D2.7 Report on Augmenting precision

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

MAIL project has received funding from the European Union’s Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie grant agreement No 823805; [H2020 MSCA RISE 2018]

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.	D2.7
Document name	MAIL_D2.7.pdf
Deliverable name	Report on Augmenting precision
Work Package	WP2
Nature[1]	R
Dissemination[2]	PU
Editor	Michal Krupinski (CBK PAN) Dzhaner Emin (IABG) Charalampos Georgiadis (AUTH)
Authors	Georgios Spanos (AUTH) Ewa Gromny (CBK PAN) Michal Krupinski (CBK PAN)
Contributors	All project partners
Date	28.12.21

[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

Aristotle University of Thessaloniki (AUTH) Greece	Industrieanlagen Betriebsgesellschaft MBH (IABG) Germany
Gounaris N. – Kontos K. OE (HOMEOTECH) Greece	Centrum Badan Kosmicznych Polskiej Akademii Nauk (CBK PAN) Poland
Universitat Politecnica de Valencia (UPV) Spain	Fundacion Centro De Servicios Y Promocion FOrestral Y de su Industria De Castilla y Leon (CESEFOR) Spain

ABBREVIATIONS

Term	Explanation
EVI	Enhanced Vegetation Index
GEE	Google Earth Engine
GRD	Ground Range Detected
IW	Interferometric Wide Swath
K	Kappa
LC	Land Cover
ML	Marginal Land
MSI	Multispectral Instrument
NDTI	Normalized Difference Tillage Index
NDVI	Normalized Difference Vegetation Index
NIR	Near Infra-Red
nML	Non-Marginal Land
OA	Overall Accuracy
OOB	Out-of-bag (error estimate)
PA	Producer's Accuracy
RF	Random Forest
S2GLC	Sentinel-2 Global Land Cover
SAR	Synthetic Aperture Radar

SAVI	Soil Adjusted Vegetation Index
SR	Surface Reflectance
UA	User's Accuracy

Contents

EXECUTIVE SUMMARY

The methodology that was developed in Task 2.3 of the ***MAIL*** project for the detection of potential marginal lands (ML) involves the search, acquisition, and processing of 36 environmental and socio-economic factors/layers from various sources. Additionally, this analysis was based on 2017-’18 data and each of these products is updated with different frequency, which can vary from yearly to every 5 or 10 years, depending on the scale of the product. The scope of this Task is to develop a methodology for the detection of MLs with freely available satellite imagery, by harnessing the high spatial and temporal resolution of these missions.

Therefore, a new tool was developed on the Google Earth Engine for the purposes of the Task 2.8 “Augment precision in MLs detection”, which is essentially a supervised classification algorithm. The classifier utilized for this study is the Random Forest supervised classifier, and the methodology exploits the Sentinel-1 (GRD) and Sentinel-2 (SR, L2A) image collections as well as some Hard Layers developed in previous Tasks. A similar with the Task 2.4 accuracy assessment workflow was also developed for this Task, since a direct comparison with the outcome of 2.3 would take place to validate the performance of the tool.

The “Enhanced Classification” tool manifested a comparable accuracy with the results of Task 2.3, although it did not manage to surpass it in terms of the accuracy metrics that were computed. This inferiority was hypothesized to be due to the quality of the training sampling data for the ML class. Thus, a case study was performed deploying a portion of the reference data of Task 2.4 for training and the rest for validation in the accuracy assessment. The accuracy assessment manifested a clear increase in all of the investigated metrics, revealing that the “Enhanced classification” tool is a reliable tool for the detection of MLs, if training data of high quality is provided. Most importantly though, it is a tool that extends the functionality of the detection methodology developed in Task 2.3 only for 2017-’18, to any point in time from 2017 onwards without the tedious process of acquiring a bundle of bulky data.

1. INTRODUCTION

The definition of marginality, or more precisely marginal land, differs according to the final goal of the study. In monetary terms marginal lands are those lands which have a negative economic output, e.g. the investment of resources into the land is higher than expected monetary gain. Ecological definitions on the other hand define marginal lands in terms of their biophysical resource composition (often the scarcity thereof). (D2.1 “Literature review and existing models report” and D2.3 “Report on Methodology development”). Because of the big variations in its definition, more than 130 different factors (variables) have been related to the identification of MLs. For the development of the detection methodology for MLs in Task 2.3, more than 30 distinct layers acquired from various sources across Europe were deemed crucial for the identification of MLs. These layers were used either as “hard” or as “soft” constraints inside a ML classification procedure. Each factor was given a specific weight based on its perceived or estimated importance for the definition of marginality. The entire marginal land mapping methodology can be seen in Deliverable 2.3. It was developed based solely on fixed in time data, for the year 2017-2018.

The scope of this Task is to develop a methodology to augment the precision of marginal lands map by combining the outcomes of the Tasks 2.3 and 2.4. Nevertheless, one key difference of the methodology that will be followed in Task 2.8 compared to the one followed in Task 2.3, is the source of data; instead of acquiring data from various sources only freely available satellite data will be used. Consequently, instead of the tedious process of acquiring a large volume of specialized and sophisticated data for the whole Europe, which also come with separate inherent uncertainties and are updated in sparse and unequal intervals, only the readily available satellite imagery will be acquired. Thus, only a single source of uncertainty is introduced in the analysis, which is also typically documented for each of these products. More importantly though, the competitive superiority of this methodology is the ability to harness the temporal capabilities of satellites, providing up-to-date thematic maps for any point in time, instead of a fixed point in time.

Having said that, the aim of this Task is to augment the precision of the detection of MLs by utilizing publicly available satellite data. More specifically the objectives of this study are to:

- Develop a methodology for an “on-the-fly” detection of MLs applying a time series analysis based on satellite imagery on the Google Earth Engine platform,
- Assess the performance of the algorithm by deploying the validation polygons of Task 2.4, and
- Evaluate the potential of the algorithm in a case study using input (training) data of higher quality.

2. LITERATURE REVIEW

To identify the most robust methods for identification of marginal lands using satellite imagery, a detailed literature review was performed. Most of the research focused on national level (e.g. (Peterson & Aunap, 1998), (Löw et al., 2015), (Kolecka, 2021)) and high resolution satellite imagery delivered by Landsat satellites (e.g. (Peterson & Aunap, 1998), (Dara et al., 2018), (Yin et al., 2018)). In most of the analysed research papers, time series of NDVI were used to identify classes related with land abandonment; mostly from agricultural lands. Sensors, locations, and algorithms applied are summarized in Table 1.

Table 1. The summary of literature review about marginal lands detection using satellite imagery.

Paper	Satellite data	Area of interest	Summary
(Peterson & Aunap, 1998)	Landsat	Estonia	Knowledge based, step by step, classification approach. Starting from classification into land and water, forest and no forest, masking out wetlands, urban areas and mines. Result presents difference in area of arable lands on district level between to moments in time: 1990 and 1993.
(Kuemmerle et al., 2008)	Landsat	Poland, Slovakia, Ukraine	Firstly forest, water and built-up areas are masked out. Secondly, supervised change detection algorithm based on SVM is applied on three classes: unchanged areas, fallow land, afforestation areas. Overall accuracy reached 91%.
(Bai et al., 2008)	AVHRR	Argentina, China, Cuba, Senegal, South Africa, and Tunisia	Analyse of NDVI time series, adjusted with the rainfall. Land degradation is indicated by a declining trend of climate-adjusted net primary productivity.
(Alcantara et al., 2013)	MODIS	Eastern Europe	Supervised SVM-based algorithm applied on NDVI product (8-days composites) and four classes: active agriculture, forest, abandoned agriculture, other. Overall accuracy reached 49%.
(Liu et al., 2015)	NOAA, GIMMS NDVI3g	Earth	Analysis of greening and browning trends on global scale, divided into three periods of time (1982–1994, 1995–2004 and 2005–2012).

(Löw et al., 2015)	Landsat	Kazakhstan	Annual classifications for period 2009-2014. Analysis of trajectory of changes to detect crop abandonment into 3 classes: • five year lasting cover of shrubs or bare soil; • transitions of bare soil to herbaceous vegetation • transitions from herbaceous vegetation to shrub (both assumed to indicate different stages of an on-going succession of vegetation on abandoned fields.
(Carlson et al., 2017)	MODIS, Landsat	Ecrins National Park (French Alps)	Identification of greening trends on a small area in high Alps based on maximum NDVI trend estimations.
(Dara et al., 2018)	Landsat	Northern Kazakhstan	Investigation of cropland and non-cropland classes, based on Random Forest classifier and temporal segmentation of LandTrend with annual time series of probabilities.
(Yin et al., 2018)	Landsat	Georgia and the North Caucasian Federal District of Russia	Spatial (in eCogniition) and temporal (LandTrend) segmentation used to estimate the probability of agricultural land calculated from Random Forest based model. Identified classes included: agricultural land abandonment, stable agricultural land, fallow and re-cultivation.
(Kolecka, 2021)	Landsat	Poland	Greening as an indicator of agricultural land abandonment. Classification based on LandTrend and aggregated annual and seasonal spectral indices.

3. DATA ACQUISITION

One of the outcomes of the ***MAIL*** project is the a web-application (MAIL Map Portal), encompassing all the tools that have been developed for the purposes of this project, in the form of thematic maps and tools. For the objectives of such a project dealing with Pan-European GIS and Earth Observation data, the *Google Earth Engine* was deemed as the most appropriate platform for a set of reasons.

