	1.50	1.50	1.50	1.50	1.75	2.00	2.25
sodicity	0.20	0.20	0.40	0.40	0.60	0.60	0.60	0.80	0.80	1.00	1.20	1.20	1.20	1.20	1.40	1.60	1.80
clay	0.17	0.17	0.33	0.33	0.50	0.50	0.50	0.67	0.67	0.83	1.00	1.00	1.00	1.00	1.17	1.33	1.50
sand	0.17	0.17	0.33	0.33	0.50	0.50	0.50	0.67	0.67	0.83	1.00	1.00	1.00	1.00	1.17	1.33	1.50
contamination	0.17	0.17	0.33	0.33	0.50	0.50	0.50	0.67	0.67	0.83	1.00	1.00	1.00	1.00	1.17	1.33	1.50
cation exchange capacity	0.17	0.17	0.33	0.33	0.50	0.50	0.50	0.67	0.67	0.83	1.00	1.00	1.00	1.00	1.17	1.33	1.50
productivity	0.14	0.14	0.29	0.29	0.43	0.43	0.43	0.57	0.57	0.71	0.86	0.86	0.86	0.86	1.00	1.14	1.29
dryness	0.13	0.13	0.25	0.25	0.38	0.38	0.38	0.50	0.50	0.63	0.75	0.75	0.75	0.75	0.88	1.00	1.13
natural toxicity	0.11	0.11	0.22	0.22	0.33	0.33	0.33	0.44	0.44	0.56	0.67	0.67	0.67	0.67	0.78	0.89	0.10



	slope	depth available to roots	acidity (pH)	texture	erosion	stoniness	soil organic matter	water capacity	flood	sodicity	clay	sand	contamination	cation exchange capacity	productivity	dryness	natural toxicity	Weights
slope	0.17	0.17	0.17	0.17	0.17	0.17	0.17	0.17	0.17	0.17	0.17	0.17	0.17	0.17	0.17	0.17	0.17	0.17
depth available to roots	0.17	0.17	0.17	0.17	0.17	0.17	0.17	0.17	0.17	0.17	0.17	0.17	0.17	0.17	0.17	0.17	0.17	0.17
acidity (pH)	0.09	0.09	0.09	0.09	0.09	0.09	0.09	0.09	0.09	0.09	0.09	0.09	0.09	0.09	0.09	0.09	0.09	0.09
texture	0.09	0.09	0.09	0.09	0.09	0.09	0.09	0.09	0.09	0.09	0.09	0.09	0.09	0.09	0.09	0.09	0.09	0.09
erosion	0.06	0.06	0.06	0.06	0.06	0.06	0.06	0.06	0.06	0.06	0.06	0.06	0.06	0.06	0.06	0.06	0.06	0.06
stoniness	0.06	0.06	0.06	0.06	0.06	0.06	0.06	0.06	0.06	0.06	0.06	0.06	0.06	0.06	0.06	0.06	0.06	0.06
soil organic matter	0.06	0.06	0.06	0.06	0.06	0.06	0.06	0.06	0.06	0.06	0.06	0.06	0.06	0.06	0.06	0.06	0.06	0.06
water capacity	0.04	0.04	0.04	0.04	0.04	0.04	0.04	0.04	0.04	0.04	0.04	0.04	0.04	0.04	0.04	0.04	0.04	0.04
flood	0.04	0.04	0.04	0.04	0.04	0.04	0.04	0.04	0.04	0.04	0.04	0.04	0.04	0.04	0.04	0.04	0.04	0.04
sodicity	0.03	0.03	0.03	0.03	0.03	0.03	0.03	0.03	0.03	0.03	0.03	0.03	0.03	0.03	0.03	0.03	0.03	0.03
clay	0.03	0.03	0.03	0.03	0.03	0.03	0.03	0.03	0.03	0.03	0.03	0.03	0.03	0.03	0.03	0.03	0.03	0.03
sand	0.03	0.03	0.03	0.03	0.03	0.03	0.03	0.03	0.03	0.03	0.03	0.03	0.03	0.03	0.03	0.03	0.03	0.03
contamination	0.03	0.03	0.03	0.03	0.03	0.03	0.03	0.03	0.03	0.03	0.03	0.03	0.03	0.03	0.03	0.03	0.03	0.03
cation exchange capacity	0.03	0.03	0.03	0.03	0.03	0.03	0.03	0.03	0.03	0.03	0.03	0.03	0.03	0.03	0.03	0.03	0.03	0.03
productivity	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.02
dryness	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.02
natural toxicity	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.02

Table 21: The normalized pairwise comparison matrix and the calculated weights.

Table 22: Terrain and Soil indicator classes, scores, and weights.

Indicator	Classes	Score	Weight
slope	[15% - 40%]	10	0.17
	[40% - 65%]	5	
	[65% - 90%]	1	
depth available to roots	[100cm – 66.7cm]	10	0.17
	[66.7cm – 33.3cm]	5	
	[33.3cm – 0cm]	1	
texture	[30% - 53.3%]	10	0.09
	[53.3% - 76.7%]	5	
	[76.7% - 100%]	1	
stoniness	[10% - 15%]	10	0.06
	[15% - 20%]	5	
water capacity	[100mm – 50mm]	10	0.04
	[50mm – 0mm]	5	
clay	[50% - 58.7%]	10	0.03
	[58.7% - 67.3%]	5	
	[67.3% - 76%]	1	
sand	[60% - 70%]	10	0.03
	[70% - 80%]	5	
	[80% - 90%]	1	

Table 23: Sustainability indicator classes, scores, and weights.

Indicator	Classes	Score	Weight
acidity (pH)	[pH>8, pH<6]	10	0.09
	[pH>8.5, pH<5.25]	5	
	[pH>9, pH<4.5]	1	
erosion	[200 – 241.7]	10	0.06
	[241.7 – 283.4]	5	
	[283.4- 325]	1	
flood	[50% - 66.7%]	10	0.04
	[66.7% - 83.3%]	5	
	[83.3% - 100%]	1	
sodicity	[6% - 36.7%]	10	0.03
	[36.7% - 67.4%]	5	
	[67.4% - 98%]	1	
contamination	[1cg/kg – 3cg/kg]	10	0.03
	[3cg/kg – 10cg/kg]	5	
	[10cg/kg – 23.5cg/kg]	1	
dryness	[0.5 – 0.34]	10	0.02
	[0.34 – 0.18]	5	
	[0.18 – 0]	1	
natural toxicity	[150g/Kg - 328g/Kg]	10	0.02
	[328g/Kg - 506g/Kg]	5	
	[506g/Kg – 684g/Kg]	1	

Table 24: Productivity indicator classes, scores, and weights.

Indicator	Classes	Score	Weight
soil organic matter	[OM < 1%, OM ≥ 20%]	10	0.06
	[OM < 0.75%, OM ≥ 30%]	5	
cation exchange capacity	[8 – 5.3]	10	0.03
	[5.3 – 2.7]	5	
	[2.7 – 0]	1	
productivity	Grasslands:		0.02
	[6-4]	10	
	[4-2]	5	
	[2-0]	1	
	Forests:		
	[3-2]	10	
	[2-1]	5	
	[1-0]	1	

6.6 Production of intermediate layers

Some of the indicators used in the analysis, namely *soil texture* and *dryness*, are calculated based on other variables. These indicators will be referred to as “synthetic” indicators and require the production of intermediate layers in the GIS modelling procedure.

More specifically, the *soil texture* layer consists of the *stoniness*, *clay* and *sand* variables (all with independent thresholds), along with an intermediate layer calculated as (Elbersen et al.) :

$$t = \text{silt} + (2 * \text{clay})$$

Where *silt* and *clay* are percentages found in soil. The threshold applied to this intermediate layer is ≥ 30%, for marginal lands.

Moreover, the *dryness* layer is created by computing the Aridity Index (AI), which consists of *precipitation* and *evapotranspiration* and is calculated as:

$$AI = \frac{P}{PET}$$

Where *P* is the total annual precipitation and *PET* is the total annual potential evapotranspiration.

Apart from the indicator intermediate layers, another intermediate layer that needs to be calculated is the *weighted overlay (WO)* layer that will be:

$$WO = S * W$$

Where *S* are the scores assigned for each indicator class and *W* are the weights assigned for each indicator (according to Table 22, Table 23 and Table 24).

7. DEVELOPMENT AND REALIZATION OF A GIS USING THE DATASETS COLLECTED IN T2.2

7.1 Datasets overview

The 201 datasets collected in T2.2, were filtered based on the indicators found in the literature review. Datasets that were closest to the indicators conceptually were chosen, resulting initially in a sub-collection of 70 datasets (Datasets v1). These datasets were then filtered again, and by concurrently examining the data and thresholds used in literature, a best match was found, resulting in the final datasets (Datasets v2) and threshold simultaneously. The Datasets v1 can be thus used as a reference for potential changes in the methodology and thresholds used.

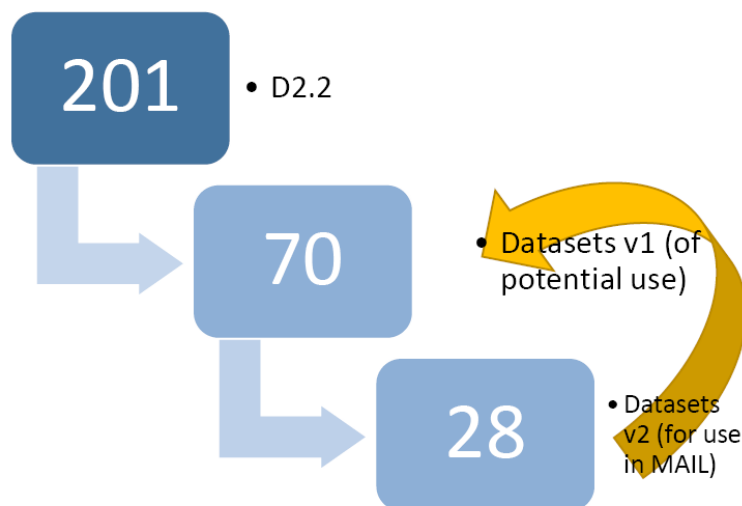


Figure 6: The elimination steps followed for selecting the final datasets used in *MAIL*.

In addition to datasets from T2.2, three new datasets were added to cover the needs for two indicators, namely the “Global Surface Water”¹³ dataset by the Joint Research Center, the “Global Forest Change”¹⁴ dataset by the University of Maryland and the “TerraClimate”¹⁵ dataset, by the University of Idaho (Table 25). The “Global Surface Water” dataset is used to

¹³ This data provides information on the location and temporal distribution of water areas on a global scale over the last 3.6 decades (1984-2019) and provides statistics on the extent and change of these water areas. The data set was produced from Landsat (TM, ETM+ and OLI). The available layers include: occurrence, occurrence change intensity, seasonality, recurrence, transitions and maximum water extent <https://global-surface-water.appspot.com/>

¹⁴ <https://earthenginepartners.appspot.com/science-2013-global-forest>

¹⁵ TerraClimate is a dataset of monthly climate and climatic water balance for global terrestrial surfaces. Some of the available layers are: Actual evapotranspiration, climate water deficit, reference evapotranspiration, precipitation accumulation, soil moisture, minimum temperature and maximum temperature <http://www.climatologylab.org/terraclimate.html>



calculate the flooding/ponding indicator and the “Global Forest Change” dataset is used as a land cover dataset, for forest areas detection and masking. Moreover, based on the datasets used in the studied literature, 7 more datasets are suggested to **MAIL**, for potential use in the future (Table 26) and are included in Datasets v1.

Table 25: Additional datasets for use in **MAIL.**

Dataset Name	Raster/ Vector	CRS	Spatial Resolution /MMU	Coverage	Product Date	Use
Global Surface Water	R	WGS84	30m	global	1984-2019	flooding indicator
Global Forest Change	R	WGS84	30m	global	2000-2018	forest mask
TerraClimate	R		2.5 arc minutes	global	1958-2015	aridity index (precipitation accumulation, evapotranspiration)

Table 26: Datasets for potential use in MAIL, as an addition to datasets from T2.2.

Dataset Name	Raster/ Vector	CRS	Spatial Resolution /MMU	Coverage	Product Date	Potential Use
SoilGrids	R	WGS84	250m	global	2016	acidity, cation exchange capacity, texture
MODIS LST	R	Sinusoidal	1km	global	2007-2019	Land surface temperature
MODIS NPP	R	Sinusoidal	1km	global	2000-2014	productivity
TRMM 3B43	R		0.25 arc degrees	global	1998-2019	monthly and annual precipitation
GHSL population	R	WGS84	250m	global	2015	population / socioeconomic indicators
K1 mountain	R	WGS84	1km	global	2000 (1996)	
K3 mountain	R	WGS84	250m	global	2017 (2010)	

Table 27 shows the final datasets used in **MAIL**.

Table 27: Final datasets used in *MAIL*.

Soil	Climate	Terrain	Sustainability	Productivity	LULC	Raster/Vector	CRS	Spatial Resolution / MMU	Coverage	Product Date	Values Range
Total available water content from PTF						R	ETRS89	1km	European	2013	
Clay content (topsoil & subsoil)						R	ETRS89	1km	European	2013	0-76
Depth available to roots						R	ETRS89	1km	European	2013	0-150
Sand content (topsoil & subsoil)						R	ETRS89	1km	European	2013	0-90
Silt content (topsoil & subsoil)						R	ETRS89	1km	European	2013	0-71
Soil pH in Europe						R	ETRS89	5km	EU25 (Romania & Bulgaria are not included,) +Norway, Switzerland, Croatia, Albania	2009	0.6-8.8
Volume of stones						R	WGS84		EU27	2008	0%-20%
		EU-DEM				R	ETRS89	25m	European	2011	
			Soil erosion by water (RUSLE2015)			R	ETRS89	100m	EU28	2015	0-325
	TerraClimate: Monthly Climate and Climatic Water Balance for Global Terrestrial Surfaces		TerraClimate: Monthly Climate and Climatic Water Balance for Global Terrestrial Surfaces			R	WGS84	~4.5km	global	1958-2018	PET: 80-239, PC: 0-1559
WISE derived soil property estimates			WISE derived soil property estimates			R	WGS84	~900m	global	2015	sodicity: ESP [-9-98] , nitrogen: TOTN [-9-23.48] , gypsum: GYPS [-9-684] , CEC: CECs [-9-128], soil types
				Organic carbon content (topsoil & subsoil)		R	ETRS89	1km	European	2013	0-39.5
				Soil biomass productivity of forest areas		R	ETRS89	1km	EU27	2016	0.17-10



Soil	Climate	Terrain	Sustainability	Productivity	LULC	Raster/Vector	CRS	Spatial Resolution / MMU	Coverage	Product Date	Values Range
				Soil biomass productivity of grasslands and pastures		R	ETRS89	1km	EU27	2016	0-10
					Coastline EU	V	ETRS89		European		thematic
					CORINE LC	V	ETRS89	25ha	EEA39	2018	thematic
					CORINE LC change	V	ETRS89	5ha	EEA39	2018	thematic
					Global Forest Change	R	WGS84	30m	global	2000-2018	thematic
					Imperviousness Classified Change (IMCC)	R	ETRS89	20m	European	2012-2015	thematic
					Imperviousness Density (IMD) HRL	R	ETRS89	20m	European	2015	thematic
			JRC Global Surface Water		JRC Global Surface Water	R	WGS84	30m	global	1984-2019	0-100%
					Nationally designated areas (CDDA)	V	WGS84		European	2019	thematic
					Natura2000	V	ETRS89		EEA33	2018	thematic
					S2GLC	R	ETRS89	10m	European	2020	thematic
					TanDEM-X Global Forest map	R	WGS84	50m	global	2011-2015	thematic
					Tree Cover Density (TCD)	R	ETRS89	20m	European	2015	thematic
					Tree Cover Density Change (TCDC)	R	ETRS89	20m	European	2012-2015	1-100%



7.2 Dataset transformation to common file formats, projection, etc.

Most of the final datasets in Datasets v2 (Table 27) were in the European Terrestrial Reference System 1989 (ETRS89), whereas others were in the World Geodetic System 1984 (WGS84). Since it was decided to use the ETRS89 as a coordinate system for all layers in **MAIL**, the datasets in WGS84 were transformed using the ArcGIS “Project Raster” tool, into ETRS89.

Resolution of all datasets varies from 10m (S2GLC) to 10km (WISE derived soil property estimates dataset).

To implement the first phase of the proposed methodology (i.e., the implementation of “hard” thresholds), all layers’ original resolution were resampled to the finest dataset resolution (i.e., 10m) of the S2GLC (Land Cover Map of Europe) which is used as Basemap. Moreover, to combine the datasets used in this first phase of “hard” thresholds, all the produced intermediate raster layers, are converted into Binary Raster (i.e., 0 = no data, 1 = data that need to be excluded and they don’t represent “marginal lands”). Furthermore, to avoid errors due to mismatches during the implementation of the proposed algorithms, the initial Basemap is set as “Snap Raster” Layer.

Finally, data will be resampled to the marginal land MMU (1ha) in the end of all processes.

7.3 GIS implementation

All datasets (vector and raster) that were collected based on the methodology development (Datasets v2), were imported into a file geodatabase (MAIL_v1.gdb) and then into a GIS project (Figure 7). There, the datasets were transformed to the ETRS89 coordinate system and the layers created, were grouped into categories (soil, climate, terrain, sustainability, productivity, LULC, socioeconomic) (Figure 8). This initial GIS project will work as a base for the workflows implemented in the next stage of the project.

The GIS software which was used in this phase and will be also used in the next steps, is the ArcGIS Pro¹⁶ software by Esri, with a license provided by IABG.

Due to the large scale of the study and to the large datasets, the tile structure is used for the raster data in the geodatabase. Raster datasets were imported in the form of Raster Catalogs¹⁷ and Raster Mosaics¹⁸. A raster catalog is a simple container for managing raster datasets,

¹⁶ <https://www.esri.com/en-us/arcgis/about-arcgis/overview>

¹⁷ <https://desktop.arcgis.com/en/arcmap/10.3/tools/data-management-toolbox/create-raster-catalog.htm>

¹⁸ <https://desktop.arcgis.com/en/arcmap/10.3/manage-data/raster-and-images/creating-a-mosaic-dataset.htm>



whereas a mosaic dataset is more advanced as overview images can be built and a processing for each raster dataset or on the entire mosaic dataset can be defined. Both are stored within a geodatabase.

Additionally, tiled processing components and parallel processing will be used during the workflow run. To improve the performance and scalability of feature overlay tools, the operational logic called *adaptive subdivision processing*¹⁹ will be used. This is a built-in feature of ArcGIS that is enabled when data cannot be processed within the available amount of physical memory, and thus, all data are subdivided in tiles, processed separately and then reconstructed again. In addition, the *Dice*²⁰ function can also be used when processing large vector layers (i.e., multipoint, lines, polygon feature layer or feature class, and more).

Moreover, it is important to stress out that, the workflows implementing the methodology developed, should have an independency in structure, so that potential changes in one step will not affect the next steps in processing. This is most important to be done in the second phase of the proposed methodology which is most prone to changing.

Finally, it should be noted, that the implementation of each developed workflow, pointed out the need for upgrading the original geodatabase type. Namely, File geodatabases can scale up to 1TB in size which can be raised to 256TB for processing extremely big data²¹. Nevertheless, in order to process big data such as large mosaic LULC datasets used in **MAIL** project, an upgrade of the existing geodatabase into an *Enterprise geodatabase*²² is recommended. More specifically, Enterprise Geodatabases can be unlimited in size and numbers of users. Moreover, this type of geodatabase can also provide the opportunity of developing advanced queries when stored in Microsoft SQL Server format.

¹⁹ <http://resources.arcgis.com/en/help/main/10.1/index.html#//01m10000000r000000>

²⁰ <https://desktop.arcgis.com/en/arcmap/10.3/tools/data-management-toolbox/dice.htm>

²¹ <https://desktop.arcgis.com/en/arcmap/latest/manage-data/geodatabases/types-of-geodatabases.htm>

²² <https://enterprise.arcgis.com/en/server/latest/manage-data/windows/enterprise-geodatabases-and-arcgis-enterprise.htm>

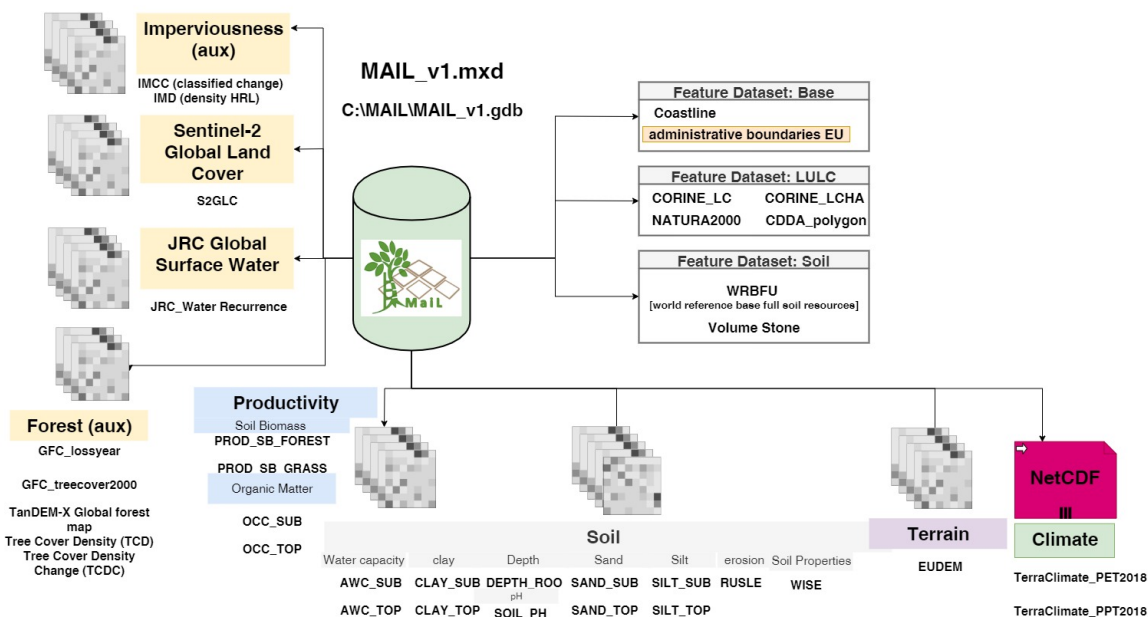


Figure 7: Overview of the GIS Geodatabase structure and contents. Source: personal compilation of Maria Tassopoulou.

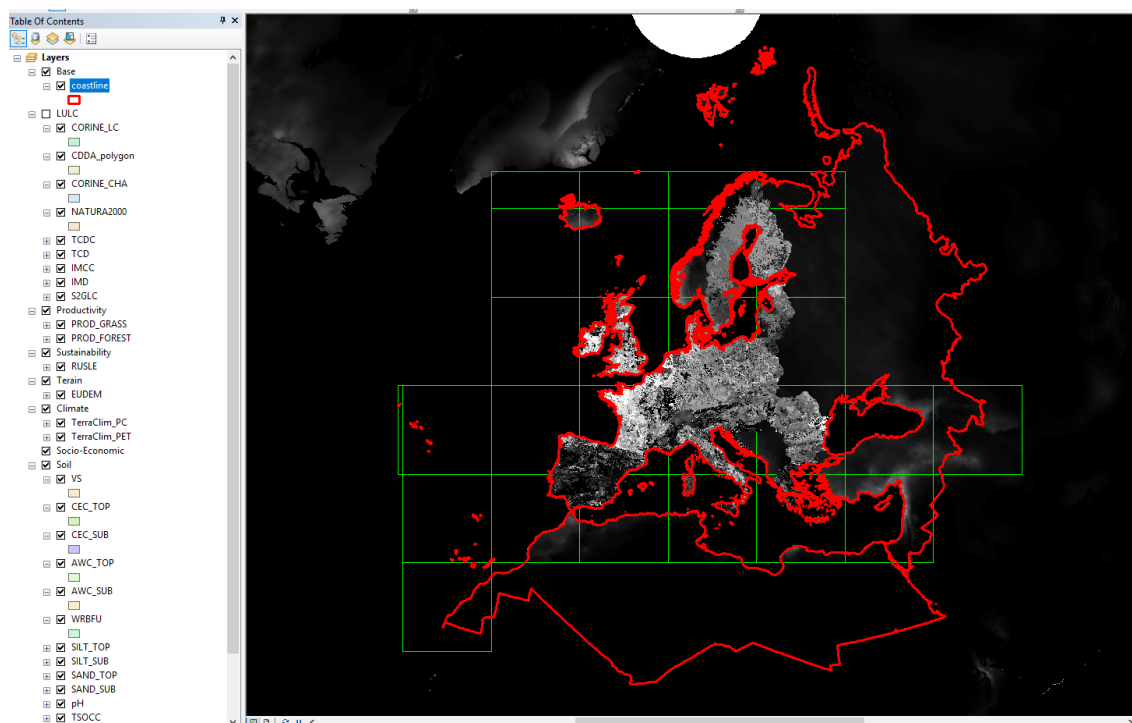


Figure 8: Datasets and layers as seen in the ArcGIS project.



7.4 Layers

	Base	Soil	Climate	Terrain	Sustainability	Productivity	LULC
Vector	Coastline	VS					CORINE_LC
		WRB					CORINE_LC_CHANGE
							CDDA_Polygon
							NATURA2000
Raster		AWC_TOP	TerraClim_PC	EU-DEM	RUSLE_EROSION	OCC_TOP	GFC_lossyear
		AWC_SUB	TerraClim_PET		JRC_WATER_RECURRENCE	OCC_SUB	GFC_treecover2000
		CLAY_SUB			WISE_Cation exchange capacity	PROD_GRASS	IMCC
		CLAY_TOP				PROD_FOREST	IMD
		DEPTH_ROO					S2GLC
		SAND_TOP					TANDEM-X
		SAND_SUB					TCDC
		SILT_TOP					TCD
		SILT_SUB					
		Soil_Ph					
		Stoniness					
		WISE_Gypsum					
		WISE_Nitrogen					
		WISE_Sodium					

Figure 9: The final names of the Datasets v2 GIS layers.

8. EVALUATION OF EXISTING AND DEVELOPMENT OF INDICATOR SETS

8.1 Selection of test sites

According to the definition of Marginal Lands, as defined for the **MAIL** project, Marginal Lands are considered lands with significant environmental or socioeconomic constraints and with potential to impact national accounting for C stock. Agricultural and impervious lands, forest areas or other valuable lands (e.g., protected areas or lands with important local uses), are excluded from this definition. Therefore, examples of potential Marginal Lands include mainly degraded and/or abandoned lands, areas with naturally low productivity due to biophysical constraints, and other lands that shall meet the requirements of additional indicators related to the terrain, soil, productivity, and other biophysical characteristics of each area.

Within the context of the **MAIL** project, several sites were tested across Europe. These areas include semi-mountainous areas, abandoned lands near urban environments, post-mining sites, post-industrial sites, and more.

The proposed methodology tested in large areas on a regional level in different European Countries including areas in Greece, Spain, Poland, and Germany. Examples of these areas are the Region of Central Macedonia, which is the largest and second-most populous region in Greece (Figure 10), the Region of Eastern Macedonia and Thrace, and more specifically, the agricultural areas of “Proskynites” and “Xylagani” adjacent to the city of Komotini, and the area of the Municipality of Nogueruelas (Teruel) in the Central Eastern part of the Iberian Peninsula, Spain (Figure 11). All sites consist of several potential sites that could be defined as Marginal Lands including semi-urban degraded lands and low productivity lands adjacent to natural parks and forest areas. It should be noted that the selected test sites also include the proposed pilot sites that will be used in the following Task 2.4 of the project for the assessment and the validation of the results of the proposed methodology. Namely, the suburban forest of “Kedrinis Lofos” located in the north and the northeast site of the city of Thessaloniki, Greece, is a selected pilot case of the **MAIL** project. Moreover, “Sierra de Espadán” in the eastern Spain province is also a site with potential Marginal Lands.

Furthermore, the federal state of Saxony was selected as test site in Germany. The aforementioned state includes lowland areas of low productivity while post-mining areas are also a distinctive land use in the region. More specifically, Nochten and Welzow sites located in the northern part of the federal state of Saxony, have been selected as representative pilot sites which include large post-mining areas that could be defined as Marginal Lands (Figure 12).

Finally, “Staszów” - the region within Staszów County (part of Świętokrzyskie Voivodeship) was the corresponding selected test site in Poland. The size and the whole region are divided into industrial (north) and rural (south) part. Local deposits of gypsum stone are one of the largest in Europe. 50% of national lime production and 25% of national cement production come from this Voivodeship. According to official data, the area of devastated land was increasing between 2004 and 2012. This region includes upland and lowland areas, strongly fragmented croplands, and low productivity lands that could be defined as potential Marginal Lands.

Concluding, the proposed methodology was implemented in 7 different test sites across Europe. All sites are areas with different geographical and physical characteristics, including different types of land uses – land cover, and variations on their soil and terrain characteristics.

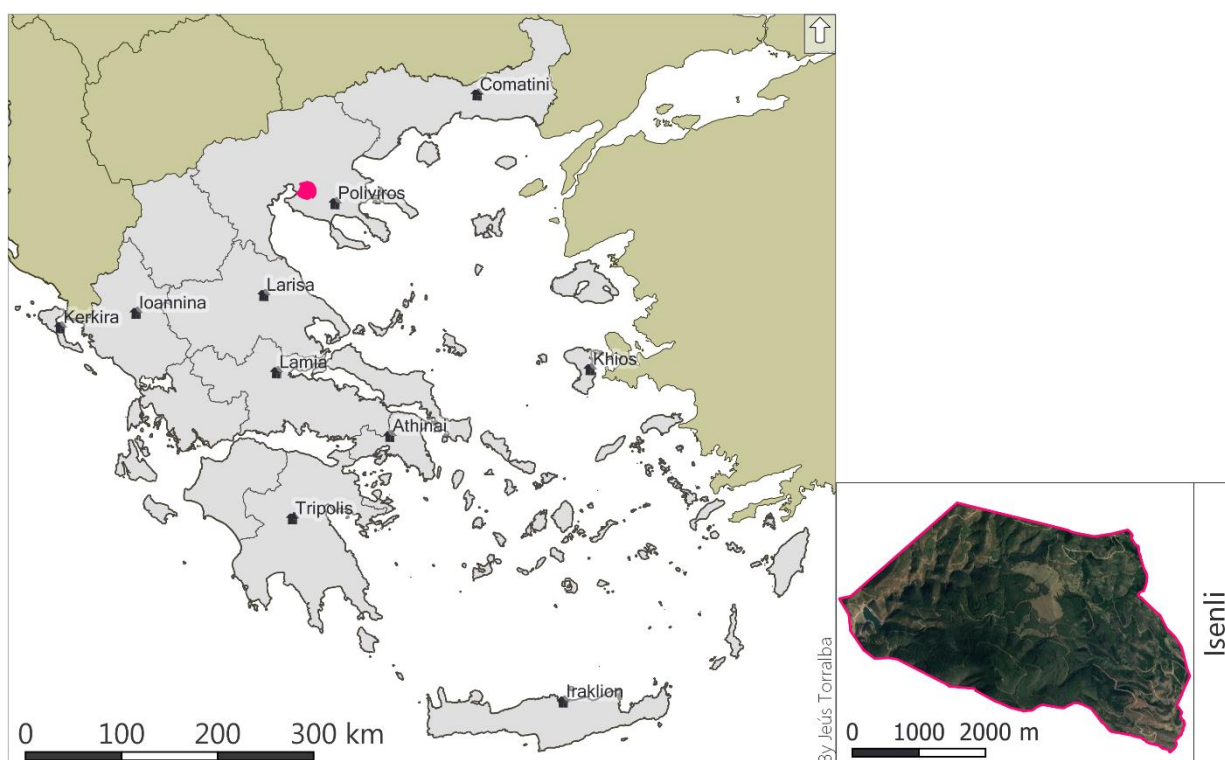


Figure 10: Greece (left) and the pilot site of the afforestation forest of “Isenli” (right image outlined with pink). Source: personal compilation of Jesús Torralba Pérez.

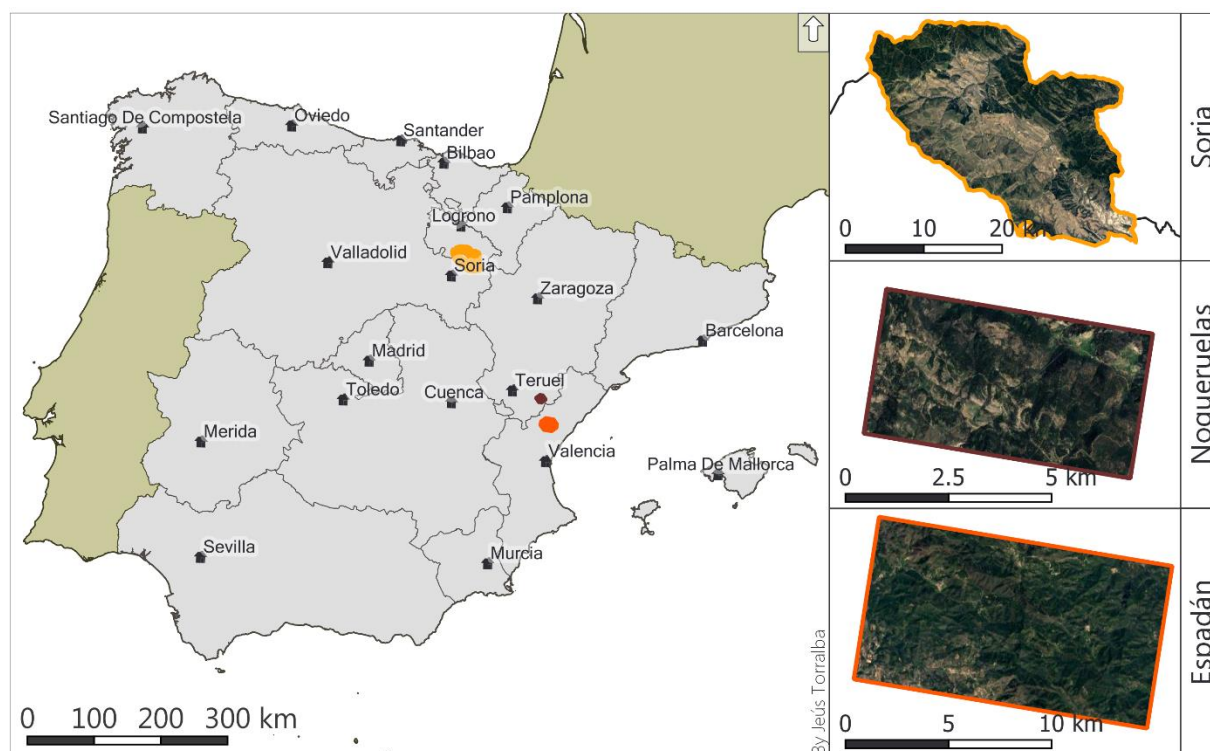


Figure 11: Spain (left) and the pilot site of “Soria” (outlined with light orange), “Nogueruelas” (outlined with dark red) and “Espadán” (right image outlined with dark orange). Source: personal compilation of Jesús Torralba Pérez.

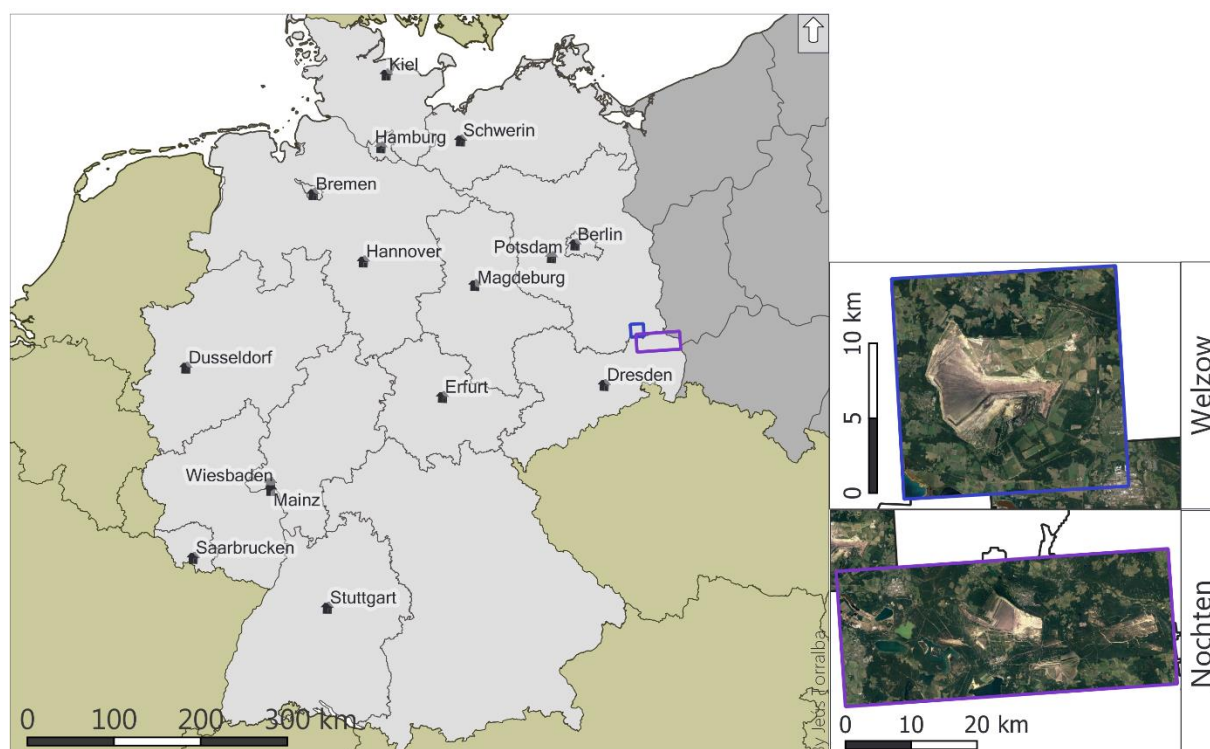


Figure 12: Germany (left) and the pilot sites of “Welzow” (outlined with blue) and “Welzow” (outlined with purple). Source: personal compilation of Jesús Torralba Pérez.

