CIGeo

Tercer Congreso en Ingeniería Geomática
Valencia,
7 – 8 de Julio de 2021
DOI: https://dx.doi.org/10.4995/CiGeo2021.2021.12729

METHODOLOGICAL PROPOSAL FOR THE IDENTIFICATION OF MARGINAL LANDS WITH REMOTE SENSING-DERIVED PRODUCTS AND ANCILLARY DATA

PROPUESTA METODOLÓGICA PARA LA IDENTIFICACIÓN DE TIERRAS MARGINALES MEDIANTE PRODUCTOS DERIVADOS DE TELEDETECCIÓN Y DATOS AUXILIARES

Jesús Torralba[a,b,*], Luis A. Ruiz[a], Charalampos Georgiadis[c], Petros Patias[c], Rodrigo Gómez-Conejo[d], Natalia Verde[c], Maria Tassopoulou[c], Fernando Bezares[d], Ewa Grommy[e], Sebastian Aleksandrowicz[e], Elke Krätzschmar[b], Michał Krupiński[e], Juan Pedro Carbonell-Rivera[a]

[a] Geo-Environmental Cartography and Remote Sensing Group (CGAT), Universitat Politècnica de València, Camino de Vera s/n, 46022 Valencia, Spain. jetorpe@upv.es; laruiz@upv.es; juacarri@upv.es

[b] Industrieanlagen-Betriebsgesellschaft MBH (IABG), Hermann-Reichelt-Str. 3, 01109 Dresden, Germany. Kraetzschmar@iabg.de

[c] School of Rural and Surveying Engineering, Laboratory of Photogrammetry and Remote Sensing Unit (PERS lab), The Aristotle University of Thessaloniki, 54124 Thessaloniki, Greece. harrisg@civil.auth.gr; patias@auth.gr; nverde@auth.gr; mtasso@live.com

[d] CESEFOR (Centro de Servicios y Promoción Forestal y de su Industria de Castilla y León), Polígono Industrial las Casas, calle C, parcela 4. 42005 Soria, Spain. rodrigo.gomez@cesefor.com; fernando.bezares@cesefor.com

[e] Centrum Badań Kosmicznych Polskiej Akademii Nauk (CBK PAN), Bartycka 18A, 00-716 Warszawa, Poland. egromny@cbk.waw.pl; saleksandrowicz@cbk.waw.pl; mkrupinski@cbk.waw.pl

Abstract:

The concept of marginal land (ML) is dynamic and depends on various factors related to the environment, climate, scale, culture, and economic sector. The current methods for identifying ML are diverse, they employ multiple parameters and variables derived from land use and land cover, and mostly reflect specific management purposes. A methodological approach for the identification of marginal lands using remote sensing and ancillary data products and validated on samples from four European countries (i.e., Germany, Spain, Greece, and Poland) is presented in this paper. The methodology proposed combines land use and land cover data sets as excluding indicators (forest, croplands, protected areas, impervious areas, land-use change, water bodies, and permanent snow areas) and environmental constraints information as marginality indicators: (i) physical soil properties, in terms of slope gradient, erosion, soil depth, soil texture, percentage of coarse soil texture fragments, etc.; (ii) climatic factors e.g. aridity index; (iii) chemical soil properties, including soil pH, cation exchange capacity, contaminants, and toxicity, among others. This provides a common vision of marginality that integrates a multidisciplinary approach. To determine the ML, we first analyzed the excluding indicators used to delimit the areas with defined land use. Then, thresholds were determined for each marginality indicator through which the land productivity progressively decreases. Finally, the marginality indicator layers were combined in Google Earth Engine. The result was categorized into 3 levels of productivity of ML: high productivity, low productivity, and potentially unsuitable land. The results obtained indicate that the percentage of marginal land per country is 11.64% in Germany, 19.96% in Spain, 18.76% in Greece, and 7.18% in Poland. The overall accuracies obtained per country were 60.61% for Germany, 88.87% for Spain, 71.52% for Greece, and 90.97% for Poland.

Key words: land use, land cover, idle land, land degradation, GIS, remote sensing, Google Earth Engine

Resumen:

El concepto de tierra marginal (ML) es dinámico y depende de factores relacionados con el entorno, el clima, la escala, la cultura y la economía. los métodos actuales de identificación de ML son también diversos y están basados en múltiples parámetros y variables derivados del uso y cobertura del suelo reflejando, en su mayoría, fines de gestión específicos. En este artículo se presenta una propuesta metodológica para la identificación de tierras marginales mediante el uso de productos derivados de teledetección y datos auxiliares, validándose sobre muestras obtenidas en cuatro países europeos: Alemania, España, Grecia y Polonia. La metodología combina datos de usos y coberturas del suelo como indicadores excluyentes (bosque, tierras de cultivo, áreas protegidas, áreas impermeables, cambios de usos del suelo, cuerpos de agua y áreas de nieve permanente) e información ambiental como indicadores de marginalidad, esto es, (i) propiedades físicas del suelo como la pendiente, profundidad de suelo, erosión del suelo, textura, porcentaje de fragmentos de textura gruesa del suelo, etc.; (ii) factores climáticos como el índice de aridez; (iii) propiedades químicas del suelo como pH, capacidad de intercambio catiónico, contaminantes y toxicidad, entre otros, con el objetivo de abordar una visión común de la marginalidad que integre un enfoque multidisciplinar. Para obtener las coberturas de ML primero se analizaron los indicadores excluyentes para delimitar las áreas con un uso del suelo establecido. En segundo lugar, se determinaron los umbrales para cada indicador de marginalidad a través de los cuales el suelo se transforma,

* Corresponding Author: Jesus Torralba, jetorpe@upv.es

disminuyendo progresivamente su aprovechamiento productivo. Finalmente, la superposición de las capas de indicadores de marginalidad se llevó a cabo con la herramienta *Google Earth Engine*. El resultado final se categorizó en 3 niveles de ML con diferente productividad: alta, baja y tierras potencialmente inadecuadas. Los resultados obtenidos indican que el porcentaje de tierras marginales sobre la extensión total de cada país analizado es de 11,64% en Alemania, 19,96% en España, 18,76% en Grecia y 7,18% en Polonia. La precisión global obtenida por país fue del 60,61% para Alemania, del 88,87% para España, del 71,52% para Grecia y del 90,97% para Polonia.