Google Earth Engine (GEE) is a cloud-based platform designed to make planetary-scale geospatial analysis possible not only for remote sensing experts, but also for a much broader audience that lacks the technical background required to deploy supercomputers or large-scale cloud computing services. By utilizing Google's massive computational capabilities, researchers are able to work on large-scale investigations that have a high impact on the society and environmental management like deforestation, drought, natural disasters, diseases, water management, climate monitoring, etc. (Gorelick et al., 2017).

Therefore, the methodology for the task of "Augmenting precision in the detection of MLs" was developed on the GEE. The data catalog of Earth Engine entails an exhaustive repository of publicly available geospatial datasets, including collections from satellite and aerial imaging systems, as well as environmental, climatic, and socio-economic variables. Since one of the objectives of this study is to develop a methodology for the detection of MLs by utilizing freely available satellite imagery, the products of two Copernicus missions will be exploited; the *Sentinel-1* and the *Sentinel-2* image collections, readily available in the GEE platform, that have already been successfully deployed in numerous land cover and time series studies (Al-Nahmi et al., 2017; Huang et al., 2016; Kuc & Chormański, 2019; Lavreniuk et al., 2017; Mitri et al., 2020; Osgouei et al., 2019; Zollini et al., 2019).

The Sentinel-1 mission is a constellation of two sun-synchronous, polar-orbiting satellites, Sentinel-1A and Sentinel-1B, launched on the 3rd of April 2014 and 25th of April 2016, respectively. Bearing an active C-band Synthetic Aperture Radar (SAR) sensor, fixed at a 5.405 GHz frequency (corresponding to a wavelength of ~5.546 cm), they gain the advantage of being able to acquire imagery day and night, regardless of weather conditions (Geudtner et al., 2014). Sentinel-1 operates in four exclusive acquisition modes and supports operation either in single (HH or VV) or dual (HH + HV or VV + VH) polarization. However, the Interferometric Wide swath (IW) mode, with VV + VH polarization, is the main operational mode, offering a balance between consistent and efficient performance and decreasing operational costs. The spatial resolution in the IW mode of Sentinel-1 is 20 m and the revisit time of the constellation is six days at the Equator (ESA, 2020a).

The Sentinel-2 mission is also a constellation of two sun-synchronous, polar-orbiting satellites, Sentinel-2A and Sentinel-2B, launched on the 23rd of June 2015 and on the 7th of March 2017, respectively. Sentinel-2 is equipped with an optical Multispectral Instrument (MSI) acquiring images in 13 spectral bands: four bands at 10 m, six bands at 20 m and three bands at 60 m spatial resolution. The full mission of the twin satellites is designed in a way that they will be able to offer global coverage and a revisiting time of five days at the Equator (ESA, 2020b).

These missions provide image products of different levels of processing. Regarding the Sentinel-1 mission, the Level-1 processing Ground Range Detected (GRD) products were utilized in this task, which consist of focused SAR data that have been detected, multi-looked and projected to ground range using the Earth ellipsoid model WGS 84 (ESA, 2020c). The image collection in GEE is updated daily and new assets are ingested within two days after they become available. Earth Engine further applies the following pre-processing steps for each scene: *Apply orbit file, GRD border noise removal, Thermal noise removal, Radiometric calibration,* and *Terrain correction*. The latter terrain-corrected values are finally converted to decibels via log scaling (10*log10(x)) (Google developers, 2021). For Sentinel-2, two datasets can be found in the catalog of Google Earth Engine: The Level-1C orthorectified top-of-atmosphere reflectance and the Level-2A orthorectified atmospherically corrected surface reflectance. In order to minimize the atmosphere's scattering and absorption effects from the scenes, the Sentinel-2 Surface Reflectance (SR) image collection was used.

Another product that was deployed in this study is the Land Cover Map of Europe 2017, which is a product resulting from the Phase 2 of the Sentinel-2 Global Land Cover (S2GLC) 2017 project. The S2GLC 2017 product demonstrates the land cover classification of most of the European continent. The classification was performed using multi-temporal Sentinel-2 data collected during the year 2017 by applying the random forest classifier and existing land cover/use databases as the source of training samples. The S2GLC 2017 dataset is delivered with 10 m spatial resolution with a thematic legend composed of 13 land cover classes and a thematic overall accuracy of 86.1% on a continental scale (Malinowski et al., 2020).

In addition, one of the outcomes of the Task 2.3 is going to be integrated in this methodology, the binary mask with 2 classes (Marginal Lands and No Marginal Lands), called as "ML_Hard thresholds" layer in D2.3 Conejo et al., (2021). Starting from the whole European area, areas that do not meet the requirements of the definition of Marginal Lands are incrementally excluded based on land cover type (i.e., urban areas, protected areas, water and forest areas, areas covered with snow, and more). The outcome of this procedure is the "ML_Hard thresholds" layer, entailing all the potentially suitable for plantation MLs. The accuracy assessment performed on the Task 2.4 for the aforementioned layer manifested an overall accuracy of 67.73% for all the testing sites merged.

Finally, in order to be able to compare the performance of this methodology with the one followed in the Task 2.3, the ML and nML validation polygons provided for the scope of the Accuracy Assessment on Task 2.4 will be utilized. This dataset will also be a valuable asset for the Case Study described in detail in Chapter 6 "Case study".

4. WORKFLOW DEVELOPMENT

In this Chapter, the steps that were followed for the development of the MLs detection algorithm will be described. This however proved to be a challenging task, since the definition of marginality is a complex matter (Peter et al., 2018). For instance, land classified as marginal in a given place or time might be considered as productive (non-marginal) in the other spatiotemporal context (Ciria et al., 2019; Sallustio et al., 2018). Hence, marginality is always relative to a certain use e.g., crop production or livestock grazing (Lewis & Kelly, 2014).Moreover, marginality is a dynamic phenomenon and spatiotemporally static characterization of marginality is unable to capture the shifting character of some of the factors that constitute marginality (Nalepa & Bauer, 2012).

As it can be seen, categorization is closely related with constrains causing marginality and the study's goals. As detected in definitions, the categorization of marginal land is usually performed focusing on a single aspect of marginality; environmental including constrains for biological production such as hazards or biophysical limits or economical performing a simple cost analysis using specific crop.

The definition of MLs under the ***MAIL*** project as it was concluded on D2.1 "Literature review and existing models report" is: *"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. (…)* ***MAIL*** *project will focus on areas in which it is possible to convert them to forest lands.*"

Consequently, marginality is driven by three main forces: environmental factors, socioeconomic factors and cultural factors. A satellite image analysis, however, is based primarily on phenological characteristics, thus the socioeconomic and cultural factors are challenging to incorporate. Having said that, this task aims to develop an algorithm that will abide by the ***MAIL*** definition and take as many as possible of the factors that drive marginality into consideration, while bearing in mind the limitations that such task entails.

4.1 Exploratory steps

A series of trials were conducted for this reason until the finalization of the MLs detection algorithm and a brief overview of the steps taken will be given in this section. A literature review on the detection of MLs using remote sensing data manifested that the Normalized Difference Vegetation Index (NDVI) is frequently used as a proxy in such investigations, since a potential deviation from the norm may indicate land degradation and improvement (Bai et al., 2008).

The NDVI is a well-established and commonly used vegetation index in earth observation studies because it is roughly correlated with green plant biomass and vegetation cover (Box et al., 1989; Tucker, 1979) The NDVI is an index based on the relative reflectance values in the red and near infrared (NIR) spectrum and is computes as:

$$NDVI = \frac{\rho_{Red} - \rho_{NIR}}{\rho_{Red} + \rho_{NIR}}$$

The amount of solar energy reflected by vegetation cover in the red wavelength depends primarily on chlorophyll content, whereas the amount of solar energy reflected by vegetation in the NIR wavelength is influenced by the amount and the condition of green biomass, leaf tissue structure, and water content (Jensen, 1996). Initially, the Normalized Difference Vegetation Index (NDVI) values, as well as the values' trends, of known ML and nML areas (Task 2.4 validation polygons) were explored for a 5 years period (2015-2020), under the hypothesis that the NDVI values of ML and nML parcels would be distinct or show a consistent temporal trend pattern. Nevertheless, the NDVI values of the ML parcels exhibited a significant variance among the test countries, but also among the different parcels residing in the same country, and no specific trend could be detected. This led to the decision to integrate a different time series analysis, the harmonic model.

Harmonic analysis, also termed spectral analysis or Fourier analysis, decomposes a time dependent periodic phenomenon into a series of cosine waves (terms) and an additive term (Davis & Sampson, 1986; Rayner, 1971). Each wave is defined by a unique amplitude and phase, where the amplitude value is half the height of a wave, and the phase is the difference between the start of the year and the peak of the wave over the range 0–2π (Figure 1). Each harmonic term accounts for a proportion of the variance in the original time series data-set and much like the principal component analysis, the first two terms (components) entail the majority of the variance in a data set (Jakubauskas et al., 2002).