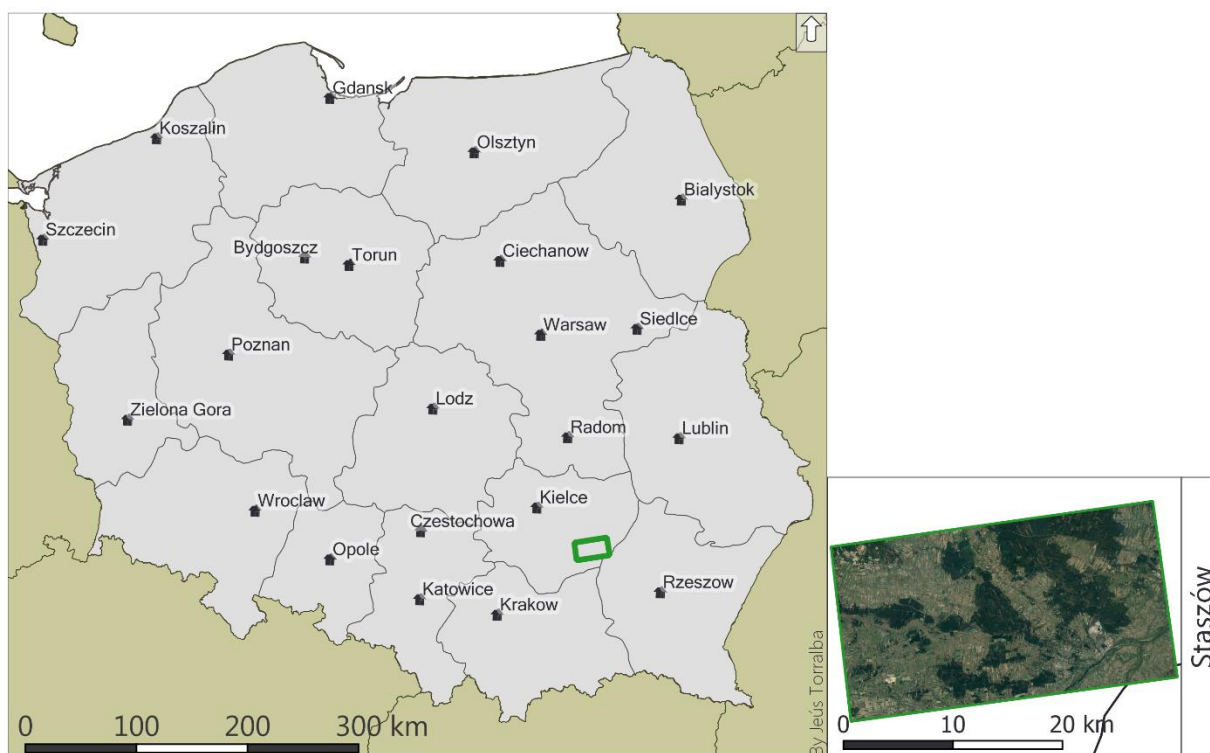


Figure 13: Poland (left) and the corresponding selected pilot site (outlined with green). Source: personal compilation of Jesús Torralba Pérez.

8.2 Testing different sets of existing indicators and thresholds

The indicators and thresholds for detecting MLs were selected taking into account the combination of the criteria used in selected ML studies from T2.1 “Literature review on Marginal Land definition” and the available datasets from T2.2 “Collection of appropriate existing European/Global datasets”.

In the first phase of the proposed methodology, different combinations of the available LULC datasets were tested to define the extent of the selected Land Uses that should be incrementally removed and that are not MLs.

In addition, the same approach of combining different indicators and thresholds of each category of “soft” constraints according to the literature review, applied during the second phase of the proposed methodology.

8.2.1. Selection of appropriate Land Use/Land Cover datasets

The first phase of the proposed methodology (i.e., the implementation of “hard” thresholds phase) suggests the delineation of specific Land Use/Land Cover (LULC) classes that should be excluded from further analysis and that are not MLs according to the literature review and the definition of Marginal Lands.

Namely, five different types of LULC should be identified and incrementally removed: Other Land Cover areas including Water bodies, Permanent Snow, Marshes and Peatbogs, Forest Areas, Croplands, Impervious Areas, Protected Areas, and Changed Areas. To identify these categories and to delineate those LULC, several combinations of the available datasets as already defined in Task 2.1 were tested.

Considering the spatial resolution, and the reference year of each available dataset, various analysis tools (e.g., Fuzzy Overlay, Weighted Overlay, etc.) were used to combine them. To perform those functions, an initial validation of each available dataset was also implemented by combining the existing datasets with historical high -resolution satellite imagery (e.g., google earth images). The final selection of each LULC dataset is described in the following pages.

8.2.1.1. Basemap

The implementation of the first phase of the proposed methodology requires the selection of a reliable Land Use – Land Cover Basemap. Taking into account the available LULC products as described in Deliverable 2.1 of the [MAIL](#) project, the Land Cover Map of Europe 2017 (S2GLC), which represents land cover classification of the European continent, has been chosen as Basemap.

The S2GLC product consists of 13 land cover classes (Table 28) and has been produced using classification algorithms for the analysis of over 15.000 Sentinel-2 images. The final product is delivered with 10 m spatial resolution and an overall accuracy of 86% (Figure 15). A quick visual assessment of the available dataset using Google earth images has revealed slight classification errors in certain areas across Europe (Figure 14). However, the overall accuracy of the product is satisfying for the purposes of the [MAIL](#) project, while the fine resolution of the final product is a key parameter for utilizing this specific dataset as Basemap.

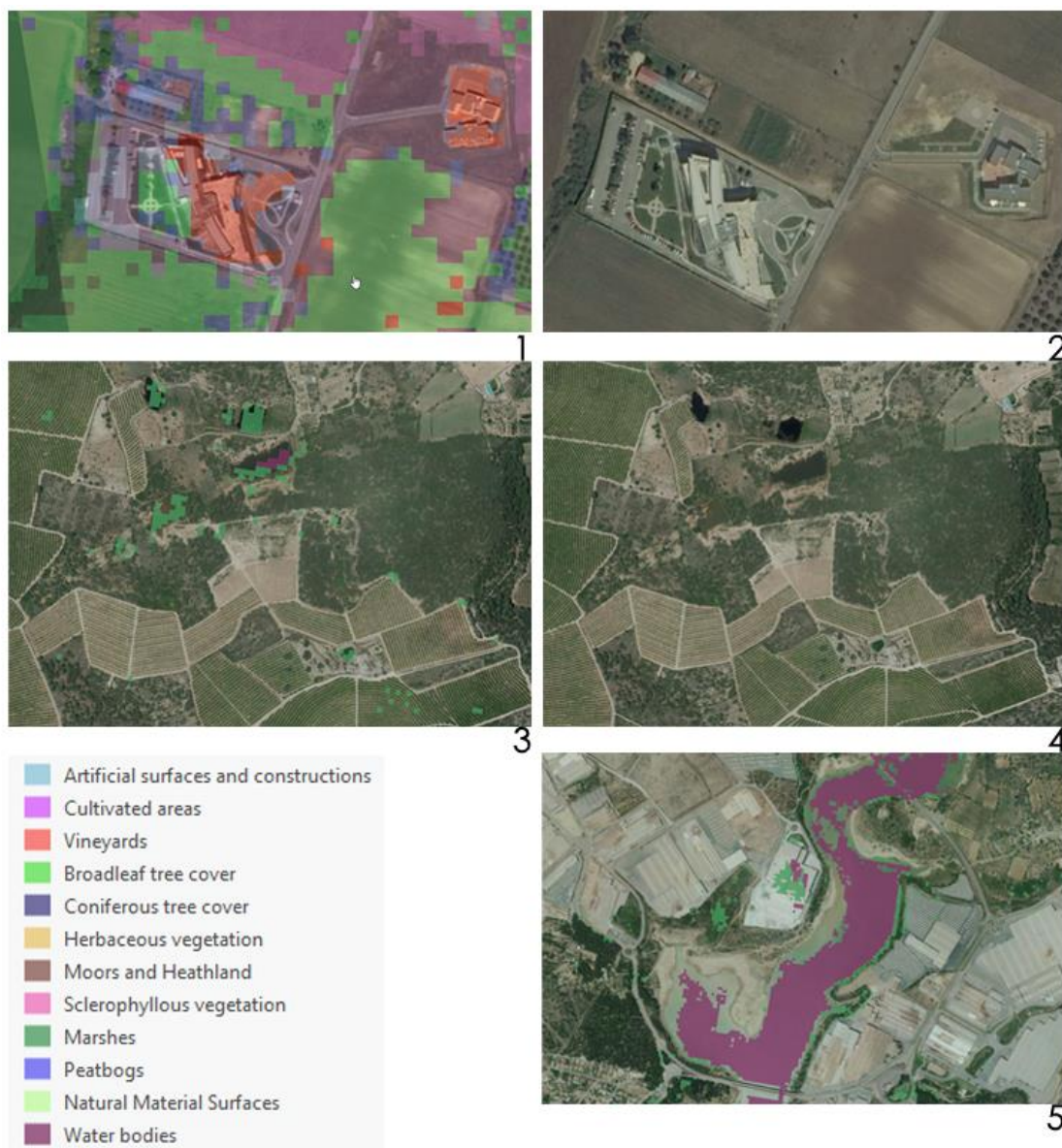


Figure 14: Overview of classification commission and omission errors of S2GLC raster layer. Artificial surfaces and constructions in images 1 and 2 are misclassified as “natural material surfaces” in areas within the boundaries of the selected test sites in Thessaloniki, Greece. In addition, croplands (3) and artificial surfaces and constructions (4) are classified as “marshes” and “water bodies” (5) in Espadán, Spain test site.

It should be noted that the Basemap of the project can be replaced with a more adequate LULC dataset in the future without affecting the implementation of the proposed methodology and the developed algorithms.

Finally, two supplementary vector files including the European Union’s boundaries and the European Union’s coastline are used for defining the clipping extent of all raster datasets. Those files are in the folder “BASE” of the MAIL geodatabase.



Table 28: Overview of the Land Cover Map of Europe 2017 technical specifications.

BASEMAP LAYER	Release Date	Spatial Resolution/MMU	Coverage	Classes	
				ID	Name
Land Cover Map of Europe 2017 (S2GLC)	2017	10m	European	0	Clouds
				62	Artificial surfaces and constructions
				73	Cultivated areas
				75	Vineyards
				82	Broadleaf tree cover
				83	Coniferous tree cover
				102	Herbaceous vegetation
				103	Moors and Heathland
				104	Sclerophyllous vegetation
				105	Marshes
				106	Peatbogs
				121	Natural material surfaces
				123	Permanent snow-covered surfaces
				162	Water bodies

Land Cover map of Europe after post-processing 13 LC classes	1.1.1. Artificial surfaces and constructions	2.1.1. Cultivated and managed areas	2.2.1. Vineyards	2.3.1. Herbaceous vegetation	3.1.1. Deciduous broadleaf tree cover	3.1.2. Evergreen coniferous tree cover	3.2.2. Moors and Heathland	3.2.3. Sclerophyllous vegetation	3.3. Natural material surfaces	3.3.5. Permanent snow and glaciers	4.1.1. Marshes	4.1.2. Peatbogs	5. 1. 1. Water bodies	Total	User's accuracy
Clouds, snow, ice, mask	11	1	0	0	0	0	0	0	28	56	0	0	1	97	
1.1.1. Artificial surfaces and constructions	1548	65	1	7	0	0	5	2	430	0	0	0	6	2064	75,00%
2.1.1. Cultivated and managed areas	37	11774	26	574	32	5	12	23	46	0	4	0	1	12534	93,94%
2.2.1. Vineyards	8	426	445	58	12	1	10	26	7	0	0	0	1	994	44,77%
2.3.1. Herbaceous vegetation	15	886	20	5592	127	6	689	289	72	7	88	13	0	7804	71,66%
3.1.1. Deciduous broadleaf tree cover	5	44	1	40	10436	202	165	113	4	0	12	5	1	11028	94,63%
3.1.2. Evergreen coniferous tree cover	8	16	2	12	171	8374	82	60	1	0	3	6	2	8737	95,85%
3.2.2. Moors and Heathland	10	65	1	249	53	24	760	50	49	0	5	58	2	1326	57,32%
3.2.3. Sclerophyllous vegetation	3	26	0	39	54	10	53	181	3	0	4	0	0	373	48,53%
3.3. Natural material surfaces	96	73	2	86	2	0	38	64	1319	7	1	2	7	1697	77,73%
3.3.5. Permanent snow and glaciers	4	0	0	0	0	0	0	0	73	9	0	0	0	86	10,47%
4.1.1. Marshes	37	76	2	55	40	3	58	7	18	0	140	83	79	598	23,41%
4.1.2. Peatbogs	5	17	0	62	16	4	197	0	5	0	61	578	55	1000	57,80%
5. 1. 1. Water bodies	42	1	0	4	2	1	2	0	63	6	6	0	3558	3685	96,55%
Total	1818	13469	500	6778	10945	8630	2071	815	2090	29	324	745	3712	51926	
Producer's accuracy	85,15%	87,42%	89,00%	82,50%	95,35%	97,03%	36,70%	22,21%	63,11%	31,03%	43,21%	77,58%	95,85%	OA	86,11%
														Kappa	0,83

Figure 15: Overview of the classification accuracy of S2GLC LULC product (Malinowski et al., 2019).

8.2.1.2. Delineation of Other Land Cover

The definition of “Other Land Cover” includes the delineation of four types of Land Cover:

1. Water bodies
2. Permanent Snow
3. Marshes
4. Peatbogs

The aforementioned types of coverage are included in the Land Cover Map of Europe 2017 (S2GLC) which we have already chosen to be used as Basemap. Within that context, the LULC classes that need to be excluded from the further analysis are represented in Table 29.

Table 29: Other Land Cover Types that are not Marginal Lands.

BASEMAP LAYER	Release Date	Spatial Resolution/MMU	Coverage	Classes	
				ID	Name
Land Cover Map of Europe 2017 (S2GLC)	2017	10m	European	105	Marshes
				106	Peatbogs
				123	Permanent snow-covered surfaces
				162	Water bodies

These LULC types are included both in the S2GLC product and the Pan-European Copernicus' High-Resolution Layers. Nevertheless, for the purposes of the **MAIL** project, we chose to delineate them using only the S2GLC product to maintain the data consistency of the process.

It should be noted, that Marshes and Peatbogs are two types of Land Use that are often hard to distinguish while it is difficult to define those classes especially when they coexist in the same regions (Government of Alberta, 2015), (Malinowski et al., 2019). Besides, the class "Inland Marshes" is not applicable for salt marshes and salt meadows under the tidal influence according to the definition of Copernicus' Corine LULC. Nevertheless, the corresponding "Marsh" class in S2GLC Layer, includes both "Inland Marshes" and "Salt Marshes" of Corine LC.

Moreover, three LULC datasets can provide information regarding the delineation of water bodies, including the S2GLC product, the Pan – European Copernicus' Corine LULC and the JRC Global Surface Water raster dataset and more specifically, the "maximum water extent" layer (Table 30). Figure 16 represents a visual interpretation of all three datasets which has revealed that S2GLC can be utilized to delineate adequately water bodies for the purposes of the **MAIL** project.

Table 30: Overview of the available LULC datasets for delineating Water bodies.

Water bodies LC DATASET	Acquisition Date	Spatial Resolution/MMU	Coverage
Land Cover Map of Europe 2017 (S2GLC)	2017	10m	European
CORINE Land Cover	2019	25 ha	European Economic Area Countries 39 (EEA 39)
Global Surface Water database provided by Joint Research Centre (JRC) – Maximum Water extent dataset	2019	30m	Global

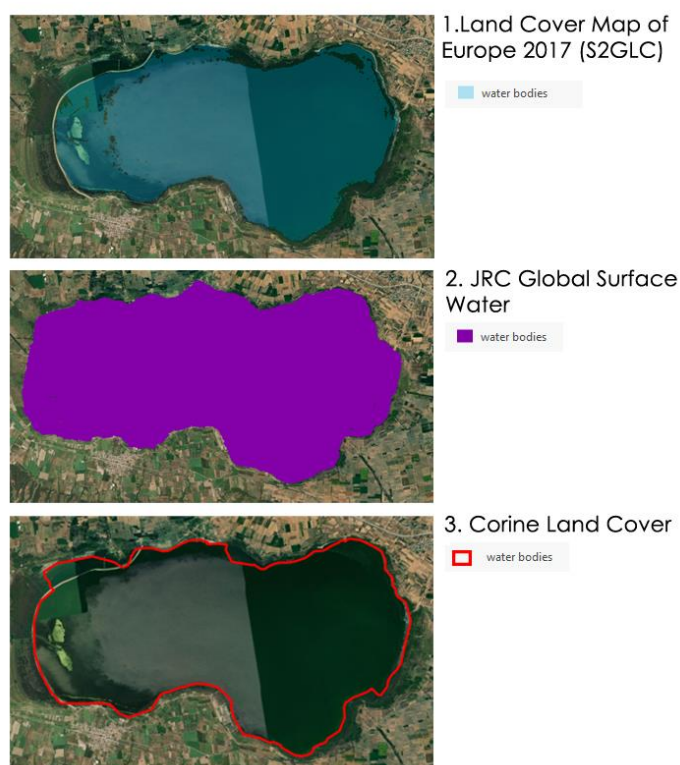


Figure 16: Visual assessment of the available datasets for the delineation of water bodies in the test site of Thessaloniki - Greece. The results of the analysis indicate that the S2GLC dataset can result to adequate delineation of water bodies.

8.2.1.3. Delineation of Forest

According to United Nations Food and Agriculture Organization (FAO, 2000) forest is characterized as the land with tree crown cover (or equivalent stocking level) of more than 10 % and area of more than 0.5 ha. The trees should be able to reach a minimum height of 5 m at maturity in situ (Chazdon et al., 2016).



The definition includes forest nurseries and seed orchards that constitute an integral part of the forest, forest roads, cleared tracts, firebreaks, and other small open areas, forest in national parks, nature reserves and other protected areas of specific scientific historical, cultural or spiritual interest, windbreaks and shelter belts of trees with an area of more than 0.5 ha and width of more than 20 m, and finally, plantations primarily used for forestry purposes (rubber wood plantations and cork oak stands).

On the contrary, the definition excludes land predominantly used for agricultural practices, tree plantations of fruit trees or olive oil trees, gardens, agroforestry areas (i.e., tree planted in agroforestry systems), and urban parks.

Following the forest definition, Table 31 represents the proposed thresholds that could be applied for the definition of forest areas.

Table 31: Overview of the proposed thresholds according to FAO forest definition.

Parameters	Threshold values
Minimum area (ha)	0.5 ha
Minimum tree height (m)	5m
Crown Cover (%)	10%
Temporary (years)	10 years appx.
Strip width (m)	20m

An important parameter that should be also considered is that the European Union (EU) has many different types of forests, and their distribution is mainly determined by climate, soil type, altitude, the topography of each area and other economic and social parameters. Thus, there is no one answer valid of “what is a forest” for all Member States (Union, 2020). Within that context, each EU country has adopted different forest definition parameters as described in ANNEX II of EU Regulation 2018/841 of the European Parliament (European Parliament, 2018), (Table 32).

Table 32: Forest definition parameters for EU countries (ANNEX II EU Regulation 2018/841).

Member State	Area (ha)	Tree crown cover (%)	Tree Height (m)
Belgium	0.50 ha	20%	5m
Bulgaria	0.10 ha	10%	5m
Czech Republic	0.05 ha	30%	2m
Denmark	0.50 ha	10%	5m
Germany	0.10 ha	10%	5m
Estonia	0.50 ha	30%	2m
Ireland	0.10 ha	20%	5m



Member State	Area (ha)	Tree crown cover (%)	Tree Height (m)
Greece	0.30 ha	25%	2m
Spain	1.00 ha	20%	3m
France	0.50 ha	10%	5m
Croatia	0.10 ha	10%	2m
Italy	0.50 ha	10%	5m
Cyprus	0.30 ha	10%	5m
Latvia	0.10 ha	20%	5m
Lithuania	0.10 ha	30%	5m
Luxembourg	0.50 ha	10%	5m
Hungary	0.50 ha	30%	5m
Malta	1.00 ha	30%	5m
Netherlands	0.50 ha	20%	5m
Austria	0.05 ha	30%	2m
Poland	0.10 ha	10%	2m
Portugal	1.00 ha	10%	5m
Romania	0.25 ha	10%	5m
Slovenia	0.25 ha	30%	2m
Slovakia	0.30 ha	20%	5m
Finland	0.50 ha	10%	5m
Sweden	0.50 ha	10%	5m
United Kingdom	0.10 ha	20%	2m

Following the 2018/841 directive and the FAO forest definition, five different datasets can provide information of the delineation of forest areas (Table 33) including the Global Forest Change Layers that represent changes in forest cover for specific periods.

Table 33: Overview of the available LULC datasets for delineating Forest Areas.

FOREST LC DATASET	Acquisition Date	Spatial Resolution/MMU	Coverage
HRL Tree Cover Density (TCD)	2012-2015	20m	Albania, Austria, Belgium, Bosnia and Herzegovina, Bulgaria, Croatia, Cyprus, Czechia, Denmark, Estonia, Finland, France, Germany, Greece, Hungary, Iceland, Ireland, Italy, Kosovo, Latvia, Liechtenstein, Lithuania, Luxembourg, Malta, Montenegro, Netherlands, North Macedonia, Norway, Poland, Portugal, Romania, Serbia, Slovakia, Slovenia, Spain, Sweden, Switzerland, Turkey, United Kingdom



FOREST LC DATASET	Acquisition Date	Spatial Resolution/MMU	Coverage
Global Forest Change – Tree Cover	2000-2018	30m	Global
Tree Cover Density Change (TCDC)	2012-2015	100m	Albania, Austria, Belgium, Bosnia and Herzegovina, Bulgaria, Croatia, Cyprus, Czechia, Denmark, Estonia, Finland, France, Germany, Greece, Hungary, Iceland, Ireland, Italy, Kosovo, Latvia, Liechtenstein, Lithuania, Luxembourg, Malta, Montenegro, Netherlands, North Macedonia, Norway, Poland, Portugal, Romania, Serbia, Slovakia, Slovenia, Spain, Sweden, Switzerland, Turkey, United Kingdom
TanDEM-X Global Forest map	2011-2015	50m	Europe
Global Forest Change Loss	2000-2018	30m	Global

These datasets were produced using different remote-sensing techniques and thus, different commission and omission errors occurred during their production. Moreover, the spatial resolution of each available product also varies. It should be noted that according to the technical specification of HRL TCD layer, the final derived forest type product allows to get as close as possible to the FAO forest definition combining both 1) a dominant leaf type product that has a MMU of 0.5 ha, as well as a 10% tree cover density threshold applied, and 2) a support layer that maps, based on the dominant leaf type product, trees under agricultural uses and urban context.

Namely, the Copernicus High-Resolution Layer (HRL) on Tree Cover Density (TCD) is produced from 20 m resolution satellite imagery through a combination of automatic processing and interactive rule-based classification. This dataset is assessed by visual interpretation on Very High-Resolution (VHR) satellite data and/or aerial ortho-imagery as reference data (Langanke, 2017).

In addition, the Global Forest Change (GFC) layer produced by the University of Maryland is derived from time-series analysis of Landsat images at a 30-meter spatial resolution. To improve the results of this LULC layer, the University of Maryland starting from 2019, provides a new updated 2.0 version of the data which is more consistent than the previously available data versions (Hansen et al., 2013).

Sannier, Pennec, and Dufourmont (2017) performed a comparative validation of the HRL-TCD Copernicus dataset and the Global Forest Change products provided by the University of Maryland. The results of the analysis show that the HRL-TCD dataset meets the thematic



classification accuracy requirements based on a 10-30% threshold of Tree Cover (TC) values, whilst the GFC layer meets only these requirements for the 30% threshold. In both datasets, results appear strongest for central Europe and somewhat weaker for western and especially southern Europe. Finally, TCD meets the minimum accuracy at pan-European level for omission errors and exceeds the target accuracy for commission errors. In addition, GFC layer meets the minimum accuracy requirement at pan-European level for omission and commission errors for the 30% thresholds.

To meet the requirements of forest definition in all EU countries, in the case of the **MAIL** project, we decided to implement a common 30% threshold of Tree Cover values to define Forest Areas. Thus, both datasets are eligible for delineating the forest extent. However, the implementation of a common threshold for all EU countries resulted in poor results in countries where forest areas are defined using lower threshold values of the area, the tree cover crown, and the tree height parameters. Nevertheless, to balance the inadequacies, different thresholds related to the terrain, the soil type, and the productivity of each area are applied during the second phase of the proposed methodology.

Besides, the German Aerospace Centre (DLR) has created the global TanDEM-X Forest/Non-Forest Map which is produced using interferometric data acquired by the German TanDEM-X radar satellite mission. This dataset shows the extent of forested areas at a resolution of 50 meters²³. (Martone et al., 2018) highlights the use of external reference data to produce TanDEM-X forest/Non forest map. More specifically, both the GFC layer and the Copernicus HRL TCD dataset were utilized as reference data during the production and the validation of the TanDEM-X product. It should be noted that resolutions and projections of those external reference data are different from each other. To tackle this issue, all the available reference data were re-projected and rescaled to match with the TanDEM-X data grid.

A quick visual assessment of the aforementioned datasets for the same reference year (2015) indicates that TanDEM-X Global Forest Map could be excluded from further analysis as it provides information that is already included in the other two datasets (Figure 17).

²³https://www.dlr.de/content/en/articles/news/2019/02/20190506_globale-tandem-x-waldkarte-verfuegbar.html

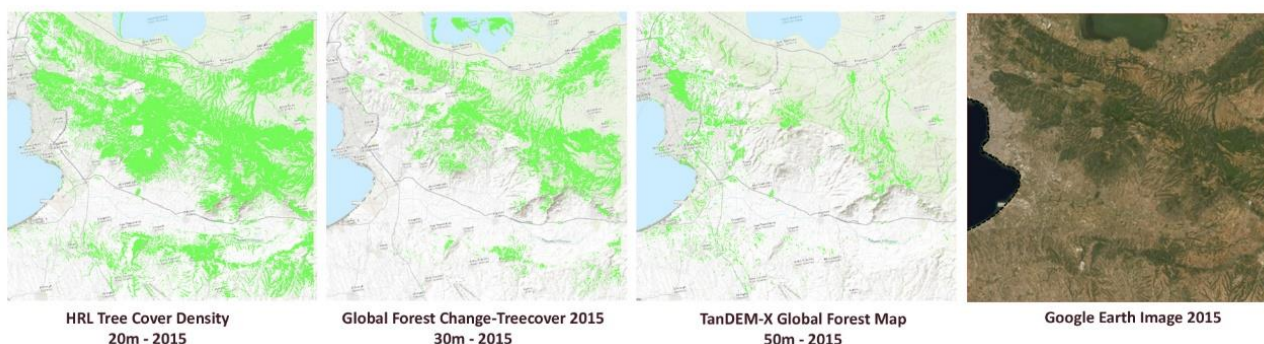


Figure 17: Visual assessment of the available forest datasets in the test site of Thessaloniki, Greece. The results of the analysis indicate that TanDEM-X dataset (3rd image left to right) could be excluded from further analysis.

Finally, the Land Cover Map of Europe 2017 (S2GLC) represents land cover classification of the European continent. The S2GLC product consists of 13 land cover classes, including two classes that represent forest areas: Broadleaf tree cover and Coniferous tree cover. This dataset has been produced using classification algorithms for the analysis of over 15.000 Sentinel-2 images. The final product is delivered with 10 m spatial resolution and an overall accuracy of 86%. More specifically, according to the error matrix for the S2GLC classification result for the whole Europe (Malinowski et al., 2019) the overall accuracy of the classification of the classes referring to forest areas are >95%. It should be noted that the Copernicus HRL TCD layer and the supplementary HRL Dominant Leaf Type were also used for the training of the classification algorithms. Thus, to delineate forest extent, the S2GLC layer is not included in the analysis.

8.2.1.4. Delineation of Croplands

“Croplands” is one land use class that should be also excluded from further analysis. To delineate this class, the S2GLC layer is combined with the CORINE Land Cover (LC) dataset. More specifically, class 73 of the S2GLC dataset represents cultivated areas while class 75 represents vineyards. These classes are corresponding to classes 2xx of Corine LC dataset of “agricultural areas” except class 231 which represent Pastures (i.e., lands that are used at least 5years permanently for fodder production including natural or sown herbaceous species, unimproved or lightly improved meadows, and grazed or mechanically harvested meadows) and are classified as “Herbaceous vegetation” in the S2GLC layer (Table 34).

Table 34: Overview of the available LULC datasets for delineating Croplands.

CROPLANDS LC DATASET	Geometry Type/format	Spatial Resolution/ MMU	CROPLANDS Class Name
Land Cover Map of Europe 2017 (S2GLC)	Raster Dataset	10m	Class 73 – Cultivated areas Class 75 - Vineyards
Corine Land Cover	Vector - Single polygon	25 Hectares MMU	Classes 2xx except for class 231

8.2.1.5. Delineation of Impervious areas

Impervious areas represent all sealed and constructed areas that are covered mainly by buildings both conventional and specific buildings, and artificial constructions (3D) or impervious surfaces (2D) including specific structures and facilities and open sealed surfaces. Therefore, according to the definition of Marginal Lands, those areas should be identified and excluded from further analysis as they do not represent Marginal Lands.

To delineate impervious areas, three different datasets, two raster layers, and one vector layer, are combined (Table 35). It should be noted that classes 131 and 132 of CORINE LC vector dataset represent Mineral extraction sites and Dumpsites accordingly, and thus are not included during the analysis.

Besides, regarding the Copernicus HRL Imperviousness product, this dataset captures the percentage of soil sealing. More specifically, the final product describes the spatial distribution of artificially sealed areas, including the level of sealing of the soil per area unit which is presented as imperviousness degree ranging from 1 to 100%.

Within the context of the **MAIL** project, to delineate impervious areas using the HRL IMD layer, a threshold of 30% is applied. Namely, to maintain the consistency between the available datasets, we utilize the definition of “continuous urban fabric” as specified in the guidelines of the CORINE LC product (Gyorgy Büttner, Kosztra, Hazeu, & Arnold, 2017). According to the proposed definition, impermeable features like buildings, roads, and artificially surfaced areas, which represent the urban fabric, range from 30 to 80% land coverage. Thus, the proposed threshold when applied in the corresponding HRL IMD layer, ensures that the output will include only the continuous urban fabric and not permanent or other built-up areas of sparse soil sealing degree.

**Table 35: Overview of the available LULC datasets for delineating impervious areas.**

IMPERVIOUS LC DATASET	Geometry Type/format	Spatial Resolution/MMU	IMPERVIOUS Class Name
Land Cover Map of Europe 2017 (S2GLC)	Raster Dataset	10m	Class 62 – Artificial surfaces and constructions
Corine Land Cover	Vector - Single polygon	25 Hectares MMU	Classes 1xx except for Class 131 – Mineral extraction sites and class 132 - dump sites
Copernicus High-Resolution Layer – Imperviousness Density (HRL – IMD)	Raster Dataset	20m	Include areas with >30% Imperviousness Value

A quick visual assessment of those datasets for the same reference year (2017), indicates that the S2GLC raster dataset is excluded from further analysis as it provides information that is already included in the remaining two datasets (Figure 18).

Concluding, the delineation of impervious areas is conducted using only the CORINE LC and the HRL IMD datasets.

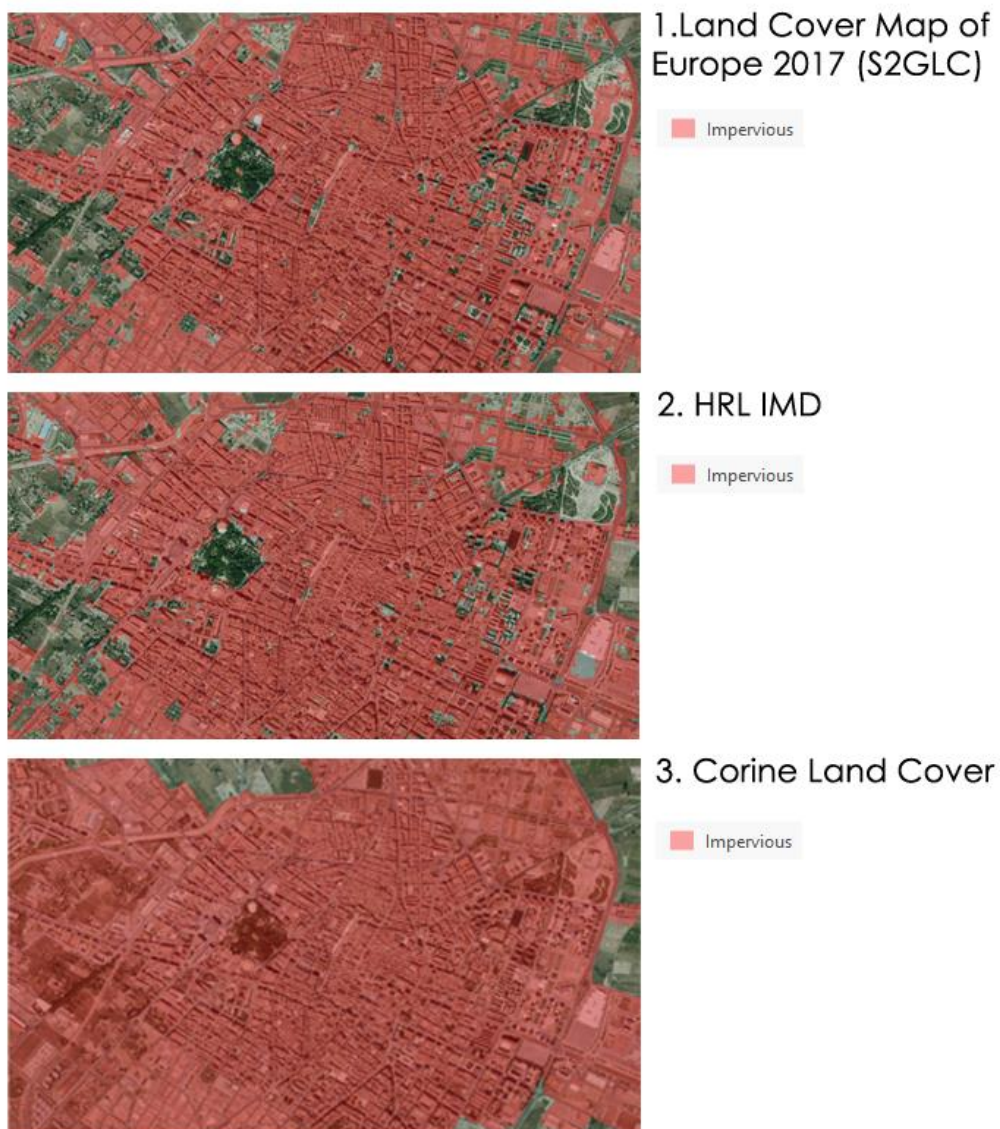


Figure 18: Visual assessment of the available datasets for the delineation of impervious areas in the city of Castello de la Plana (Spain). The results of the analysis indicate that the S2GLC dataset can be excluded from further analysis.

8.2.1.6. Delineation of Protected areas

The International Union for Conservation of Nature (IUCN), defines protected areas as areas that are clearly defined geographical spaces, recognized, dedicated, and managed, through legal or other effective means, to achieve the long-term conservation of nature with associated ecosystem services and cultural values (IUCN Definition 2008).

To delineate these areas in Europe, two different vector datasets are utilized as shown in the following Table 36 provided from the European Environment Agency and the Natura 2000 network of protected areas.

Table 36: Overview of the available datasets for the delineation of protected areas.

DATASET NAME	Geometry Type/format	Spatial Resolution/MMU	Coverage
Nationally designated protected areas (CDDA) polygons	Vector	5ha	European
Natura2000 protected sites	Vector	-	European

8.2.1.7. Delineation of Changed areas

To incorporate the dynamic aspect of MLs, changed areas are also included in the proposed methodology. Namely, within the context of the **MAIL** project, two main types of changes are considered: 1) changes related to forest activities such as afforestation and reforestation and 2) changes in urban fabric.

The delineation of changes in forest areas is implemented with the utilization of the available HRL TCD change products provided by Copernicus' Land Monitoring Service. This product shows increase or decrease of real TCD changes (%) in 2012-2015. To identify such changes in forest areas, a threshold of 50% is applied. This threshold ensures that the output will include certain changes due to reforestation or deforestation and not sparse or random changes.

Besides, changes in urban fabric are recorded in the corresponding HRL IMD change 2012-2015 product. Namely, this layer is produced as a difference between the reference years 2012-2015 and additionally 2006-2012, to fully match the CORINE LC production cycle. Changes are presented as degree of imperviousness change (-100% to +100%), in 20 m and 100 m pixel size and a supplementary classified (categorical) 20m change product. To delineate changes we extract classes "increased IMD" (value = 11) and "new cover" (value = 1). It should be noted that according to the HRL IMD change 2012-2015 product, the class "loss of cover" (value = 2) represents a total area of 0.5728 km² for the whole Europe. Thus, it could be excluded from further analysis, as these changes are unlikely to exist, while impervious areas are rarely changed into natural material surfaces or other areas that could be considered as MLs.

Besides, changes in impervious LC can be also identified in the CORINE CHA product. Namely, in the case of the **MAIL** project, changes are delineated using the 2012-2018 product, and more specifically, we extract classes that correspond to changes between 1xx to other 1xx classes except for classes 131 and 132 which represent Mineral extraction sites and Dumpsites accordingly.

8.2.2. Selection of appropriate Marginality indexes

Literature review on MLs suggests the utilization of different combinations of various indicators and criteria considering the special geophysical characteristics of each selected test site. Namely, those indicators are categorized mainly into four types: Terrain, Soil, Productivity, and Sustainability indicators. The aforementioned categories include a total of 17 indicators that have already been used in various projects for the identification of MLs (Table 37).

Table 37: Overview of the different types and categories of indicators used in the MAIL project.

Category	Indicator	Rank according to times found in bibliography
Terrain	slope	1
Soil	depth available to roots	1
Sustainability	acidity (pH)	2
Soil	texture	2
Sustainability	erosion	3
Soil	stoniness	3
Productivity	soil organic matter	3
Soil	water capacity	4
Sustainability	flood	4
Sustainability	sodicity	5
Soil	clay	6
Soil	sand	6
Sustainability	contamination	6
Productivity	cation exchange capacity	6
Productivity	productivity	7
Sustainability	dryness	8
Sustainability	natural toxicity	9

Within the context of the **MAIL** project several combinations of these 17 main indicators tested in the selected test areas during the second phase of the proposed methodology. The marginality indexes can be defined using several combinations of analysis ready datasets. Table 38 outlines the available datasets that can provide information on the selected marginality indicators and were finally used for this purpose.