Palabras clave: uso de suelo, cobertura de suelo, tierra abandonada, degradación del suelo, SIG, teledetección, Google Earth Engine

1. Introduction

The concept of marginal land (ML) has evolved across time, location, discipline (Kang *et al.* 2013b) and management objectives. Traditionally, the term "marginal lands" has been used to refer, from a purely economic perspective, to those agricultural areas that have a limited production potential (Hollander 1895; Strijker 2005; Ciria *et al.* 2019). Later, its meaning changed to include areas with biophysical, climatic, and socioeconomic constraints (Eliasson *et al.* 2010; Elbersen *et al.* 2018; Gerwin *et al.* 2018). In this regard, only in recent decades the concept of marginal land has been used to define abandoned lands physically inaccessible, with high environmental risk or providing fragile ecosystem services (Kang *et al.* 2013a).

In addition, the definition has changed as a result of the dynamics of ML themselves. Under specific and transitory ML circumstances (e.g., policies, land regulations, economic incentives, land use benefits, market profitability), humans have claimed or abandoned these lands (Strijker 2005). These circumstantial dynamics have placed the ML in a transitional state of land resources, very sensitive to natural processes, economic impacts, and diverse management. The latter has generated the mentioned recent changes in the ML definition, as a consequence of the search of land to achieve a variety of management objectives, such as the increase the bioenergy crops (Ciria *et al.* 2019; Mellor *et al.* 2021), food production land (Zhang *et al.* 2018) or carbon sequestration through reforestation (Sauer *et al.* 2012).

Analogous to the ML definition, a single identification and classification method does not exist, and the available methods only reflect management goals. These methods range from approaches focused on physical characteristics (i.e., environmental factors) to purely socioeconomic factors. In general, biophysical constraints related to agricultural productivity or bioenergy are the most commonly used for the ML identification. For example, Cai *et al.* (2011) applied the Soil Rating for Plant Growth Index (SRPG) developed by the US Department of Agriculture, where they combined sixteen soil properties related to productivity, slope, soil temperature regimes, and moisture index. Using a multi-criteria decision approach based on Geographic Information Systems (GIS) and remote sensing, Zolekar & Bhagat (2015) combined data on land use/land cover (LULC), slope, soil depth, erosion, moisture, water holding capacity, texture, and availability of nutrient to study the land suitability for agriculture in hilly zones. A similar approach based on an indicator of suitability for agricultural activity was applied by Li *et al.* (2017) using eight indicators (slope, soil erosion, soil organic carbon, texture, pH, cation exchange capacity, soil depth, and drainage) in areas where LULC types, such as water bodies, protected areas, or human settlements had previously been excluded. The crop sustainability concern and economic focus were integrated into the ML identification by Gopalakrishnan *et al.* (2011). These authors identified ML based on soil health criteria (erosion, frequently flooded, poorly drained, steeply sloped, and low productivity), current land use (includes land categories such as idle and fallow), and environmental degradation criteria (contaminated land, contaminated water resources, and water-constrained areas).

In Europe, Bertaglia *et al.* (2007) applied a slightly different approach, since they targeted areas for extensive grazing. The main difference was to consider LULC as an aggregate of biophysical constraints and socioeconomic trends. In Germany, Reger *et al.* (2007) used satellite data and historical information on land cover dynamics to detect the trend of cropland abandonment and, in addition, to identify ML. Ivanina *et al.* (2016) and Gerwin *et al.* (2018) in the Sustainable Exploitation of Biomass for Bioenergy on Land (Seemla) European project and Elbersen *et al.* (2018) in the European project Marginal Lands for Growing industrialists (Magic) have assessed and quantified the area of ML in Europe by applying biophysical criteria on agricultural and forest lands using GIS tools. These projects also considered socio-economic constraints to classify ML (accessibility, status of infrastructure, demographic parameters, and economic density (income/km2)). In this same bioenergy context, Ciria *et al.* (2019) applied a holistic approach for the identification of arable marginal lands under rainfed conditions in Spain, combining biophysical constraints with the economic performance of crops and other sustainability aspects.

Most of these methodologies mainly employed soil analysis and agricultural production indicators for ML identification and, to a lesser degree, aspects of environmental quality and sustainability were considered. The criteria to define ML should cover the needs and constraints of each time and region, and integrate a multidisciplinary approach to reflect the synergy of multiple land functions, management objectives and ecosystem services. Consequently, a single index or criterion cannot fully satisfy these needs. This paper presents a methodological proposal for the identification, mapping and classification of ML without a defined management objective, combining the use of remote sensing derived products and ancillary data.

2. Datasets and Methodology

The study area was determined according to the location of the validation samples, which included different areas throughout four European countries: Germany, Greece, Poland, and Spain.

To identify and classify ML, several data sets were combined (Tables 1 and 2) in two phases (Fig. 1) (Georgiadis *et al.* 2021). In the first phase, binary exclusion factors ("Hard" constraints) based on LULC were analyzed in a top-down stepwise approach, excluding those areas with a given land use or belonging to land cover types excluded from the general definition of ML. These "Hard" factors were used to identify and map the ML in the four countries.