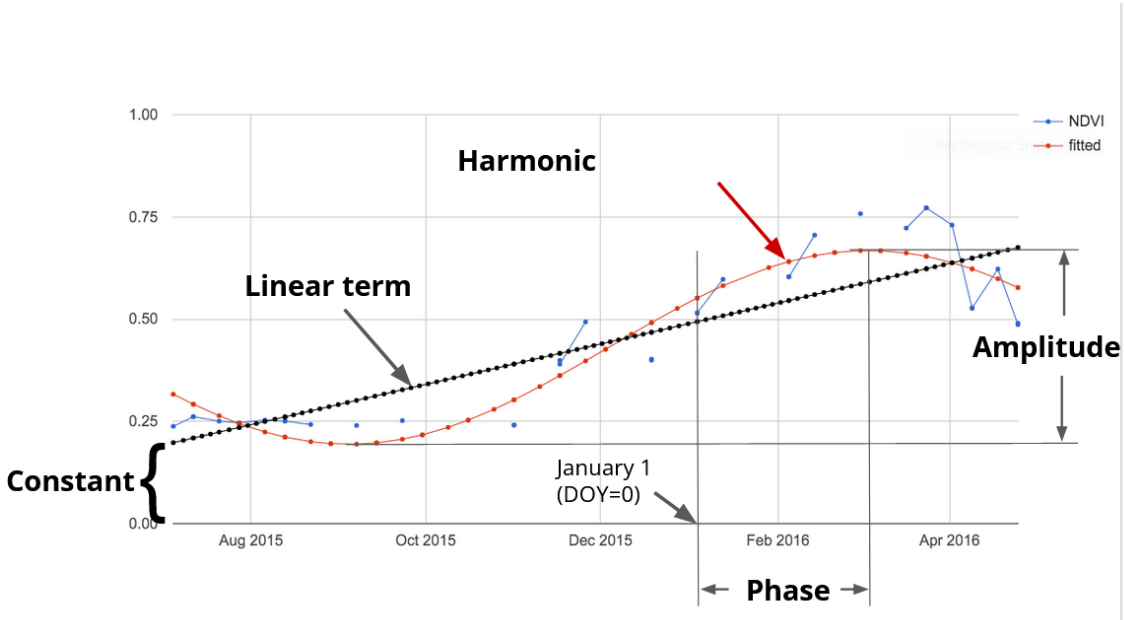


Figure 1. Harmonic model (Clinton, 2017)

In this step of the exploratory analysis, the hypothesis was that a scatterplot of amplitude vs phase values for a given area of various distinct land cover types, would manifest distinct clusters, representing each land cover (LC). In other words, it was expected that the urban LC for example would take over the lower part of the plot since no significant vegetation exists so the amplitude would be close to zero. Forest LC should exhibit a cluster in the high amplitude and medium phase part of the plot, and so on. This hypothesis was not rejected by visual interpretation of the exploratory results; however, human interpretation would fail to identify all the underlying patterns and aspects in such dataset. Hence, a machine learning classification algorithm was deemed necessary for an in-depth exploitation of the NDVI harmonic model time series analysis.

4.2 Classification Algorithm

The machine learning classification algorithm chosen for this study is the Random Forest (supervised classifier). Random Forest (RF) has been applied in many recent studies and real-life applications in a variety of domains, since it is a computationally efficient method capable of operating promptly over large datasets (Oshiro et al., 2012). Another reason that contributed in the popularity of this technique is the fact that they require a minimum input from the user, having only few parameters to tune, which makes it an attractive tool for people that are not experts on machine learning algorithms (Biau & Scornet, 2016). The RF classifier, also comes with a neat feature, the *out-of-bag* error estimate, which is computed on the observations set aside by the resampling before the tree building, which offers a quick way to check the performance of the model and fine tune the parameters, without the need of a validation set (Kruppa et al., 2013).

Random Forest is an unweighted ensemble classifier, meaning that it combines the decision of a set of classifiers by unweighted voting, to classify a given sample (Pal, 2007). As defined by Breiman, (1999) "*A random forest is a classifier consisting of a collection of tree- structured classifiers {h(x,Θk), k=1, ...} where the {Θk} are independent identically distributed random vectors and each tree casts a unit vote for the most popular class at input x*". Frequently, the user sets the number of trees by a trial and error manner, since this algorithm can work efficiently with large datasets and has been proven not to suffer from overfitting problems, although if the number of trees is significantly increased computational performance can take a toll (Oshiro et al., 2012).

After experimenting with various parameter values, number of variables and at different regions, a relatively high number of trees (1,000) was deemed appropriate for the needs of this investigation. This was mainly due to the big extent of the area that would be classified, which means that significant variations in landscape are expected to occur, but also because a smaller number of trees, like 100 – 500, exhibited higher variance in the out-of-bag error estimates, depending on the region. The rest of the parameters for the RF classifier in GEE were left in their default values.

4.3 Training data

For any supervised classification to perform as intended and reach its true potential in accuracy measures, attention needs to be paid to the samples that will serve as training data. Since the primary objective of this Task is the detection of MLs, a simple classification approach identifying the major prevailing land covers, like Forest, Croplands, Impervious, Water and MLs was followed. To delineate the regions and acquire the training samples for the Forest, Croplands and Impervious LC the Hard Layers generated in Task 2.3 will be utilized, while for the Water bodies LC, the "Water" class from the S2GLC product will be extracted, since these layers provide very high accuracy. For the classification of MLs, though, the 2.3 "ML_Hard_Thresholds" mask, and more specifically the classes that include all the potentially suitable MLs, will be sampled. This is the best and only source of data for the whole Europe that could be obtained for sampling MLs training data, even though this product manifested an ~68% overall classification accuracy in T2.4.

4.4 Trials and Errors

For the testing phase, the RF classifier was applied on Sentinel-2 imagery and a region in north Greece was selected as the testing site for a quick assessment of the classifier's performance, since most of the validation polygons offered for the T2.4 Accuracy Assessment reside in this area. Initially, the Amplitude and Phase values, extracted from the first two terms of the Harmonic analysis of the NDVI values, were used for the classification of the Sentinel-2 imagery, which resulted in a 49.3% out-of-bag (OOB) error estimate, a 51% Overall Accuracy (OA) and a Kappa (K) value of 0.02. This triggered the hypothesis that adding the mean NDVI values for each pixel would be an additional information for the algorithm, since the value of Amplitude, essentially describes the variance and not the actual NDVI values. Adding the NDVI as a classification variable improved the performance of the classifier (OOB: 35.9%) but not the accuracy (OA: 47.5%, K: -0.05).

For the next trial, two more vegetation indices were introduced to the algorithm, the Enhanced Vegetation Index (EVI) and the Soil Adjusted Vegetation Index (SAVI). EVI is similar to NDVI and is also used to quantify vegetation greenness. However, EVI corrects for some atmospheric conditions and canopy background noise and is more sensitive in areas with dense vegetation. In addition to the Red and NIR bands used in NDVI, it also incorporates an "L" value to adjust for canopy background, "C" values as coefficients for atmospheric resistance, and values from the blue band and is calculated from the following formula (A. Huete et al., 2002):

$$EVI = 2.5 * \frac{\rho_{NIR} - \rho_{Red}}{\rho_{NIR} + C_1 * \rho_{Red} - C_2 * \rho_{Blue} + L}$$

Where $C_1 = 6$, $C_2 = 7.5$ and L = 1.

SAVI is used to correct NDVI for the influence of soil brightness in areas where vegetative cover is low. SAVI is calculated as a ratio between the Red and NIR band values with a soil brightness correction factor (L) defined as 0.5 to accommodate most land cover types (A. R. Huete, 1988):

$$SAVI = \frac{\rho_{NIR} - \rho_{Red}}{\rho_{NIR} + \rho_{Red} + L} * (1 + L)$$

The classification including the Amplitude, Phase, mean NDVI, EVI and SAVI achieved an OBB error of 34.4%, an OA of 50.2% and a K of 0.004. For the next step, it was understood that since the focus of the classification is mainly on the "greenness", generating mean NDVI values for three different times in the year, namely a mean NDVI for May, July and September, could potentially introduce valuable information to the algorithm. For this reason, the following 5 variables were utilized by the RF classifier, Amplitude, Phase, NDVI_May, NDVI_July, NDVI_September, which then achieved an OOB of 33%, OA of 50.5% and K of 0.01.

Close observation to the results of the classification revealed that there was a constant misclassification of bare lands and mines with the Impervious LC. This is a common problem in Remote Sensing Land Cover Mapping applications since bare lands and build-up areas have similar spectral responses. Few indices have been developed to circumvent this challenge, however most of them, since they were developed for the latest Landsat missions, exploit the thermal band of the Landsat ETM+ sensor (As-syakur et al., 2012; Zha et al., 2003). In a recent study based on Sentinel-2 imagery Osgouei et al., (2019) experimented with incorporating the Normalized Difference Tillage Index (NDTI), which makes use of the SWIR bands of Sentinel-2, aiming to enhance the contrast between bare land and build-up areas. NDTI is computed as follows:

$$NDTI = \frac{\rho_{SWIR\,1} - \rho_{SWIR\,2}}{\rho_{SWIR\,1} + \rho_{SWIR\,2}}$$

Even though NDTI was first proposed by Van Deventer et al., (1997) for soil practices, tillage management, and crop residue mapping and has then been applied for agricultural practices and soil management, it exhibited high potential in separating impervious surfaces and build-up areas from bare lands in the multi-index classification study of Osgouei et al., (2019). Incorporating NDTI in this study further improved both the classifier’s performance and the respective classification accuracy, achieving an OOB error of 30%, an OA of 58.6% and a K of 0.17. Finally, the VV and VH polarization bands of the Sentinel-1 GRD IW mode were included to provide more information and further aid the classification capabilities of this machine learning algorithm.