**Table 38: Overview of the used datasets for the delineation of additional marginality indexes.**

Category	Indicator	Dataset Name in MAIL.gdb	Explanation
Terrain	slope	EU-DEM	Digital surface model based on SRTM and ASTER GDEM data
Soil	depth available to roots	DEPTH_ROO	Depth available to roots provided by JRC included in the European Soil Database Derived data
Sustainability	acidity (pH)	SOIL_PH	Soil pH in Europe provided by JRC - European Soil Data Centre (ESDAC)
Soil	Texture SILT + (2 X CLAY)	CLAY_TOP CLAY_SUB SILT_TOP SILT_SUB	Clay Content and Silt (mud) content in topsoil (TOP) and subsoil (SUB) provided by JRC - European Soil Data Centre (ESDAC) – European Soil Database & Soil Properties
Sustainability	erosion	RUSLE	Soil Loss by Water Erosion in Europe provided by JRC - European Soil Data Centre (ESDAC) – European Soil Database & Soil Properties
Soil	stoniness	VolumeStone	Volume of stones (%) - provided by JRC - European Soil Data Centre (ESDAC) – European Soil Database & Soil Properties PTRDB Attributes
Productivity	soil organic matter	OCC_SUB OCC_TOP	Organic Carbon content Subsoil and Topsoil provided by JRC - European Soil Data Centre (ESDAC) – using LUCAS Soil-DayCent for EU
Soil	water capacity	AWC_SUB AWC_TOP	Available water capacity (Topsoil and Subsoil) provided by JRC - European Soil Data Centre (ESDAC) – European Soil Database & Soil Properties PTRDB Attributes
Sustainability	flood	JRC_water_recurrence	JRC Global Surface water
Sustainability	sodicity	WISE	WISE derived soil property estimates provided by ISRIC
Soil	clay	CLAY_SUB CLAY_TOP	Clay content (subsoil and topsoil) – provided by JRC based on LUCAS topsoil data
Soil	sand	SAND_SUB SAND_TOP	Sand Content (subsoil and topsoil) – provided by JRC based on LUCAS topsoil data



Category	Indicator	Dataset Name in MAIL.gdb	Explanation
Sustainability	contamination	WISE	Contamination is based on Nitrogen levels. We select attribute "TOTN" – total nitrogen in WISE database
Productivity	cation exchange capacity	WISE	Cation exchange capacity is defined using class "CECS" in WISE database
Productivity	productivity	PROD_SB_FOREST PROD_SB_GRASS	Soil biomass productivity of forest areas and of grasslands and pastures. The final layer includes both datasets.
Sustainability	Dryness $AI = \frac{P}{PET}$	TerraClimate_pet_2018 TerraClimate_ppt_2018	Dryness is known as Aridity Index which is calculated using the total annual precipitation (P) and the total annual potential evapotranspiration (PET). Those layers are extracted from the TerraClimate multidimensional file (netcdf)
Sustainability	natural toxicity	WISE	Natural toxicity is defined based on Gypsum levels. We select attribute "GYPS" in WISE database

Regarding the use of Topsoil and Subsoil datasets, according to (Hiederer, 2013), since the soil textures are in many cases different in the topsoil and the subsoil, respectively, due to the material of which the soil was originally formed, to the soil development, and to its use, those layers have been examined separately. Thus, during the analysis, we examine both layers (topsoil and subsoil) separately.

Besides, regarding the use of layer of "soil organic matter" (Matteodo et al., 2018) suggests that even though considerably high proportions (between 30% and 63%) of carbon (C) are stored in the subsoil, most of the studies on soil have focused on the topsoil which is the main zone of activity for crop roots, and where the need to understand nutrient and water use efficiency is paramount. Although considerable concentrations of Soil organic matter occur in the topsoil, there can be equal or greater in some cases in the subsoil. Thus, during the analysis, we select to include both layers.

Two different methodological approaches were tested. In the first approach, we proceed with the implementation of all 17 thresholds as described in literature review. All indicators thresholds applied in all test sites across Europe and a final weighted overlay was also implemented for the classification of the results.

The results of the analysis indicated that after the implementation of approximately more than 5 indicators at the same time, the potential areas that could be characterized as MLs were way limited than expected. Namely, the implementation of a set of indicators including “slope”, “depth available to roots”, “acidity”, “texture”, “erosion” and “stoniness” returned almost no suitable areas of MLs in the test site of Thessaloniki, Germany and Poland, when we applied the “erosion” threshold (i.e., $RUSLE > 200$ ton/ha/yr). More specifically, soil erosion in the selected test sites has values less than 200 ton/ha/yr (Figure 19).

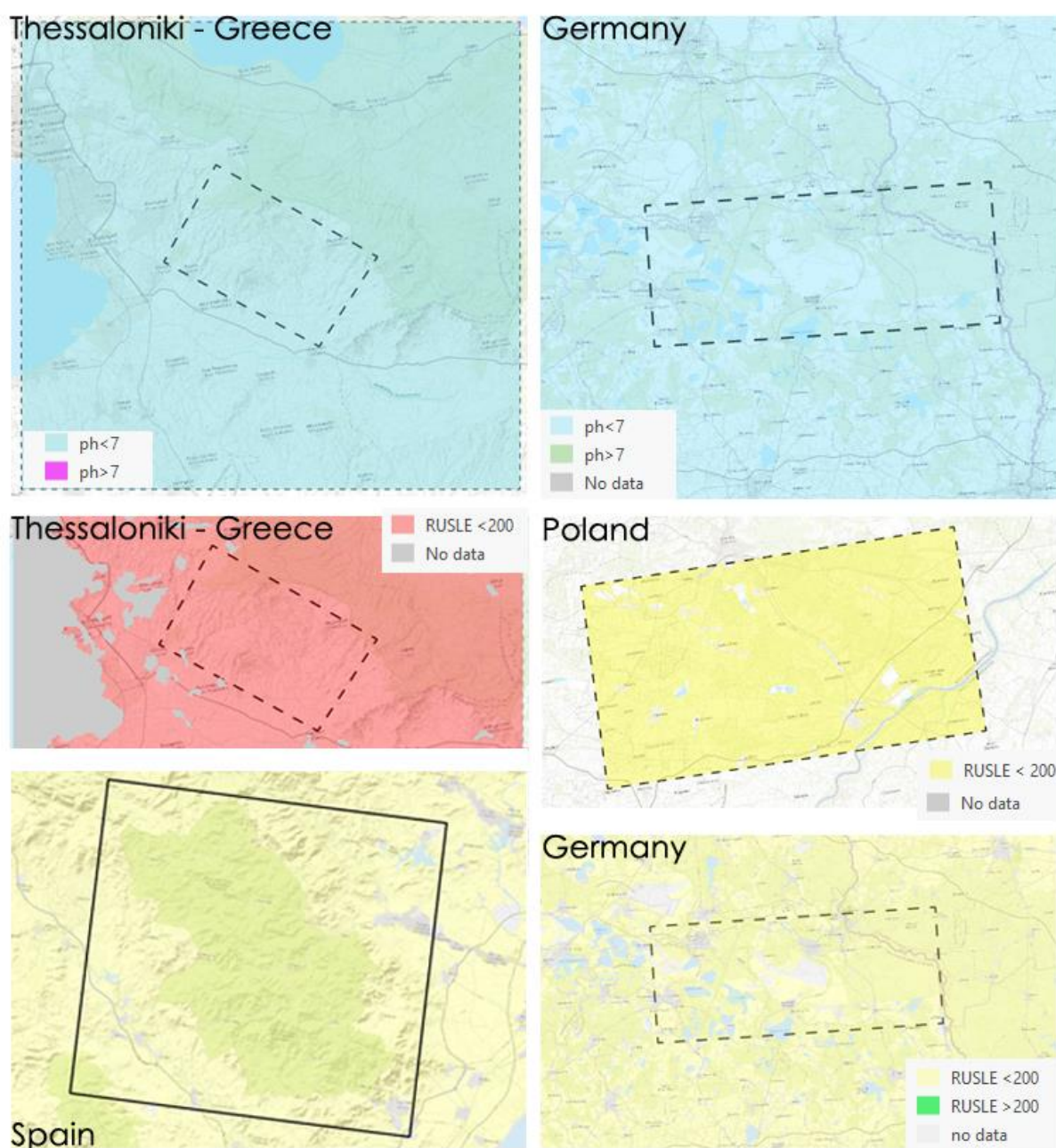


Figure 19: Values of Soil pH and Erosion (i.e. RUSLE) in the selected test sites.

On the other hand, the impact of several indicators including “soil ph”, on the final results were negligible. Namely, according to the literature review, MLs cannot be areas with soil ph = 7. In the case of Thessaloniki and Germany, the maximum value of soil ph = 6.6 and 5.8 correspondingly (Figure 19). Thus, this indicator had no impact in the final results and could be excluded. Besides, the implementation of the indicator of “natural toxicity” or the indicator of “dryness” which are used in specific ML studies, had barely any effect when applied in the selected test sites which are large areas with few variations on their productivity and sustainability characteristics.

Concluding, taking into account the proposed thresholds and the fact that the selected test sites cover a wide range of areas with different geophysical characteristics across Europe, the implementation of a weighted overlay of all 17 indicators – thresholds at the same time, cannot provide adequate results. In addition, the intermediate results after the implementation of each threshold indicated that marginality indexes related to the Terrain and the Soil characteristics of each area were the ones with the most significant impact on the analysis. Thus, an alternative approach was finally followed.

In the second methodological approach, the applied thresholds were related only to selected marginality indexes per category (i.e., terrain, soil, sustainability and productivity), according to their importance as ranked based on literature review (Table 37). Thus, the marginality indexes that were finally applied within the context of the **MAIL** project are outlined in the following Table 39.

Table 39: Overview of the finally selected marginality indexes.

Indicator	Rank according to literature review	Dataset Name in MAIL.gdb	Explanation
Slope (Terrain)	1	EUDEM	Digital surface model based on SRTM and ASTER GDEM data
depth available to roots (Soil)	1	DEPTH_ROO	Depth available to roots provided by JRC included in the European Soil Database Derived data
Texture SILT + (2 X CLAY) (Soil)	2	CLAY_TOP CLAY_SUB SILT_TOP SILT_SUB	Clay Content and Silt (mud) content in topsoil (TOP) and subsoil (SUB) provided by JRC - European Soil Data Centre (ESDAC) – European Soil Database & Soil Properties



Indicator	Rank according to literature review	Dataset Name in MAIL.gdb	Explanation
soil organic matter (Productivity)	3	OCC_SUB OCC_TOP	Organic Carbon content Subsoil and Topsoil provided by JRC - European Soil Data Centre (ESDAC) – using LUCAS Soil-DayCent for EU

The results of the analysis indicated that the implementation of the aforementioned set of indicators and their corresponding threshold values, returned adequate results when applied in semi-mountainous areas. Nevertheless, the same methodology was not acceptable for lowland or semi-urban areas. For instance, slope is a main marginality index for detecting MLs. Namely, (Gopalakrishnan, Cristina Negri, & Snyder, (2011), suggests that marginally suitable lands are areas with **slope > 15%**.

Nevertheless, the implementation of the aforementioned threshold in lowland areas such as the area of Nochten and Welzow in Saxony, Germany, and the region of “Świętokrzyskie” in Poland resulted in the exclusion of nearly all potential sites that could be identified as MLs and they satisfy the requirements of the remaining constraints. To overcome this barrier, we suggest adapting the slope threshold accordingly, taking into account the geophysical characteristics of three potential types of sites (Table 40).

Moreover, regarding the remaining marginality indexes, the applied threshold is common for all three different types of areas.

Table 40: Slope thresholds according to each type of test sites.

INDICATOR	Proposed threshold	Type of area
SLOPE	MLs are areas where slope is $\geq 15\%$	urban and semi-urban areas
	MLs are areas where slope is $\geq 8\%$	Lowland areas
	MLs are areas where slope is $\geq 15\%$	Semi-Forest areas

8.3 Development of new indicators sets

Finally, three different sets of indicators of “soft” constraints are developed within the **MAIL** project (Table 41).

Table 41: Overview of the new thresholds and marginality indexes developed within the MAIL project.

INDICATOR	Type of area		
	Urban & Semi-urban areas	Lowland areas	Semi-Forest areas
Slope (Terrain)	MLs are areas where slope is $\geq 15\%$	MLs are areas where slope is $\geq 8\%$	MLs are areas where slope is $\geq 15\%$
depth available to roots (Soil)	MLs are areas where depth available to roots is $< 100\text{cm}$		
Texture SILT + (2 X CLAY) (Soil)	MLs are areas where Texture is $\leq 30\%$		
soil organic matter (Productivity)	MLs are areas where Soil Organic Matter is between 1-19% (topsoil and subsoil)		

8.4 Initial results according to the proposed methodology

8.4.1 Initial results after the implementation of “hard” thresholds

As already stated, the first phase of the proposed methodology includes the “hard” thresholds approach (Figure 2). The results of the analysis of this first phase create an intermediate layer for each test site, named “ML HARD” including all potential MLs after the exclusion of all LULC types that are not fulfilling the definition of MLs.

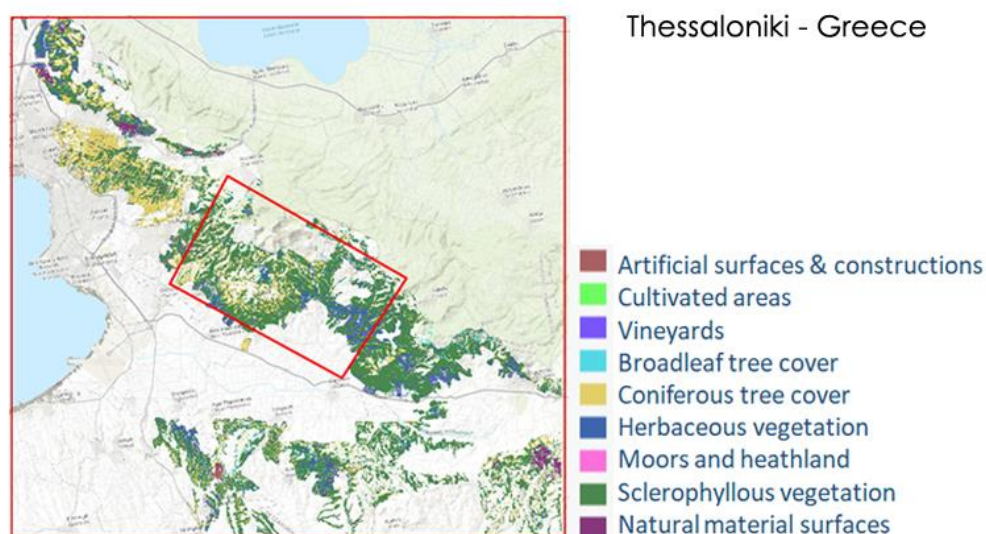


Figure 20: Overview of the intermediate results of the first phase of the proposed methodology (ML_HARD thresholds) for the test site of Thessaloniki – Greece.

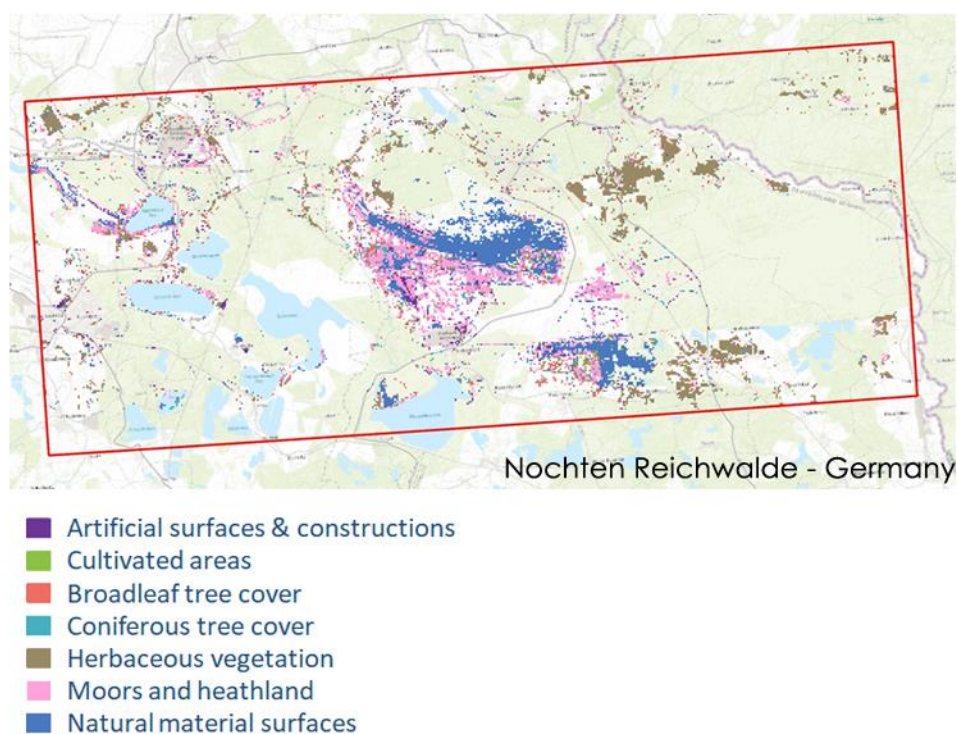


Figure 21: Overview of the intermediate results of the first phase of the proposed methodology (ML_HARD thresholds) for the test site of Nochten Reichwalde – Germany.

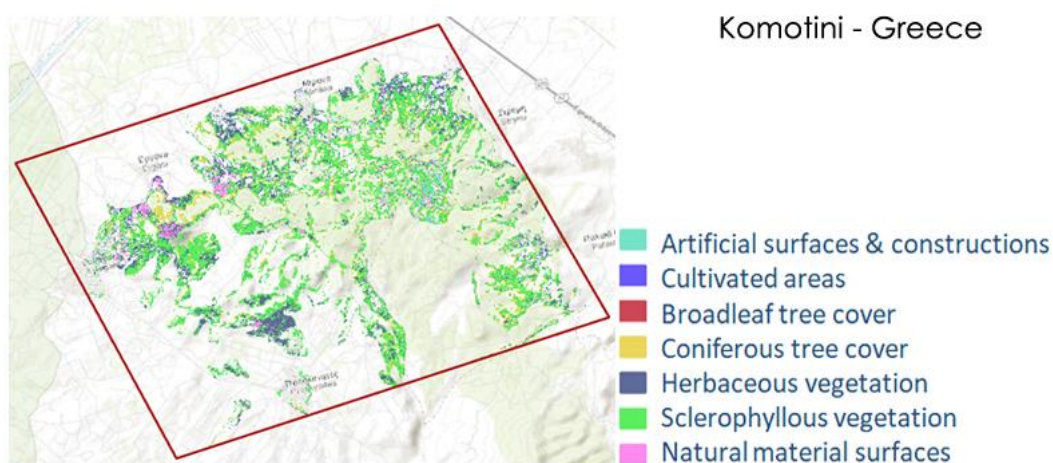


Figure 22: Overview of the intermediate results of the first phase of the proposed methodology (ML_HARD thresholds) for the test site of Komotini – Greece.

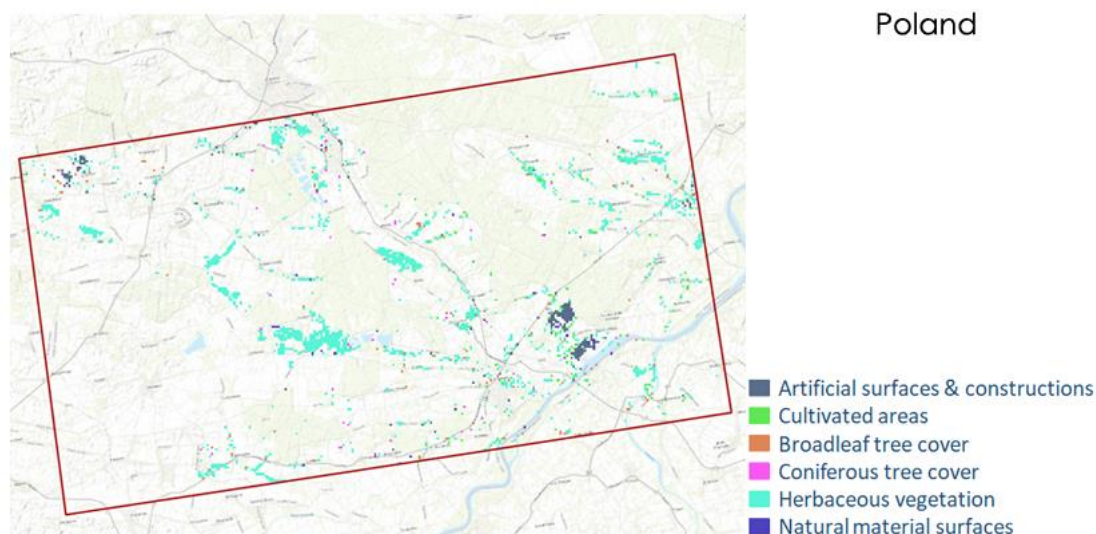


Figure 23: Overview of the intermediate results of the first phase of the proposed methodology (ML_HARD thresholds) for the test site of Poland.

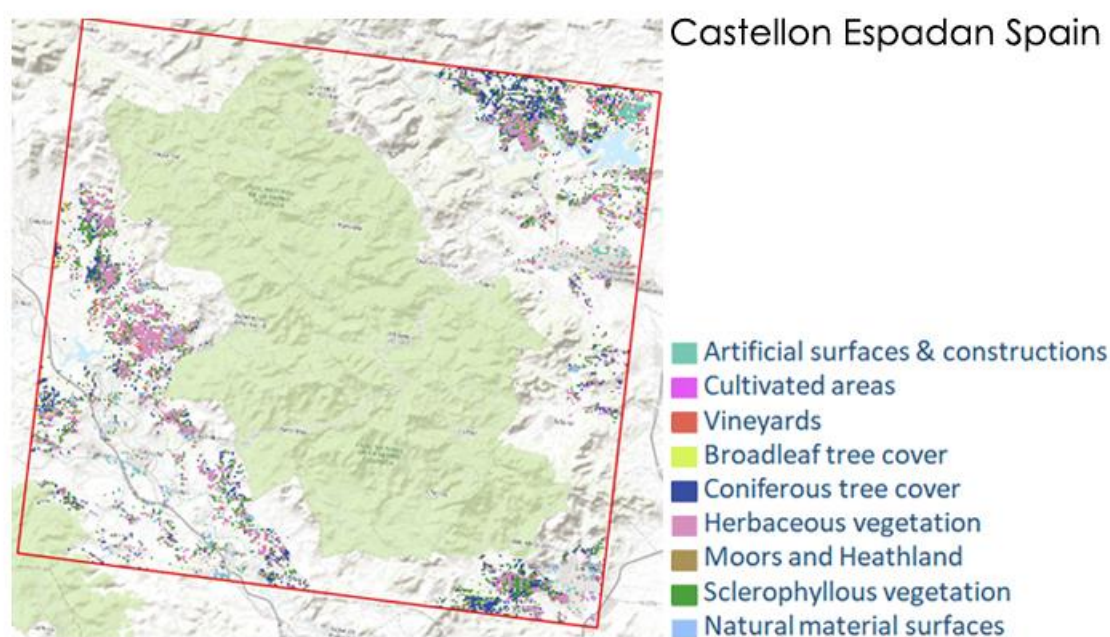


Figure 24: Overview of the intermediate results of the first phase of the proposed methodology (ML_HARD thresholds) for the test site of Castellon Espadán – Spain.

The results of the analysis indicated that after the exclusion of the aforementioned LULC datasets, the remaining values that are classified as potential MLs include areas that according to the selected S2GLC basemap are classified as “artificial surfaces & constructions”, “cultivated areas” and “vineyards” and “broadleaf tree cover” and “coniferous tree cover”. More specifically, these errors are mainly due to classification commission and omission errors of



the original selected basemap and the selected LULC datasets that were utilized to delineate forest, croplands, and impervious areas. It should be noted though, that during the second phase of the implementation of additional marginality indexes, these errors are minimized as these areas do not meet the requirements of the additionally applied thresholds. Besides, the utilization of a more reliable and updated LULC dataset as basemap could improve the results of the analysis.

8.4.2 Results after the implementation of additional marginality indexes

After the implementation of the “hard” thresholds, we proceed to the additional thresholds phase. The final remaining layer is re-classified into tree classes as shown in Table 42.

Table 42: The classes of the results of MLs.

	FINAL LAYER OF MLS		
	Class Name		
	Marginal lands with high plantation suitability	Marginal lands with low plantation suitability	Potentially unsuitable lands
Class name according to S2GLC dataset	<ul style="list-style-type: none"> Herbaceous vegetation Moors and heathland Sclerophyllous vegetation Natural Material Surfaces 	<ul style="list-style-type: none"> Cultivated areas Vineyards 	<ul style="list-style-type: none"> Artificial surfaces and constructions Broadleaf tree cover Coniferous tree cover

The results of the proposed methodology and the corresponding areas of “Marginal lands with high plantation suitability”, “Marginal lands with low plantation suitability” and “Potentially unsuitable lands” in the selected test sites, are the following:

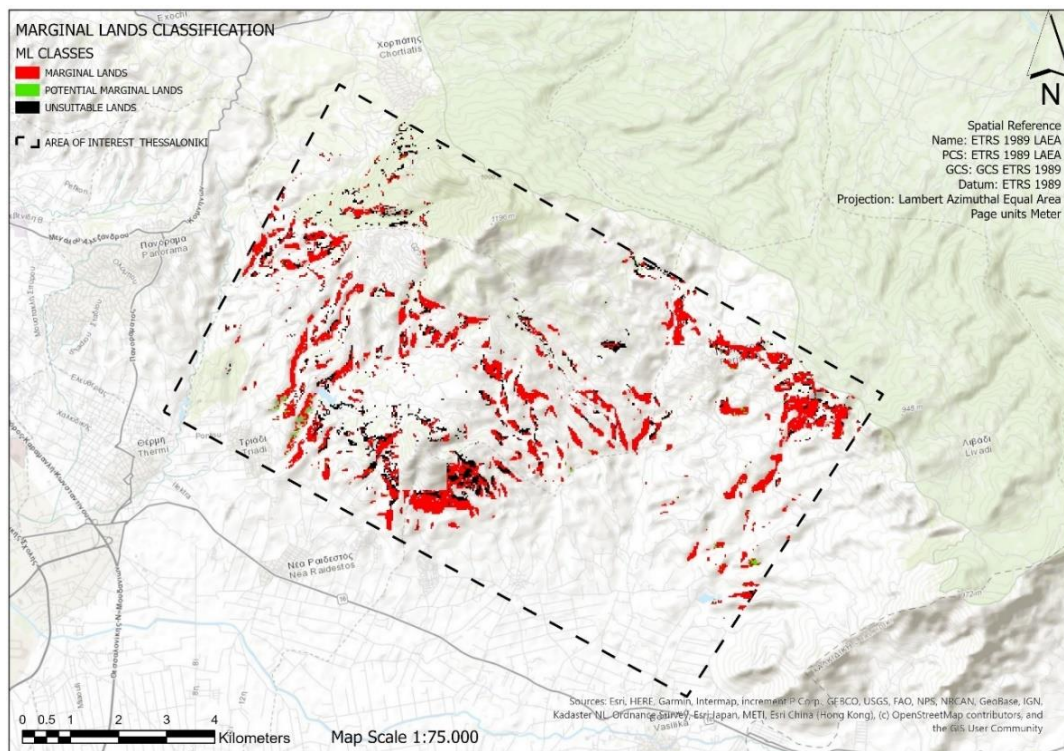


Figure 25: Overview of the results of the proposed methodology for the test site of Thessaloniki – Greece.

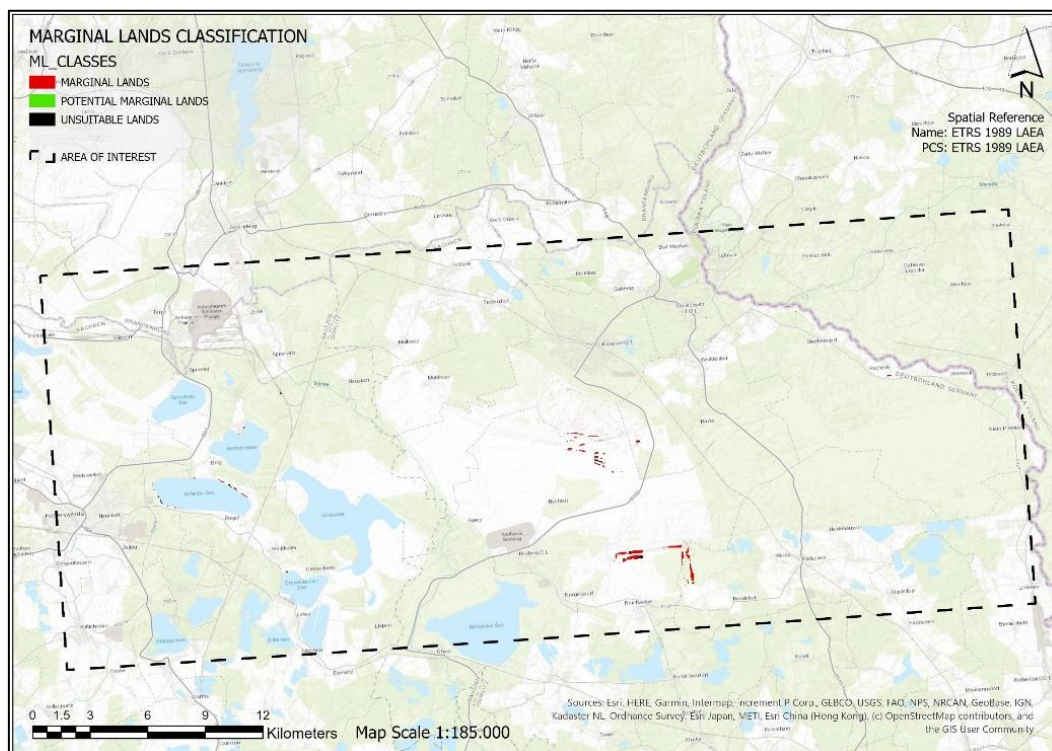


Figure 26: Overview of the results of the proposed methodology for the test site of Nochten Reichwalde Germany.

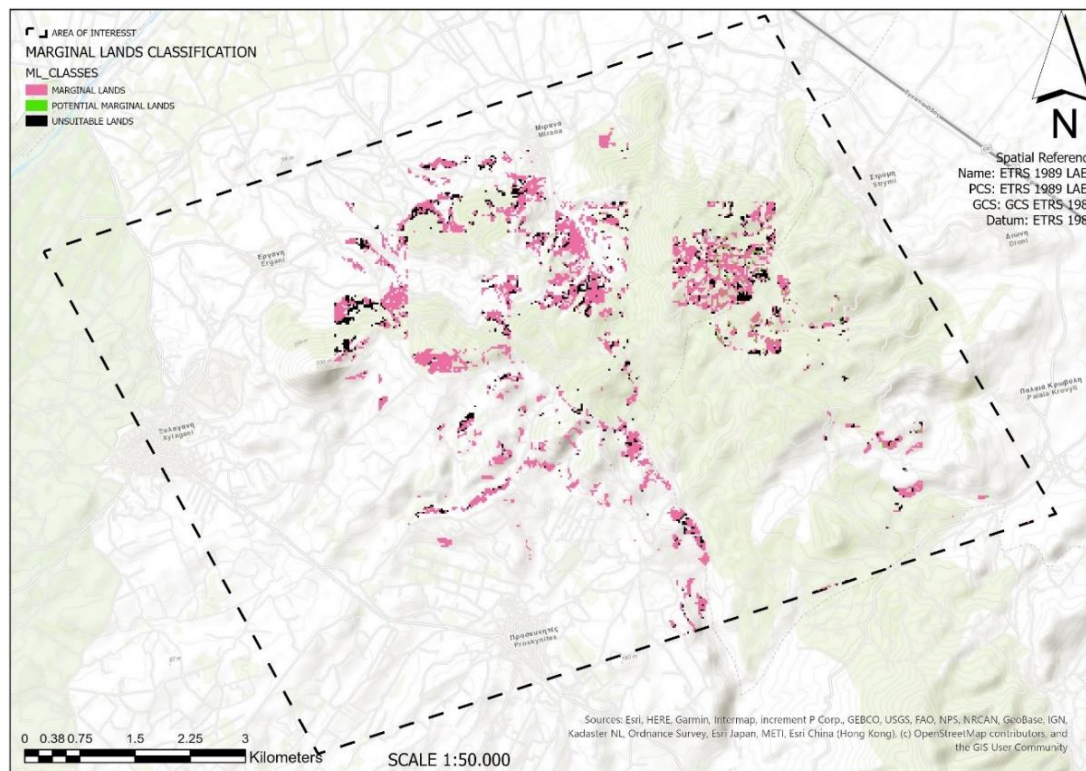


Figure 27: Overview of the results of the proposed methodology for the test site of Komotini – Greece.

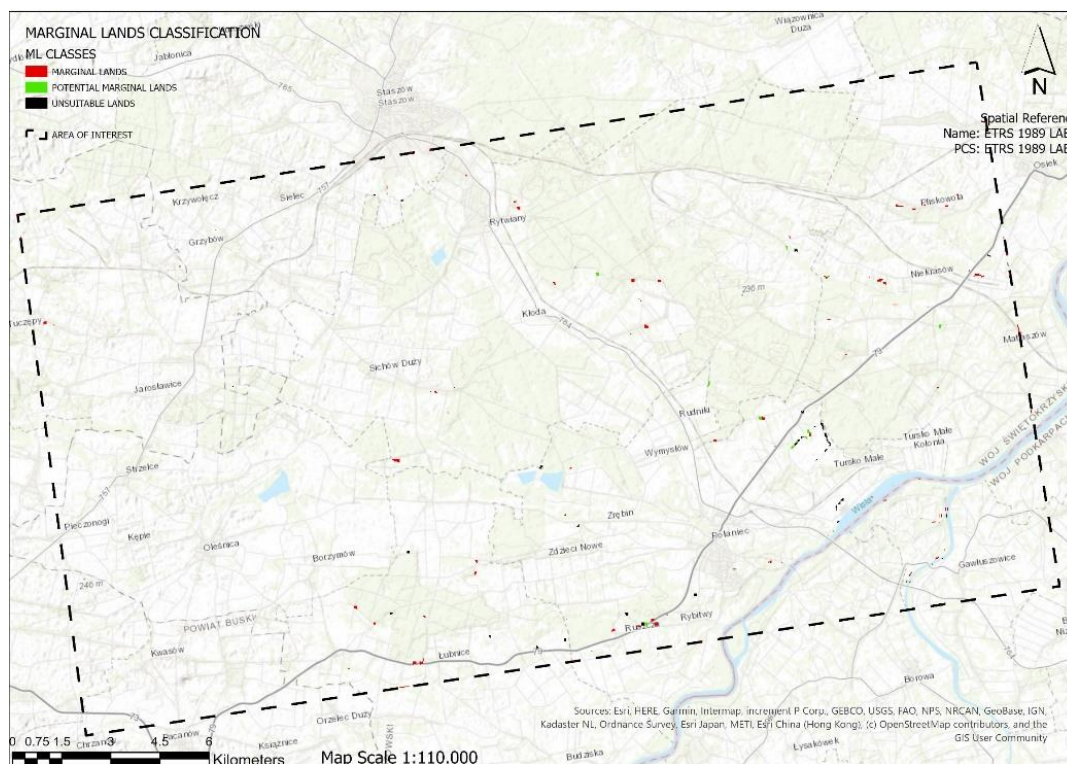


Figure 28: Overview of the results of the proposed methodology for the test site of Poland.

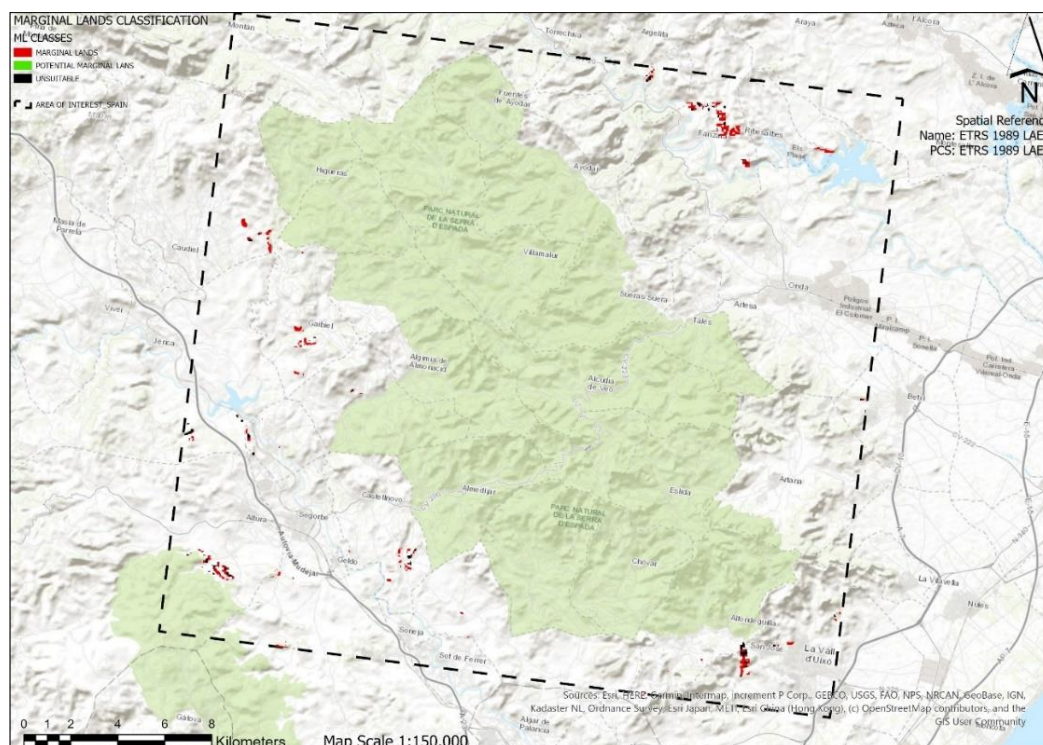


Figure 29: Overview of the results of the proposed methodology for the test site of Castellon Espadán – Spain.

The corresponding areas calculated in hectares of each class for the selected test sites are outlined in Table 43.

Table 43: Areas in hectares of each type of MLs for the selected test sites.

		FINAL LAYER OF MLS		
		Class Name and area (ha)		
		Marginal lands with high plantation suitability	Marginal lands with low plantation suitability	Potentially unsuitable lands
Test sites	Thessaloniki – Greece	702 ha	14ha	206ha
	Nochten Reichwalde – Germany	91ha	2ha	15ha
	Komotini – Greece	309ha	4ha	70ha
	Poland	36ha	9ha	16ha
	Castellon Espadán - Spain	173ha	0ha	55ha

9. DESIGN AND DEVELOPMENT OF COMPUTATION WORKFLOWS FOR ALL EUROPE.

9.1 Workflow design for indicator sets and algorithm development to produce new Hard layers

The computation workflows for the implementation of the proposed methodology were developed independently to ensure that in case of possible changes, the structure of the workflows to produce each remaining intermediate layer, will not be affected. Thus, in the first top-down stepwise approach of hard thresholds, six individual workflows were created while an additional final workflow combines all the intermediate resulting layers to calculate the intermediate layer of “ML_HARD” which contains all the remaining LULC classes that meet the requirements of the definition of Marginal Lands.