In the second phase, the resulting potential ML were classified based on their marginality degree with variable thresholds ("Soft" constraints), including soil condition and biophysical factors, such as inherent properties of the soil or land, and climatic data as a transitory and restrictive property. Finally, three classes of ML were obtained according to their level of productivity: high, low and potentially unsuitable land.

2.1. Hard constraints datasets

The LULC selected in the bibliography review as excluding indicators to determine ML were: marshes, peatbogs, permanent snow-covered surfaces, water bodies, forest, croplands, impervious, protected areas and changed areas. Table 1 describes the datasets used to determine each LULC, their spatial resolution, the dates of each dataset and the sources in which they are described in detail. All datasets downloaded (Fig. 1.1) were free and open access. Products 1,3 and 4 were available on a global scale and the rest on a European scale.

2.2. Soft constraints datasets

A literature review, based on studies focused specifically on the methodological aspects of ML mapping, was carried out to determine the indicators that constitute the "Soft" constraints with non-thematic data (numerical data). In particular, those studies detailing soil, climate, terrain, sustainability, productivity and LULC constraints were considered. For each study, we registered the extent, minimum mapping unit (MMU), technology used, datasets, indicators used, the indicators thresholds and the ML classification scheme.

Table 1: Datasets used to determine the "Hard" constraints in ML identification.

Hard Constraints	*Source*	*Abbreviation*	*Resolution*	*Year*	*Reference*
Marshes, Peatbogs, Permanent snow-covered surfaces, Water bodies	(1) Sentinel-2 Global Land Cover	S2GLC	10 m	2017	(Malinowski *et al.* 2020)
Forest	(2) Copernicus High-Resolution Layer - Tree Cover Density	HRL-TCD	20 m	2015	(European Environment Agency 2018)
	(3) Global Forest Change - Tree Cover	GFC-TC	30 m	2000	(Hansen *et al.* 2013)
	(4) Global Forest Change - Loss	GFC-LSS	30 m	2015 & 2018	(Hansen *et al.* 2013)
Croplands	(5) Sentinel-2 Global Land Cover	S2GLC	10 m	2017	(Malinowski *et al.* 2020)
	(6) CORINE Land Cover	CORINE LC	25 ha	2018	(European Environment Agency 2019a)
Impervious	(7) Copernicus High-Resolution Layer – Imperviousness Density	HRL–IMD	20 m	2017	(European Environment Agency 2018)
	(8) CORINE Land Cover	CORINE LC	25 ha	2018	(European Environment Agency 2019b)
Protected Areas	(9) EU Nationally designated protected areas inventory	CDDA	-	2018	(European Environment Agency 2019c)
	(10) Natura2000 Network	Natura2000	20 m	2018	(European Environment Agency 2019d)
Changed Areas	(11) Copernicus High-Resolution Layer - Tree Cover Density Change	HRL-TCDC	20 m	2012-2015	(European Environment Agency 2018)
	(12) Copernicus High-Resolution Layer - Impervious Classified Change	HRL-IMCC	20 m	2012-2015	(European Environment Agency 2018)
	(13) Corine Land Cover Change	CORINE LC CHA	25 ha	2012-2018	(European Environment Agency 2019b)

In order to define the land marginality, the maximum and minimum thresholds for each indicator were determined (Fig. 1.7). These thresholds refer to the value ranges that could be reached by a particular indicator and were defined based on the examined literature. Each indicator was divided into 3 ranges: a) representing the best indicator values (score 10) and corresponding to suitable, fertile, or productive land; b) representing the average values of the indicator (score 5), related to low fertile land and low productivity; and c) representing the restrictive indicator values (score 1) for land that could potentially be unsuitable or incompatible with any activity or management. In the case that more than one threshold was found in the literature, the threshold was established based on the maximum and minimum values for all Europe. During this phase, we ensured that each indicator had a data set available for the four countries to be analyzed. Consequently, when an indicator could not be matched to a dataset (conceptually and in units) for the four countries, the indicator was omitted. The thresholds selected for each indicator, as well as the reference of the document from which it was obtained, are shown in Table 2.

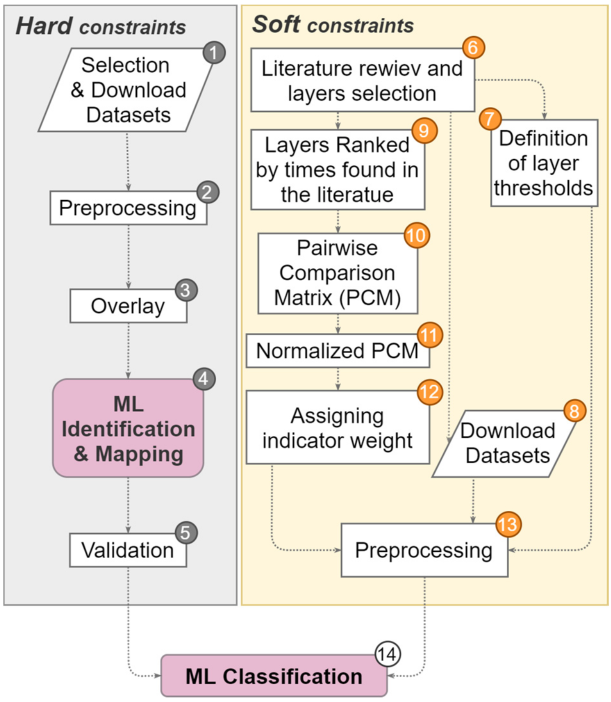

Figure 1: General methodological workflow for identification, mapping and classification of ML

Furthermore, the "Soft" indicators were ranked according to the number of times that they were found in the literature (Fig. 1.9), as a way to analyze their importance. Once ranked, the Pairwise Comparison Matrix (PCM) of the ranks (Zolekar & Bhagat 2015) was calculated. Afterward, the normalized PCM was performed and a weight was obtained for each indicator (Fig. 1.11 and 1.12). To normalize the values, each PCM value was divided by the sum of the values of its column, then the weights of each indicator (Table 2) were calculated using the average of the values in its row. These weights were scaled from 0 to 1 in ascending order to preserve the hierarchy according to their importance in the marginality, and their sum is equal to 1.