These trials indicate the fact that no single index would be adequate for a large extent LC classification, like the one intended in this Task, but on the other hand not all indices are equally important. Moreover, it was realized that sufficient and relevant information was required for the classifier to meet its potential. More trials were conducted with various combinations of the aforementioned variables, aiming to conclude on the most efficient set of variables for this study and the one which yielded better and more stable results across the various environments (Mediterranean and Central European) is described on the following Chapters.

4.5 European training image

As it has been discussed, apart from diverse in spectral response, MLs are a dynamic and ever-changing “land cover”, therefore it would not be accepted to sample over the region that was classified as ML in T2.3, based on 2017-2018 data, in any point in time in the future. In other words, the area that has been delineated as potential ML for 2017-2018, will most probably change (either expand or shrink) in e.g., 10, 20 or 30 years, as well as the spectral response of the various features on earth. Consequently, a Pan-European image encompassing all the variables, based on 2017-2018 observations, required for the upcoming classification had to be generated.

The workflow that was followed for the generation of this Pan-European Basemap image is shown graphically on Figure 2 and is described in detail in this section. Two sets of imagery were exploited for the extraction of the required layers; the Sentinel-1 and Sentinel-2 image collections, readily available from GEE’s dataset catalogues. All processes that will be described later in this section, are based on imagery acquired by the sensors from 1st of April 2017 until 1st of April 2019. This way only data that are in accordance with the analysis made on T2.3 are included.

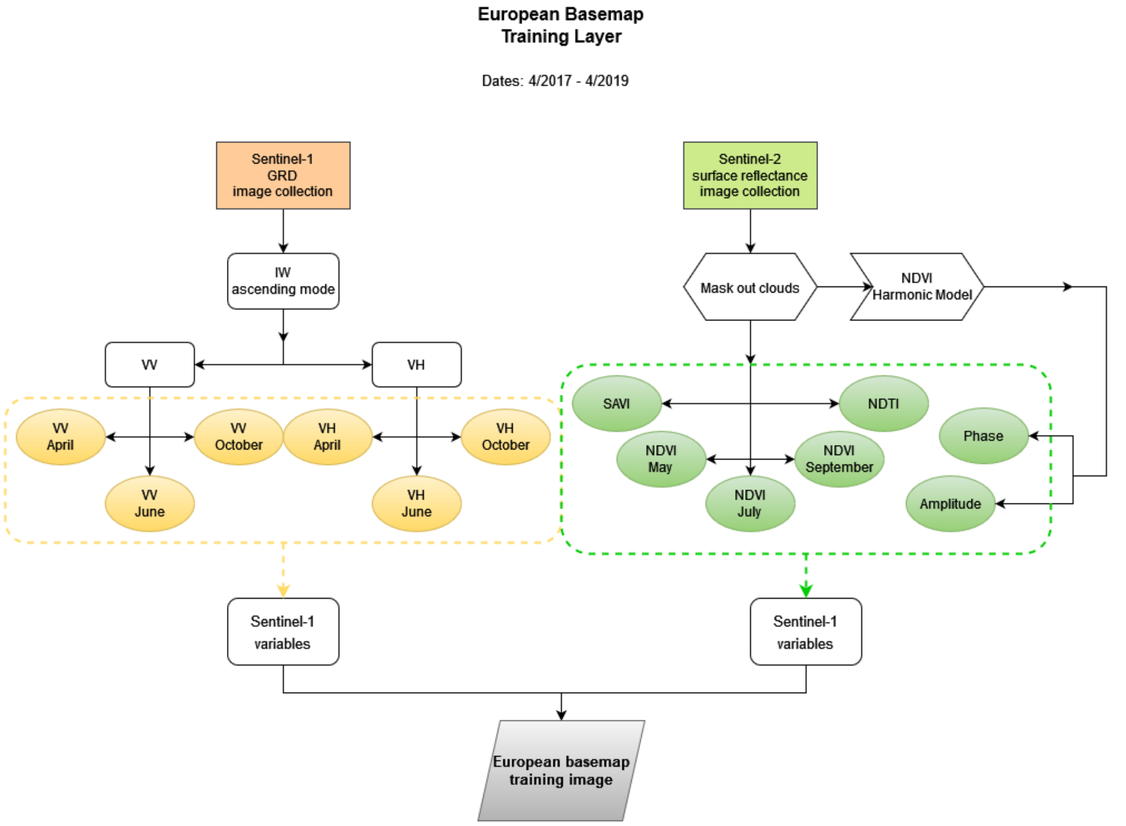


Figure 2. Pan-European Basemap training image workflow. Source: personal compilation of Georgios Spanos.

The Sentinel-1 imagery were initially filtered based on the instrument's mode and the orbit's direction pass and only the IW mode and the ascending pass images were acquired. Then, the mean values for every pixel for April, June and October were computed separately for the VV and the VH polarization acquisitions. All six layers that have been calculated, were ultimately concatenated in a single six-band image.

For the Sentinel-2 images a standard cloud masking script was applied before any processing to remove pixels that are influenced by clouds. Then the Phase and Amplitude parameters from the Harmonic Model based on NDVI values were computed for each pixel, as demonstrated by Nicholas Clinton in the Earth Engine User Summit of 2017. In addition, the mean NDVI value for each pixel was calculated for the months of May, July and September, since these months are related with the season that vegetation is at its highest in all of the regions of Europe. Finally, the SAVI and the NDTI also showed to improve the classifier's performance, hence mean values for the whole time range were computed for each pixel. In the end, these seven layers were added in a single seven band image.

After the Sentinel-1 and the Sentinel-2 images encompassing all the required layers were generated, they were concatenated in a single 13-band image. The procedure described above was iteratively applied for each country of the European Union (EU) and exported as an asset on GEE, but because of the memory restrictions in GEE, the spatial resolution was reduced to 100 meters. Nonetheless, this is not expected to affect the quality of this study, since the minimum area under the ***MAIL*** project identified as ML has already been defined in D2.3 as 1 hectare (Conejo et al., 2021). Finally, a mosaic of the 39 countries including all 13 bands (variables), namely: *Amplitude, Phase, NDVI_May, NDVI_July, NDVI_September, SAVI, NDTI, VV_April, VV_June, VV_October, VH_April, VH_June, VH_October*, was created and exported as a GEE asset. Having this image exported already as an asset on GEE provides a further advantage to the end-user of reducing the tool's computation time in the ***MAIL*** web application.

4.6 Classification Tool Development

Moving on to the classification workflow the user firstly draws a polygon to specify the area of interest and specifies the dates' range for which the analysis will take place. For this classification the user needs to define a starting and an ending date, in a way that at least two years of satellite imagery acquisitions are included. This means that the starting date cannot be less than two years prior to the current date. All the processing steps applied by the algorithm for the LC classification and ultimately for the detection of MLs are depicted on Figure 3.

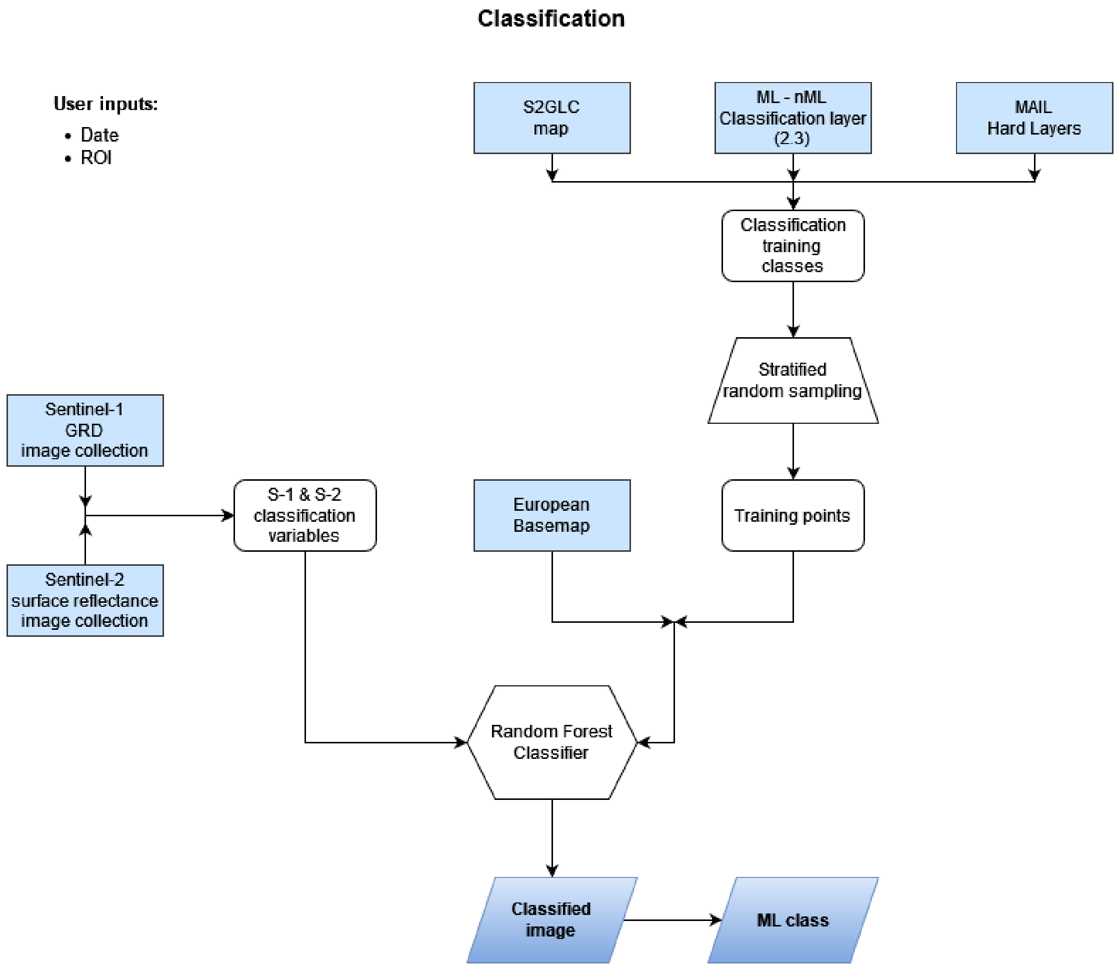


Figure 3. MLs Classification algorithm workflow. Source: personal compilation of Georgios Spanos.