It should be noted that the developed workflows were developed using mainly the Spatial Analyst Toolbox of the software ArcGIS Pro and all intermediate layer contain only the remaining LULC classes that fulfill the applied thresholds. The main idea is to produce intermediate layers containing LULC classes that will finally be excluded, and they do not contain potential MLs areas.

9.1.1 Workflow design to produce intermediate hard layer “Other Land Cover”

The "Sentinel2 Global Land Cover (S2GLC)" layer was used as the starting point and basis for making the MLs layer. First, the land uses 162 (water bodies), 123 (permanent snow), 106 (peatbogs), and 105 (marshes) were eliminated (Figure 30). Over the resulting layer, the remaining intermediate layers (forest, croplands, protected areas, impervious and changed areas) were removed.



Figure 30: Workflow design to produce an intermediate layer (red box) Other Land Cover.

Where:

S2GLC: Land Cover Map of Europe 2017

The intermediate layer “S2GLCotherLC” contains all the S2GLC classes except for “water bodies”, “permanent snow”, “marshes” and “peat bogs”.

9.1.2 Workflow design to produce intermediate hard layer “Forest”

This workflow is designed for the delineation of “forest areas” (Figure 31). As already stated, the forest extent is outlined using a combination of different datasets that can provide this information. It should be noted, that to overcome the barrier of variations in the reference year of each dataset, for the purposes of the project, we defined forest extent setting 2018 as the reference year. Thus, in cases where the available datasets were produced before 2018, we utilized the available “change” layers between these reference years and 2018, to define changes in forest cover and to incorporate these changes in the final forest extent of 2018.

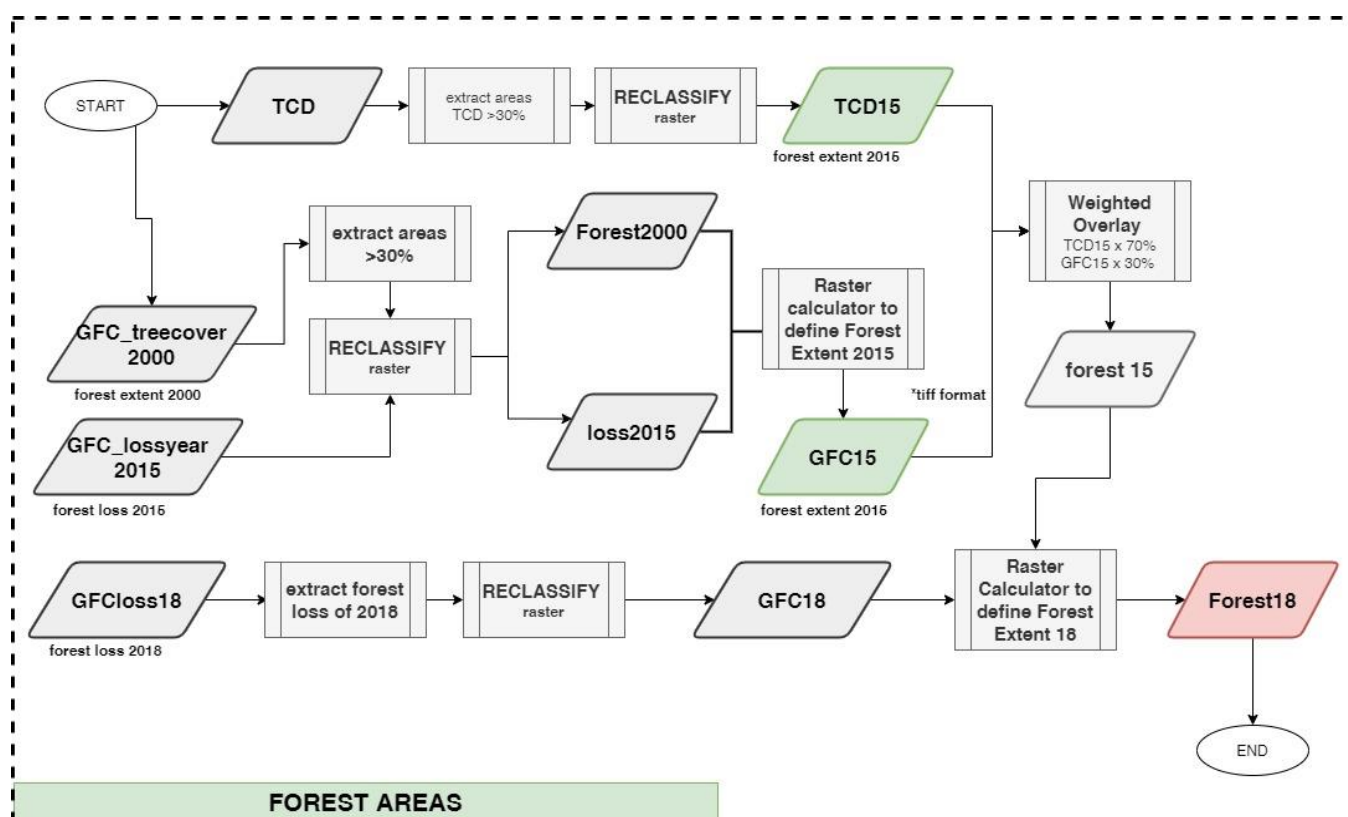


Figure 31: Workflow design to produce an intermediate layer (red box) Forest Extent of 2018.

Where:

TCD: HRL Tree Cover Density (TCD)

GFC_treecover2000: Global Forest Change Tree cover of 2000



GFC_lossyear 2015: Global Forest Change Tree cover of 2015

GFC_lossyear 2018: Global Forest Change Tree cover of 2018

TCD15: forest extent of 2015 according to the HRL Tree Cover Density dataset after the extraction of all potential sites with Tree Cover density > 30%

GFC15: Forest extent of 2015 according to the Global Forest Change Tree Cover dataset

GFC18: Forest extent of 2015 according to the Global Forest Change Tree Cover dataset

The scope of this workflow is to define Forest Extent for the reference year of 2018 (intermediate layer: Forest18). Starting from the HRL TCD dataset we extract all areas (extract by attributes) with Tree Cover Density >30%. The produced raster layer is reclassified into two classes representing Forest/Non-Forest areas (Class 0 and Class 1 respectively). The final intermediate layer is the Forest Extent of 2015 (TCD15) according to the HRL TCD dataset.

The same procedure applies to the second available Forest dataset, the Global Forest Change which represents the forest extent of 2000 and the corresponding changes between 2000-2015. The two intermediate layers “Forest2000” and “Loss2015” are further combined to define the forest extent of 2015 according to the Global Forest Change – Tree Cover dataset (Raster Calculator). Besides, a weighted overlay is applied to combine these two forest datasets taking into account the comparative validation of HRL-TCD and the GFC products which indicated that the HRL-TCD layer represents forest extent more accurately. The final intermediate layer represents forest extent for the reference year of 2015.

Finally, to delineate forest extent for the reference year of 2018, the produced layer is combined with the corresponding “loss of 2018”.

9.1.3 Workflow design to produce intermediate hard layer “Croplands”

This workflow is designed for the delineation of “croplands” (Figure 32). As already stated, information on croplands can be acquired using a combination of the Land Cover Map of Europe (S2GCL) and the corresponding Corine LC.

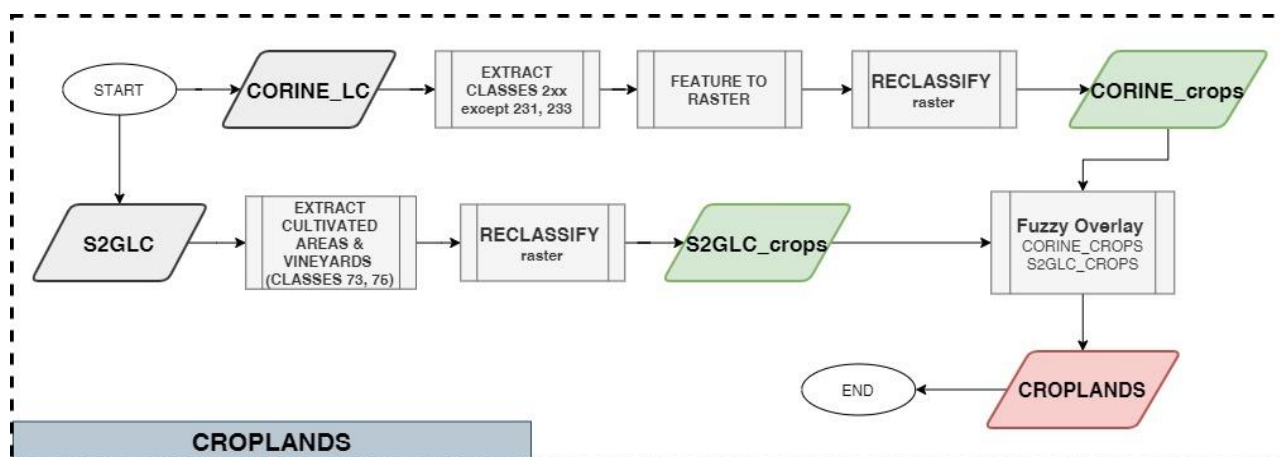


Figure 32: Workflow design to produce an intermediate layer (red box) Croplands.

Where:

Corine_LC: Corine Land Cover

S2GLC: Land Cover Map of Europe

Corine_Crops: Croplands as identified in Corine_LC dataset

S2GLC_crops: Croplands as identified using S2GLC dataset

More specifically, Class 73 of the S2GLC layer represents cultivated areas, while class 75 represents vineyards. In addition, the corresponding classes are classes 2xx of Corine LC dataset including “agricultural areas” except class 231 which represent Pastures (i.e., lands that are used at least 5years permanently for fodder production including natural or sown herbaceous species, unimproved or lightly improved meadows and grazed or mechanically harvested meadows). Both datasets are reclassified into two classes representing Croplands/Non-Cropland areas. Finally, a Fuzzy Overlay is performed to combine these layers. The final intermediate layer “Croplands” includes information on cropland areas.

In detail, the function of “Fuzzy Overlay” allows the analysis of multiple datasets using multicriteria overlay analysis. Namely, in the case of [MAIL](#) project, the method of “Fuzzy And” and “Fuzzy or” were utilized to produce the intermediate layers containing information of the forest extent, croplands, Impervious and changed areas as a combination of all the available datasets. The “Fuzzy And” overlay, returns the minimum value of the sets the cell location belongs to. Thus, we can identify the least common denominator of all the available datasets according to the suggested criteria. In addition, the corresponding “Fuzzy Or” overlay, is useful for identifying the highest membership values for any of the input criteria. Namely, to produce the intermediate layer “croplands” we utilized the “Fuzzy or” function which allowed the

identification of all locations that had at least one of the criteria being fully in the suitable datasets, a value of 1 which represents croplands.

9.1.4 Workflow design to produce intermediate hard layer “Protected areas”

This workflow is designed for the delineation of protected areas (Figure 33). The intermediate vector layer of “Protected areas” is derived from a combination of the “Nationally designated protected areas” (CDDA_polygon) and the Natura2000 protected sites vector dataset.

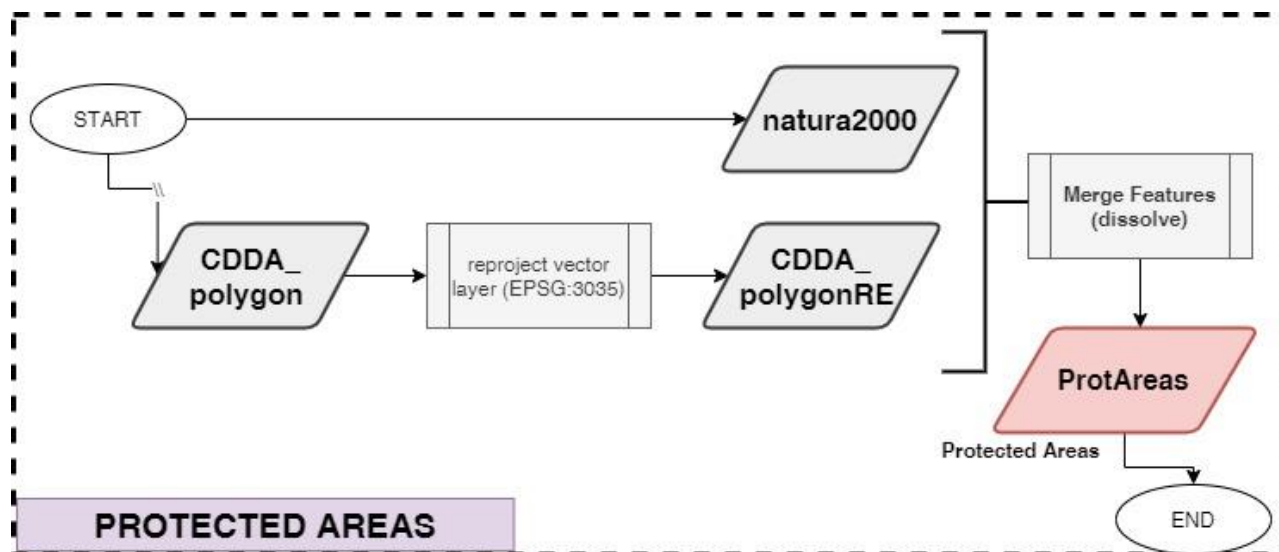


Figure 33: Workflow design to produce intermediate layer (red box) Protected areas.

Where:

CDDA_polygon: Nationally designated protected areas vector dataset

Natura2000: Natura2000 protected sites

9.1.5 Workflow design to produce intermediate hard layer “Impervious”

This workflow is designed for the delineation of Impervious areas (Figure 34). The intermediate vector layer “Impervious” is derived from a combination of two different LULC datasets: The Corine Land Cover and the HRL IMD raster dataset.

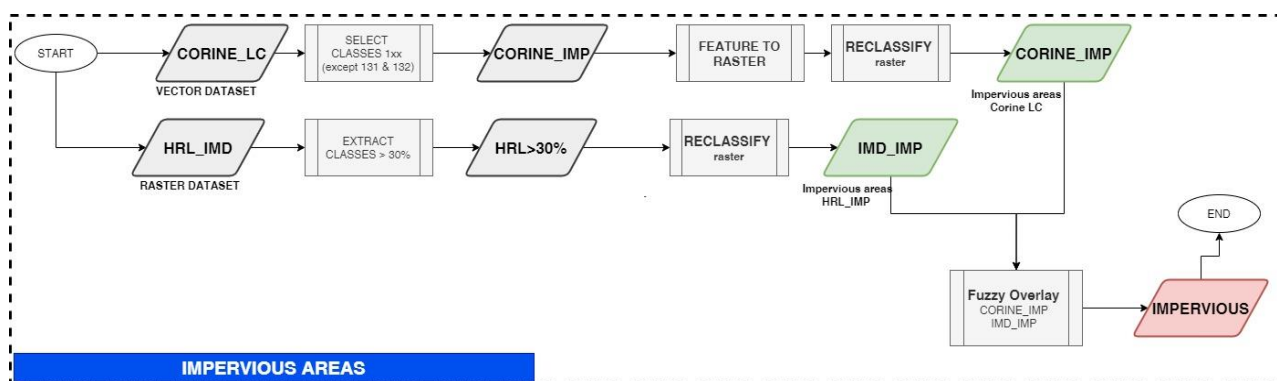


Figure 34: Workflow design to produce intermediate layer (red box) Impervious areas.

Where:

Corine_LC: Corine Land Cover

HRL_IMD: Copernicus High- Resolution Layer – Imperviousness Density

More specifically, Classes 1xx of CORINE LC represent “impervious areas” except for classes 131 and 132 which represent Mineral extraction sites and Dumpsites accordingly and thus are not included during the analysis. The final vector layer is converted to raster and both datasets are reclassified into two classes representing Impervious/Non-Impervious areas. Finally, a Fuzzy Overlay and more specifically “Fuzzy or” function is performed to combine these layers. The final intermediate layer “Impervious” includes information on impervious areas.

9.1.6 Workflow design to produce intermediate hard layer “Changed Areas”

This workflow is designed for the delineation of changed areas including changes in impervious surfaces and changes due to forest activities (Figure 35). The intermediate vector layer “Changed” is derived from a combination of three different LULC datasets: The Corine Land Cover Changes, the Tree Cover Density Changes, and the HRL Impervious changes dataset.

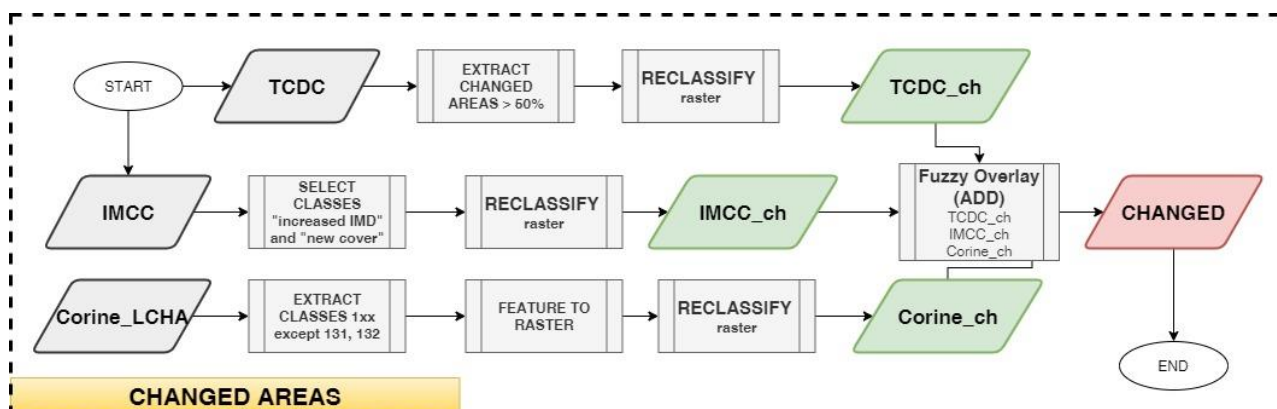


Figure 35: Workflow design to produce an intermediate layer (red box) Changed areas.



Where:

TCDC: High-Resolution Tree Cover change product

IMCC: High-Resolution Impervious change product

Corine_LCHA: Corine Land Cover Change product

TCDC_ch: All changed areas according to the TCDC dataset

IMCC_ch: All changed areas according to the IMCC dataset

Corine_ch: All changed areas according to Corine LC dataset

More specifically, to delineate changes in forest areas, we extract changes, where the Tree Cover Density changed, is more than 50%. Besides, changes in the urban fabric and impervious areas, are outlined using the classes “increased IMD” and “new cover” from the IMCC dataset (extract by attributes). Finally, changes in impervious areas are also extracted using classes 1xx except for classes 131 and 132 which represent changes in Mineral extraction sites and Dumpsites accordingly. All intermediate layers are reclassified into two classes representing “Changed/Non-Changed areas” and finally a fuzzy overlay using the function “Fuzzy AND” is performed to produce the final intermediate layer “CHANGED” which represents all the occurred changes. In the case of the delineation of “changed areas,” we utilized function “Fuzzy AND” to include all the occurred changes in each dataset.

9.1.7 Workflow design to produce intermediate hard layer “ML_HARD”

This workflow is designed to produce the Intermediate layer “ML_HARD” which represents all potential Marginal Lands after the implementation of the first phase of the proposed methodology (Figure 36). More specifically, all the previously produced intermediate layers that include Land Uses that cannot be Marginal Lands are now combined with the S2GLC basemap. The final “ML_HARD” layer is a raster dataset based on the S2GLC LULC basemap including all remaining classes that could represent MLs.

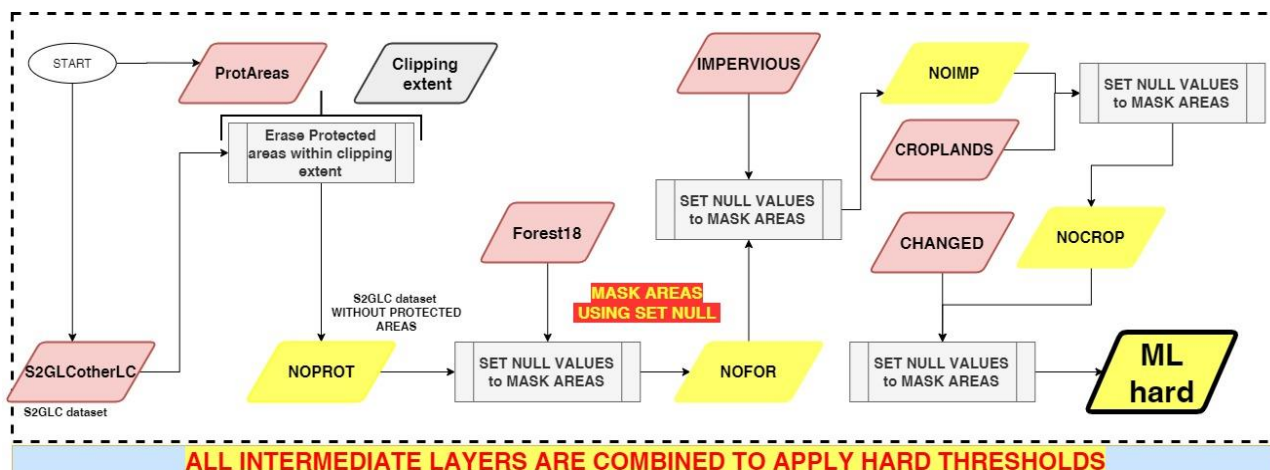


Figure 36: Workflow design for the production of an intermediate layer (red box) “ML HARD”.

More specifically, the workflow aims to combine all previously produced intermediate layers, by using the function “Set Null Values” which results in the extraction from the original S2GLC basemap of all Land Uses that do not meet the requirements of the definition of MLs.

The final combination of all workflows of this first phase of the proposed methodology is outlined in Figure 37.

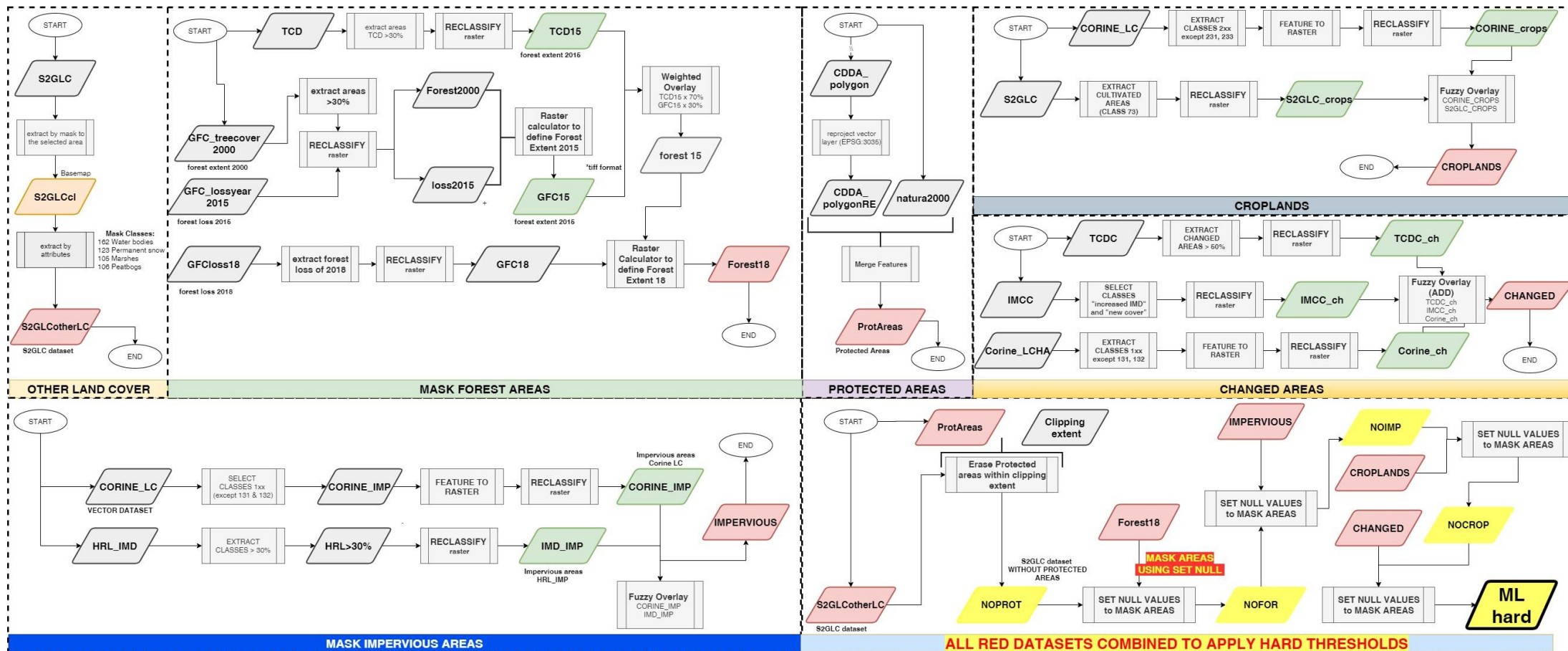


Figure 37: Workflow design for the implementation of the first phase (Hard layers) of the proposed methodology.



9.2 Workflow design for indicator sets and algorithm development to produce new SOFT layers

The computation workflows for the implementation of the proposed methodology were developed independently to ensure that in case of possible changes, the structure of the workflows to produce each remaining intermediate layer, will not be affected. This workflow is designed to produce the "Soft" intermediate layers described in chapters 6.4.2, 6.4.3, and 6.4.4. Table 44 summarizes the calculated soft layers, whether they were for topsoil (T), subsoil (S) or without differentiating the soil profile (-), the number of layers involved in each indicator (1 or 2), the different scores (based on different thresholds) and weights per indicator described extensively in section 6.5. Furthermore, there are a total of 17 indicators and 24 soft layers, where the theoretical maximum of marginality to be obtained is 9.98. Those pixels close to this value would be the most favorable marginal areas and those close to zero would be the most unfavorable marginal areas for implementing any type of forest plantation. It is a theoretical maximum because it is not probable that a pixel will result where all the indicators of marginality have a score of 10. To elaborate the division between "Marginal lands with high plantation suitability", "Marginal lands with low plantation suitability" and "Unsuitable MLs" once all the soft layers have been processed and overlapped, the maximum and minimum values will be analyzed and the division into the three categories of marginality will be elaborated.

It should be noted that in the column "weight" it is possible to observe gray cells, these weights correspond to half of the weight calculated in section 6.5 since originally it was not considered to use the topsoil and subsoil layers.

Table 44: Soft layer summary.

Type	ID	Layer name	Topsoil/ Subsoil	Nº of layers	Weight	Scores		
						Marginal lands with high plantation suitability	Marginal lands with low plantation suitability	Unsuitable MLs
						10	5	1
Terrain and Soil	1	Slope	-	1	0.17	1.7	0.85	0.17
	2	Depth Available Roots	-	1	0.17	1.7	0.85	0.17
	3	Stoniness	T	2	0.03	0.3	0.15	-
			S		0.03	0.3	0.15	-
	4	Texture	T	2	0.045	0.45	0.225	0.045
			S		0.045	0.45	0.225	0.045
	5	Clay	S	2	0.015	0.15	0.075	0.015
			T		0.015	0.15	0.075	0.015
	6	Sand	S	2	0.015	0.15	0.075	0.015



Type	ID	Layer name	Topsoil/ Subsoil	Nº of layers	Weight	Scores		
						Marginal lands with high plantation suitability	Marginal lands with low plantation suitability	Unsuitable MLs
						10	5	1
			T		0.015	0.15	0.075	0.015
	7	Total Available Water	S	2	0.02	0.2	0.1	-
			T		0.02	0.2	0.1	-
Sustainability	8	Soil Acidity	-	1	0.09	0.9	0.45	0.09
	9	Soil Erosion	-	1	0.06	0.6	0.3	0.06
	10	Flooding	-	1	0.04	0.4	0.2	0.04
	11	Sodicity	-	1	0.03	0.3	0.15	0.03
	12	Toxicity Contamination	-	1	0.03	0.3	0.15	0.03
	13	Natural Toxicity	-	1	0.02	0.2	0.1	0.02
	14	Dryness	-	1	0.02	0.2	0.1	0.02
Productivity	15	Soil Organic Matter	S	2	0.03	0.3	0.15	-
			T		0.03	0.3	0.15	-
	16	Caption Exchange Capacity	-	1	0.03	0.3	0.15	0.03
	17	Productivity	-	1	0.01	0.14	0.07	0.014
			-	1	0.01	0.14	0.07	0.014
	Total			24	1.00	9.98	4.99	0.838

9.2.1 Workflow design to produce intermediate soft layer related to “Terrain and Soil”

This workflow is designed to produce the Intermediate layer: “Slope”, “Stoniness”, “Water available content”, “Depth Available to roots”, “Clay”, “Sand” and “Texture” which has been used during the second phase of the proposed methodology (Figure 38).

The slope has been calculated using the Digital Surface model based on SRTM and ASTER GDEM data (EU-DEM). According to the proposed methodology, three different thresholds have been applied for the delineation of the slopes (Table 22).

These intermediate raster layers were reclassified according to the thresholds defined in Table 22. More specifically, the final intermediate layers “Depth available roots”, “Clay”, “Sand” and “Textura” contain 4 classes: 0, areas outside the reference thresholds; 1, “unsuitable lands”; 5, “Marginal lands with low plantation suitability”; and 10, “Marginal lands with high plantation suitability”. The intermediate layers “Stoniness” and “Water available content” contain 3 classes: 0, areas outside the reference thresholds; 5, “Marginal lands with low plantation suitability”; and 10, “Marginal lands with high plantation suitability”. All these layers were based on European Soil Database Derived data layer downloaded from the European Commission Joint Research Centre (JRC) and more specifically from the European Soil Data Center

(ESDAC). Once the reclassification of the layers is done, each one will be multiplied with its corresponding weight. For the layers “Stoniness”, “Water available content”, “Clay”, “Sand” and “Texture” where exists information about the topsoil and subsoil, the initial weight defined in Table 22 was divided by two to maintain the same proportion and representativeness of the layer over the whole, as indicated at the beginning of this subchapter.

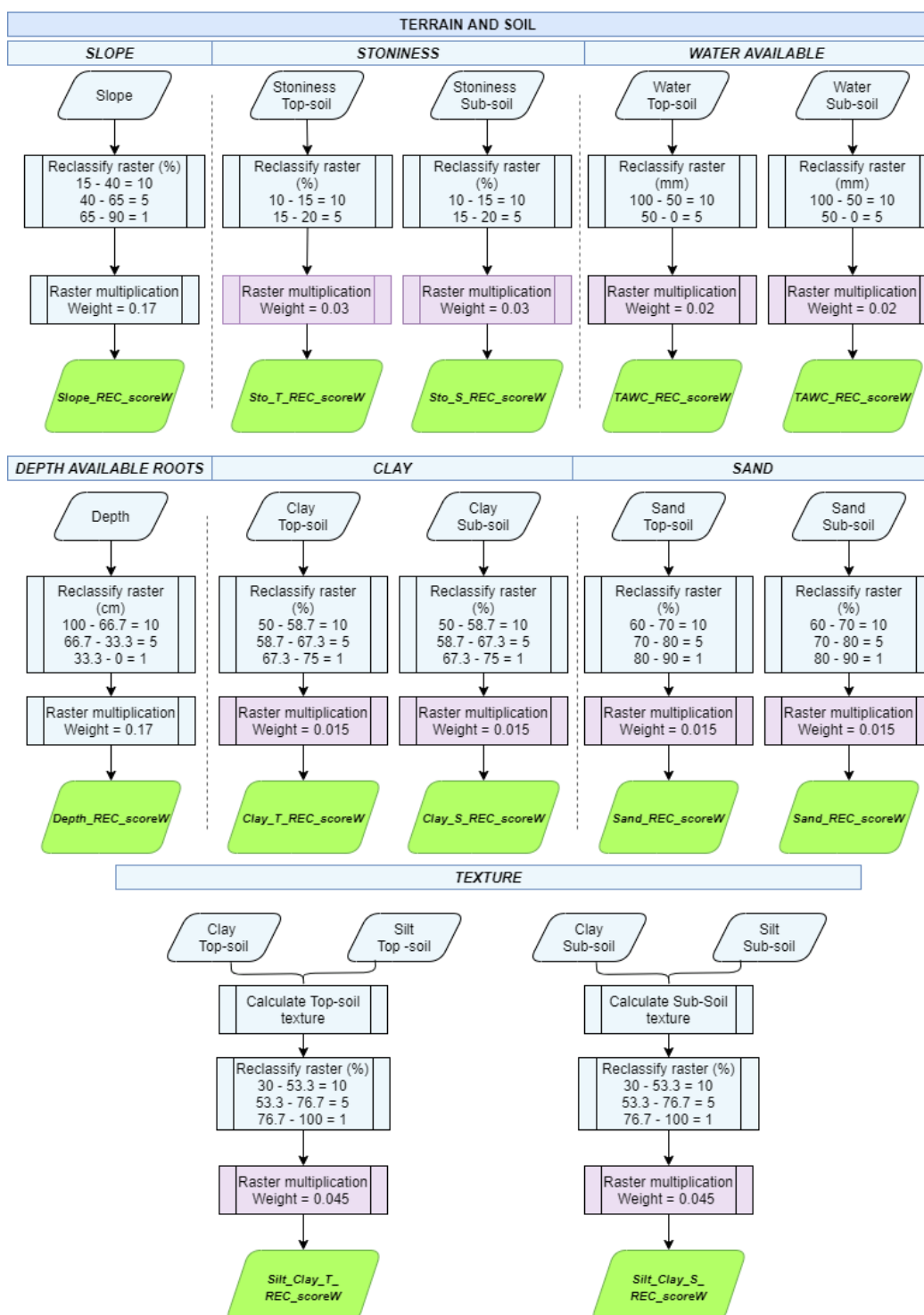


Figure 38: Workflow design for the computation of terrain and soil indicators. Source: personal compilation of Jesús Torralba Pérez.



9.2.2 Workflow design to produce intermediate soft layer related to “Sustainability”

This workflow is designed to produce the Intermediate layer: “Sodicity”, “Contamination”, “Natural toxicity”, “Acidity”, “Erosion”, “Dryness” and “Flooding” which has been used during the second phase of the proposed methodology (Figure 39).

“Sodicity”, “Contamination” and “Natural toxicity” these variables were obtained from the harmonized dataset of derived soil properties for the world (WISE30sec) downloaded from the ISRIC — World Soil Information. “Acidity” will be based on the Soil pH in Europe layer and “Erosion” will be based on the soil erosion by water (RUSLE2015) layer. Both were downloaded from the European Commission Joint Research Centre (JRC) and more specifically from the European Soil Data Center (ESDAC). “Flooding” has been based on the Global Surface Water (Pekel, Cottam, Gorelick, & Belward, 2016) that was downloaded from the European Commission Joint Research Centre (JRC). “Dryness” has been based on the TerraClimate, a high-resolution global dataset of monthly climate and climatic water balance from 1958–2015 (Abatzoglou, Dobrowski, Parks, & Hegewisch, 2018) for the year 2018 and available at the Climatology Lab from the University of California Merced.

These intermediate raster layers were reclassified according to the thresholds defined in Table 23. More specifically, these intermediate layers contain 4 classes: 0, areas outside the reference thresholds; 1, unsuitable lands; 5, potential MLs; and 10, MLs. Once the reclassification of the layers is done, each one is multiplied with its corresponding weight.

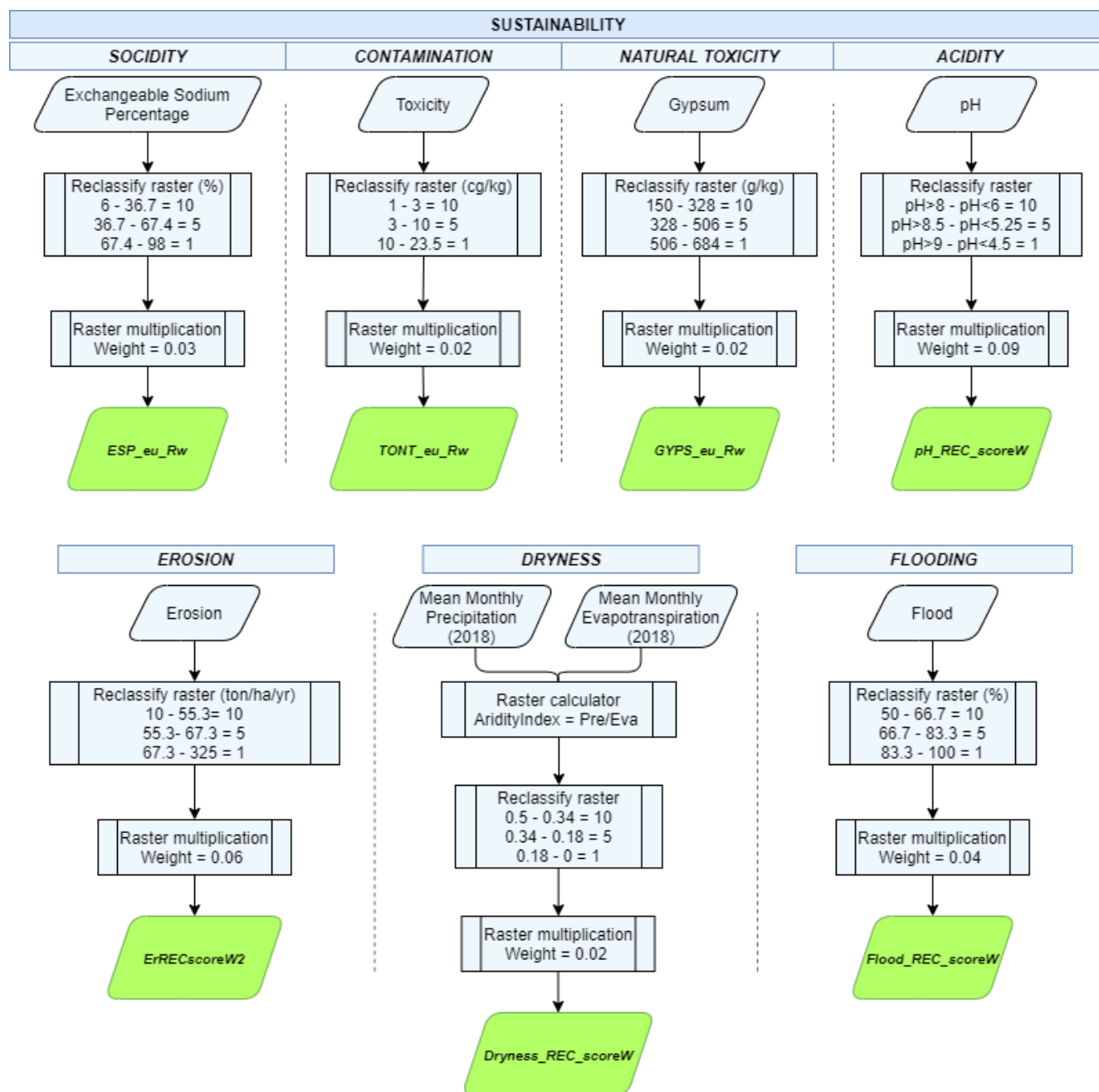


Figure 39: Workflow design for the computation of sustainability indicators. Source: personal compilation of Jesús Torralba Pérez.