2.3. Intermediate layer production

The different data sets for identifying and classifying ML were processed independently. In regard to the "Hard" indicators, six workflows were performed to determine LULC where different datasets were combined to produce binary intermediate layers.

1) The S2GLC product was selected to determine the land cover of marshes, peatbogs, permanent snow-covered surfaces and water bodies. The S2GLC product consists of thirteen land cover classes (overall accuracy = 86%) and was developed using classification algorithms for the analysis of more than 15,000 Sentinel-2 images (Malinowski *et al.* 2020).

2) The Copernicus HRL-TCD 2015, GFC-TC 2000, and GFC-LSS 2015 and 2018 layers were combined to delimit the forest zones (defined as those areas with a tree cover higher than 30% and a minimum area of 0.5 ha). All pixels that were selected as forest in 2000, 2015, and 2018 were considered as forest land cover.

3) To determine croplands, we selected S2GLC class 73 (cultivated areas) and 75 (vineyards) and CORINE LC class 2 where agricultural areas are included (except 231 which represents pastures). Both datasets were reclassified into two classes representing croplands/non-cropland areas. Then, a fuzzy overlay was performed to identify all pixels considered as crops in both datasets.

4) Impervious areas represent all sealed and constructed areas that are primarily covered by buildings or impermeable surfaces. To delimit impervious areas, CORINE LC class 1 ("Artificial areas", except classes 131 (mineral extraction sites) and 132 (dumpsites)) and the Copernicus HRL-IMD product were used. In the latter, those areas with a threshold above 30% were defined as impervious. The final vector layer was converted to raster and both datasets were reclassified and combined in the same way that for croplands.

5) Protected areas were delimited by merging Natura2000 and CDDA from the European Environment Agency. The final vector layer was converted to raster and both datasets were reclassified into two classes representing protected/non-protected areas.

6) To incorporate the dynamic aspect of ML, changed areas were also included in the proposed methodology. Specifically, two main types of changes were considered: i) changes related to forest activities such as afforestation and reforestation and ii) changes in the urban fabric. The delineation of changes in forest areas was implemented using the Copernicus HRL-TCDC change product. This product shows real tree cover density (TCD) changes (%) between 2012 and 2015. To identify such changes in forest areas, a threshold of 50% was applied. This threshold ensured that the output included certain changes due to reforestation or deforestation and not sparse or random changes. On the other hand, changes in urban fabric and impervious areas were outlined using the classes "increased IMD" and "new cover" from the HRL–IMCC 2012-2015 product and the CORINE LC CHA 2012-2018 product. Specifically, we extracted class 1 except classes 131 and 132. All intermediate layers were reclassified into two classes (changed/non-changed areas) and finally a fuzzy overlay was performed to produce the final intermediate layer "changed", representing all the occurred changes. In this case, the fuzzy overlay was performed to identify all pixels that were registered as changed in one of the three datasets.

Table 2: Summary of the “Soft” indicators used in the ML classification methodology. The indicators are grouped by data type: terrain and soil, sustainability, and productivity. "Threshold-based on" refers to the scientific publication from which the threshold was obtained. "Source" refers to the repository from which the data was freely downloaded. ”Rank” refers to the position in which each indicator appeared in the bibliographic ranking. “Weight” refers to the weights obtained in the normalized PCM. “Score” refers to the value given to each of the thresholds in order to categorize them according to their marginality. The data sets with two weights were the indicators with information for Topsoil (T) and Subsoil (S) and they were treated as independent indicators.