Initially, the variables based on the Sentinel-1 GRD and the Sentinel-2 SR image collections are calculated, as they have been described in the previous section ("4.5 European training image") this time for the user-specified region and dates. This user-delineated image entailing all the classification variables will be referred from this point onwards as "*Classification Variables*" image. Then, an image containing all the relevant for this classification LCs is created, including the delineated regions of Forests, Croplands, Impervious, Water bodies and MLs as distinctive bands, from the ***MAIL*** D2.3 "Hard Layers", the S2GLC map and the "ML_Hard_Thresholds", as mentioned in Chapter "4.3 Training data". From here on this image is going to be referred as "*Training Raster*".

For the training process of the RF classifier, a stratified random sampling per class is performed on the Training Raster, creating 1,000 random points (training points) for each band (LC), in a 10 m. scale, which coincides with the resolution of the Training Raster, since the parent layers also have a 10 m spatial resolution. Afterwards, these training points are superimposed on the pre-developed Pan-European Basemap image, and the properties of each band are extracted to the training points.

The RF classifier then uses these points to train itself for the desired classes (LCs) based on the extracted properties (classification variables), by planting 1,000 decision trees. Now, the classification is performed on the Classifications Variables image and map depicting each LC in a distinctive colour is displayed followed by the map containing only the MLs class. It is in this stage that the RF classifier displays all information that is relevant with the classification, including the number of trees planted, the number of classes produced, the importance of each variable for the classification process and OOB error estimate. For the implementation of this classification tool in the MAIL WebApp and to distinguish it from the MLs detection methodology developed in Task 2.3, this tool will be referred as "*Enhanced Classification*".

4.7 Accuracy Assessment

In order to evaluate the performance of the developed MLs detection methodology in this Task, an accuracy assessment methodology was developed as well, and applied on the resulting maps. The fundamental objective of this study is to augment the precision in the detection of MLs. Thus, the primary goal of this assessment methodology was that it had to be as similar as possible with the one developed in T2.4 of the ***MAIL*** project, so that the resulting values of this Task are directly comparable with the ones from D2.4 "Report on Accuracy Assessment". The reason this is mentioned, is because the analysis of T2.4 was performed on the ArcGIS and R Studio software, while this Task is developed on the GEE platform. The series of steps followed for the realization of this assessment are shown as a flowchart on Figure 4.

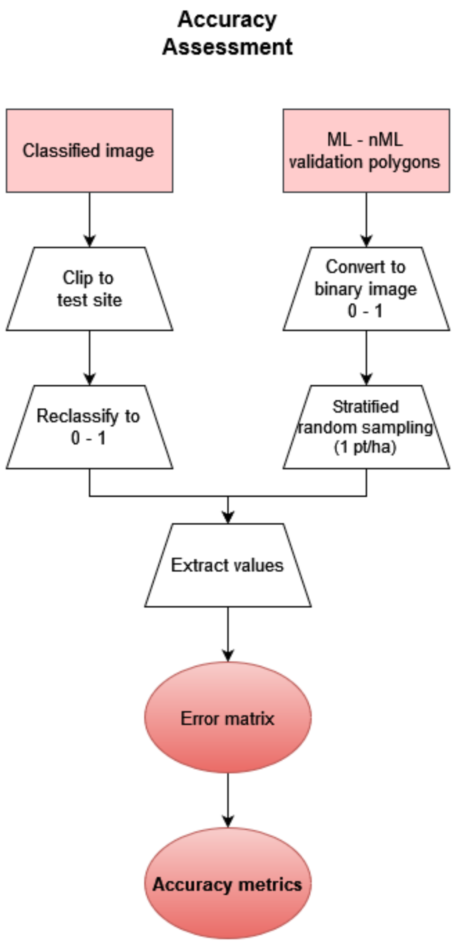


Figure 4. Accuracy Assessment methodology workflow. Source: personal compilation of Georgios Spanos.

In Task 2.4 the respective project partners from each country (Greece, Spain, Germany, and Poland) provided reference polygons designating ML and nML areas. The same polygons are used in this study as well and the assessment was carried out for test sites of each country separately. As a first step the classified image is clipped to the extent of the provided polygons for the according country and is then reclassified into 1 for the ML class and 0 for the rest of the classes (Forest, Croplands, Impervious, Water).

Later, the “ML – nML validation polygons” feature collection is converted to a binary image based on the ML/nML field (ML=1, nML=0) for a stratified random sampling to be performed setting a specified number of points (validation points) to be generated for each class. In D2.4 the authors concluded on a 1 point/ha sampling, so the resulting number of points allocated to each class depends on the total area of the provided validation polygons for each country, as shown in Table 2. In the case of Greece, the specific shape of the ML reference polygons resulted in a lower number of points than the area suggests. The validation points are then superimposed on the clipped, reclassified (0-1) image, and the values of classification product, as well as the ones from the validation polygons are extracted to the validation points.

Table 2. Sample points allocation per country

Country	ML		nML	
	Area [ha]	Allocated points	Area [ha]	Allocated points
Greece	7988	7988	5274	4966
Spain	1649	1649	2199	2199
Germany	352	352	20,914	20,914
Poland	539	539	2463	2463

Finally, an error matrix is generated based on which the accuracy metrics can be computed for each country; namely: the Overall Accuracy, the User's and Producer's Accuracy and the Kappa value, as well as the OOB Error Estimate generated by the RF classifier.

Overall Accuracy (OA) is the ratio between the correctly classified samples to the total number of samples. It essentially tells us what percentage of the reference sites were correctly mapped out of all of them (Congalton, 1991b).

User's accuracy (UA) is the proportion of the area mapped as a particular category that is actually that category "on the ground" (Congalton, 1991a). If a user employs the final map in order to locate a particular spatial unit, the user's accuracy gives the conditional probability of that map location actually representing the mapped unit. It is calculated by dividing the correct classified pixels in a class by the total number of pixels that were classified in that class (row total) and multiplying by 100 (Banko, 1998).

Producer's accuracy (PA) is the proportion of the area that is a particular LC class on the ground and it is also mapped as that class (Congalton, 1991a). The producer's accuracy measures how well a given area is classified and provides the producer of the final classification map with the conditional probability of a particular location of spatial unit appearing as that on the map. It is computed by dividing the number of correct pixels in one class by the total number of reference pixels for this class (column total) and multiplying by 100 (Banko, 1998).

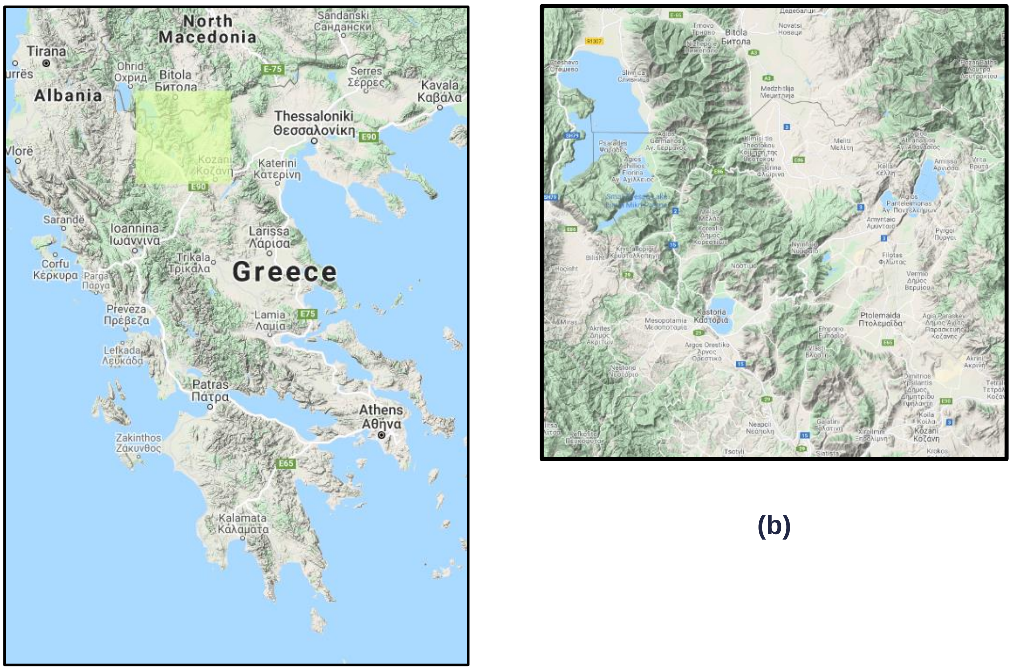
Kappa (K) is a measure of agreement between the predictions and the actual class. It can also be a comparison of the overall accuracy to the expected random chance accuracy. (Jensen, 1996). The statistical significance of any given classification matrix can also be determined by utilizing the Kappa coefficient as a basis. According to Cohen (1960), Kappa can be considered as the chance-corrected proportional agreement and takes values from +1 (perfect agreement) to -1 (complete disagreement).