9.2.3 Workflow to produce intermediate soft layer related to “Productivity”

This workflow is designed to produce the Intermediate layer: “Soil Organic Matter”, “Productivity” and “Cation Exchange Capacity” which has been used during the second phase of the proposed methodology (Figure 40).

“Soil Organic Matter” for the top-soil and sub-soil has been based on European Soil Database Derived data layer and “Productivity” will be calculated for the forest and grassland with the

Soil Biomass Productivity maps of grasslands and pasture, of croplands and of forest areas in the European Union (EU27) layer. Both datasets were downloaded from the European Commission Joint Research Centre (JRC) and more specifically from the European Soil Data Center (ESDAC). "Cation Exchange Capacity" has been based on the SoilGrids250m 2.0 Cation exchange capacity at pH7 layer was downloaded from the ISRIC — World Soil Information.

The final intermediate layers "Productivity" and "Cation Exchange Capacity" contain 4 classes: 0, areas outside the reference thresholds; 1, unsuitable lands; 5, potential MLs; and 10, MLs. The intermediate layer "Soil Organic Matter" contains 3 classes: 0, areas outside the reference thresholds; 5, potential MLs; and 10, MLs. Once the reclassification of the layers is done, each one is multiplied with its corresponding weight. For the layers "Soil Organic Matter" and "Productivity" where exists information about the topsoil and subsoil, the initial weight defined in Table 24 was rectified as indicated above.

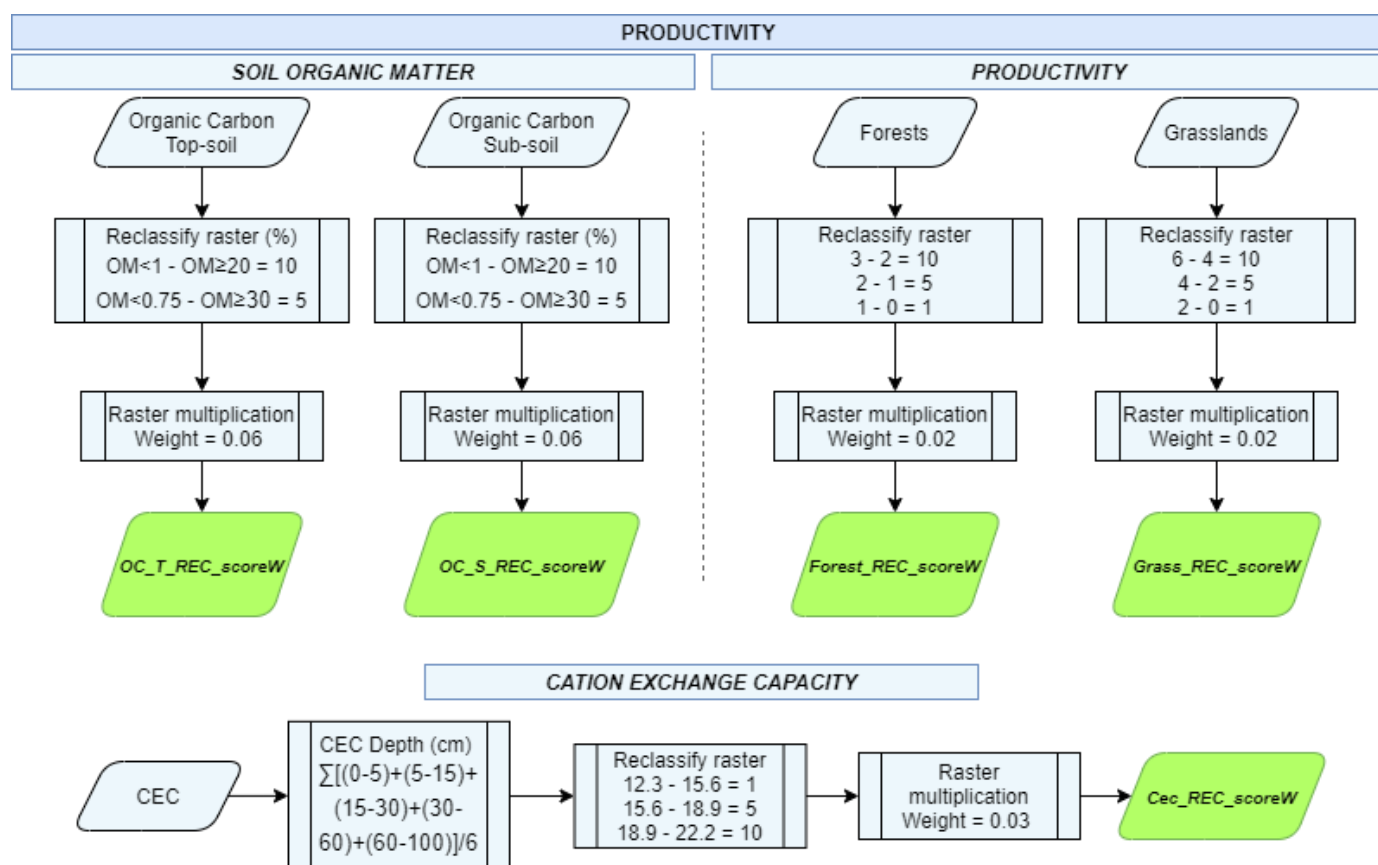


Figure 40: Workflow design for the computation of productivity indicators. Source: personal compilation of Jesús Torralba Pérez.



9.3 Workflow design to produce final layer “ML_FINAL”

This workflow is designed to produce the final layer “ML_FINAL” after the implementation of the additional marginality indexes and the final classification of the remaining areas into three classes: “Marginal lands with high plantation suitability”, “Marginal lands with low plantation suitability” and “Potentially unsuitable lands”. More specifically, the previously produced “ML_HARD” raster layer after the finalization of the first phase of the proposed methodology, is now combined with all the remaining marginality indexes containing additional “soft” thresholds (Figure 41).

It should be noted that the second phase of the proposed methodology is prone to change. Thus, additional marginality indexes can also be added during the finalization of the proposed methodology. Nevertheless, the design and implementation of the proposed methodology was structured in a way that changes that may occur will not affect the already implemented workflows.

The final combination flow of all layers can be done in Google Earth Engine. To do this, all layers must be uploaded as assets created in ArcGIS Pro. Finally, special consideration should be taken regarding the spatial resolution of each produced layer. Thus, prior to the implementation of any calculations, all the intermediate layers need resampling according to the intermediate layer “ML_HARD”. The final “ML_FINAL” layer should be a raster layer with 10 m spatial resolution.

To carry out the subdivision in the three types of marginal lands (“Marginal lands with high plantation suitability”, “Marginal lands with low plantation suitability”, “Potentially unsuitable lands”) three methodological approaches were developed:

- a) The first is to calculate the maximum and minimum value obtained in the MLs layer and divide the difference by three. This will obtain three ranges of equal magnitude.
- b) The second methodological approach is to calculate the 25th and 75th percentiles of the values obtained in the MLs layer. This approach would penalize the “Marginal lands with high plantation suitability” and the “Potentially unsuitable lands” and increasing the number of pixels belonging to the middle layer “Marginal lands with low plantation suitability”.
- c) The third one is to calculate the 33rd and 66th percentile of the values obtained in the layer MLs. In this case, the total number of pixels is divided equally into the three categories. But with the difference with respect to the a) method that in this case the thresholds are established with the maximum number of pixels and not with the maximum and minimum values.



Figure 41: Workflow design for the computation of final MLs. Source: personal compilation of Jesús Torralba Pérez.

10. IMPLEMENTATION OF WORKFLOWS TO THE GIS

The workflows developed were implemented in the GIS using the model builder and the spatial analysis toolbox of the ArcGIS Pro software. In some cases, it was necessary to use Google Earth Engine to process all layers at the European level due to layer accessibility or hardware limitations. Each workflow was programmed independently so that it could be modified at any time. It should be noted that the developed workflows were tested independently for each test site using only the raster data set representing the selected area. On the other hand, the procedures and the application of the proposed methodology were automated at the European level.

10.1 Automatization of procedures and tools description for hard layers

The following indicative models applied in the ArcGIS pro model builder were developed for the elaboration of the hard layers. The functions of ArcGIS are explained as they appear.

Figure 42 shows the workflow for "other land cover" in the ArcGIS Pro model builder.

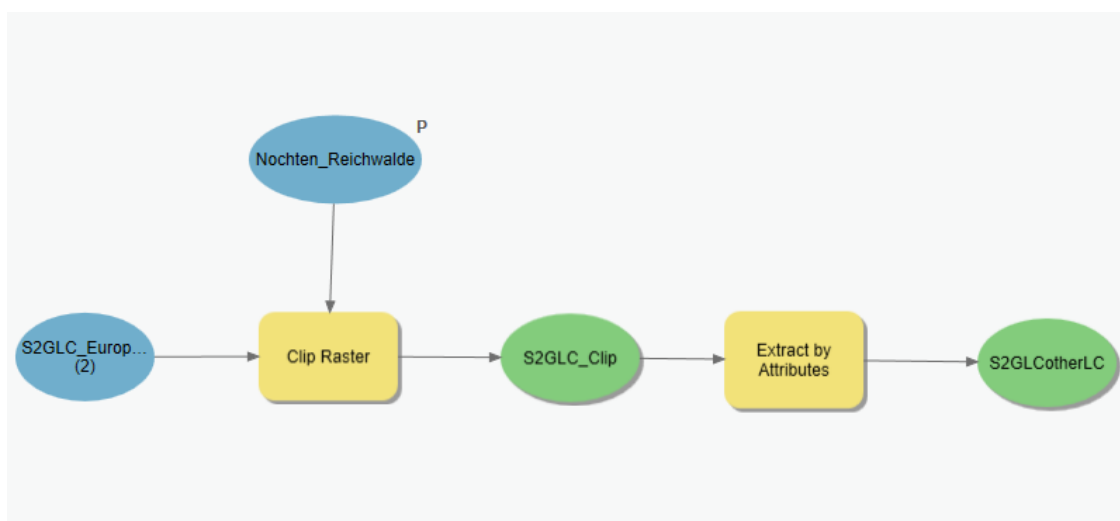


Figure 42: Implementation of workflow for the delineation of “other land cover” in ArcGIS Pro model builder.

Where:

Clip raster: this tool allows you to extract a portion of a raster dataset based on a template extent. The clip output includes any pixels that intersect the template extent. An existing raster or vector layer can be used as the clip extent.

Extract by Attributes: this tool extracts the cells of a raster based on a logical query.

Figure 43 shows the workflow for "forest areas" in the ArcGIS Pro model builder.

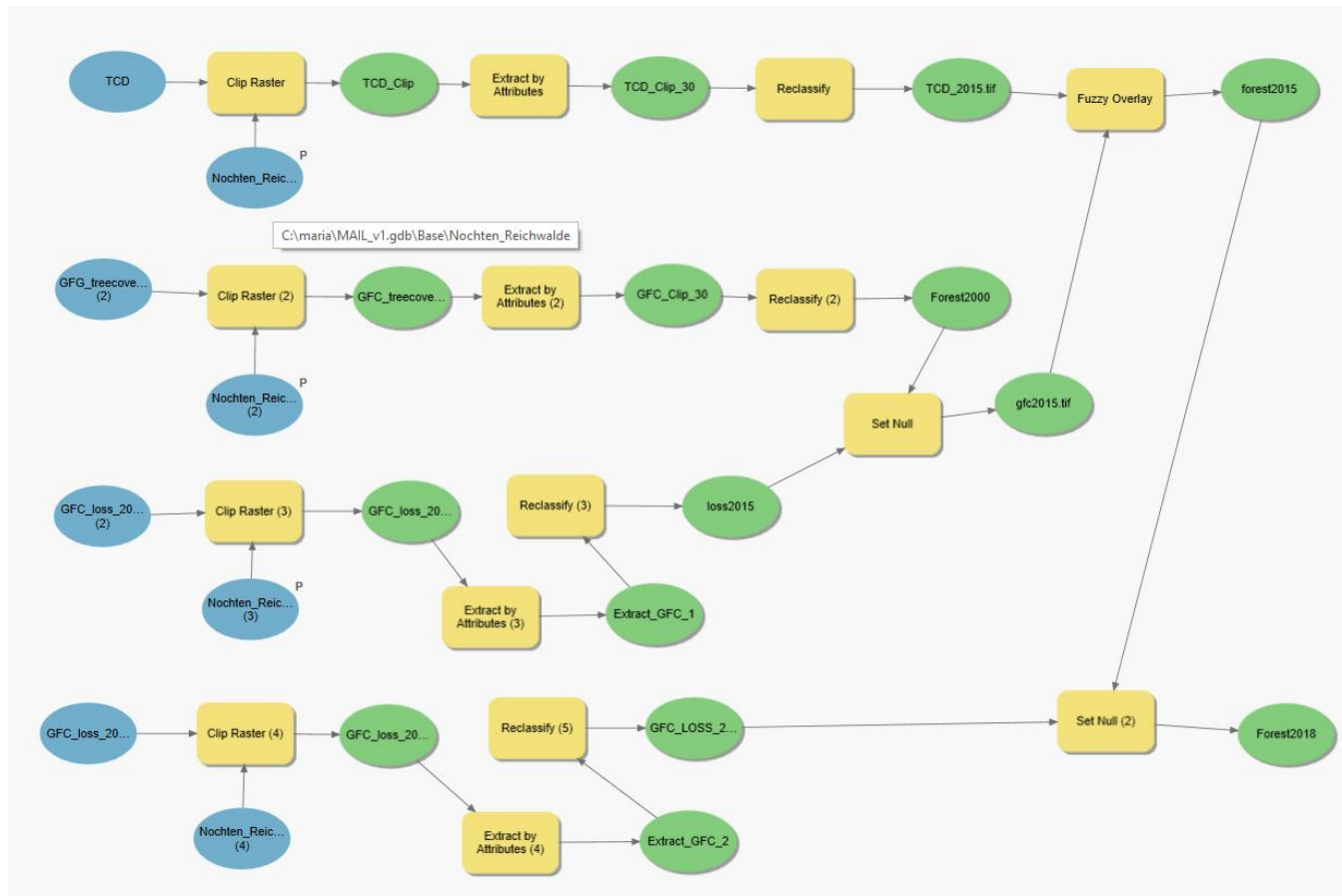


Figure 43: Implementation of workflow for the delineation of “Forest areas” in ArcGIS Pro model builder.

Where:

Reclassify: this tool Reclassifies or changes the values in a raster. When using the Reclassify tool as part of a model: i) If the input to the tool is derived data from a tool that is not already run, the remap parameter in the Reclassify tool will be empty until the preceding tool is run, and the model is validated. To avoid this, always run preceding tools before connecting their output variables as input to the Reclassify tool. Alternatively, you can create a custom reclassification table by adding entries. ii) If exposing the reclassification table as a model parameter, the reclass field must be exposed as a variable; however, it does not need to be set as a model parameter. If the field is not exposed as a variable, the classify and unique values buttons will be disabled in the model tool dialog box.

Fuzzy overlay: this tool allows the analysis of the possibility of a phenomenon belonging to multiple sets in a multicriteria overlay analysis. Not only does Fuzzy Overlay determine what sets the phenomenon is possibly a member of, it also analyzes the relationships between the membership of the multiple sets. The Overlay type lists the methods available to combine the data based on set theory analysis. Each method allows the exploration of the membership of

each cell belonging to various input criteria. The available methods are fuzzy And, fuzzy Or, fuzzy Product, fuzzy Sum, and fuzzy Gamma. Each approach provides a different aspect of each cell's membership to the multiple input criteria.

Set null: sets identified cell locations to NoData based on specified criteria. It returns NoData if a conditional evaluation is true, and returns the value specified by another raster if it is false.

Figure 44 shows the workflow for "croplands" in the ArcGIS Pro model builder.

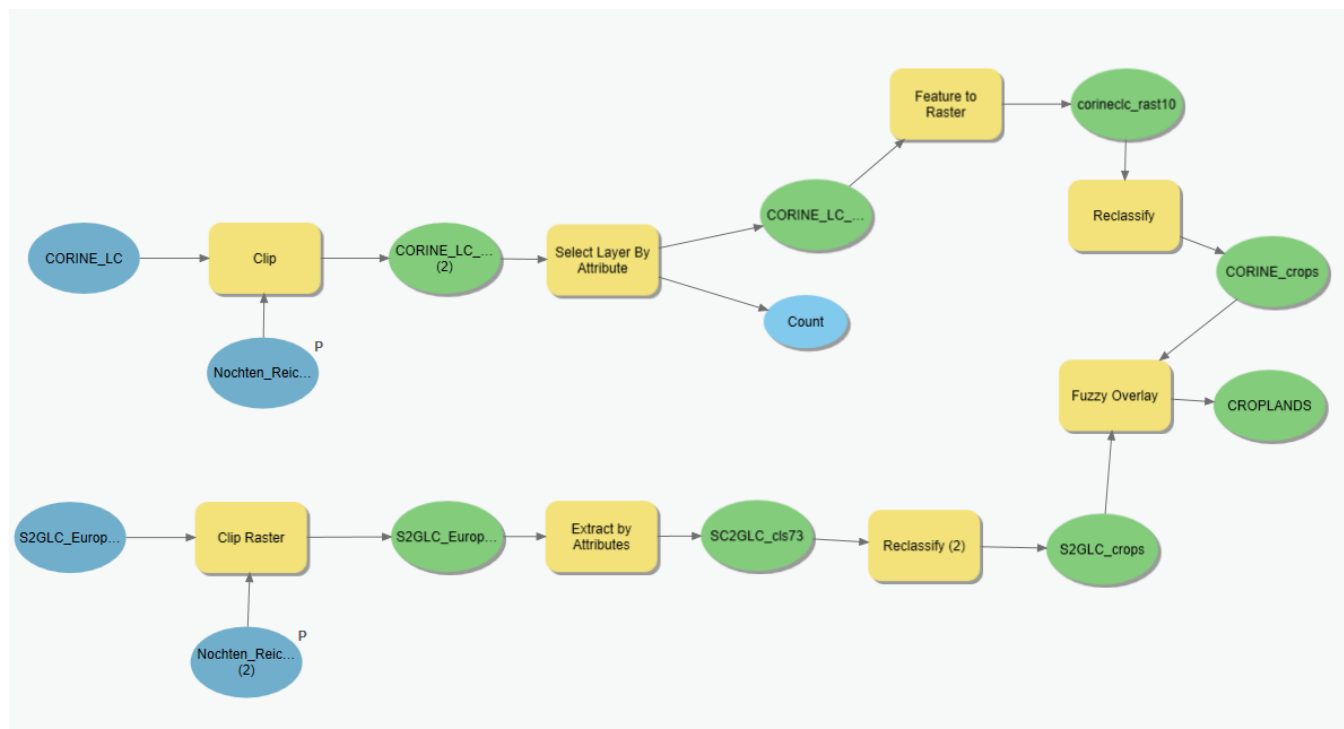


Figure 44: Implementation of workflow for the delineation of “Croplands” in ArcGIS Pro model builder.

Where:

Select layer by attribute: adds, updates, or removes a selection based on an attribute query.

Feature to raster: converts features to a raster dataset. Any feature class (geodatabase, shapefile, or coverage) containing point, line, or polygon features can be converted to a raster dataset.

Figure 45 shows the workflow for "changed areas" in the ArcGIS Pro model builder.

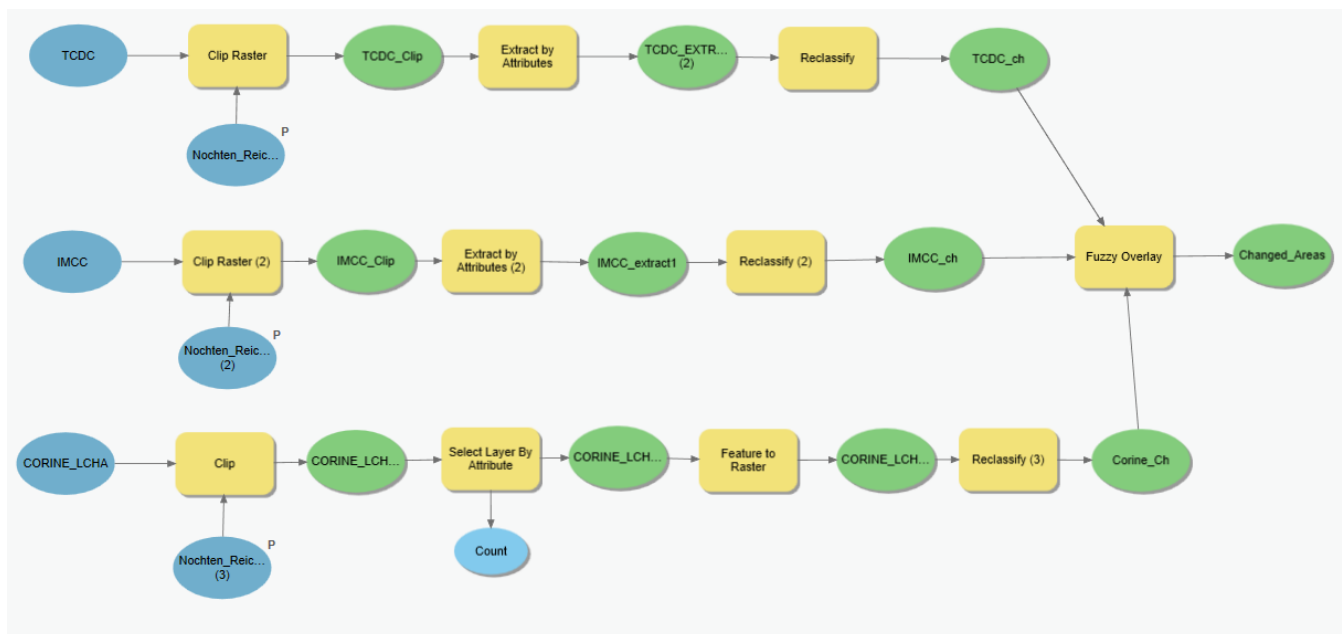


Figure 45: Implementation of workflow for the delineation of “Changed Areas” in ArcGIS Pro model builder.

Figure 46 shows the workflow for "impervious areas" in the ArcGIS Pro model builder.

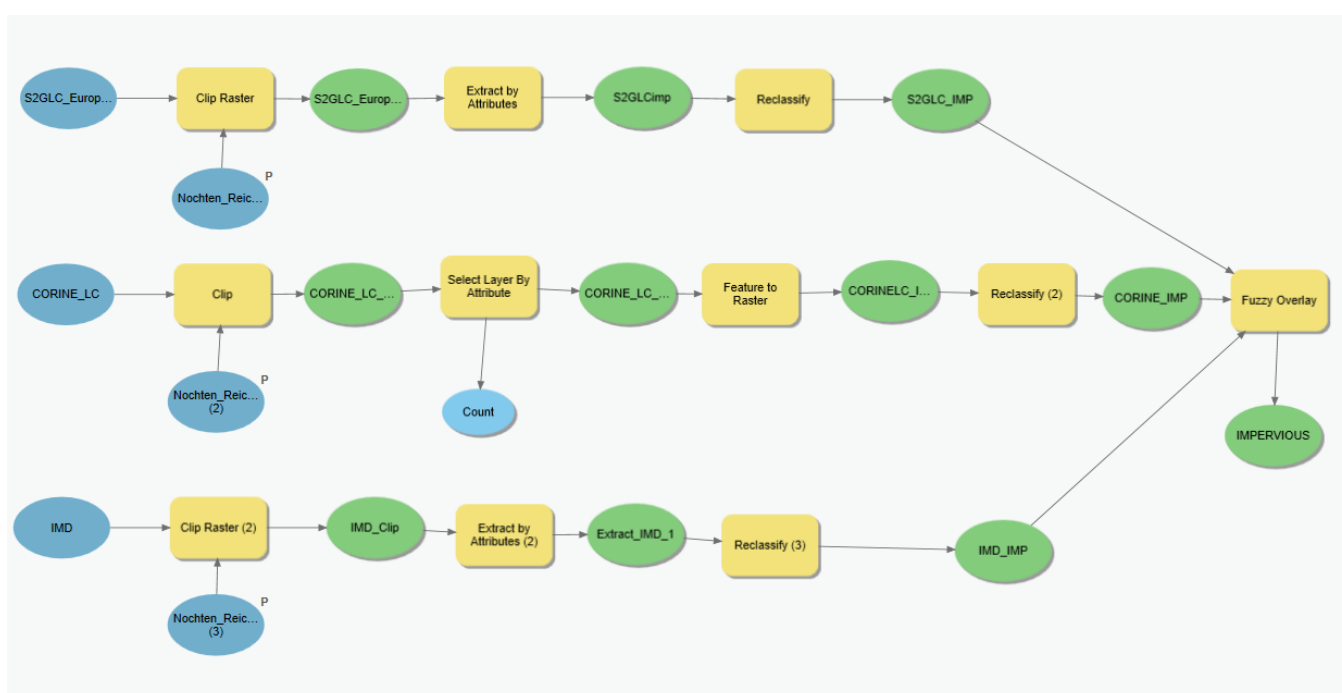


Figure 46: Implementation of workflow for the delineation of “Impervious Areas” in ArcGIS Pro model builder.

To elaborate the final layer “ML_HARD” with the hard layers it was necessary to implement a workflow in two sequences (Figure 47 and Figure 50).

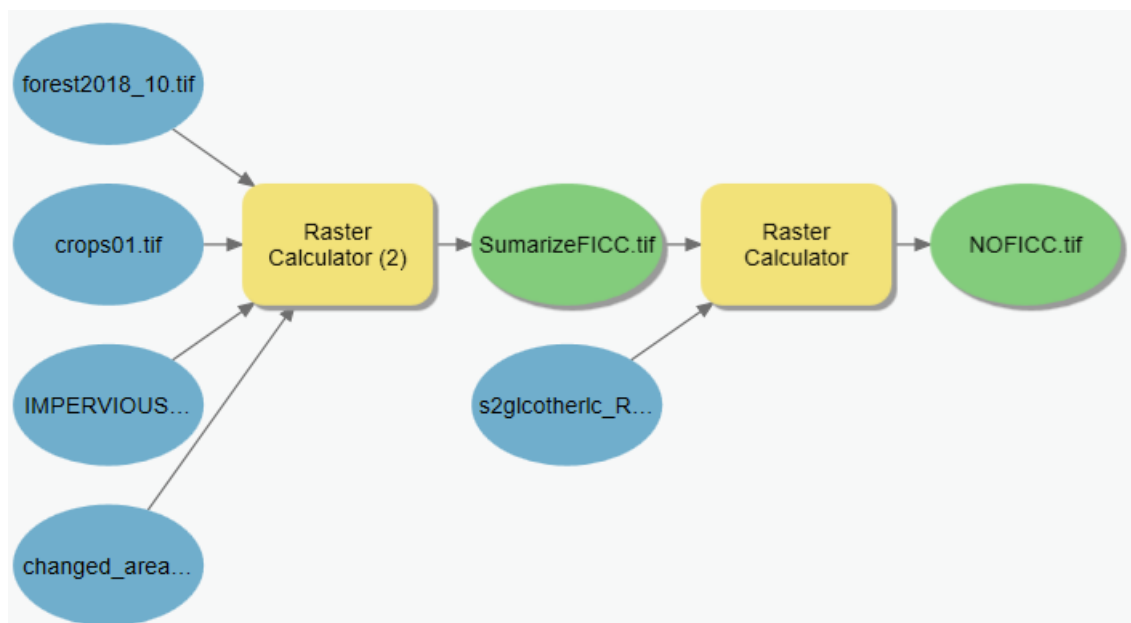


Figure 47: First step for the implementation of the workflow to produce the final “ML_HARD”.

First the layers "Forest", "Croplands", "Impervious" and "Changed areas" were reclassified to values 0 and 1, where the values 1 represented the value of the layer under study. For example, for the “Forest” layer, pixels with a value of 1 represent areas of forest and 0 represent areas of non-forest, i.e., pixels with a value of 0 are the pixels that will potentially form the “Marginal Land” layer. Once the layers have been reclassified, the sum of all of them was done with the raster calculator as shown in Figure 48. The result is a raster named “SumarizedFICC.tif”.

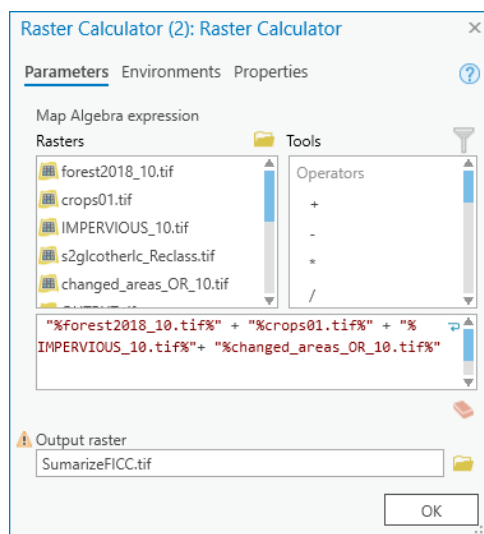


Figure 48: Summarize of "Forest", "Croplands", "Impervious" & "Changed areas" layers.

This way, in the SumarizeFICC.tif layer we will have values from 0 to 4, where 1 means that one of the 4 layers is present, 2 means that 2 of the 4 layers are present, 3 means that 3 of the 4 layers are present, 4 means that all 4 layers are present, and 0 would be the value corresponding to marginal land. Then, using the raster calculator tool and the use of conditionals as shown in Figure 49, we remove from the base layer "S2GLCotherLC" (which will serve as a processing extension) those pixels with a value greater than or equal to 1 in the "SumarizeFICC.tif" layer. At the same time with this tool, the final values are reclassified, resulting in a NOFICC.tif (No - Forest, Impervious Croplands and Changed areas) layer where the value zero corresponds to those pixels of the layers "Forest", "Croplands", "Impervious" and "Changed areas" and the pixels with value 1 will be those corresponding to Marginal Lands.

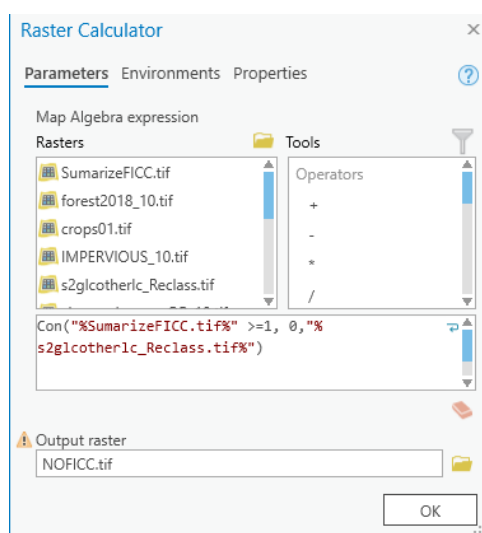


Figure 49: Selection and reclassification by conditionals in ArcGIS pro raster calculator tool.

The resulting layer "NOFICC.tif" is then subtracted from the protected areas as shown in Figure 50. That step was worked at the end since it was necessary to convert the "protected areas" layer from vector to raster, which process required a lot of processing time and hardware because of the number of vertices of the vector layer for all of Europe.

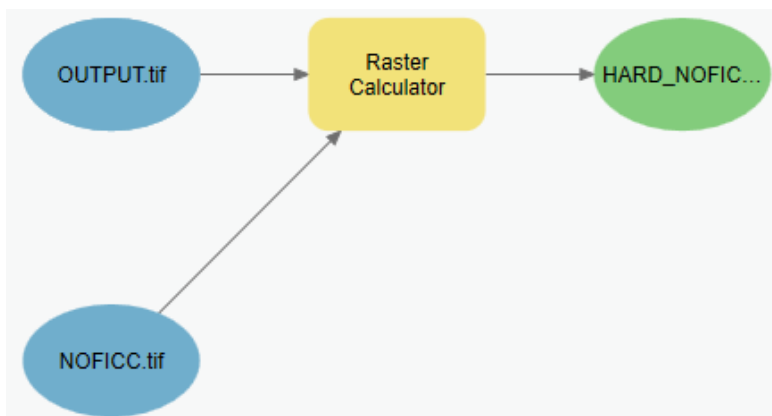


Figure 50: Second step for the implementation of the workflow to produce the final “ML_HARD”.

Finally, the “ML_HARD” layer is obtained with a spatial resolution of 10 by 10 meters, divided into tiles and in raster format with values of 1 and 0, where the value 1 represents the marginal areas. This layer already represents the marginal areas, and it would be only necessary to calculate the soft layers to be able to subdivide the layer according to its marginality.

10.2 Automatization of procedures and tools description for soft layers

The following indicative models applied in the ArcGIS pro model builder were developed for the elaboration of the soft layers. The functions of ArcGIS pro used were 2: i) *Reclassify*: To convert the original values to scores 10 (Marginal lands with high plantation suitability), 5 (Marginal lands with low plantation suitability), 1 (Potentially unsuitable lands), and 0 (Out of defined ranges); ii) *Raster calculator*: To multiply the reclassified rasters by the weights determined in section 6.5. In the case that the layers have been processed with GEE the functions used will be specified in the corresponding layer.

Figure 51 shows the workflow for "Depth available roots" in the ArcGIS Pro model builder. The ranges for the reclassification of this layer in centimeters of depth were: 0-33.3 = 1, 33.3-66.7 = 5, 66.7-100 = 10 and >100 = 0. The weight for this layer was 0.17, resulting in a final layer with 4 values, 0.17, 0.85, 1.7 and 0.

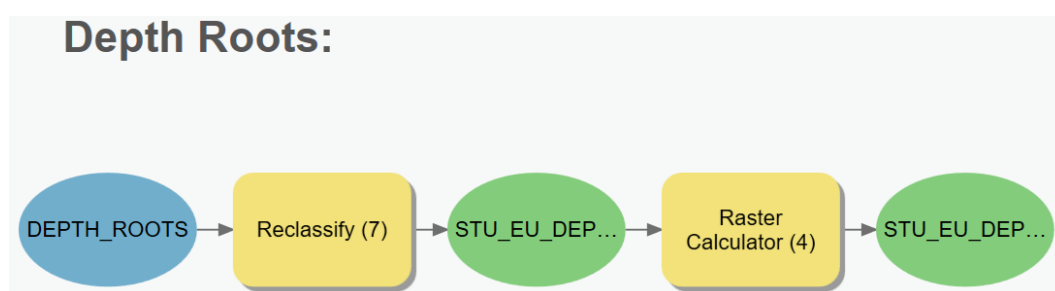


Figure 51: Implementation of workflow for the delineation of “Depth available roots” in ArcGIS Pro model builder.

Figure 52 shows the workflow for "texture" in the ArcGIS Pro model builder. This layer is processed independently for topsoil and subsoil. First, the layers of silt and clay are related according to the formula of Elbersen et al. (2018) where $\text{texture} = \text{silt}(\%) + (2 * \text{clay}(\%))$. Then, the ranges for the reclassification of this layer in percentage were: $100-76.7 = 1$, $76.7-53.3 = 5$, $53.3-30 = 10$ and $<30 = 0$. The weight for this layer was 0.09, but since there are two layers the weight is divided into 2, resulting in a final weight of 0.045 for each layer. The final rasters have 4 values: 0.045, 0.225, 0.45 and 0.

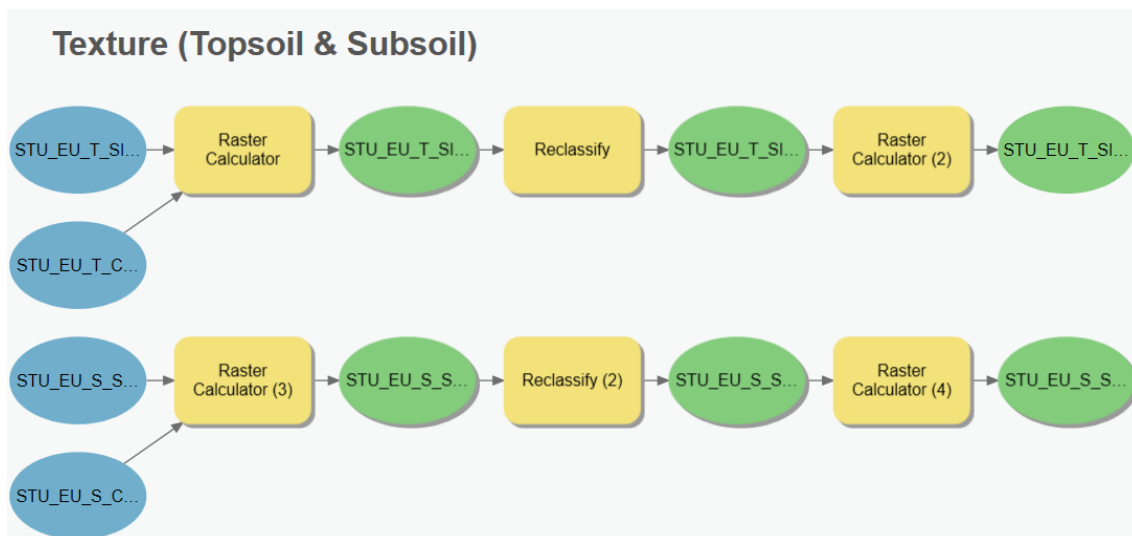


Figure 52: Implementation of workflow for the delineation of “texture” in ArcGIS Pro model builder.

Figure 53 shows the workflow for "stoniness" in the ArcGIS Pro model builder. This layer is processed independently for topsoil and subsoil. The ranges for the reclassification of this layer in percentage of stoniness were: $20-15 = 5$, $15-10 = 10$, $<10 = 0$ and $>20 = 0$. The weight for this layer was 0.06, but since there are two layers the weight is divided into 2, resulting in a final weight of 0.03 for each layer. The final rasters have 3 values: 0.15, 0.3 and 0.