Type	*Rank*	*Layer name*	*Thresholds*	*Threshold-based on*	*Source*	*Weight*	*Score*
Terrain and Soil	1	Slope (%)	[15 - 40] [40 - 65] [65 - 90]	(Gopalakrishnan *et al.* 2011)	(EuropeEnvironment Agency 2017)	0.17	10 5 1
	1	Depth Available to Roots (cm)	[100 – 66.7] [66.7 – 33.3] [33.3 – 0]	(Ciria *et al.* 2019)	European Soil Data Centre (ESDAC), esdac.jrc.ec.europa.eu, European Commission, Joint Research Centre	0.17	10 5 1
	3	Coarse fragments (T/S) (%)	[10 - 15] [15 - 20]	(Ciria *et al.* 2019)		0.03 0.03	10 5
	2	Texture (T/S) (%)	[30 - 53.3] [53.3 - 76.7] [76.7 - 100]	(Elbersen *et al.* 2018)		0.045 0.045	10 5 1
	6	Clay (T/S) (%)	[50 - 58.7] [58.7 - 67.3] [67.3 - 76]	(Eliasson *et al.* 2010)		0.015 0.015	10 5 1
	6	Sand (T/S) (%)	[60 - 70] [70 - 80] [80 - 90]	(Eliasson *et al.* 2010)		0.015 0.015	10 5 1
	4	Total Available Water (T/S) (mm)	[100 – 50] [50 – 0]	(Zolekar & Bhagat 2015)		0.02 0.02	10 5
Sustainability	2	Soil Acidity	[pH>8, pH<6] [pH>8.5, pH<5.25] [pH>9, pH<4.5]	(Ciria *et al.* 2019)		0.09	10 5 1
	3	Soil Erosion (t/ha/year)	[10 – 55.3] [55.3 – 67.3] [67.3- 325]	(Eurostat 2020)		0.06	10 5 1
	4	Flooding (%)	[50 - 66.7] [66.7 - 83.3] [83.3 - 100]	(Gopalakrishnan *et al.* 2011)	(Pekel *et al.* 2016)	0.04	10 5 1
	5	Socidity (%)	[6 - 36.7] [36.7 - 67.4] [67.4 - 98]	(Eliasson *et al.* 2010)	(Batjes 2016)	0.03	10 5 1
	6	Toxicity Contamination (cg/kg)	[1– 3] [3 – 10] [10 – 23.5]	(Gopalakrishnan *et al.* 2011; Ivanina *et al.* 2016)		0.03	10 5 1
	9	Natural Toxicity (g/kg)	[150 - 328] [328 - 506] [506 – 684]	(Eliasson *et al.* 2010)		0.02	10 5 1
	2	Dryness (Aridy Index)	[0.5 – 0.34] [0.34 – 0.18] [0.18 – 0]	(Ivanina *et al.* 2016; Elbersen *et al.* 2018)	(Abatzoglou *et al.* 2018)	0.02	10 5 1
Productivity	6	Caption Exchange Capacity (cmol(+)/kg)	[22.2 - 18.9] [18.9 - 15.6] [15.6 - 12.3]	-	(ISRIC-World Soil Information 2020)	0.03	10 5 1
	3	Soil Organic Matter (T/S) (%)	[OM < 1%, OM ≥ 20%] [OM < 0.75%, OM ≥ 30%]	(Elbersen *et al.* 2018; Ciria *et al.* 2019)	European Soil Data Centre (ESDAC), esdac.jrc.ec.europa.eu, European Commission, Joint Research Centre	0.03 0.03	10 5
	7	Productivity Grasslands	[6-4] [4-2] [2-0]	-		0.02	10 5 1
		Productivity Forests	[3-2] [2-1] [1-0]	-		0.02	10 5 1

A single workflow was followed to obtain the "Soft" indicators' intermediate layers. First, the raster values were reclassified according to the thresholds described in Table 2, where layers with two or three values (scores) were obtained. Then, the pixel values were multiplied by the weight calculated from the PCM. Finally, we obtained raster layers with 3 values representing the 3 ranges of marginality. The indicators "Coarse fragments", "Texture", "Clay", "Sand", “Total Available Water”, and “Soil organic matter” information from Topsoil (T) and Subsoil (S) were available and processed as independent indicators. T and S layers, were processed independently since they originally formed and developed differently, even their characteristics could have been influenced by land use (Hiederer 2013). There were two exceptions in the preprocessing of the "Soft" indicators, involving the preparation of a dataset before applying the described workflow: (1) the soil indicator "Texture" was obtained by applying the equation proposed by Elbersen *et al.* (2018) where the silty texture is added to twice the clayey texture, and silt (T/S) layers were downloaded from ESDAC; and (2) the Aridity Index (AI), which relates accumulated precipitation (mm) and reference evapotranspiration (Penman 1948), the high-resolution monthly dataset "Terraclima" for 2018 (Abatzoglou *et al.* 2018) available from Google Earth Engine (GEE) was used. To obtain a single value per pixel and year, the twelve images of 2018 were filtered, then the precipitation and evapotranspiration bands were selected, followed by the division between the precipitation and evapotranspiration bands. Finally, the yearly mean value was calculated for every pixel.

All intermediate layers were resampled using the nearest neighbor method to a pixel size of 10 x 10 meters, which is the spatial resolution of the highest resolution product (S2GLC), used as a base-map in the next steps. In addition, all layers were projected to the horizontal coordinate system European Terrestrial Reference System 1989 (ETRS89) using Lambert Azimuthal Equal-Area projection (LAEA).

2.4. Potential ML identification

The potential ML were identified and mapped by combining all the "Hard" intermediate layers, which include LULC types that cannot be ML, with the S2GLC base-map. In particular, raster layers were mathematically combined and new values were assigned to the output layer. In this process, we co-registered every layer so that the output raster cells were aligned with raster cells of the S2CLC base-map. The final layer was a binary raster file, based on the S2GLC base-map, including all remaining classes that are potentially ML.

To properly assess the classification of ML and non-marginal lands (nonML), experts from the four countries with previous knowledge of land use, landscape, terrain, and general knowledge of the countries included in the study, provided polygons of reference ML and nonML areas. On the reference polygons, stratified random sampling of points with a sample size of 1 pt./ha was carried out. The classification accuracy was quantified by country through the confusion matrix by contrasting the reference values with the classification results. The performance of the classification was also measured with other indexes including the overall accuracy, Kappa index, recall (%) and F-measure (%). Recall corresponds with the fraction of ML validation samples classified as positive, among the total number of positive ML. While F-measure is the harmonic mean of the model's precision and recall (Carbonell-Rivera *et al.* 2020).