5. Results

This tool can be applied on spatially different areas, either in size or geographical latitude and longitude, and on different time moments. For comparison, with the Task 2.3 outcomes, and assessment purposes, the tool was executed for the same years as the T2.3 methodology, 2017-2018, therefore, its performance can be visually and quantitively assessed and commented. For the qualitative assessment the resulting map from the LC classification output of the tool will be displayed for a given location, while for the quantitative assessment of the tool, the results of the accuracy assessment methodology will be presented for the four test countries.

5.1 Land Cover classification output

Regarding the LC classification and mapping, here we present as an example the resulting map of an arbitrarily selected site encompassing all desired classes, lying mostly in the territory of north Greece but also including some fractions from North Macedonia and Albania (Figure 5).



(a)

(b)

Figure 5. Location of example area

In Figure 6, the five classes of interest are depicted as produced by the Enhanced Classification tool. The Forest LC in green colour, the Croplands LC in orange, the Impervious in purple, Water in blue and MLs in red.

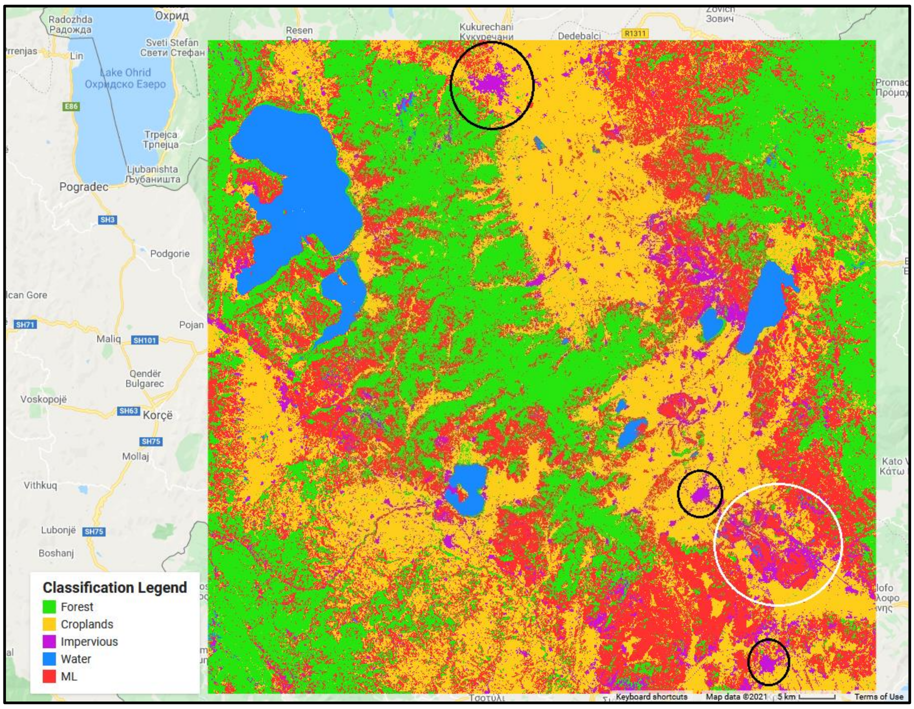
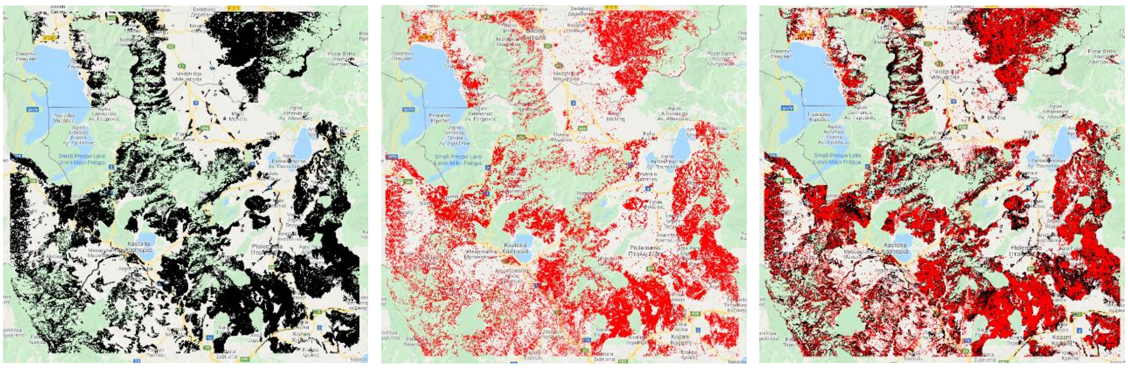


Figure 6. Output of enhanced classification algorithm

Starting with the Water class, which is a distinctively delineated feature, we can see that the water bodies are mapped very accurately following the boundaries of the lakes. The rest of the classes are more challenging to assess visually. However, by overlaying the classified map on the Satellite view of GEE, a match of the classes appended with reality is indeed present for the classes of Forest and Agriculture, even though the last one is an admittedly a diverse class in signature response. As mentioned in a previous chapter, a frequent source of confusion for the algorithm was the separation of Impervious areas from MLs. The region highlighted at the bottom right in a white circle is the coal mine of DEI (electricity provider of Greece), one of the biggest in Greece, and the confusion mentioned before is evident there. This is very much expected though, since the spectral signatures of these features are very similar. The typical Impervious features, nevertheless, like the city of Bitola, Ptolemaida and Kozani (highlighted in black) are mapped accurately as well.



(a) **(b)** **(c)**

Figure 7. Comparison of ML class based on (a) Task 2.3 (black), and (b) Task 2.8 (red) and (c) 2.8 overlaid on 2.3

In Figure 7(b) the ML class is extracted from the classified image in order to compare it with the 2.3 “ML_Hard_Threholds” image Figure 7(a). A general agreement between the two products is observed, since the ML class extent is on accord with the “ML_Hard_Thresholds” mask, covering relatively the same area with some minor differences. In order to provide a more accurate comparison for the ML class, the area of the MLs, based on the task 2.3 and the task 2.8 methodologies was computed. Out of the 856,179 ha of the total example area, 196,784 ha were identified as ML from the 2.3 detection methodology, and 229,323 ha were identified as ML from the tool developed in this task, 2.8 (Table 3).

Table 3. ML area comparison for example area

	Total Area	MLs based on 2.3	MLs based on 2.8
Area (ha)	856,179	196,784	229,323

5.2 MLs detection performance

A quantitative assessment of the classification algorithm's performance was also carried out as mentioned in chapter 4.7 Accuracy Assessment, the results of which are presented on the following Table 4. The UA and PA are produced for each class of interest, however since in this study we are mainly interested in the ML class, the values presented on the Table are the User's and Producer's Accuracy for the ML class.

Table 4. Accuracy Assessment comparison table (2.3 vs 2.8)

Country	**Overall Accuracy**		**User's Accuracy**		**Producer's Accuracy**		**Kappa**		**OOB Error Estimate**
Task	*2.3*	*2.8*	*2.3*	*2.8*	*2.3*	*2.8*	*2.3*	*2.8*	
Greece	71.5	61.4	77.7	73.7	73.9	58.1	0.41	0.23	0.26
Spain	82.8	73.1	77.4	74.3	84.6	57.2	0.65	0.43	0.23
Germany	60.6	54.3	3.6	41.9	90.0	88.5	0.04	0.03	0.21
Poland	90.9	94.2	92.4	94.3	54.2	71.0	0.64	0.77	0.25

The tool's highest accuracy was achieved for the test site of Poland, where the OA reached a value of 94.2%, UA 94.3%, PA 71% and K a value of 0.77, which denotes substantial agreement. In addition, the test site of Poland was the only test site that the Enhanced Classification tool achieved higher values in all of the accuracy metrics compared with the methodology of 2.3. The test site of Poland is followed by the test site of Spain, in performance accuracy with an OA of 73.1%, UA 74.3%, PA 57.2% and K 0.43. Third in rank comes the test site of Greece for which an OA, UA, PA and K of 61.4%, 73.7%, 58.1% and 0.23 was achieved respectively. The tool's manifested accuracy in the test sites of Greece and Spain is evidently lower than what was achieved from the 2.3 methodology. Finally, in the test site of Germany a slight reduction in OA (54.3% from 60.6%) and PA (88.5% from 90 %) is observed, but also a significant increase in UA (41.9% from 3.6%).

Since one of the objectives of this Task was to create a tool that would augment the precision in the detection of MLs, these values may not seem inspiring at first.

Nevertheless, two things ought to be mentioned; Firstly, the temporal advantage of this tool should not be underestimated. Contrary to the methodology developed in Task 2.3, with this tool, MLs can be mapped for any place in the European boundaries and for any point in time from 2017, that the Sentinel-2 SR image collection is available, onwards. Secondly, it should be noted that this is not an assessment on the performance of the whole LC classification algorithm, rather than an assessment on the mapping accuracy of the MLs class, which is sampled from the “ML_Hard_Thresholds” layer which achieved an ~68% OA.