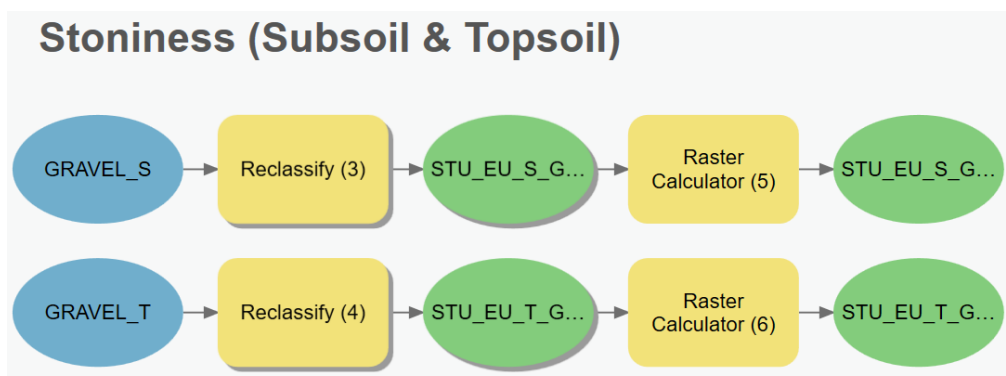


Figure 53: Implementation of workflow for the delineation of “stoniness” in ArcGIS Pro model builder.

Figure 54 shows the workflow for "Water capacity" in the ArcGIS Pro model builder. This layer is processed independently for topsoil and subsoil. The ranges for the reclassification of this layer in mm were: 0-50 = 5, 50-100 = 10, >100 = 0. The weight for this layer was 0.04, but since there are two layers the weight is divided into 2, resulting in a final weight of 0.02 for each layer. The final rasters have 3 values: 0.1, 0.2 and 0.

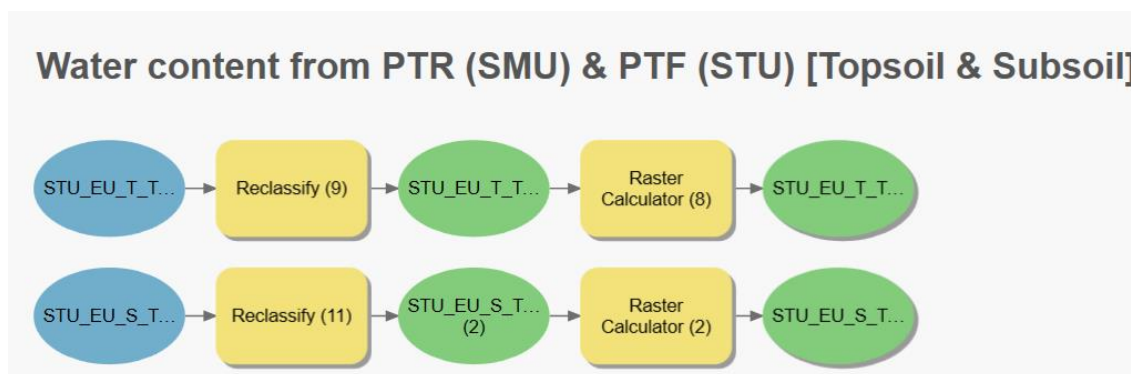


Figure 54: Implementation of workflow for the delineation of "Water capacity" in ArcGIS Pro model builder.

Figure 55 shows the workflow for "clay" in the ArcGIS Pro model builder. This layer is processed independently for topsoil and subsoil. The ranges for the reclassification of this layer in percentage of clay were: 76-67.3 = 1, 67.3-58.7 = 5, 58.7-50 = 10, <50 = 0. The weight for this layer was 0.03, but since there are two layers the weight is divided in 2, resulting in a final weight of 0.015 for each layer. The final rasters have 4 values: 0.015, 0.075, 0.15 and 0.

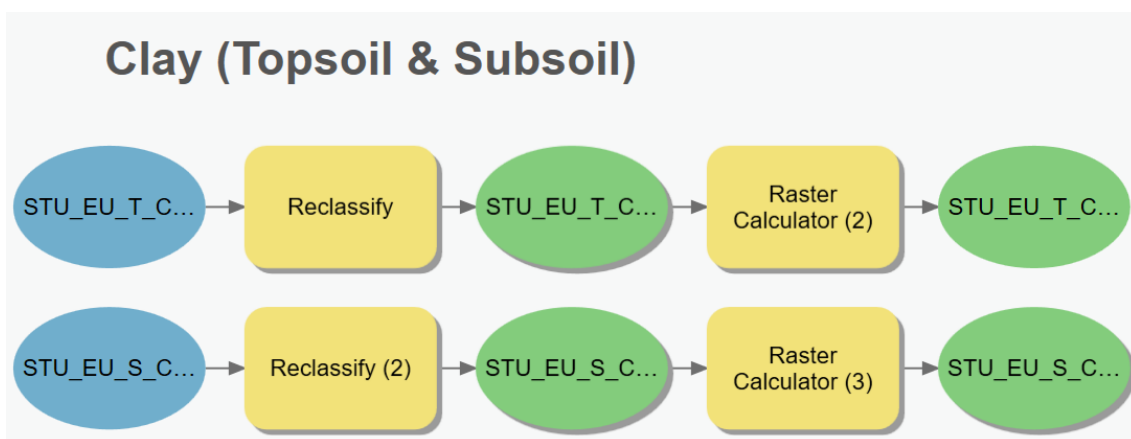


Figure 55: Implementation of workflow for the delineation of "clay" in ArcGIS Pro model builder.

Figure 56 shows the workflow for "sand" in the ArcGIS Pro model builder. This layer is processed independently for topsoil and subsoil. The ranges for the reclassification of this layer in percentage of sand were: 90-80 = 1, 80-70 = 5, 70-60 = 10, <60 = 0. The weight for this

layer was 0.03, but since there are two layers the weight is divided in 2, resulting in a final weight of 0.015 for each layer. The final rasters have 4 values: 0.015, 0.075, 0.15 and 0.

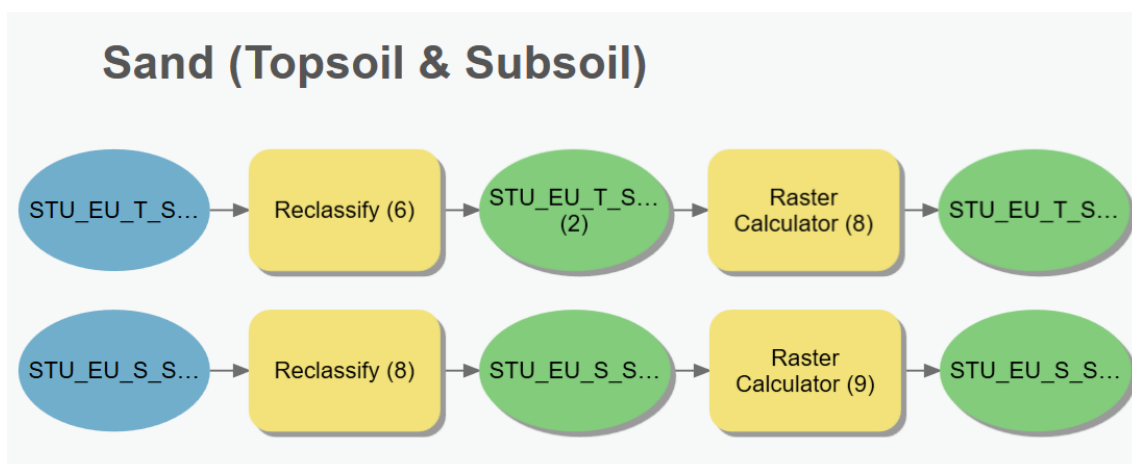


Figure 56: Implementation of workflow for the delineation of “sand” in ArcGIS Pro model builder.

Figure 57 shows the workflow for "acidity (pH)" in the ArcGIS Pro model builder. The ranges for the reclassification of this layer in pH values were: $\text{pH} > 9 - \text{pH} < 4.5 = 1$, $\text{pH} > 8.5 - \text{pH} < 5.25 = 5$, $\text{pH} > 8 - \text{pH} < 6 = 1$ and $6 < \text{pH} < 8 = 0$. The weight for this layer was 0.09, but since there are two layers the weight is divided into 2, resulting in a final weight of 0.045 for each layer. The final rasters have 4 values: 0.09, 0.45, 0.9 and 0.

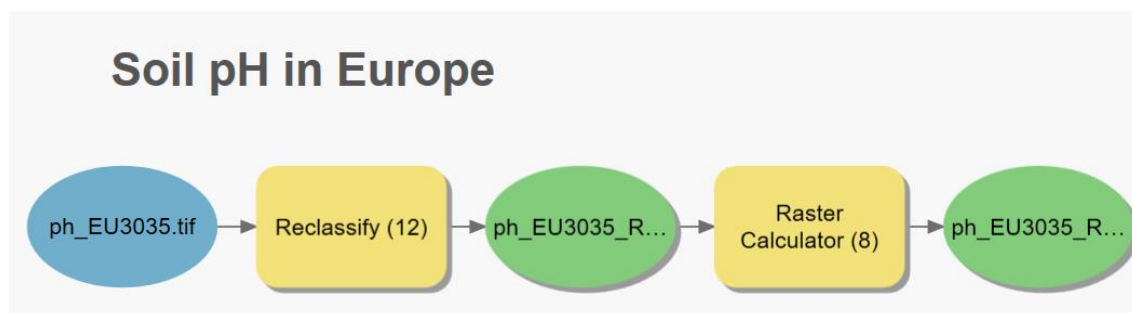


Figure 57: Implementation of workflow for the delineation of “acidity (pH)” in ArcGIS Pro model builder.

Figure 58 shows the workflow for "erosion" in the ArcGIS Pro model builder. The ranges for the reclassification of this layer in tons per hectare and year (ton/ha/year) were $325 - 283.4 = 1$, $283.4 - 241.7 = 5$, $241.7 - 200 = 10$ and $< 200 = 0$. The weight for this layer was 0.06, resulting in a final layer with 4 values: 0.06, 0.3, 0.6 and 0.

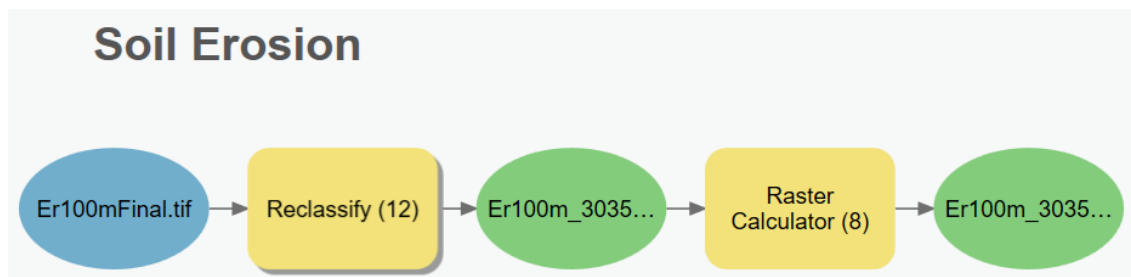


Figure 58: Implementation of workflow for the delineation of “erosion” in ArcGIS Pro model builder.

The variables layers "sodicity", "toxicity contamination" and "natural toxicity" will be obtained from the harmonized dataset of derived soil properties for the world (WISE30sec). This is a raster layer with an associated attribute table, where each column in the table represents a soil indicator. The pixels in this raster can acquire the values of each of the columns. In this sense, 3 rasters were created, one for each study layer reclassified according to the following values:

- The ranges for the reclassification of "sodicity" layer whose values are in exchangeable sodium percentage were: $98-67.4 = 1$, $67.4-36.7 = 5$, $36.7-6 = 10$, $<6=0$ and $>100= 0$.
- The ranges for the reclassification of "toxicity contamination" layer whose values are total nitrogen in cg/kg were: $23.5-10 = 1$, $10-3 = 5$, $3-1 = 10$, $<1=0$ and $>23.5= 0$.
- The ranges for the reclassification of "natural toxicity" layer whose values are gypsum content in g/kg were: $684-506 = 1$, $506-328 = 5$, $328-150 = 10$, $<150=0$ and $>684= 0$.

Figure 59 shows the raster calculator tool for the WISE in the ArcGIS Pro model builder for the layers "sodicity", "toxicity contamination" and "natural toxicity". The weight for the "sodicity" and "toxicity contamination" layers was 0.03, resulting in two final layers with 4 values: 0.03, 0.15, 0.3, and 0. The weight for the "natural toxicity" layer was 0.02, resulting in a final layer with 4 values: 0.02, 0.01, 0.2, and 0.

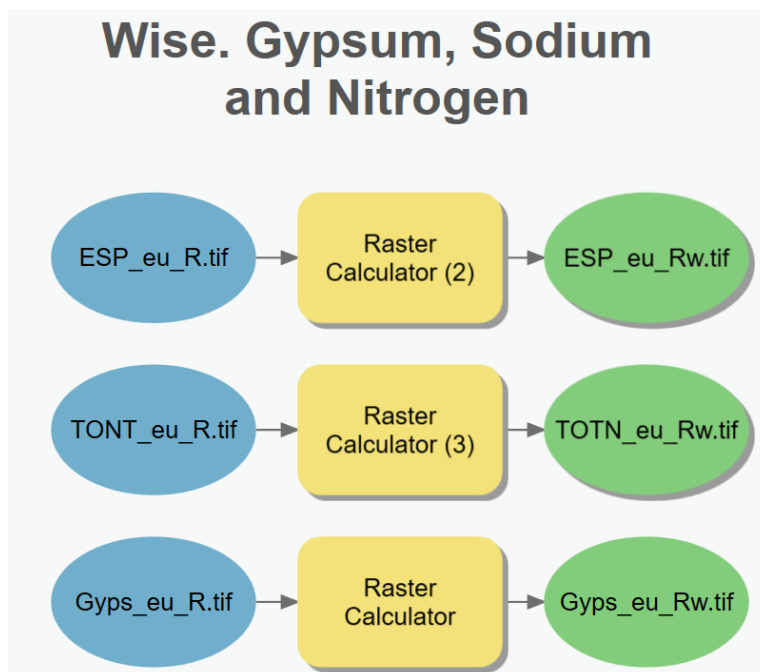


Figure 59: Implementation of workflow for the delineation of “sodicity”, “toxicity contamination” and “natural toxicity” in ArcGIS Pro model builder.

Figure 60 shows the workflow for "soil organic matter" in the ArcGIS Pro model builder. This layer is processed independently for topsoil and subsoil. The ranges for the reclassification of this layer in percentage of organic matter were: $OM \geq 30 - OM < 0.75 = 5$, $OM \geq 20 - OM < 1$, and $1 < OM < 20 = 0$. The weight for this layer was 0.06, but since there are two layers the weight is divided into 2, resulting in a final weight of 0.03 for each layer. The final rasters have 3 values: 0.15, 0.3 and 0.

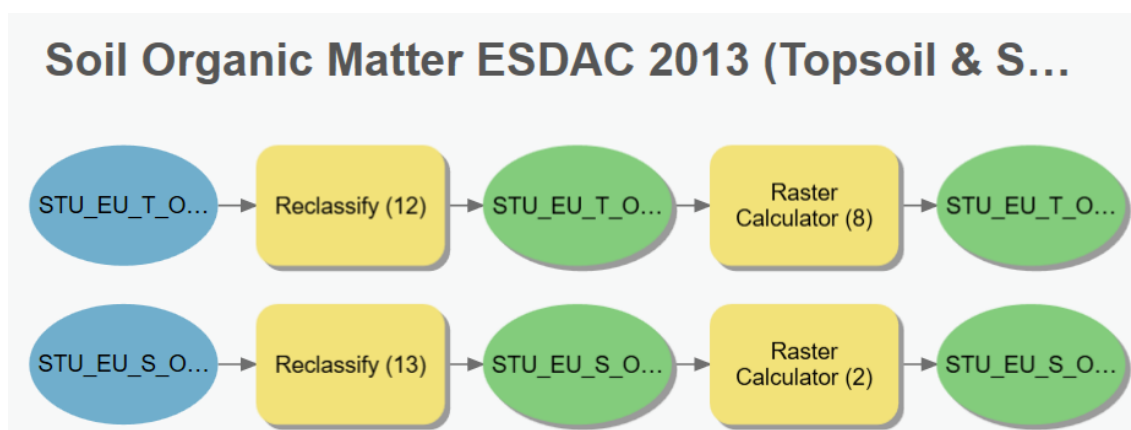


Figure 60: Implementation of workflow for the delineation of “soil organic matter” in ArcGIS Pro model builder.

Figure 61 shows the workflow for "productivity" in the ArcGIS Pro model builder. This layer is processed independently for grasslands and forests. The ranges for the reclassification of

grasslands layer in units were: 0-2 = 1, 2-4 = 5, 4-6 = 10 and >10=0, and the ranges for the reclassification of forest layer in units were: 0-1 = 1, 1-2 = 5, 2-3 = 10 and >3=0. The weight for these layers was 0.02, but since there are two layers the weight is divided into 2, resulting in a final weight of 0.01 for each layer. The final rasters have 4 values: 0.014, 0.07, 0.14 and 0.

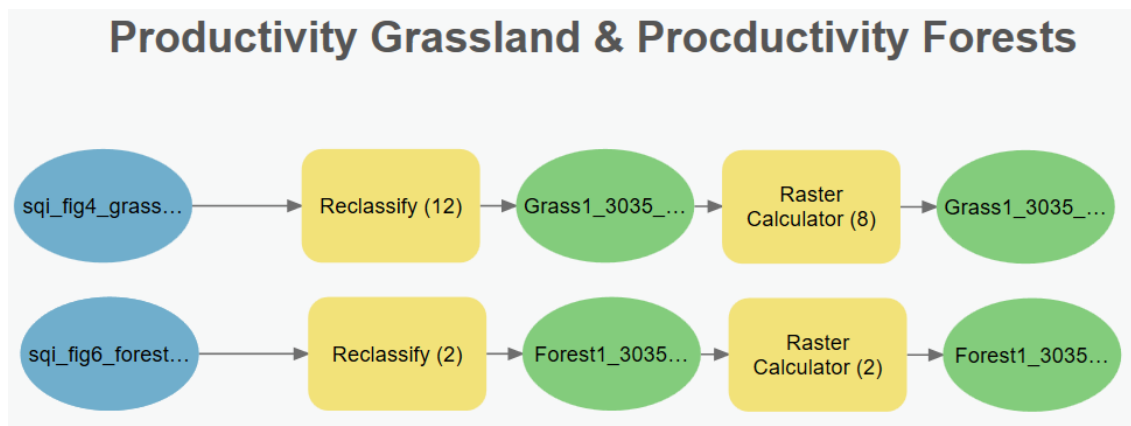


Figure 61: Implementation of workflow for the delineation of “productivity: grasslands & forest” in ArcGIS Pro model builder.

The indicators "Flood", "Cation Exchange Capacity (CEC)" and "Dryness" were processed in GEE. “Flood” (Figure 62) is based on the Global Surface Water (Pekel et al., 2016) from the European Commission Joint Research Centre (JRC) that is available in GEE. “Cation Exchange Capacity” is based on the SoilGrids250m 2.0 Cation exchange capacity at pH7 layer was downloaded from the ISRIC — World Soil Information. Access to this information is complex because it is based on a folder structure and small raster tiles, which makes it very difficult to download the information for the whole of Europe. This layer of ISRIC is available in GEE, so we work on this platform to save time and processing (Figure 63). “Dryness” is based on the TerraClimate, from the University of California Merced. This layer is also available in GEE which facilitates processing (Figure 64). Three main functions were used to manipulate these layers, in case more functions were used they are described in each workflow:

- `.where()` [from `ee.Image`]: It was used to make the classification. Performs conditional replacement of values. For each pixel in each band of 'input', if the corresponding pixel in 'test' is nonzero, output the corresponding pixel in value, otherwise output the input pixel. If at a given pixel, either test or value is masked, the input value is used. If the input is masked, nothing is done.
- `.multiply()` [from `ee.Number`]: It was used to multiply the reclassified image by its corresponding weight.

- `.reproject()` [from `ee.Image`]: Force an image to be computed in a given projection and resolution.

Flooding

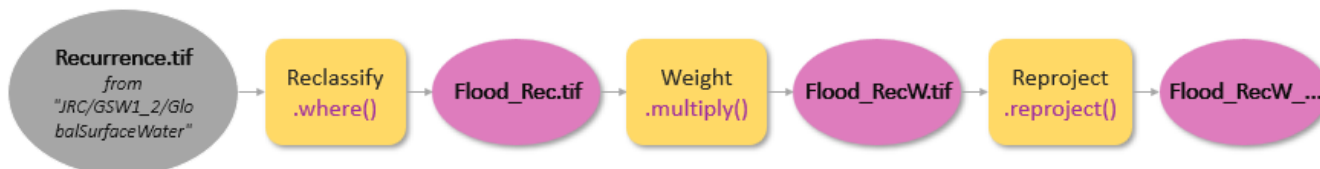


Figure 62: Implementation of workflow for the delineation of “Flooding” in GEE.

For the CEC layer, before applying the described functions, it was necessary to calculate the average for the 1-meter-deep soil profile since CEC measurements by depth ranges are available. The following function was used for this purpose:

- `ee.Reducer.mean()` [from `ee.Reducer`]: Returns a Reducer that computes the (weighted) arithmetic mean of its inputs.

Cation Exchange Capacity (CEC)

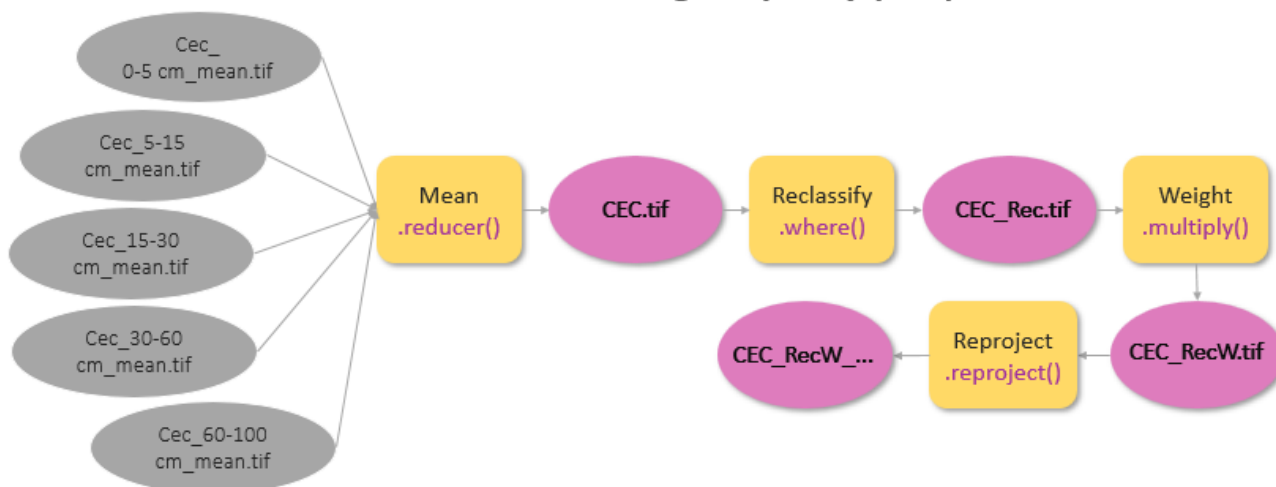


Figure 63: Implementation of workflow for the delineation of “Cation Exchange Capacity (CEC)” in GEE.

Four steps (Figure 64) were necessary to calculate the aridity index previous to the reclassification and the multiplication by its weight. First, the images were filtered only for the year 2018 with the `.filter` function, obtaining a set of 12 images. Then the precipitation and evapotranspiration bands were selected from the available bands with the `.select` function. Thirdly, the division between precipitation and evapotranspiration was made with the `.divide()`

function for each month. And finally, the average was calculated for each pixel for the whole year with the function `.mean()`. The functions that were not previously explained are described below.

- `ee.Filter.date()` [from `ee.filter`]: Filter a collection by date range.
- `.select()` [from `ee.Image`]: Selects bands from an image. Returns an image with the selected bands.
- `.divide()` [from `ee.Image`]: Divides the first value by the second, returning 0 for division by 0 for each matched pair of bands in `image1` and `image2`.
- `.mean()` [from `ee.ImageCollection`]: Reduces an image collection by calculating the mean of all values at each pixel across the stack of all matching bands. Bands are matched by name.

Dryness (Aridity Index)

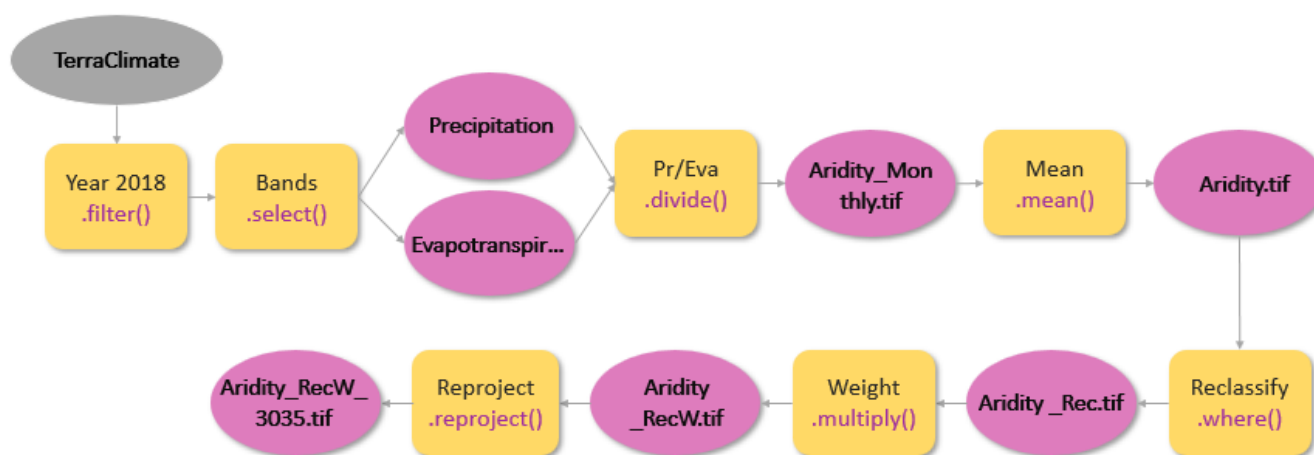


Figure 64: Implementation of workflow for the delineation of “Dryness or Aridity Index” in GEE.

10.3 Automatization of procedures and tools description for Final MLs

The final layer of MLs classified into three categories “Marginal lands with high plantation suitability”, “Marginal lands with low plantation suitability” and “Potentially unsuitable lands” was obtained with the use of GEE. For this purpose, first, all the “Soft” layers described in Table 44 and the “ML_hard ” layer were loaded as “Assets” (data set loaded in the GEE platform manually) in GEE.

The `.add()` function was used to perform the “soft” layers overlap. This function considers the snap of the pixels and allows to combine of the images generated for each indicator into a single image. Then the `.updateMask()` function was chosen to clip the “soft” raster with the “ML Hard”. In other words, the ML Hard layer establishes the limits of the MLs and the “soft” layer the values of the pixels. Once the clip was done, the three subdivisions of the values described

in section 9.1.7 were carried out. Table 45 shows the ranges of values obtained when using the three segmentation methods: a) computing the maximum and minimum and dividing the range of values by 3, b) computing the 25th and 75th percentile and setting these values as threshold limits, c) computing the 33rd and 66th percentile to keep the same number of pixels in each category.

Table 45: Methods to subdivide by type of marginality (“Marginal lands with high plantation suitability”, “Marginal lands with low plantation suitability” and “Potentially unsuitable lands”).

Method		a)		b)		c)	
		Max.	Min.	P25	P75	P33	p66
Value		6.862	0.120	1.390	2.890	1.579	2.516
		Thresholds					
1	<i>Marginal lands with high plantation suitability</i>	6.862	4.615	6.862	2.890	6.862	2.516
2	<i>Marginal lands with low plantation suitability</i>	4.615	2.367	2.890	1.390	2.516	1.579
3	<i>Potentially unsuitable lands</i>	2.367	0.120	1.390	0.120	1.579	0.120

Figure 65 shows the workflow in GEE.

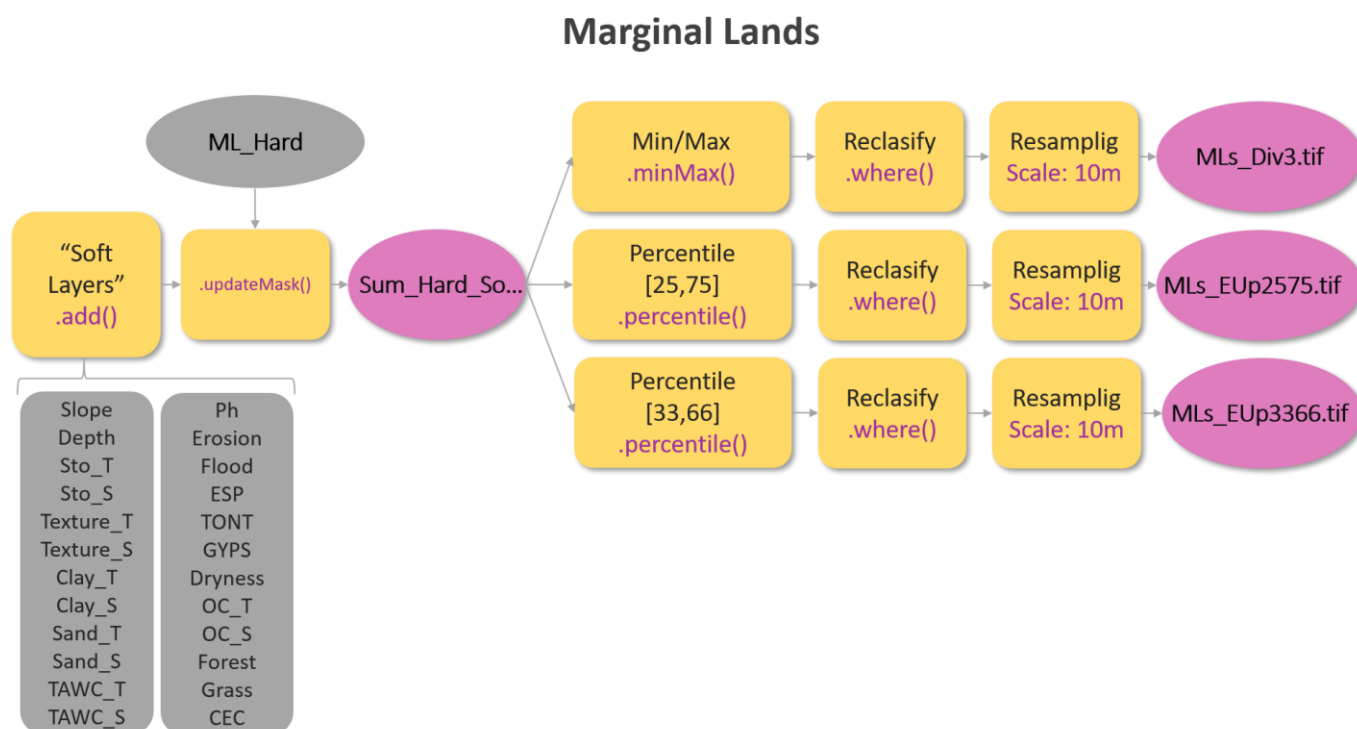


Figure 65: Implementation of workflow for the delineation of “MLs” in GEE.



where the GEE functions are:

- `ee.Reducer.mean()` [from `ee.Reducer`]: Returns a Reducer that computes the (weighted) arithmetic mean of its inputs.
- `.add()` [from `ee.Image`]: Adds the first value to the second for each matched pair of bands in `image1` and `image2`.
- `.updateMask()` [from `ee.Image`]: Updates an image's mask at all positions where the existing mask is not zero. The output image retains the metadata and footprint of the input image.
- `ee.Reducer.minMax()` [from `ee.Reducer`]: Returns a Reducer that computes the minimum and maximum of its inputs.
- `ee.Reducer.percentile()` [from `ee.Reducer`]: Create a reducer that will compute the specified percentiles, e.g. given `[0, 25, 75, 100]`.
- `.where()` [from `ee.Image`]: It was used to make the classification. Performs conditional replacement of values. For each pixel in each band of 'input', if the corresponding pixel in 'test' is nonzero, output the corresponding pixel in value, otherwise output the input pixel. If at a given pixel, either test or value is masked, the input value is used. If the input is masked, nothing is done.
- `.scale()` [from `Export.image.toDrive`]: Resolution in meters per pixel.



11. DEVELOPMENT OF GUIDELINES FOR THE USAGE OF INDICATORS SET AND GIS WORKFLOWS

This chapter describes recommendations for the use of GIS indicators and workflows for multilayer analysis of MLs in specific areas or the whole of Europe. These guidelines are recommendations that are intended to help users when selecting an ML indicator, a processing software, between tools within the same software and recommendations in specific processes. The recommendations can also help users to select and prioritize among some possible workflows.

11.1 General guidelines

- Ideally, the data layers will be obtained from free platforms. The EU Commission and the European Space Agency provide a large number of data sets concerning specific soil characteristics, as well as a large number of remote sensing products derived from the Copernicus Program.
- When the layers are requested from the European Commission, the purpose for which the information is requested must be indicated. For the case of the MAIL project an example of application could be: "This layer is one of the inputs in the project H2020 'Identifying Marginal Lands in Europe and strengthening their contribution potentialities in a CO2 sequestration strategy'".
- The formats for storage of geographic information will be preferably raster, more specifically in ASCII Grid (.asc), TIFF/GeoTIFF (.tif, .tiff, .tff Y .tffw), Raster Geodatabase (.gdb), ENVI (.hdr, .img, .dat, .bsq, etc.), ERDAS (.raw, .img, .ige, .GIS, etc.), LiDAR (.las, .laz, .sid, .view, etc.), SAR (.raw, .hgt), Terrain (.dem, .dt0, .dt1, .dt2, .hfz). In case there is no raster information will work with vector data such as *.shp, *.gpkg, *.csv, *.kml, *.kmz, *.gpx, *.json and *.geojson. The use of vector files for all Europe requires high computational performance.
- There are two types of software to manipulate the information layers: open-source and licensed. As open-source software, we recommend the use of QGIS and as licensed software, ArcGIS PRO. For the latter software, the MAIL project has the licenses enabled by the IABG partner in its offices in Dresden. Both software allow all the proposed workflows to be carried out. As an alternative to working with all the images at the European level is possible to use the free GEE platform but requires minimal knowledge of programming.
- For the storage of the intermediate and final layers, it is recommended to work with a storage cloud in which it is possible to share with all the members of the project. Within



the MAIL project, the partner AUTH has enabled an FTP for the storage and management of the information generated. Access must be granted by AUTH personnel.

- To keep the marginal land indicators updated, it is necessary to carry out a continuous analysis of the bibliographic references as well as a periodic update of the land use base layers.

11.2 Specific guidelines for the software and GIS workflows

- The workflows are independent for each layer so they can be modified for a layer without altering the rest of the results.
- It is recommended before automating the processes in GIS to define the objective of the analysis and design the workflow or process diagram. In this diagram, it is important to identify the process, know the input layers, the output layers, and the tools with which we will manipulate the layers.
- Before implementing the flows in "model builder" it is important to perform a desktop test. This means that before implementing in the model builder, a manual test should be performed with the tools and layers or a subset of the layers that we are going to use.
- The different workflows have been implemented in the ArcGIS model builder since a license is available.
- When the computational requirements exceeded the capacity of the personal computer, GEE was used, since this platform uses the GOOGLE servers to process, this improves the performance of the algorithms and reduces the processing time.
- It is important to know the tools and the different parameters of the tools before implementing them in the model builder.
- The following ArcGIS Pro tools have been mainly used in the MAIL project workflows: *"Clip raster", "Extract by Attributes", "Reclassify", "Fuzzy overlay", "Set null", "Select layer by attribute", "Feature to raster" and "Raster calculator"* which were described in chapters 10.1 and 10.2.
- The following GEE functions have been mainly used in the MAIL project workflows: *"ee.Reducer.mean()", "ee.Filter.date()", ".divide()", ".mean()", ".add()", ".updateMask()", "ee.Reducer.minMax()", "ee.Reducer.percentile()", ".where()", and ".scale()"* which were described in chapters 10.2 and 10.3
- Identify the variables, constants, and parameters.
- Small ArcGIS model builder flows are recommended. If you intend to make large models, it is recommended that you create sub-models.



- In the GIS tools, it is recommended to configure the environment parameters. These parameters are:
 - Output Coordinate System: Tools that honor the Output Coordinate System environment will create output geodatasets with the specified coordinate system. Processing (calculation of geometric relationships and modification of geometries) occurs in the same coordinate system as the output geodataset. This environment overrides the default coordinate system on output geodatasets created by geoprocessing tools (Esri, 2020).
 - Processing extent: Tools that honor the Extent environment will only process features or rasters that fall within the extent specified in this setting. The Extent environment setting defines the features or rasters that will be processed by a tool. It is useful when you need to process only a portion of a larger dataset (Esri, 2020).
 - Raster Analysis – Cell Size: Tools that honor the Cell size environment set the output raster cell size, or resolution, for the operation. The default output resolution is determined by the coarsest of the input raster datasets. when specifying a cell size finer than the input raster datasets. No new data is created; cells are interpolated using nearest neighbor resampling. The result is only as precise as the coarsest input (Esri, 2020).
 - Raster Analysis – Snap Raster: Tools that honor the Snap Raster environment will adjust the extent of output rasters so that they match the cell alignment of the specified snap raster. A snap raster is typically used when inputs to tools: (1) Have different cell alignments, (2) Have different cell alignments, and (3) Have different coordinate systems (Esri, 2020).
 - Raster Storage – Compression. The primary benefits of compressing data are that compressed data requires less storage space, and data display times will be quicker because there is less information to transmit. Data compression can be lossy (JPEG and JPEG 2000) or lossless (LZ77, PackBits, CCITT). Lossless compression means that the values of cells in the raster dataset are not changed or lost. You should choose lossless or no compression if the pixel values of the raster dataset will be used for analysis or deriving other data products. The amount of compression will depend on the data and compression quality. The more homogeneous the data, the higher the compression ratio. The lower the compression quality, the higher the compression ratio. Lossy compression normally results in higher compression ratios when compared to lossless compression (Esri, 2020).



The base layer to establish these environment parameters in the MAIL project was the S2GLC layer to maintain the same output characteristics for all layers.