2.5. ML classification

To classify the potential ML, the intermediate raster layers of the "Soft" indicators were overlaid in GEE, with special attention to the pixel alignment of the different layers. The last step in the ML mapping was the reclassification of the product resulting from the weighted superposition of the "Soft" indicators into three ML productivity categories: high, low and potentially unsuitable land. For this purpose, the minimum (ML_m) and maximum (ML_M) values obtained by the marginality layer were calculated. Then, the range of values was divided into three intervals using three different approaches to establish the upper and lower limits of each category:

a) *Equal magnitude.* The class interval was set by dividing the range of values into 3 equal parts.

b) *25^{th}-75^{th} percentiles.* The 25^{th} and 75^{th} percentiles were calculated to establish them as class limits. The interval [ML_m, $P25^{th}$) represented "potentially unsuitable land" category, [$P25^{th}$, $P75^{th}$) was the "low productivity ML" category, and [$P75^{th}$, ML_M) was the "high productivity ML" category.

c) *33^{rd}-66^{th} percentiles.* The 33^{rd} and 66^{th} percentiles were calculated to establish them as class limits. The interval [ML_m, $P33^{rd}$) characterized "potentially unsuitable land" category, [$P33^{rd}$, $P66^{th}$) was the "low productivity ML" category, and [$P66^{th}$, ML_M) was the "high productivity ML" category.

3. Results and Discussion

The percentage of ML per country was higher in Mediterranean countries. In particular, Spain was the country where most ML were identified (Fig. 2), with 20.0% (100,983 km^2), followed by Greece with 18.8% (24,770 km^2). In addition to a climate characterized by a prolonged summer drought, these countries also have an abrupt and varied topography that restricts the use of some lands for agriculture and forests production. In contrast, the countries where fewer ML were identified were Germany, with 11.6% (41,606 km^2), and Poland, with 7.2% (22,442 km^2), both countries have a relief dominated by flatlands and a continuous rainfall regime throughout the year, facilitate agriculture and forest (natural or plantations).

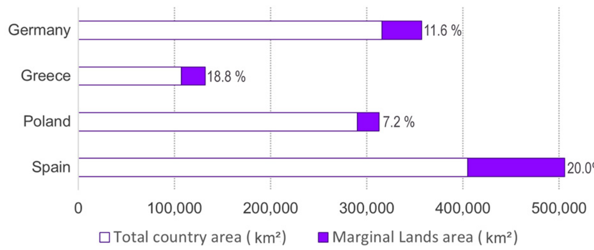

Figure 2: Percentage of area identified as ML per country.

From the Gerwin *et al.* (2018) study in the Seelman project, only the results from Greece and Germany were available, and they identified 14.6% ML and 9.4% ML per country, respectively. This differed from our results by 4.16% for Greece and 2.24% for Germany. This is mainly due to their indicators and thresholds choice as they adjusted them to the site requirements of certain forest species, in order to find ML for biomass production and bioenergy purposes. The results were also compared with those obtained by Elbersen *et al.* (2018) in the Magic project for all European countries. Elbersen *et al.* (2018) found percentages of ML areas per country similar to our results (Greece 23.108 km2 (17.5%), Germany 33.896 km2 (9.5%), and Poland 27.372 km2 (8.8%)). However, for Spain they found 167.680 km2 (33.1%). These differences could be due to the use of land irrigation indicators in the project. This fact, in addition to the desertification of some transitional areas could explain this increase in the area of ML by 13.2%.

The distribution of the potential marginal lands can be observed in Figure 3a In the case of Spain, the presence of ML was distributed throughout the territory, and there was a tendency for marginal lands to increase around the mountain ranges. These areas, if not covered by forest or under protection, had no clear land use and were covered by grassland, herbaceous vegetation, moorland, heathland, sclerophyllous vegetation, marshes and, to a lesser extent, natural material surface (i.e., bare rock, hard pan, mineral fragments, bare soils and natural deposits). The results for Greece were similar to Spain, with a dispersion of ML throughout the country but mostly clustered in the mountainous ranges or in vegetation areas without protection, generally covered by herbaceous and sclerophyllous vegetation. In Greece, also noteworthy was the number of ML identified on the islands, mainly along rocky coastal shores. In the case of Germany, a generalized absence was observed in the central and flat part of the country, concentrating the ML around the mountainous region located in the south, dominated by rocky surfaces, and in the northwest of the country, where several river valleys are located, and moorland and heathland covers are predominant. Poland was the country with the least area identified as ML, being scattered throughout the country and with no remarkable clusters. This could be due to the fact that ML in Poland are related to unmanaged areas mainly covered by herbaceous vegetation without a defined land use.