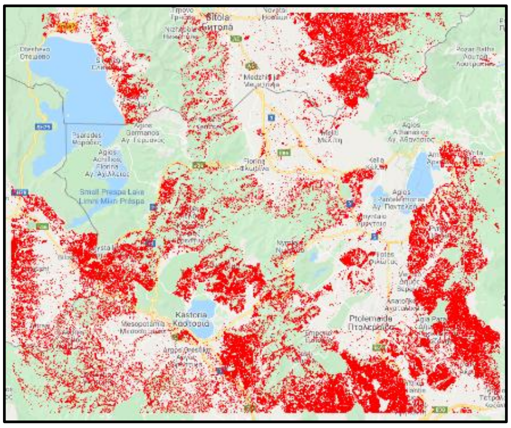
6. CASE STUDY

This dependency on the “ML_Hard_Thresholds” layer raises the following question: *“What is the tool’s potential in detecting MLs, if training data of higher quality for the ML class are available?”*

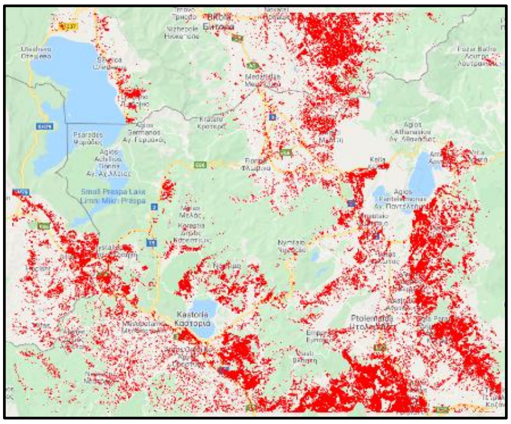
In this Chapter the answer to this question will be explored, under the hypothesis that with accurate training data, the Enhanced Classification tool is able to provide augmented precision results. Therefore, in this case study the same methodology as described in Chapter 4.6 Classification Tool Development will be followed, with the only difference that the ML class will be sampled from a portion of the ML reference polygons, and the rest will be used for validation.

This case study will be performed only for the test sites of Greece and Spain, since for these test sites an abundance of reference polygons was provided, for the realization of the Task 2.4, hence sufficient samples will be available both for the training and the validation of the algorithm. The workflow that was followed to split the samples into training and validation includes a random sampling on the reference polygons and the addition of an extra column with random numbers from 0 to 1. Then a threshold value is applied on the latter column of the ML samples and the points below the specified threshold are used for training of the ML class. The rest ML and all of the nML samples are used later for the accuracy assessment. After many trials, the appropriate threshold value for Greece was decided to 0.1, while for Spain 0.5.

The same example area will be used here to compare visually how the distribution of the ML class changes depending on the training data input Figure 8. There is a discernible change between the two methods in the distribution of the ML class, and in the second method, the ML class is clearly more constrained. Computation-wise the occupied area of the ML class in the example area based on the general methodology was 196,784 ha, while based on the case study methodology it has been reduced to 156,567 ha.



(a)



(b)

Figure 8. ML class in example area (a) General methodology, (b) Case study methodology

Moving on to the results of the primary aim of this case study, the accuracy assessment, as it was performed for the test sites of Greece and Spain, the results are summarised in the following table (Table 5). For the test site of Greece, the OA increased from 71.5% in Task 2.3, and 61.4% in the general methodology of the Task 2.8, to 77.7%, while the K value from 0.41, and 0.23 to 0.54 respectively. For the test site of Spain OA was 82.8% and 73.1% in Task 2.3 and the previous methodology respectively and reached the 86.1% with the current one and the K value from 0.65 and 0.43 increased as well to 0.66. Moreover, the OOB Error Estimate, which is the generalization error as the forest-building progresses, reduced in both cases to 0.16.

Table 5. Accuracy assessment comparison table (2.3 vs 2.8 case study)

Country	**Overall Accuracy (%)**			**Kappa**			**OOB Error Estimate**	
Task	*2.3*	*2.8*	*2.8 (cs)*	*2.3*	*2.8*	*2.8 (cs)*	*2.8*	*2.8 (cs)*
Greece	71.5	61.4	77.7	0.41	0.23	0.54	0.26	0.16
Spain	82.8	73.1	86.1	0.65	0.43	0.66	0.23	0.16

7. Conclusions

Under the framework of the ***MAIL*** project Task 2.8 “Augment precision in MLs detection”, the “Enhanced Classification” tool, based on the RF classifier and a time series analysis, was developed and is going to be incorporated in the Final ***MAIL*** Web application. The tool achieved a 61.4%, 73.1%, 54.3% and 94.2% OA for the test sites of Greece, Spain, Germany, and Poland respectively. Apart from running efficiently thanks to the RF classifier, the tool’s major advantage is that, thanks to the methodology developed in Task 2.3, it can be applied and identify potentially suitable for afforestation MLs in any point in time; it is therefore a futureproof tool.

In a case study that the reference ML and nML polygons of Task 2.4 were split to training and validation data, the tool achieved higher accuracy in the test sites of Greece and Spain than the methodology of Task 2.3, proving that the tool is able to detect MLs with good accuracy, as long as high-quality training data for MLs are available. Furthermore, it is proven once again, that the RF classifier is a very efficient and accurate classifier, able to handle big amounts of data, as well as the fact that good input results in good output.

Finally, this case study suggests that in the future the tool’s functionality should be extended in order the user to be able and provide its own reference data for the training of the ML class. Until then, the user is advised to take into consideration the OA documented here and utilize all the other options available in the Decision Support System of the ***MAIL*** web application before concluding on the final area on which the afforestation project will take place.

REFERENCES

[1] Al-Nahmi, F., Saddiqi, O., Hilali, A., Rhinane, H., Baidder, L., El Arabi, H., & Khanbari, K. (2017). Application of remote sensing in geological mapping, case study Al Maghrabah area - Hajjah region, Yemen. *ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, *4*(4W4), 63–71. https://doi.org/10.5194/isprs-annals-IV-4-W4-63-2017

[2] Alcantara, C., Kuemmerle, T., Baumann, M., Bragina, E. V., Griffiths, P., Hostert, P., Knorn, J., Müller, D., Prishchepov, A. V., Schierhorn, F., Sieber, A., & Radeloff, V. C. (2013). Mapping the extent of abandoned farmland in Central and Eastern Europe using MODIS time series satellite data. *Environmental Research Letters*, *8*(3). https://doi.org/10.1088/1748-9326/8/3/035035

[3] As-syakur, A. R., Adnyana, I. W. S., Arthana, I. W., & Nuarsa, I. W. (2012). Enhanced Built-Up and Bareness Index (EBBI) for Mapping Built-Up and Bare Land in an Urban Area. *Remote Sensing*, *4*(10), 2957–2970. https://doi.org/10.3390/rs4102957

[4] Bai, Z. G., Dent, D. L., Olsson, L., & Schaepman, M. E. (2008). Global Assessment of Land Degradation and Improvement 1. Identification by remote sensing. In *GLADA Report 5*.

[5] Banko, G. (1998). *A Review of Assessing the Accuracy of Classifications of Remotely Sensed Data and of Methods Including Remote Sensing Data in Forest Inventory*. IR-98-081.

[6] Biau, G., & Scornet, E. (2016). A random forest guided tour. *Test*, *25*(2), 197–227. https://doi.org/10.1007/S11749-016-0481-7/FIGURES/4

[7] Box, E. O., Holben, B. N., & Kalb, V. (1989). Accuracy of the AVHRR vegetation index as a predictor of biomass, primary productivity and net CO2 flux. *Vegetatio*, *80*(2), 71–89. https://doi.org/10.1007/BF00048034

[8] Breiman, L. (1999). *Random Forests-Random Features*.

[9] Carlson, B. Z., Corona, M. C., Dentant, C., Bonet, R., Thuiller, W., & Choler, P. (2017). Observed long-term greening of alpine vegetation - A case study in the French Alps. *Environmental Research Letters*, *12*(11). https://doi.org/10.1088/1748-9326/aa84bd

[10] 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

[11] Clinton, N. (2017). EEUS Time Series. *Earth Engine User Summit*.

[12] Cohen, J. (1960). A Coefficient of Agreement for Nominal Scales. *Educational and Psychological Measurement*, *20*(1), 37–46. https://doi.org/10.1177/001316446002000104

[13] Conejo, G. R., Verde, N., Georgiadis, C., Tassopoulou, M., Krupinski, M., Grommy, E., Aleksandrowicz, S., & Pérez, T. J. (2021). D2.3 | Report on Methodology development. In *MAIL: Identifying Marginal Lands in Europe and strengthening their contribution potentialities in a CO2 sequestration strategy*.

[14] Congalton, R. G. (1991a). A review of assessing the accuracy of classifications of remotely sensed data. *Remote Sensing of Environment*, *37*(1), 35–46.