11.3 Specific guidelines on databases

- It is recommended not to work from the database, instead to download a copy of the information to the personal computer.
- The files in the database are in compressed (.zip) format. By compressing the raster and vector files, the information is maintained, and the weight of the files is reduced, which speeds up the upload and download processes in the database.
- Folders have been created in the database with descriptive names to identify the layers within them.
- A folder has been created for the "Hard" layers and another one for the "Soft" layers. Within each folder, it is subdivided into different folders.
- Annex IV: Database structure describes in detail the structure of the database.

11.4 Specific guidelines for nomenclature, projections, and spatial resolution

- For all layers, the horizontal coordinate system shall be the European Terrestrial Reference System 1989 (ETRS89) using Lambert Azimuthal Equal-Area projection (LAEA).
- The file format will be raster data (discrete for the Hard layers and float for the soft layers) because those are very useful for analysis and for storing data that varies continuously.
- The extent used will be the 28 European Member States.
- The soft layers "Flooding", "Cation Exchange Capacity", "Dryness" and the heavier layers such as "ML_Hard" and "MIs_Div3", "MIs_EUp2575" and "MIs_EUp3366" are divided into tiles of 655.4 km by 655.4 km.

12. EXPERIMENTAL RESULTS FOR MLS IDENTIFICATION

This chapter presents the results of the mapping of marginal lands for the MAIL project in Europe and the 7 pilot areas following the methodology described in the previous chapters.

First, the results obtained in the process of generating the "ML_Hard" layer are shown where all the areas of "impervious", "croplands", "forest", "protected areas" and "changed" were excluded. Then the results are presented applying the restrictions described as soft layers on the areas defined as marginal with the layer "ML_HARD".

12.1 Experimental results for Hard marginal land indicators

Figure 66 shows the map of marginal lands for all of Europe. This layer has been generated in raster format (.tif) in tiles of 655.4km by 655.4 km with binary values, 0 (non-marginal) and 1 (marginal) in the coordinate system European Terrestrial Reference System 1989 (ETRS89) using Lambert Azimuthal Equal-Area projection (LAEA). The coverage of the mosaic is for all the countries of the European Union and consists of 17 tiles.

Figure 67 shows the marginal lands for each of the countries that are partners in the MAIL project. In Table 46 the marginal surface by country is presented, where it is observed that the Mediterranean countries are those that have more marginal lands, achieving Spain 19.96% and Greece 18.76% of the territory as marginal lands according to the proposed methodology. The total country area was obtained from the Eurostat area by NUTS 3 region update on 25-08-2020.

Table 46: Total surface of the countries that are part of the MAIL project, surface of marginal lands by country, and percentage that represent the marginal lands within the surface of the country.

	Total country area (km ²)	ML area (km ²)	ML over the total (%)
Spain	505,970	100,983	19.96
Greece	132,049	24,770	18.76
Germany	357,340	41,606	11.64
Poland	312,679	22,442	7.18



Figure 66: Layer "ML_Hard" for Europe where the orange colors represent the areas defined as marginal excluding land use "impervious", "croplands", "forest", "protected areas" and "changed". Source: personal compilation of Jesús Torralba Pérez.

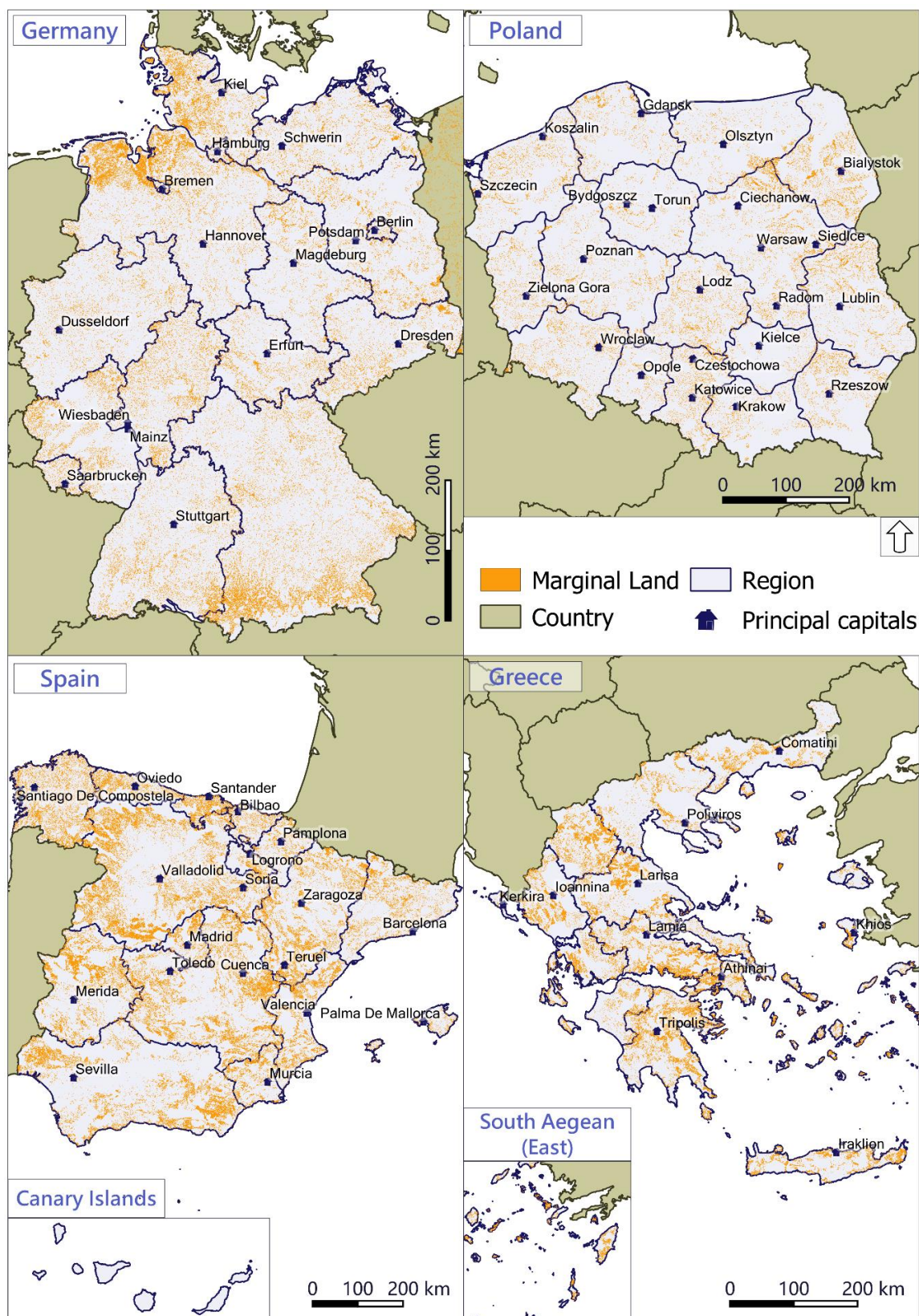
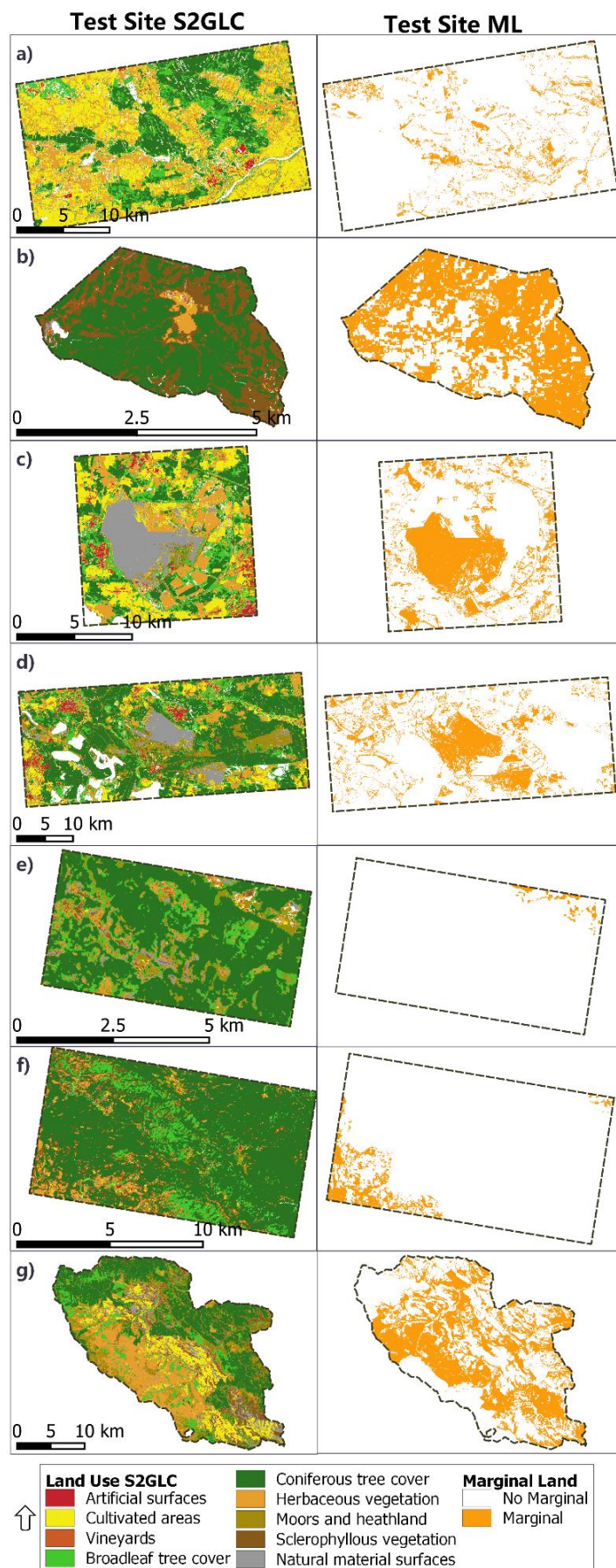


Figure 67: Layer "ML_Hard" for Germany, Poland, Spain, and Greece where the orange colors represent the areas defined as marginal excluding land use "impervious", "croplands", "forest", "protected areas" and "changed". Source: personal compilation of Jesús Torralba Pérez.



The marginal lands of the 7 pilot sites were also obtained and the land uses under the areas identified as marginal lands were analyzed (Figure 68). The land uses were obtained from the Sentinel-2 Global Land Cover (S2GLC) 2017 layer.

Figure 69 describes in percentage the land uses, where the pilot sites with the highest percentage of marginal lands are Isenli (38.6), Welzow (26.3), Soria (26.1), Nochten (16.9), where Staszów (6.3), Espadán (3.9) and Noguieruelas (0.8) are the pilot sites with the lowest percentage of marginal lands. These differences are due to the interspecific variation of the study sites, for example, the test site of Espadán is classified mostly as a protected area, the pilot area of Noguieruelas is a conifer forest zone, the site of

Figure 68: Representation of the land uses of the S2GLC layer (left column) and the areas classified as marginal lands according to the methodology proposed in the MAIL project (right column). The pilot sites are: a) Staszów (Poland), b) Isenli (Greece), c) Welzow (Germany), d) Nochten (Germany), e) Noguieruelas (Spain), f) Espadan (Spain) and g) Soria (Spain). Personal source: Jesús Torralba Pérez.

Soria is covered mostly by a conifer forest, herbaceous vegetation and moors and heathland, Welzow is an area where there was previously an open-cast mine, the Isenli pilot area is conifer afforestation and the Staszów area is mostly covered by farmland, herbaceous vegetation, and conifer forests.

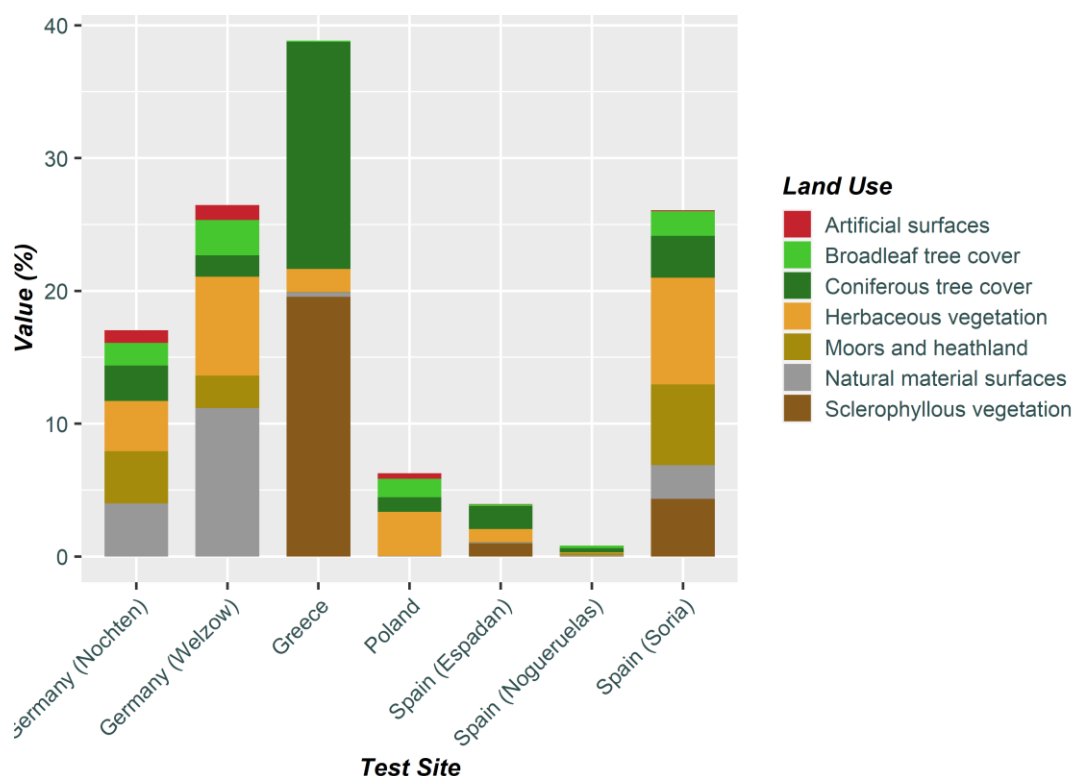


Figure 69: Percentage of marginal land for each land use over the total marginal land per test site. Land use has been obtained from the S2GLC layer. Source: personal compilation of Jesús Torralba Pérez.

Table 47 Table 47: Hectares and percentages of land use and marginal land by pilot site and land use. The land use has been obtained from S2GLC layer.summarizes per pilot site: the area in hectares per land use category of the S2GLC layer, the area of marginal land per land use category, the percentage of the area of a land use classified as marginal, and the percentage of marginal land for each land use over the total marginal land per test site. From dark orange to light orange, the percentage of land use classified as marginal according to the methodology proposed in the MAIL project is remarked. Dark orange represents when more than 50% of the area in that land use category has been classified as marginal. Salmon orange when it has been classified as marginal between 25 and 50 % of the land use category, and light orange when it has been classified between 10 and 25 % of the area of a particular land use category as marginal. It is possible to observe that each pilot site has a different nature, yet it is observed that the land uses considered as marginal for the set of pilot sites are those of the categories of "herbaceous vegetation", "moors and heathland" and "natural material



surfaces". It is also shown that the categories "Cultivated areas" and "Vineyards" are not considered as marginal in any pilot site.



	Poland (Staszów)				Greece (Isenli)				Germany (Welzow)				Germany (Nochten)			
	Land Use Area (ha)	Area MLs (ha)	MLs for Land Use (%)	MLs over total ML (%)	Land Use Area (ha)	Area MLs (ha)	MLs for Land Use (%)	MLs over total ML (%)	Land Use Area (ha)	Area MLs (ha)	MLs for Land Use (%)	MLs over total ML (%)	Land Use Area (ha)	Area MLs (ha)	MLs for Land Use (%)	MLs over total ML (%)
Artificial surfaces	823.0	273.6	33.2	6.7	0.1	0.1	100.0	0.0	992.9	278.2	28.0	4.2	4248.7	1169.6	27.5	5.5
Cultivated areas	15210.5	0.0	0.0	0.0	2.7	0.0	0.0	0.0	4284.7	0.0	0.0	0.0	9136.7	0.0	0.0	0.0
Vineyards	0.0	0.0	0.0	0.0	12.3	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
Broadleaf tree cover	6826.9	914.8	13.4	22.3	2.9	2.0	70.8	0.3	3349.7	660.1	19.7	10.1	12712.7	2144.0	16.9	10.1
Coniferous tree cover	9891.5	718.2	7.3	17.5	686.2	337.5	49.2	44.0	3836.7	398.8	10.4	6.1	42367.7	3305.8	7.8	15.6
Herbaceous vegetation	11662.2	2160.1	18.5	52.7	34.3	34.2	99.9	4.5	4337.8	1843.6	42.5	28.1	10070.9	4723.8	46.9	22.2
Moors and heathland	0.0	0.0	0.0	0.0	0.6	0.6	87.3	0.1	1101.6	605.3	54.9	9.2	10986.2	4909.8	44.7	23.1
Sclerophyllous vegetation	0.0	0.0	0.0	0.0	415.7	386.4	93.0	50.3	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
Natural material surfaces	217.3	29.9	13.7	0.7	7.0	7.0	99.9	0.9	3381.2	2767.0	81.8	42.2	6349.2	4978.7	78.4	23.4
Total	44631.4	4096.5			1161.7	767.7			21284.5	6553.0			95872.0	21231.6		
	Spain (Nogueruelas)				Spain (Espadán)				Spain (Soria)							
Artificial surfaces	0.0	0.0	0.0	0.0	17.3	1.2	6.9	0.2	339.1	92.4	27.3	0.4				
Cultivated areas	7.7	0.0	0.0	0.0	127.1	0.0	0.0	0.0	6034.2	0.0	0.0	0.0				
Vineyards	5.7	0.0	0.0	0.0	44.7	0.0	0.0	0.0	305.4	0.0	0.0	0.0				
Broadleaf tree cover	323.1	6.5	2.0	24.3	1142.1	20.5	1.8	3.3	5613.9	1837.9	32.7	7.1				
Coniferous tree cover	1488.8	9.8	0.7	36.5	8022.4	273.8	3.4	43.8	16393.7	3084.4	18.8	12.0				
Herbaceous vegetation	147.9	2.5	1.7	9.2	721.8	157.8	21.9	25.2	11348.9	7934.0	69.9	30.8				
Moors and heathland	238.2	5.0	2.1	18.6	237.7	4.5	1.9	0.7	9017.9	5992.7	66.5	23.3				
Sclerophyllous vegetation	45.9	0.4	1.0	1.6	1102.7	157.5	14.3	25.2	4879.4	4292.6	88.0	16.7				
Natural material surfaces	73.0	2.6	3.6	9.8	97.3	10.5	10.8	1.7	3780.4	2485.7	65.8	9.7				
Total	2330.2	26.8			11513.3	625.7			57712.9	25719.7						

Table 47: Hectares and percentages of land use and marginal land by pilot site and land use. The land use has been obtained from S2GLC layer.



12.2 Experimental results for marginal land identification applying soft indicators

In this section, the results are presented to classify marginal lands according to their marginality: "MLs with high plantation suitability", "MLs with low plantation suitability" and "Potentially unsuitable lands". Adding all the rasters of the soft indicators (17 indicators and 24 layers described in section 9.2) we obtained a final raster where the minimum value was 0.120 and the maximum value 6.862, although the theoretical maximum value would have been 9.98 in the hypothetical case that all indicators on the same pixel would have had the maximum value. As explained in section 10.3, three methods were used to divide up marginal lands according to their marginality: a) computing the maximum and minimum and dividing the range of values by 3, b) computing the 25th and 75th percentile and setting these values as threshold limits, c) computing the 33rd and 66th percentile to keep the same number of pixels in each category.

Figures 70, 71, and 72 show the marginal lands for Europe for the three methods. Figure 70 shows the results of divided by 3 the range of values. This method increases the "Potentially unsuitable lands" because around the value 2 is where the majority of the data is. Figure 71 shows the results obtained by using method b, where percentiles 25th (value = 1.390) and 75th (value = 2.890) are calculated and these values are established as class limits. This method penalizes the classes "Potentially unsuitable lands" and "MLs with high plantation suitability" and increasing the number of pixels belonging to the middle layer "MLs with low plantation suitability". Figure 72 shows the results when applying the c method, which consists of computing the 33rd percentile (value = 1.579) and 66th percentile (value = 2.516) and establishing these values as the limits of the marginal classes. This method distributes the total number of pixels in three equal groups, but with the difference to the first method that in this case the thresholds are established with the maximum number of pixels and not with the maximum and minimum values. Method c benefits the "MLs with high plantation suitability" class because the lower threshold for this class is fixed at 2.516 (Table 48).

These three layers have been generated in raster format (.tif) in tiles of 655.4km by 655.4 km with float values in the coordinate system European Terrestrial Reference System 1989 (ETRS89) using Lambert Azimuthal Equal-Area projection (LAEA). The coverage of the mosaic is for all the countries of the European Union and consists of 17 tiles.

Figure 73 shows the results obtained by implementing the three methods in the seven pilot sites. It can be seen that there are great differences in the results if method a) is applied or if methods b) and c) are applied.



Figure 70: Final layer of MLs classified into three categories “Marginal lands with high plantation suitability”, “Marginal lands with low plantation suitability” and “Potentially unsuitable lands” with the method a - computing the maximum and minimum and dividing the range of values by 3. Source: personal compilation of Jesús Torralba Pérez.



Figure 71: Final layer of MLs classified into three categories “Marginal lands with high plantation suitability”, “Marginal lands with low plantation suitability” and “Potentially unsuitable lands” with the method b - computing the 25th and 75th percentile and setting these values as threshold limits. Source: personal compilation of Jesús Torralba Pérez.



Figure 72: Final layer of MLs classified into three categories “Marginal lands with high plantation suitability”, “Marginal lands with low plantation suitability” and “Potentially unsuitable lands” with the method c - computing the 33rd and 66th percentile to keep the same number of pixels in each category. Source: personal compilation of Jesús Torralba Pérez.

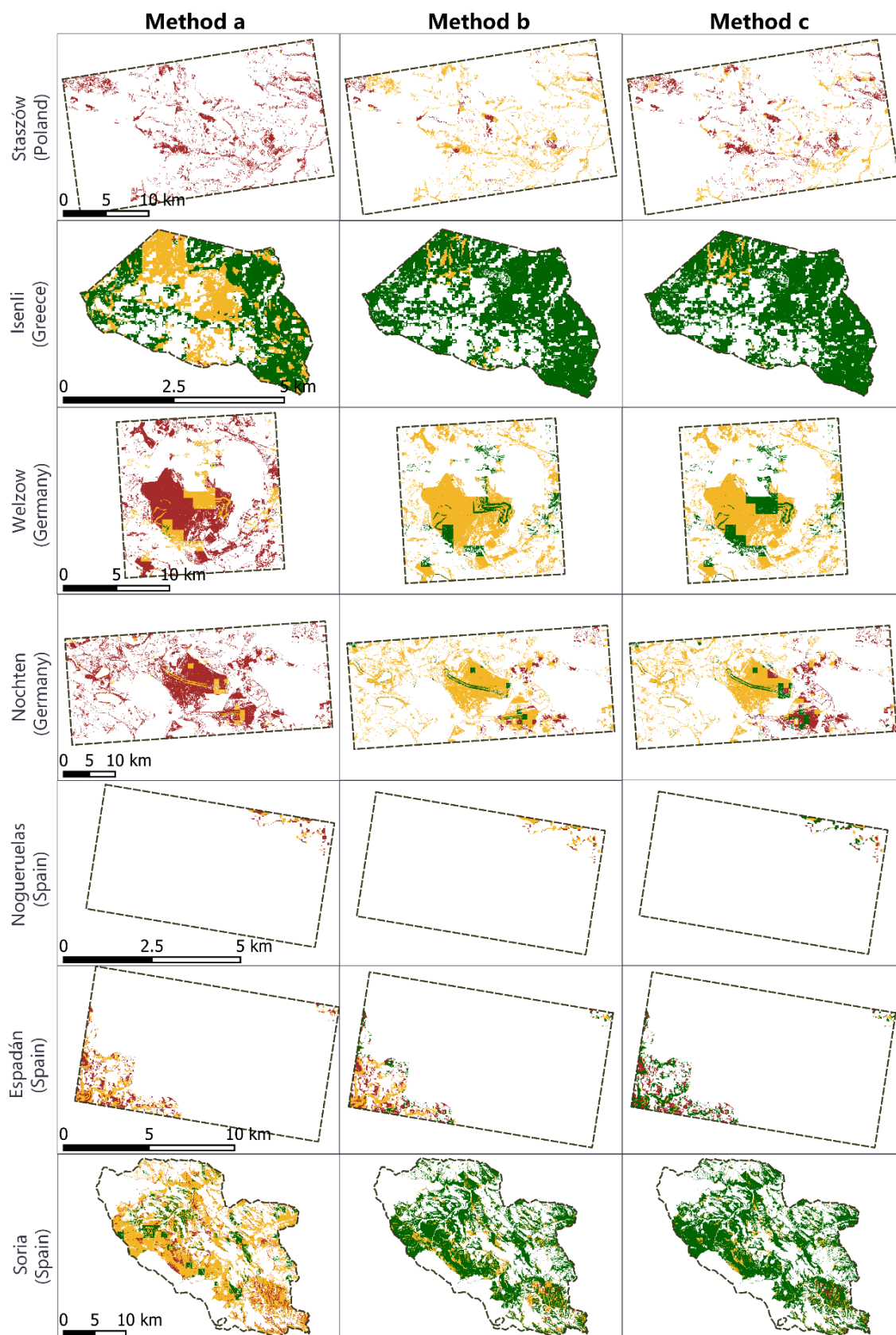


Figure 73: Final layer of MLs classified with 3 methods into 3 categories “Marginal lands with high plantation suitability” (green), “Marginal lands with low plantation suitability” (gold), and “Potentially unsuitable lands” (dark red) in the pilot sites. Source: personal compilation of Jesús Torralba Pérez.



Table 48 shows for every method the percentage of individual marginality categories per pilot site.

Table 48: Hectares and percentage of total area for each of the marginal land categories according to marginality (“Marginal Lands with high plantation suitability”, “Marginal Lands with low plantation suitability”, and “Potentially unsuitable lands”) for each pilot site and method.

		Poland (Staszów)		Greece (Isenli)		Germany (Welzow)		Germany (Nochten)		Spain (Nogueruelas)		Spain (Espadán)		Spain (Soria)	
Method	Type	ha	%	ha	%	ha	%	ha	%	ha	%	ha	%	ha	%
A	<i>MLs with high plantation suitability</i>	0.0	0.00	450.6	22.64	27.1	0.11	6.6	0.01	0.0	0.00	0.0	0.00	3611.5	3.66
	<i>MLs with low plantation suitability</i>	19.6	0.03	315.2	15.84	1554.8	6.26	2023.3	1.62	12.7	0.39	341.8	2.14	19194.2	19.45
	<i>Potentially unsuitable lands</i>	4076.4	6.23	2.2	0.11	4951.2	19.95	19090.6	15.31	14.6	0.45	282.0	1.77	2913.7	2.95
	<i>Total MLs</i>	4096.0	6.3	768.1	38.6	6533.0	26.3	21120.5	16.9	27.3	0.8	623.9	3.9	25719.3	26.1
B	<i>MLs with high plantation suitability</i>	17.8	0.03	742.0	37.28	1095.4	4.41	1255.5	1.01	0.3	0.01	103.0	0.64	20367.3	20.64
	<i>MLs with low plantation suitability</i>	3487.6	5.33	26.0	1.31	5437.6	21.91	17891.5	14.35	21.0	0.65	272.1	1.70	5082.2	5.15
	<i>Potentially unsuitable lands</i>	590.6	0.90	0.0	0.00	0.0	0.00	1973.5	1.58	6.0	0.18	248.7	1.56	269.9	0.27
	<i>Total MLs</i>	4096.0	6.3	768.1	38.6	6533.0	26.3	21120.5	16.9	27.3	0.8	623.9	3.9	25719.3	26.1
C	<i>MLs with high plantation suitability</i>	18.9	0.03	746.4	37.50	1581.1	6.37	1933.4	1.55	12.7	0.39	341.2	2.14	22587.9	22.89
	<i>MLs with low plantation suitability</i>	1682.5	2.57	21.6	1.09	4952.0	19.95	14043.5	11.26	8.3	0.25	33.7	0.21	2751.0	2.79
	<i>Potentially unsuitable lands</i>	2394.6	3.66	0.0	0.00	0.0	0.00	5143.6	4.13	6.4	0.20	249.0	1.56	380.5	0.39
	<i>Total MLs</i>	4096.0	6.3	768.1	38.6	6533.0	26.3	21120.5	16.9	27.3	0.8	623.9	3.9	25719.3	26.1

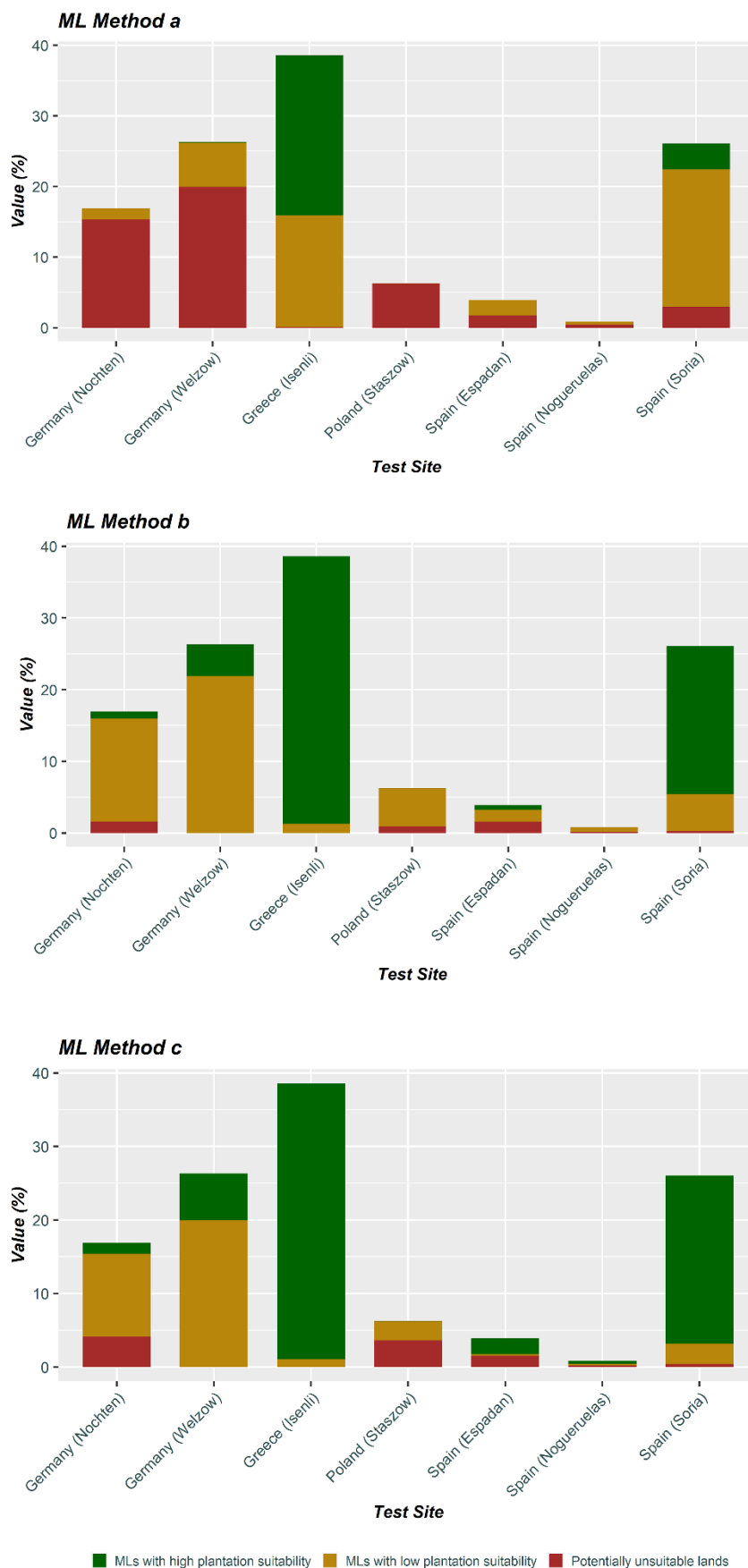


Figure 74 is a representation of the data in Table 48. A stacked bar chart was made for every method. The 3 categories of marginal lands for each pilot site are stacked on the bars.

Figure 74: Percentage for each of the marginal land categories according to marginality (“Marginal lands with high plantation suitability”, “Marginal lands with low plantation suitability”, and “Potentially unsuitable lands”) for each pilot site and approach. Source: personal compilation of Jesús Torralba Pérez.



13 1ST METHODOLOGY REFINEMENT (INDICATOR FINE-TUNING BASED ON THE EXPERIMENTAL RESULTS)

This chapter explains the adjustments of the soft layers to achieve the results shown in chapter 12 where these changes have already been implemented. The first phase of the refinement consisted of processing the soft layers of "Stoniness", "Texture", "Clay", "Sand", "Total Available Water" and "Soil Organic Matter" for two soil depths: topsoil and subsoil. This change implied the modification of the weights assigned in tables 22, 23, and 24 of chapter 6.5 and the reprocessing of these layers at the European level.

The proposed thresholds for the soft indicator "erosion" were revised. These thresholds were not appropriate for Europe, because the bibliographic reference used to establish the thresholds referred to a study in Malawi. Malawi has one of the highest erosion rates in the world mainly caused by a combination of deforestation, high rainfall (tropical climate), and high slopes (Great Rift Valley). For which the thresholds proposed by (Li et al., 2017) would not be representative of Europe. It was proposed following Eurostat (2020) to modify the erosion values since for Europe it is considered severe erosion in agriculture when the annual soil loss is greater than 10 t/ha/year, very different from the 200 t/ha/yr proposed for Malawi. The 3 intervals are performed following the Eurostat severe erosion tables. The first interval from 10 to 55.3 equal to a score of 10, with 55.3 being the maximum erosion where agriculture is carried out in the EU. The second interval from 55.3 to 67.3 equal to a score of 5, being the maximum for certain crops, forest, and semi-natural areas (excluding beaches, dunes, sand plains, bare rock, glaciers, and perpetual snow) and the third interval, from 67.3 to 325 equal to score 1, being the maximum of the layer "erosion" (European Soil Data Centre (ESDAC), Soil erosion by water (RUSLE2015), 2015).

In the case of the "cation exchange capacity" of the soil occurred the same as for the "erosion" layer. The initial thresholds for the "cation exchange capacity" layer will be modified as they were defined for Malawi (Li et al., 2017) where there are major differences from the CEC in Europe. To reformulate the CEC thresholds, the SoilGrids250m 2.0 - Cation exchange capacity at pH7 layer was downloaded from the ISRIC — World Soil Information. This data set was analyzed in GEE and the mean of the CEC layer in the 1 m deep soil profile was obtained. Then, the values for all Europe were revised, and the outliers values were eliminated by the percentiles 0.1 and 99.9. Fixing the minimum value of 12.3 (cmol/kg). To establish the maximum value, the 50th percentile was chosen, thus eliminating the areas with the best CEC in Europe, and restricting the marginality, resulting in a CEC maximum value of 22.2 (cmol/kg). Knowing these values, the subdivision in three categories of marginality was carried out: 12.3



– 15.6 (cmol/kg) very marginal (1), 15.6 – 18.9 (cmol/kg) medium marginal (5) and 18.9 – 22.2 (cmol/kg) less marginal (10).

The "Dryness" soft layer was also modified. This layer was based on the TerraClimate, a high-resolution global dataset of monthly climate and climatic water balance from 1958–2015 (Abatzoglou et al., 2018) for the year 2018 and available at the Climatology Lab from the University of California Merced. To obtain the aridity index it was necessary to work with the accumulated precipitation (mm) and the reference evapotranspiration. In this case, the FAO Penman-Monteith method is recommended for determining the reference evapotranspiration. The Penman-Monteith approach includes all parameters that govern energy exchange and corresponding latent heat flux (evapotranspiration) from uniform expanses of vegetation. Most of the parameters are measured or can be readily calculated from weather data (Penman, 1948).

Several workflows were modified during the generation of the resulting layers, but the final workflows were already shown in chapters 9 and 10.



14 CONCLUSIONS

Deliverable D2.3 performed an in-depth review of existing datasets, methodologies and approaches for the detection of marginal lands using freely available data. During the work performed in T2.3 a detailed methodology based on GIS classification was developed for the detection of marginal lands. Detailed workflows were created that describe every step of the processing chain to create the “hard” and “soft” layers and combine them for the final classification. Marginal lands were classified in three categories “Marginal lands with high plantation suitability”, “Marginal lands with low plantation suitability”, and “Potentially unsuitable lands”. The developed methodology follows an “Open Architecture” approach in which new indicators, criteria and layers can be easily incorporated in the final classification scheme. The developed methodology was applied in selected test sites and in European level. Currently the achieved results from the different classification schemes are being evaluated through the work performed at T2.4 “Accuracy assessment of m/sm MLs detection”. Furthermore T2.8 “Augment precision of MLs detection” will develop a methodology to augment the precision of marginal land detection using free and open access satellite data in selected sites. It must be noted that the results of T2.4 and T2.8 will be used to enhance the developed methodology for marginal land detection and will be incorporated in the final system. Finally, since new free and open access datasets are becoming available in a timely manner, new indicators or criteria that can be developed based on the availability of these datasets will also be examined and integrated in the final system.