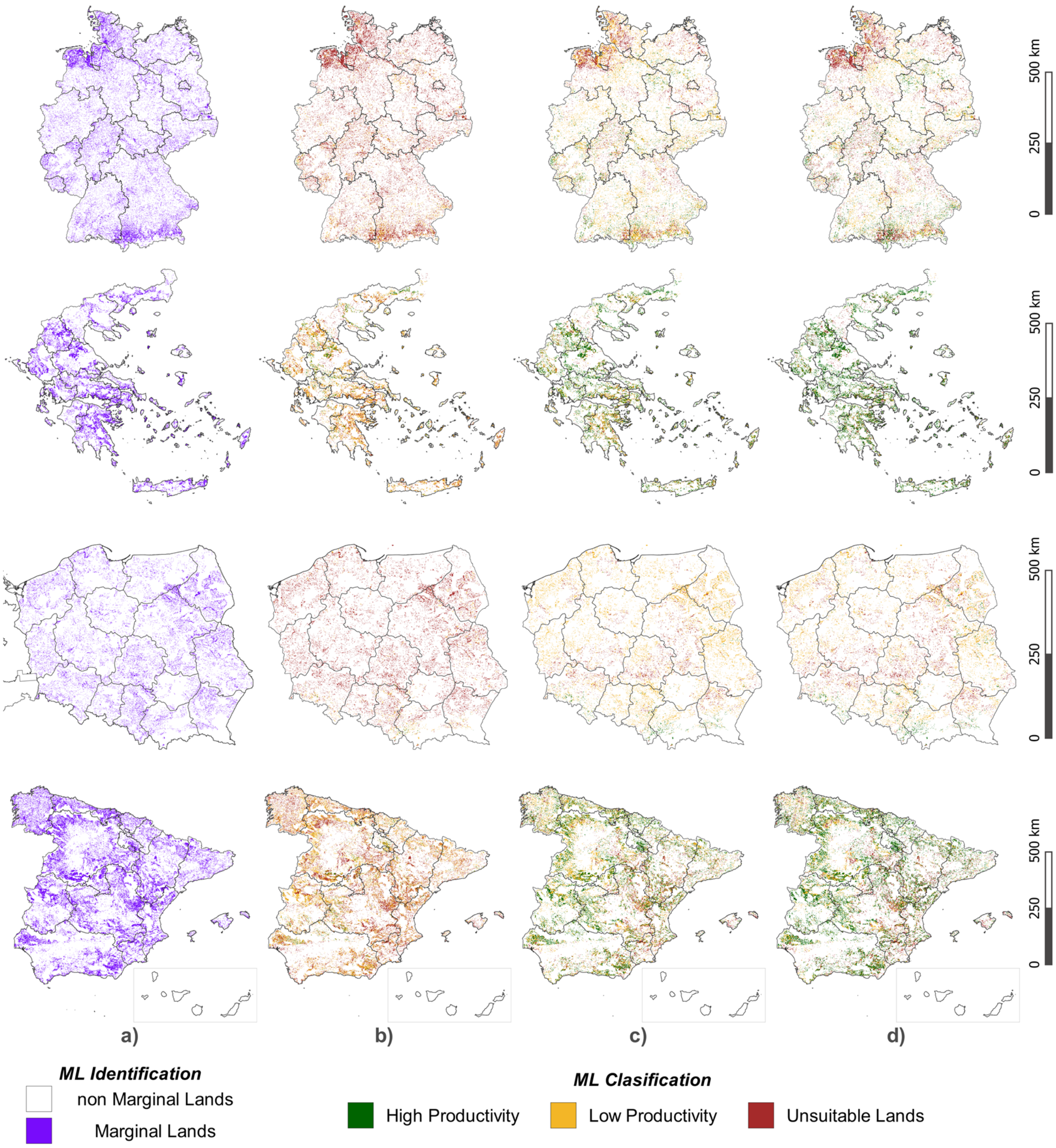

Figure 3: Graphical representation of the potential ML identification (a) and the ML classification for the three categorization approaches (b = Equal magnitude, c = 25th-75th percentiles, and d = 33rd-66th percentiles). The countries represented from top to bottom are Germany, Greece, Poland, and Spain.

Table 3 shows the confusion matrix of the identification of ML organized by country. The best accuracy was achieved in Poland, with an overall accuracy of 90.97%, then in Spain with 82.87%. In both cases, the concordance measured with the kappa index was above 0.6, and the overall quality of the classification (ML/nonML) measured with the F-scores was above 80%. For Germany, the overall accuracy was the lowest with 60.61% and a very poor concordance of 0.04. The error in Germany was due to the validation sample selection, since most of the ML samples were taken in protected areas with marginal characteristics (identified by the state of the soil and the present vegetation).

Table 3: Confusion matrices, global precision, F-score, and Kappa index by country for ML identification.

Germany	*ML*	*nonML*	*Total*	*Recall (%)*	*Overall Accuracy*
ML	317	8,436	8,753	90.06	0.61
nonML	35	12,477	12,512	59.66	
Total	352	20,913	21,265		
F-Score (%)	61.64				
Kappa	0.04				

Greece	*ML*	*nonML*	*Total*	*Recall (%)*	*Overall Accuracy*
ML	5,902	1,691	7,593	73.89	0.72
nonML	2,086	3,583	5,669	67.94	
Total	7,988	5,274	13,262		
F-Score (%)	70.69				
Kappa	0.41				

Poland	*ML*	*nonML*	*Total*	*Recall (%)*	*Overall Accuracy*
ML	292	24	316	54.17	0.91
nonML	247	2,439	2,686	99.03	
Total	539	2,463	3,002		
F-Score (%)	83.43				
Kappa	0.63				

Spain	*ML*	*nonML*	*Total*	*Recall (%)*	*Overall Accuracy*
ML	1,396	406	1,802	84.66	0.83
nonML	253	1,793	2,046	81.54	
Total	1,649	2,199	3,848		
F-Score (%)	80.90				
Kappa	0.65				

Figure 4 shows a map representing the numerical gradient of the "Soft" constraints for a detail area in Spain. In the classified layer, a minimum value of 0.12 (high marginality) and a maximum value of 6.86 (low marginality) were obtained, whereas the maximum theoretical value with respect to the sum of all the "Soft" indicators was 9.98. This shows that all zones identified as marginal with "Hard" layers were assigned a value obtained from some "soft" constraint, as no area obtained the theoretical minimum of zero. The values obtained as a result of the sum of all soft constraints and the three classification approaches with their respective value ranges are shown in Table 4.

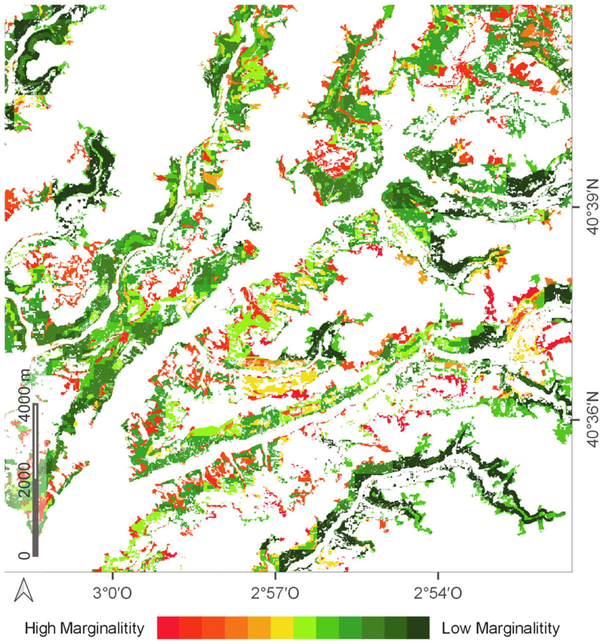

Figure 4: Representation of the gradient obtained in the application of all the "Soft" constraints on an area of Guadalajara (Spain). White areas represent nonML.