[15] Congalton, R. G. (1991b). A review of assessing the accuracy of classifications of remotely sensed data. *Remote Sensing of Environment*, *37*(1), 35–46. https://doi.org/10.1016/0034-4257(91)90048-B

[16] Dara, A., Baumann, M., Kuemmerle, T., Pflugmacher, D., Rabe, A., Griffiths, P., Hölzel, N., Kamp, J., Freitag, M., & Hostert, P. (2018). Mapping the timing of cropland abandonment and recultivation in northern Kazakhstan using annual Landsat time series. *Remote Sensing of Environment*, *213*(May), 49–60. https://doi.org/10.1016/j.rse.2018.05.005

[17] Davis, J. C., & Sampson, R. J. (1986). *Statistics and data analysis in geology* (Vol. 646). Wiley New York.

[18] ESA. (2020a). *Missions - Sentinel-1* . Sentinel Online. https://sentinel.esa.int/web/sentinel/missions/sentinel-1

[19] ESA. (2020b). *Missions - Sentinel-2*. Sentinel Online. https://sentinel.esa.int/web/sentinel/missions/sentinel-2

[20] Geudtner, D., Torres, R., Snoeij, P., Davidson, M., & Rommen, B. (2014). Sentinel-1 System capabilities and applications. *2014 IEEE Geoscience and Remote Sensing Symposium*, 1457–1460. https://doi.org/10.1109/IGARSS.2014.6946711

[21] Google developers. (2021). *Sentinel-1 Algorithms*. Google Earth Engine. https://developers.google.com/earth-engine/guides/sentinel1

[22] Gorelick, N., Hancher, M., Dixon, M., Ilyushchenko, S., Thau, D., & Moore, R. (2017). Google Earth Engine: Planetary-scale geospatial analysis for everyone. *Remote Sensing of Environment*, *202*, 18–27. https://doi.org/10.1016/j.rse.2017.06.031

[23] Huang, H., Roy, D. P., Boschetti, L., Zhang, H. K., Yan, L., Kumar, S. S., Gomez-

Dans, J., & Li, J. (2016). Separability analysis of Sentinel-2A Multi-Spectral Instrument (MSI) data for burned area discrimination. *Remote Sensing*, *8*(10). https://doi.org/10.3390/rs8100873

[24] Huete, A., Didan, K., Miura, T., Rodriguez, E. ., Gao, X., & Ferreira, L. . (2002). Overview of the radiometric and biophysical performance of the MODIS vegetation indices. *Remote Sensing of Environment*, *83*(1–2), 195–213. https://doi.org/10.1016/S0034-4257(02)00096-2

[25] Huete, A. R. (1988). A soil-adjusted vegetation index (SAVI). *Remote Sensing of Environment*, *25*(3), 295–309. https://doi.org/10.1016/0034-4257(88)90106-X

[26] Jakubauskas, M. E., Legates, D. R., & Kastens, J. H. (2002). Crop identification using harmonic analysis of time-series AVHRR NDVI data. *Computers and Electronics in Agriculture*, *37*(1–3), 127–139. https://doi.org/10.1016/S0168-1699(02)00116-3

[27] Jensen, J. R. (1996). *Introductory digital image processing: a remote sensing perspective.* (Issue Ed. 2). Prentice-Hall Inc.

[28] Kolecka, N. (2021). Greening trends and their relationship with agricultural land abandonment across Poland. *Remote Sensing of Environment*, *257*(January), 112340. https://doi.org/10.1016/j.rse.2021.112340

[29] Kruppa, J., Schwarz, A., Arminger, G., & Ziegler, A. (2013). Consumer credit risk: Individual probability estimates using machine learning. *Expert Systems with Applications*, *40*(13), 5125–5131. https://doi.org/10.1016/j.eswa.2013.03.019

[30] Kuc, G., & Chormański, J. (2019). Sentinel-2 Imagery for Mapping and Monitoring Imperviousness in Urban Areas. *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, *XLII-1/W2*(Evaluation and Benchmarking Sensors, Systems and Geospatial Data in Photogrammetry and Remote Sensing), 43–47. https://doi.org/10.5194/isprs-archives-XLII-1-W2-43-2019

[31] Kuemmerle, T., Hostert, P., Radeloff, V. C., Van Der Linden, S., Perzanowski, K., & Kruhlov, I. (2008). Cross-border comparison of post-socialist farmland abandonment in the Carpathians. *Ecosystems*, *11*(4), 614–628. https://doi.org/10.1007/s10021-008-9146-z

[32] Lavreniuk, M., Kussul, N., Meretsky, M., Lukin, V., Abramov, S., & Rubel, O. (2017). Impact of SAR data filtering on crop classification accuracy. *2017 IEEE 1st Ukraine Conference on Electrical and Computer Engineering, UKRCON 2017 - Proceedings*, 912–917. https://doi.org/10.1109/UKRCON.2017.8100381

[33] Lewis, S., & Kelly, M. (2014). Mapping the Potential for Biofuel Production on Marginal Lands: Differences in Definitions, Data and Models across Scales. *ISPRS International Journal of Geo-Information*, *3*(2), 430–459. https://doi.org/10.3390/ijgi3020430

[34] Liu, Y., Li, Y., Li, S., & Motesharrei, S. (2015). Spatial and temporal patterns of global NDVI trends: Correlations with climate and human factors. *Remote Sensing*, *7*(10), 13233–13250. https://doi.org/10.3390/rs71013233

[35] Löw, F., Fliemann, E., Abdullaev, I., Conrad, C., & Lamers, J. P. A. (2015). Mapping abandoned agricultural land in Kyzyl-Orda, Kazakhstan using satellite remote sensing. *Applied Geography*, *62*, 377–390. https://doi.org/10.1016/j.apgeog.2015.05.009

[36] Malinowski, R., Lewiński, S., Rybicki, M., Gromny, E., Jenerowicz, M., Krupiński, M., Nowakowski, A., Wojtkowski, C., Krupiński, M., & Krätzschmar, E. (2020). Automated Production of a Land Cover/Use Map of Europe Based on Sentinel-2 Imagery. *Remote Sensing*, *12*(21), 3523.

[37] Mitri, G., Nader, M., Abou Dagher, M., & Gebrael, K. (2020). Investigating the performance of sentinel-2A and Landsat 8 imagery in mapping shoreline changes. *Journal of Coastal Conservation*, *24*(3). https://doi.org/10.1007/s11852-020-00758-4

[38] Nalepa, R. A., & Bauer, D. M. (2012). Marginal lands: the role of remote sensing in constructing landscapes for agrofuel development. *Journal of Peasant Studies*, *39*(2), 403–422. https://doi.org/10.1080/03066150.2012.665890

[39] Osgouei, P. E., Kaya, S., Sertel, E., & Alganci, U. (2019). Separating Built-Up Areas from Bare Land in Mediterranean Cities Using Sentinel-2A Imagery. *Remote Sensing 2019, Vol. 11, Page 345*, *11*(3), 345. https://doi.org/10.3390/RS11030345

[40] Oshiro, T. M., Perez, P. S., & Baranauskas, J. A. (2012). How Many Trees in a Random Forest? *Machine Learning and Data Mining in Pattern Recognition*, *7376 LNAI*, 154–168. https://doi.org/10.1007/978-3-642-31537-4_13

[41] Pal, M. (2007). Random forest classifier for remote sensing classification. *International Journal of Remote Sensing*, *26*(1), 217–222. https://doi.org/10.1080/01431160412331269698

[42] Peter, B. G., Messina, J. P., & Snapp, S. S. (2018). A 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

[43] Peterson, U., & Aunap, R. (1998). Changes in agricultural land use in Estonia in the 1990s detected with multitemporal Landsat MSS imagery. *Landscape and Urban Planning*, *41*(3–4), 193–201. https://doi.org/10.1016/S0169-2046(98)00058-9

[44] Rayner, J. N. (1971). *Introduction to spectral analysis*.

[45] 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*, 526–534. https://doi.org/10.1016/j.landusepol.2018.02.033

[46] Sim, J., & Wright, C. C. (2005). The Kappa Statistic in Reliability Studies: Use, Interpretation, and Sample Size Requirements. *Physical Therapy*, *85*(3), 257–268. https://doi.org/10.1093/ptj/85.3.257

[47] Tucker, C. J. (1979). Red and photographic infrared linear combinations for monitoring vegetation. *Remote Sensing of Environment*, *8*(2), 127–150. https://doi.org/10.1016/0034-4257(79)90013-0

[48] Van Deventer, A. P., Ward, A. D., Gowda, P. H., & Lyon, J. G. (1997). Using thematic mapper data to identify contrasting soil plains and tillage practices. *Photogrammetric Engineering and Remote Sensing*, *63*, 87–93.

[49] Yin, H., Prishchepov, A. V., Kuemmerle, T., Bleyhl, B., Buchner, J., & Radeloff, V. C. (2018). Mapping agricultural land abandonment from spatial and temporal segmentation of Landsat time series. *Remote Sensing of Environment*, *210*(January), 12–24. https://doi.org/10.1016/j.rse.2018.02.050

[50] Zha, Y., Gao, J., & Ni, S. (2003). Use of normalized difference built-up index in automatically mapping urban areas from TM imagery. *International Journal of Remote Sensing*, *24*(3), 583–594. https://doi.org/10.1080/01431160304987

[51] Zollini, S., Alicandro, M., Cuevas-González, M., Baiocchi, V., Dominici, D., & Buscema, P. M. (2019). Shoreline Extraction Based on an Active Connection Matrix (ACM) Image Enhancement Strategy. *Journal of Marine Science and Engineering*, *8*(1), 9. https://doi.org/10.3390/jmse8010009

ANNEX II: TABLE OF FIGURES

ANNEX III: LIST OF TABLES