REFERENCES

- [1] Abatzoglou, J. T., Dobrowski, S. Z., Parks, S. A., & Hegewisch, K. C. (2018). TerraClimate, a high-resolution global dataset of monthly climate and climatic water balance from 1958-2015. *Scientific Data*. <https://doi.org/10.1038/sdata.2017.191>
- [2] Bai, Z., Dent, D., Olsson, L., & Schaepman, M. (2008). Global Assessment of Land Degradation and Improvement 1. Identification by remote sensing. *ISRIC – World Soil Information, Wageningen*, (November), 69. <https://doi.org/report/2008/01>
- [3] Bandaru, V., Izaurralde, R. C., Manowitz, D., Link, R., Zhang, X., & Post, W. M. (2013). Soil Carbon Change and Net Energy Associated with Biofuel Production on Marginal Lands: A Regional Modeling Perspective. *Journal of Environment Quality*, 42(6), 1802. <https://doi.org/10.2134/jeq2013.05.0171>
- [4] Bertaglia, M., Joost, S., & Roosen, J. (2007). Identifying European marginal areas in the context of local sheep and goat breeds conservation: A geographic information system approach. *Agricultural Systems*, 94(3), 657–670. <https://doi.org/10.1016/j.agsy.2007.02.006>
- [5] Büttner, G, & Kosztra, B. (2017). *CLC2018 Technical Guidelines*. Retrieved from https://land.copernicus.eu/user-corner/technical-library/clc2018technicalguidelines_final.pdf
- [6] Büttner, Gyorgy, Kosztra, B., Hazeu, G., & Arnold, S. (2017). *Updated CLC illustrated nomenclature guidelines*. Brussels.
- [7] Cai, X., Zhang, X., & Wang, D. (2011). Land Availability for Biofuel Production. *Environmental Science & Technology*, 45(1), 334–339. <https://doi.org/10.1021/es103338e>
- [8] Chazdon, R. L., Brancalion, P. H. S., Laestadius, L., Bennett-Curry, A., Buckingham, K., Kumar, C., ... Wilson, S. J. (2016). When is a forest a forest? Forest concepts and definitions in the era of forest and landscape restoration. *Ambio*, 45(5), 538–550. <https://doi.org/10.1007/s13280-016-0772-y>
- [9] Ciria, C., Sanz, M., Carrasco, J., & Ciria, P. (2019). Identification of Arable Marginal Lands under Rainfed Conditions for Bioenergy Purposes in Spain. *Sustainability*, 11(7), 1833. <https://doi.org/10.3390/su11071833>
- [10] Clark, D. A., Brown, S., Kicklighter, D. W., Chambers, J. Q., Thomlinson, J. R., & Ni, J. (2001). Measuring net primary production in forest: concepts and field methods. *Ecological Society of America*, 11(2), 356–370.
- [11] Dzieszko, P. (2014). Land-Cover modelling using CORINE land cover data and multi-layer perceptron. *Quaestiones Geographicae*, 33(1), 5–22. <https://doi.org/10.2478/quageo-2014-0004>



- [12] Elbersen, B., van Eupen, M., Mantel, S., Verzandvoort, S., Boogaard, H., Mucher, S., ... Eleftheriadis, I. (2018). Methodological approaches to identify and map marginal land suitable for industrial crops in Europe (Deliverable D2.6). In *MAGIC project reports*.
- [13] Eliasson, Å., Jones, R. J. A., Nachtergaele, F., Rossiter, D. G., Terres, J. M., van Orshoven, J., ... Le Bas, C. (2010). Common criteria for the redefinition of Intermediate Less Favoured Areas in the European Union. *Environmental Science and Policy*, 13(8), 766–777. <https://doi.org/10.1016/j.envsci.2010.08.003>
- [14] Esri. (2020). ArcGIS PRO 2.7. What is a geoprocessing environment setting? Retrieved December 25, 2020, from Resources: Tool Reference website: <https://pro.arcgis.com/en/pro-app/latest/tool-reference/environment-settings/what-is-a-geoprocessing-environment.htm>
- [15] European Environment Agency. (2015). Copernicus Land Monitoring Services - High Resolution Layers (HRL), 2012 to 2015. In *Datasets*. Retrieved from <https://land.copernicus.eu/pan-european/high-resolution-layers>
- [16] European Environment Agency. (2017). Copernicus Land Monitoring Service - Reference Data: EU-DEM. In *Datasets*. Retrieved from <https://www.eea.europa.eu/data-and-maps/data/copernicus-land-monitoring-service-eu-dem>
- [17] European Environment Agency. (2020). Natura 2000 data - the European network of protected sites. In *Datasets*. Retrieved from <https://www.eea.europa.eu/data-and-maps/data/natura-11>
- [18] European Parliament. (2018). Regulation (EU) 2018/841 Inclusion of greenhouse gas emissions and removals from land use, land use change and forestry in the 2030 climate and energy framework, and amending Regulation (EU) 525/2013 and Decision 529/2013/EU. *Official Journal of the European Union*, 156(1), 1–25. Retrieved from <http://data.europa.eu/eli/reg/2018/841/oj>
- [19] Eurostat. (2020). Agri-environmental indicator - soil erosion. Retrieved November 10, 2020, from https://ec.europa.eu/eurostat/statistics-explained/index.php/Agri-environmental_indicator_-_soil_erosion#Data_sources
- [20] FAO. (2000). *Global Forest Resources Assessment 2000*.
- [21] Gelfand, I., Sahajpal, R., Zhang, X., Izaurralde, R. C., Gross, K. L., & Robertson, G. P. (2013). Sustainable bioenergy production from marginal lands in the US Midwest. *Nature*, 493(7433), 514–517. <https://doi.org/10.1038/nature11811>
- [22] Gerwin, W., Repmann, F., Galatsidas, S., Vlachaki, D., Gounaris, N., Baumgarten, W., ... Freese, D. (2018). Assessment and quantification of marginal lands for biomass production in Europe using soil-quality indicators. *Soil*, 4(4), 267–290. <https://doi.org/10.5194/soil-4-267-2018>



- [23] Gopalakrishnan, G., Cristina Negri, M., & Snyder, S. W. (2011). A Novel Framework to Classify Marginal Land for Sustainable Biomass Feedstock Production. *Journal of Environment Quality*, 40(5), 1593. <https://doi.org/10.2134/jeq2010.0539>
- [24] Government of Alebrta. (2015). *Alberta Wetland Identification and Delineation Directive*. Edmonton, Alberta.
- [25] Grassi, G., Cescatti, A., Matthews, R., Duveiller, G., Camia, A., Federici, S., ... Vizzarri, M. (2019). On the realistic contribution of European forests to reach climate objectives. *Carbon Balance and Management*. <https://doi.org/10.1186/s13021-019-0123-y>
- [26] Hamad, R., Balzter, H., & Kolo, K. (2018). Predicting Land Use/Land Cover Changes Using a CA-Markov Model under Two Different Scenarios. *Sustainability*, 10(10), 3421. <https://doi.org/10.3390/su10103421>
- [27] Hansen, M. C., Potapov, P. V, Moore, R., Hancher, M., Turubanova, S. A., Tyukavina, A., ... Townshend, J. R. G. (2013). High-Resolution Global Maps of 21st-Century Forest Cover Change. *Science*, 342(6160), 850 LP – 853. <https://doi.org/10.1126/science.1244693>
- [28] Hiederer, R. (2013). *Mapping Soil Properties for Europe - Spatial Representation of Soil Database Attributes*. <https://doi.org/10.2788/94128>
- [29] Ivanina, V., Roik, M., & Hanzhenko, O. (2016). Report on MagL concepts, debate and indicators (D2.3). In *D2.3 of SEEMLA project*. Ukraine.
- [30] James, L. (2010). *Theory and identification of marginal land and factors determining land use change*.
- [31] JRC. (2018). *Global Surface Water - Data Users Guide (v2) Background*.
- [32] Kang, S., Post, W., Wang, D., Nichols, J., Bandaru, V., & West, T. (2013). Hierarchical marginal land assessment for land use planning. *Land Use Policy*, 30(1), 106–113.
- [33] Langanke, T. (2017). *Copernicus Land Monitoring Service - High Resolution Layer Forest: Product Specifications Document*. Belgium.
- [34] Li, G., Messina, J. P., Peter, B. G., & Snapp, S. S. (2017). Mapping Land Suitability for Agriculture in Malawi. *Land Degradation and Development*, 28(7), 2001–2016. <https://doi.org/10.1002/ldr.2723>
- [35] Liu, T., Ma, Z., McConkey, B., Kulshreshtha, S., Huffman, T., Green, M., ... Shang, J. (2012). Bioenergy production potential on marginal land in Canada. *2012 1st International Conference on Agro-Geoinformatics, Agro-Geoinformatics 2012*. <https://doi.org/10.1109/Agro-Geoinformatics.2012.6311729>
- [36] Malinowski, R., Lewinski, S., Rybicki, M., Jenerowicz, M., Gromny, E., Krupinski, M., ... Kratzschmar, E. (2019). *Deliverable 1.2 Final Report*.
- [37] Martone, M., Rizzoli, P., Wecklich, C., González, C., Bueso-Bello, J.-L., Valdo, P., ...



- Moreira, A. (2018). The global forest/non-forest map from TanDEM-X interferometric SAR data. *Remote Sensing of Environment*, 205, 352–373. <https://doi.org/https://doi.org/10.1016/j.rse.2017.12.002>
- [38] Matteodo, M., Grand, S., Sebag, D., Rowley, M. C., Vittoz, P., & Verrecchia, E. P. (2018). Decoupling of topsoil and subsoil controls on organic matter dynamics in the Swiss Alps. *Geoderma*, 330, 41–51. <https://doi.org/https://doi.org/10.1016/j.geoderma.2018.05.011>
- [39] Milbrandt, A., & Overend, R. P. (2009). Assessment of Biomass Resources from Marginal Lands in APEC Economies. In *APEC Energy Working Group* (p. 52).
- [40] Mueller, L., Schindler, U., Behrendt, A., Eulenstein, F., & Dannowski, R. (2007). *The Muencheberg Soil Quality Rating (SQR): FIELD MANUAL FOR DETECTING AND ASSESSING PROPERTIES AND LIMITATIONS OF SOILS FOR CROPPING AND GRAZING* (Vol. 2014). Muencheberg, Germany.
- [41] Niu, X., & Duiker, S. W. (2006). Carbon sequestration potential by afforestation of marginal agricultural land in the Midwestern U.S. *Forest Ecology and Management*, 223(1–3), 415–427. <https://doi.org/10.1016/j.foreco.2005.12.044>
- [42] Pekel, J. F., Cottam, A., Gorelick, N., & Belward, A. S. (2016). High-resolution mapping of global surface water and its long-term changes. *Nature*. <https://doi.org/10.1038/nature20584>
- [43] Penman, H. L. (1948). Natural evaporation from open water, bare soil and grass. *Proceedings of the Royal Society of London. Series A, Mathematical and Physical Sciences*. <https://doi.org/10.1098/rspa.1948.0037>
- [44] Peter, B. G., Messina, J. P., & Snapp, S. S. (2018). Multiscalar approach to mapping marginal agricultural land: Smallholder agriculture in Malawi. *Annals of the American Association of Geographers*, 108(4), 989–1005. <https://doi.org/10.1080/24694452.2017.1403877>
- [45] Roehrig, J., & Menz, G. (2005). The Determination of Natural Agricultural Potential in Western Africa Using the Fuzzy Logic Based Marginality Index. *EARSeL EProceedings*, 9–17.
- [46] Sallustio, L., Pettenella, D., Merlini, P., Romano, R., Salvati, L., Marchetti, M., & Corona, P. (2018). Assessing the economic marginality of agricultural lands in Italy to support land use planning. *Land Use Policy*, 76(February), 526–534. <https://doi.org/10.1016/j.landusepol.2018.02.033>
- [47] Sannier, C., Pennec, A., & Dufourmont, H. (2017). *Comparative Validation of HRL-TCD and University of Maryland Global Forest Change Products*. EU.
- [48] Schweers, W., Campbell, E., Hennenberg, K., Fritsche, U., Mang, H. P., Lucas, M., ... Zhang, N. (2011). Identification of potential areas for biomass production in China:



- Discussion of a recent approach and future challenges. *Biomass and Bioenergy*, 35(5), 2268–2279. <https://doi.org/10.1016/j.biombioe.2011.02.034>
- [49] Union, E. (2020). The European Union and forests.
- [50] Vlachaki, D., Gounaris, E., Dimitriadis, E., & Galatsidas, S. (2018). *Final guidelines for the sustainable exploitation of MagLs for bioenergy. Sustainable exploitation of biomass for bioenergy from marginal lands in Europe*. Retrieved from https://www.seemla.eu/wp-content/uploads/2019/01/D6.8_Guidelines_draft_v3.pdf
- [51] Zolekar, R. B. (2018). Integrative approach of RS and GIS in characterization of land suitability for agriculture: a case study of Darna catchment. *Arabian Journal of Geosciences*, 11(24). <https://doi.org/10.1007/s12517-018-4148-4>
- [52] Zolekar, R. B., & Bhagat, V. S. (2015). Multi-criteria land suitability analysis for agriculture in hilly zone: Remote sensing and GIS approach. *Computers and Electronics in Agriculture*, 118, 300–321. <https://doi.org/10.1016/j.compag.2015.09.016>



ANNEX I: MARGINALITY INDEX CALCULATION AND CLC RECLASSIFICATION

The land use class chosen²⁴ for less productive and non-productive land is class 3.2 (“Low vegetation”). Class 1.1.1 (“Artificial surfaces and constructions”) and class 2.1 (“Cultivated and managed areas”) are used as an indicator of a less marginal land use. Thus, the variable computed is the ratio of the percent of the neighborhood area (for further explanation see below) classified as land use class 3.2 to the percent of area in land use class 1.1.1 and class 2.1, that is:

$$\text{MARINDEX} = (\% \text{class 3.2}) / (\% \text{class 1.1.1} + \% \text{class 2.1})$$

The standard CLC nomenclature includes 44 land cover classes²⁵. These are grouped in a three-level hierarchy. The five main (level-one) categories are: 1) artificial surfaces, 2) agricultural areas, 3) forests and semi-natural areas, 4) wetlands, 5) water bodies. All national teams had to adopt this standard nomenclature according to their landscape conditions. Although the 44 categories have not changed since the implementation of the first CLC inventory (1986-1998), the definition of most of the nomenclature elements was significantly improved.

Table 49: CORINE land cover categories. Source: Copernicus Land Monitoring Service web page

CLC Level 1	CLC Level 2	CLC Level 3
1 Artificial surfaces	1.1 Urban fabric	1.1.1 Continuous urban fabric
		1.1.2 Discontinuous urban fabric
	1.2 Industrial, commercial and transport units	1.2.1 Industrial and commercial units
		1.2.2 Road and rail networks and associated land
		1.2.3 Port areas
		1.2.4 Airports
		1.3.1 Mineral extraction sites
		1.3.2 Dump sites

²⁴ Classes to be included into the final ratio will depend on the final definition of each land cover utilized by CLC project for the compilation of the European land use dataset.

²⁵ For more information regarding the definition of each class, please consult “CORINE land cover nomenclature illustrated guide”



CLC Level 1	CLC Level 2	CLC Level 3
	1.3 Mine, dump and construction sites	1.3.3 Construction sites
	1.4 Artificial non-agricultural vegetated areas	1.4.1 Green urban areas
		1.4.2 Sport and leisure facilities
2. Agricultural areas	2.1 Arable lands	2.1.1 Non-irrigated arable land
		2.1.2 Permanently irrigated land
		2.1.3 Rice fields
	2.2 Permanent crops	2.2.1 Vineyards
		2.2.2 Fruit trees and berry plantation
		2.2.3 Olive groves
	2.3 Pastures	2.3.1 Pastures
	2.4 Heterogeneous agricultural areas	2.4.1 Annual crops associated with permanent crops
		2.4.2 Complex cultivation patterns
		2.4.3 Land principally occupied by agriculture with significant areas of natural vegetation
3 Forests and semi-natural areas	3.1 Forest	3.1.1 Broad-leaved forest
		3.1.2 Coniferous forest
		3.1.3 Mixed forest
	3.2 Shrub and/or herbaceous vegetation association	3.2.1 Natural grassland
		3.2.2 Moors and heathland
		3.2.3 Sclerophyllous vegetation
		3.2.4 Transitional woodland shrub
	3.3. Open spaces with little or no vegetation	3.3.1 Beaches, dunes, and sand plains
		3.3.2 Bare rock
		3.3.3 Sparsely vegetated areas
		3.3.4 Burnt areas
		3.3.5 Glaciers and perpetual snow
4 Wetlands	4.1 Inland wetlands	4.1.1 Inland marshes
		4.1.2 Peatbogs



CLC Level 1	CLC Level 2	CLC Level 3
	4.2. Coastal wetlands	4.2.1 Salt marshes
		4.2.2 Salines
		4.2.3 Intertidal flats
5. Water bodies	5.1 Inland waters	5.1.1. Water courses
		5.1.2. Water bodies
	5.2 Marine waters	5.2.1 Coastal lagoons
		5.2.2 Estuaries
		5.2.3 Sea and ocean



ANNEX II: TABLE OF FIGURES

Figure 1: Transition between marginality and the definition on marginal lands in the framework of MAIL project.	12
Figure 2: The <i>MAIL</i> methodology overview of the first phase of the top-down stepwise approach, Source: personal compilation of Maria Tassopoulou.....	28
Figure 3: The <i>MAIL</i> methodology overview of the second phase of the proposed methodology regarding the “Soft” constraints and indicators, Source: personal compilation of Maria Tassopoulou.....	29
Figure 4: Workflow of the methodology for mapping and classifying MLs in <i>MAIL</i> . Yellow parallelograms depict datasets used in the weighted overlay, while pink parallelograms depict “synthetic” indicators/layers.	30
Figure 5: Overview of the indicator and criteria selection in the <i>MAIL</i> project.	31
Figure 6: The elimination steps followed for selecting the final datasets used in <i>MAIL</i>	52
Figure 7: Overview of the GIS Geodatabase structure and contents. Source: personal compilation of Maria Tassopoulou.	58
Figure 8: Datasets and layers as seen in the ArcGIS project.	58
Figure 9: The final names of the Datasets v2 GIS layers.	59
Figure 10: Greece (left) and the pilot site of the afforestation forest of “Isenli” (right image outlined with pink). Source: personal compilation of Jesús Torralba Pérez.....	61
Figure 11: Spain (left) and the pilot site of “Soria” (outlined with light orange), “Nogueruelas” (outlined with dark red) and “Espadán” (right image outlined with dark orange). Source: personal compilation of Jesús Torralba Pérez.	62
Figure 12: Germany (left) and the pilot sites of “Welzow” (outlined with blue) and “Welzow” (outlined with purple). Source: personal compilation of Jesús Torralba Pérez.	62
Figure 13: Poland (left) and the corresponding selected pilot site (outlined with green). Source: personal compilation of Jesús Torralba Pérez.	63



Figure 14: Overview of classification commission and omission errors of S2GLC raster layer. Artificial surfaces and constructions in images 1 and 2 are misclassified as “natural material surfaces” in areas within the boundaries of the selected test sites in Thessaloniki, Greece. In addition, croplands (3) and artificial surfaces and constructions (4) are classified as “marshes” and “water bodies” (5) in Espadán, Spain test site.....	65
Figure 15: Overview of the classification accuracy of S2GLC LULC product (Malinowski et al., 2019).....	66
Figure 16: Visual assessment of the available datasets for the delineation of water bodies in the test site of Thessaloniki - Greece. The results of the analysis indicate that the S2GLC dataset can result to adequate delineation of water bodies.	68
Figure 17: Visual assessment of the available forest datasets in the test site of Thessaloniki, Greece. The results of the analysis indicate that TanDEM-X dataset (3rd image left to right) could be excluded from further analysis.	73
Figure 18: Visual assessment of the available datasets for the delineation of impervious areas in the city of Castello de la Plana (Spain). The results of the analysis indicate that the S2GLC dataset can be excluded from further analysis.....	76
Figure 19: Values of Soil pH and Erosion (i.e. RUSLE) in the selected test sites.....	81
Figure 20: Overview of the intermediate results of the first phase of the proposed methodology (ML_HARD thresholds) for the test site of Thessaloniki – Greece.	84
Figure 21: Overview of the intermediate results of the first phase of the proposed methodology (ML_HARD thresholds) for the test site of Nochten Reichwalde – Germany.....	85
Figure 22: Overview of the intermediate results of the first phase of the proposed methodology (ML_HARD thresholds) for the test site of Komotini – Greece.	85
Figure 23: Overview of the intermediate results of the first phase of the proposed methodology (ML_HARD thresholds) for the test site of Poland.	86
Figure 24: Overview of the intermediate results of the first phase of the proposed methodology (ML_HARD thresholds) for the test site of Castellon Espadán – Spain.....	86
Figure 25: Overview of the results of the proposed methodology for the test site of Thessaloniki – Greece.	88



Figure 26: Overview of the results of the proposed methodology for the test site of Nochten Reichwalde Germany.	88
Figure 27: Overview of the results of the proposed methodology for the test site of Komotini – Greece.	89
Figure 28: Overview of the results of the proposed methodology for the test site of Poland.	89
Figure 29: Overview of the results of the proposed methodology for the test site of Castellon Espadán – Spain.	90
Figure 30: Workflow design to produce an intermediate layer (red box) Other Land Cover.	91
Figure 31: Workflow design to produce an intermediate layer (red box) Forest Extent of 2018.	92
Figure 32: Workflow design to produce an intermediate layer (red box) Croplands.	94
Figure 33: Workflow design to produce intermediate layer (red box) Protected areas.	95
Figure 34: Workflow design to produce intermediate layer (red box) Impervious areas.	96
Figure 35: Workflow design to produce an intermediate layer (red box) Changed areas.	96
Figure 36: Workflow design for the production of an intermediate layer (red box) “ML HARD”.	98
Figure 37: Workflow design for the implementation of the first phase (Hard layers) of the proposed methodology.	99
Figure 38: Workflow design for the computation of terrain and soil indicators. Source: personal compilation of Jesús Torralba Pérez.	102
Figure 39: Workflow design for the computation of sustainability indicators. Source: personal compilation of Jesús Torralba Pérez.	104
Figure 40: Workflow design for the computation of productivity indicators. Source: personal compilation of Jesús Torralba Pérez.	105
Figure 41: Workflow design for the computation of final MLs. Source: personal compilation of Jesús Torralba Pérez.	107



Figure 42: Implementation of workflow for the delineation of “other land cover” in ArcGIS Pro model builder.....	108
Figure 43: Implementation of workflow for the delineation of “Forest areas” in ArcGIS Pro model builder.	109
Figure 44: Implementation of workflow for the delineation of “Croplands” in ArcGIS Pro model builder.	110
Figure 45: Implementation of workflow for the delineation of “Changed Areas” in ArcGIS Pro model builder.....	111
Figure 46: Implementation of workflow for the delineation of “Impervious Areas” in ArcGIS Pro model builder.....	111
Figure 47: First step for the implementation of the workflow to produce the final “ML_HARD”.	112
Figure 48: Summarize of "Forest", "Croplands", "Impervious" & "Changed areas" layers. ..	112
Figure 49: Selection and reclassification by conditionals in ArcGIS pro raster calculator tool.	113
Figure 50: Second step for the implementation of the workflow to produce the final “ML_HARD”.....	114
Figure 51: Implementation of workflow for the delineation of “Depth available roots” in ArcGIS Pro model builder.	114
Figure 52: Implementation of workflow for the delineation of “texture” in ArcGIS Pro model builder.	115
Figure 53: Implementation of workflow for the delineation of “stoniness” in ArcGIS Pro model builder.	115
Figure 54: Implementation of workflow for the delineation of “Water capacity” in ArcGIS Pro model builder.....	116
Figure 55: Implementation of workflow for the delineation of “clay” in ArcGIS Pro model builder.	116



Figure 56: Implementation of workflow for the delineation of “sand” in ArcGIS Pro model builder.	117
Figure 57: Implementation of workflow for the delineation of “acidity (pH)” in ArcGIS Pro model builder.	117
Figure 58: Implementation of workflow for the delineation of “erosion” in ArcGIS Pro model builder.	118
Figure 59: Implementation of workflow for the delineation of “sodicity”, “toxicity contamination” and “natural toxicity” in ArcGIS Pro model builder.	119
Figure 60: Implementation of workflow for the delineation of “soil organic matter” in ArcGIS Pro model builder.....	119
Figure 61: Implementation of workflow for the delineation of “productivity: grasslands & forest” in ArcGIS Pro model builder.	120
Figure 62: Implementation of workflow for the delineation of “Flooding” in GEE.	121
Figure 63: Implementation of workflow for the delineation of “Cation Exchange Capacity (CEC)” in GEE.....	121
Figure 64: Implementation of workflow for the delineation of “Dryness or Aridity Index” in GEE.	122
Figure 65: Implementation of workflow for the delineation of “MLs” in GEE.	123
Figure 66: Layer "ML_Hard" for Europe where the orange colors represent the areas defined as marginal excluding land use "impervious", "croplands", "forest", "protected areas" and "changed". Source: personal compilation of Jesús Torralba Pérez.	130
Figure 67: Layer "ML_Hard" for Germany, Poland, Spain, and Greece where the orange colors represent the areas defined as marginal excluding land use "impervious", "croplands", "forest", "protected areas" and "changed". Source: personal compilation of Jesús Torralba Pérez. .	131
Figure 68: Representation of the land uses of the S2GLC layer (left column) and the areas classified as marginal lands according to the methodology proposed in the MAIL project (right column). The pilot sites are: a) Staszów (Poland), b) Isenli (Greece), c) Welzow (Germany), d)	



Nochten (Germany), e) Nogueruelas (Spain), f) Espadan (Spain) and g) Soria (Spain).
 Personal source: Jesús Torralba Pérez.132

Figure 69: Percentage of marginal land for each land use over the total marginal land per test site. Land use has been obtained from the S2GLC layer. Source: personal compilation of Jesús Torralba Pérez.133

Figure 70: Final layer of MLs classified into three categories “Marginal lands with high plantation suitability”, “Marginal lands with low plantation suitability” and “Potentially unsuitable lands” with the method a - computing the maximum and minimum and dividing the range of values by 3. Source: personal compilation of Jesús Torralba Pérez.137

Figure 71: Final layer of MLs classified into three categories “Marginal lands with high plantation suitability”, “Marginal lands with low plantation suitability” and “Potentially unsuitable lands” with the method b - computing the 25th and 75th percentile and setting these values as threshold limits. Source: personal compilation of Jesús Torralba Pérez.138

Figure 72: Final layer of MLs classified into three categories “Marginal lands with high plantation suitability”, “Marginal lands with low plantation suitability” and “Potentially unsuitable lands” with the method c - computing the 33rd and 66th percentile to keep the same number of pixels in each category. Source: personal compilation of Jesús Torralba Pérez.139

Figure 73: Final layer of MLs classified with 3 methods into 3 categories “Marginal lands with high plantation suitability” (green), “Marginal lands with low plantation suitability” (gold), and “Potentially unsuitable lands” (dark red) in the pilot sites. Source: personal compilation of Jesús Torralba Pérez.140

Figure 74: Percentage for each of the marginal land categories according to marginality (“Marginal lands with high plantation suitability”, “Marginal lands with low plantation suitability”, and “Potentially unsuitable lands”) for each pilot site and approach. Source: personal compilation of Jesús Torralba Pérez.....142



ANNEX III: LIST OF TABLES

Table 1: Factors for marginal land identification under <i>MAIL</i> scope. Source: Personal compilation.....	13
Table 2: Constrains for marginal land identification under <i>MAIL</i> scope Source: Personal compilation.	13
Table 3: Technical specifications of CLC dataset. Source: <i>MAIL</i> Deliverable D2.2.	14
Table 4: Technical specifications of Pan-European High-Resolution Layers (HRL) dataset. Source: <i>MAIL</i> Deliverable D2.2.	16
Table 5: Technical specifications of Digital European elevation model dataset. Source: <i>MAIL</i> Deliverable D2.2.....	17
Table 6: Technical specifications of Environmental dataset. Source: <i>MAIL</i> Deliverable D2.2.	17
Table 7: Technical specifications of the European Soil Database Derived data dataset. Source: <i>MAIL</i> Deliverable D2.2.....	19
Table 8: Technical specifications of the ISRIC World Soil Information dataset.	20
Table 9: Technical specifications of the S2CLG layer.....	21
Table 10: Main specifications of the datasets used. Source: Personal compilation.....	22
Table 11: The subset of 24 studies on which the indicator selection was based on.	31
Table 12: Soil and terrain marginality indicators used in <i>MAIL</i>	33
Table 13: Sustainability marginality indicators used in <i>MAIL</i>	33
Table 14: Productivity marginality indicators used in <i>MAIL</i>	33
Table 15: Terrain and soil indicators and thresholds, sorted by importance (times used in the literature review).....	38
Table 16: Sustainability indicators and thresholds, sorted by importance (times used in the literature review).....	41



Table 17: Productivity indicators and thresholds, sorted by importance (times used in literature review).	43
Table 18: Climate indicators and thresholds, sorted by importance (times used in the literature review).	45
Table 19: The ranks given to each indicator, in order to calculate the weights.....	46
Table 20: The pairwise comparison matrix of the marginality indicators, based on the ranks.	48
Table 21: The normalized pairwise comparison matrix and the calculated weights.....	49
Table 22: Terrain and Soil indicator classes, scores, and weights.	50
Table 23: Sustainability indicator classes, scores, and weights.	50
Table 24: Productivity indicator classes, scores, and weights.....	51
Table 25: Additional datasets for use in <i>MAIL</i>	53
Table 26: Datasets for potential use in <i>MAIL</i> , as an addition to datasets from T2.2.	53
Table 27: Final datasets used in <i>MAIL</i>	54
Table 28: Overview of the Land Cover Map of Europe 2017 technical specifications.	66
Table 29: Other Land Cover Types that are not Marginal Lands.....	67
Table 30: Overview of the available LULC datasets for delineating Water bodies.....	68
Table 31: Overview of the proposed thresholds according to FAO forest definition.	69
Table 32: Forest definition parameters for EU countries (ANNEX II EU Regulation 2018/841).	69
Table 33: Overview of the available LULC datasets for delineating Forest Areas.	70
Table 34: Overview of the available LULC datasets for delineating Croplands.	74
Table 35: Overview of the available LULC datasets for delineating impervious areas.	75
Table 36: Overview of the available datasets for the delineation of protected areas.....	77



Table 37: Overview of the different types and categories of indicators used in the MAIL project.	78
Table 38: Overview of the used datasets for the delineation of additional marginality indexes.	79
Table 39: Overview of the finally selected marginality indexes.	82
Table 40: Slope thresholds according to each type of test sites.....	83
Table 41: Overview of the new thresholds and marginality indexes developed within the MAIL project.	84
Table 42: The classes of the results of MLs.....	87
Table 43: Areas in hectares of each type of MLs for the selected test sites.	90
Table 44: Soft layer summary.....	100
Table 45: Methods to subdivide by type of marginality (“Marginal lands with high plantation suitability”, “Marginal lands with low plantation suitability” and “Potentially unsuitable lands”).	123
Table 46: Total surface of the countries that are part of the MAIL project, surface of marginal lands by country, and percentage that represent the marginal lands within the surface of the country.	129
Table 47: Hectares and percentages of land use and marginal land by pilot site and land use. The land use has been obtained from S2GLC layer.	135
Table 48: Hectares and percentage of total area for each of the marginal land categories according to marginality (“Marginal Lands with high plantation suitability”, “Marginal Lands with low plantation suitability”, and “Potentially unsuitable lands”) for each pilot site and method.	141
Table 49: CORINE land cover categories. Source: Copernicus Land Monitoring Service web page.....	151



ANNEX IV: DATABASE STRUCTURE

The following data structure is available in the ftp provided by AUTH:

a) Mail_WP2²⁶

b) Task2_3

c) Hard Reference to “Hard” layers

d) Hard_final_layer

NO_FICC.zip: No forest, no impervious, no crops and no changes areas.

Hard_final_layer_tiles.zip: Final layer for EU.

d) Changed_areas_1_0.zip: Final “Changed areas” layer with 1 and 0 values.

d) Crops_0_1.zip: Final “Crops” layer with 1 and 0 values.

d) forest2018_1_0.zip: Final “forest” layer with 1 and 0 values.

d) Impervious_1_0.zip: Final “impervious” layer with 1 and 0 values.

d) Protected_areas_1_0_tiles.zip: Final “protected areas” layer with 1 and 0 values.

d) s2glcotherlc: “S2GLCotherLC” contains all the S2GLC classes except for “water bodies”, “permanent snow”, “marshes” and “peat bogs”.

d) Test_sites_HARD

e) AOI Area of interest

Germany

Notchen.shp

Welzow.shp

Greece *Isenli_Forest_bondary_WGS84.shp*

Poland *Test_site_poland2.shp*

Spain

Soria_Pilot_Area.shp

Study_area_Noguruelas_WGS84.shp

Study_area_espadan_buffer.shp

e) Hard

Espana_HARD_tile.tif: hard tiles for Spain

²⁶ Bold is used to name the folders. In blue and *italics*, the different files are named.



Espadan_tSite_HARD.tif: Clipping of the hard layer for Espadán

Nogueruelas_tSite_HARD.tif: Clipping of the hard layer for Nogueruelas

Soria_tSite_HARD.tif: Clipping of the hard layer for Soria

Germany_HARD_tile.tif: hard tiles for Germany

Germany_tSite_HARD.tif: Clipping of the hard layer for Welzow

Germany_tSiteNochten_HARD.tif: Clipping of the hard layer for Nochten

Greece_HARD_tile.tif: hard tiles for Greece

Greece_tSite_HARD.tif: Clipping of the hard layer for Isenli

Poland_HARD_tile.tif: hard tiles for Poland

Poland_tSite_HARD.tif: Clipping of the hard layer for Poland test site

e) Hard_cut_s2gl Clipping of the S2GLCotherLC layer with the Hard layer for test sites

Espadan_tSite_s2gl.tif

Nogueruelas_tSite_s2gl.tif

Soria_tSite_s2gl.tif

Germany_tSite_s2gl.tif

Germany_tSiteNochten_s2gl.tif

Greece_tSite_s2gl.tif

Poland_tSite_s2gl.tif

e) s2gl Clipping of the S2GLCotherLC layer for test sites

s2gl_espadan.tif

s2gl_nogueruelas.tif

s2gl_Soria.tif

s2gl_Germany.tif

s2gl_Germany_Nochten2.tif

s2gl_Greece.tif

s2gl_Poland.tif

e) *Area_test_site.xlsx*

e) *Test_sites_hard_s2gl.aprx*



c) 3_Soft.zip

1_Depth_available_roots

STU_Eu_DEPTH_ROOTS.tif

STU_EU_DEPTH_ROOTS_REC_score.tif

STU_EU_DEPTH_ROOTS_REC_scoreW.tif

2_Stoniness

STU_EU_T_GRAVEL_REC_scoreW03.tif

STU_EU_T_GRAVEL_REC_score.tif

STU_EU_S_GRAVEL_REC_scoreW03.tif

STU_EU_S_GRAVEL_REC_score.tif

3_Silt

STU_EU_T_SILT.rst

3_Texture

STU_EU_T_SILT_CLAY_REC_ScoreW045.tif

STU_EU_T_SILT_CLAY_REC_Score.tif

STU_EU_T_SILT_CLAY.tif

STU_EU_S_SILT_CLAY_REC_ScoreW045.tif

STU_EU_S_SILT_CLAY_REC_Score.tif

STU_EU_S_SILT_CLAY.tif

4_Clay

STU_EU_T_CLAY_REC_ScoreW015.tif

STU_EU_T_CLAY_REC_Score.tif

STU_EU_S_CLAY_REC_scoreW015.tif

STU_EU_S_CLAY_REC_score.tif

5_Sand

STU_EU_T_SAND_REC_scoreW015.tif

STU_EU_T_SAND_REC_score.tif

STU_EU_S_SAND_REC_score.tif

STU_EU_S_SAND_RE_scoreW015.tif

6_Orders_STU_EU_Layers_NOSTUDY

STU_EU_T_TEXT_CLS.rst

STU_EU_T_BD.rst

STU_EU_S_TEXT_CLS.rst



STU_EU_S_BD.rst

STU_EU_ALLOCATION.rst

6_Total_Available_water

STU_EU_T_TAWC_REC_scoreW02.tif

STU_EU_T_TAWC_REC_score.tif

STU_EU_S_TAWC_REC_scoreW02.tif

STU_EU_S_TAWC_REC_score.tif

SMU_EU_T_TAWC_REC_scoreW02.tif

SMU_EU_T_TAWC_REC_score.tif

SMU_EU_S_TAWC_REC_scoreW02.tif

SMU_EU_S_TAWC_REC_score.tif

7_Soil_Drainage

wrb_EU_r.tif

wrb_EU.tif

wrb.tif

Soil_Dranage_WRB_REC.tif

8_Soil_Moisture

HWSD.tif

10_Soil_acidity

ph_EU3035_REC_scoreW.tif

11_Soil_erosion

ErRECscoreW2_Compress.tif

ErRECscoreW2.tif

12_Flooding

Original_flooding_GEE.zip

13_Sodicity

ESP_eu_Rw.tif

ESP_eu_R.tif

14_Toxicity_Contamination

TOTN_eu_Rw.tif

TONT_eu_R.tif

15_Natural_Toxicity

Gyps_eu_Rw.tif



Gyps_eu_R.tif

16_Dryness

Dryness.zip

17_SoilOrganicMatter

STU_EU_T_OC_REC_scoreW03.tif

STU_EU_T_OC_REC_score.tif

STU_EU_S_OC_REC_scoreW03.tif

STU_EU_S_OC_REC_score.tif

18_Caption_Exchange_Capacity

18_Caption_Exchange_Capacity.zip

19_Productivity

Grass1_3035_REC_scoreW01.tif

Grass1_3035_REC_score.tif

Forest1_3035_REC_scoreW01.tif

Forest1_3035_REC_score.tif

2Approach

Sum.tif

2Approach_Sum_REC.tif

2A_STU_EU_T_SILT_CLAY_REC.tif

2A_STU_EU_T_OC_REC2.tif

2A_STU_EU_T_OC_REC.tif

2A_STU_EU_S_SILT_CLAY_REC.tif

2A_STU_EU_S_OC_REC2.tif

2A_STU_EU_S_OC_REC.tif

2A_STU_EU_DEPTH_ROOTS_REC_score.tif

c) Final_MLs_Layer

d) MLs_Div3

MLs_Europe_Div3-0000458752-0000131072.tif: There are 45 tiles for all Europe

e) Test_sites_Div3

MLs_Welzow_div3.tif

MLs_Soria_div3.tif

MLs_Poland_div3.tif



MLs_Nogueruelas_div3.tif

MLs_Nochten_div3.tif

MLs_Grecia_div3.tif

MLs_Espadan_div3.tif

d) MLs_Europep2575

MLs_Europe_p2575-0000458752-0000131072.tif: There are 44 tiles for all Europe

e) Test_sites_2575

MLs_Welzow_p2575.tif

MLs_Soria_p2575.tif

MLs_Poland_p2575.tif

MLs_Nogueruelas_p2575.tif

MLs_Nochten_p2575.tif

MLs_Greece_p2575.tif

MLs_Espadan_p2575.tif

d) MLs_Europep3366

MLs_Europe_p3366-0000458752-0000065536.tif: There are 42 tiles for all Europe

e) Test_sites_3366

MLs_Welzow_p3366.tif

MLs_Soria_p3366.tif

MLs_Poland_p3366.tif

MLs_Nogueruelas_p3366.tif

MLs_Nochten_p3366.tif

MLs_Greece_p3366.tif

MLs_Espadan_p3366.tif