Table 4: Methods to subdivide the ML types.

Method	*Equal Magnitude*		*25th-75th percentiles*		*33rd-66th percentiles*	
	Max.	*Min.*	*P25*	*P75*	*P33*	*P66*
Value	6.86	0.12	1.39	2.89	1.58	2.52
	Thresholds					
High productivity ML	6.86	4.62	6.86	2.89	6.86	2.52
Low productivity ML	4.62	2.37	2.89	1.39	2.52	1.58
Potentially unsuitable land	2.37	0.12	1.39	0.12	1.58	0.12

Figure 3b-d shows the result of mapping the three categories of ML in terms of productivity using the three different classification approaches for the four countries. There is no optimal methodology to classify these three types of ML. According to the final use or application of the map, different thresholds should be defined, since there is no definition that clearly limits the 3 classes. These classifications give us a vision of the limitations and opportunities of the territory.

4. Conclusions

The definition of marginal land is ambiguous and does not have a spatial representation or a spectral response directly measurable using remote sensing data. ML are mainly marginal due to their exclusion from management for not being operationally effective or productive for agriculture, or for not having a minimum soil depth, or have extreme chemical-toxic soil values, all these are factors that block forest growth. The identification methodology presented in this paper quantified areas with undefined or non-specific LULC and with a clear potential to be considered ML. From the area covered in this study, Mediterranean countries, with more extreme temperatures and rainfall regimes, and prolonged drought, present more extension corresponding to potential marginal areas than central-northern European countries.

In the classification methodology, the qualitative physical functions of soil, soil restrictions, landscape and productivity have been evaluated. In addition, a value was assigned to each ML patch, representing the marginality level attending to the soft constraints or factors considered in the study. The interspecific differences between countries indicate that, in order to improve the classification and provide applicability to the method, it would be convenient to adjust the thresholds of the indicators for each biogeographic zone.

Processing the information in GEE improved greatly the time performance and computer processing capacity, in addition to the direct availability of several data sets in the cloud.

For future work, it is recommended to include social and economic factors that influence the consideration of ML (i.e., demographic parameters, level of industrialization of the country, per capita income, agrarian policies, forest policies, distance to roads). Socio-economic factors can provide information about why an area of land has become marginal over time and location. Besides, understanding the socio-economic characteristics of an area can be a key factor in the successful management of the activities to be implemented on the ML. On the other hand, it would be desirable to include in future classification methodologies constraints that identify the ecosystem services of the land (i.e., flora and fauna protection, hydrological balance, prevention of exotic plants invasion, erosion control, prevention of eutrophication). In this context, all areas identified as marginal cannot be used for productive purposes in terms of their environmental impacts on biodiversity, water resources and landscape. Nevertheless, mapping the location of potentially marginal lands can help to identify areas with irreversible erosion risk, land degradation, promote biodiversity in isolated populations and contribute to climate change mitigation with sustainable reforestation actions.

Acknowledgements

This research has been funded by the European Commission through the H2020-MSCA-RISE-2018 MAIL project (grant 823805) and by the Fondo de Garantía Juvenil en I+D+i from the Spanish Ministry of Labour and Social Economy.

References

ABATZOGLOU, J. T., DOBROWSKI, S. Z., PARKS, S. A. and HEGEWISCH, K. C., 2018. TerraClimate, a high-resolution global dataset of monthly climate and climatic water balance from 1958-2015. Scientific Data.

BATJES, N. H., 2016. Geoderma Harmonized soil property values for broad-scale modelling (WISE30sec) with estimates of global soil carbon stocks. *Geoderma*, **269**, pp. 61–68.

BERTAGLIA, M., JOOST, S. and 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**, pp. 657–670.

CAI, X., ZHANG, X. and WANG, D., 2011. Land Availability for Biofuel Production. *Environmental Science & Technology*, **45**, pp. 334–339.

CARBONELL-RIVERA, J. P., ESTORNELL, J., RUIZ, L. A., TORRALBA, J. and CRESPO-PEREMARCH, P., 2020. Classification of UAV-based photogrammetric point clouds of riverine species using machine learning algorithms: a case study in the Palancia river, Spain. *ISPRS Archives*, **43**, pp. 659–666.

CIRIA, C., SANZ, M., CARRASCO, J. and CIRIA, P., 2019. Identification of Arable Marginal Lands under Rainfed Conditions for Bioenergy Purposes in Spain. *Sustainability*, **11**, 1833.

ELBERSEN, B., VAN EUPEN, M., MANTEL, S., VERZANDVOORT, S., BOOGAARD, H., MUCHER, S., CICARRELI, T., ELBERSEN, W., BAI, Z., IQBAL, Y., COSSEL, M., MCCALLUM, I., CARRASCO, J., CIRIA RAMOS, C., MONTI, A., SCORDIA, D. and ELEFTHERIADIS, I., 2018. Methodological approaches to identify and map marginal land suitable for industrial crops in Europe (Deliverable D2.6).

ELIASSON, Å., JONES, R. J. A., NACHTERGAELE, F., ROSSITER, D. G., TERRES, J.-M., VAN ORSHOVEN, J., VAN VELTHUIZEN, H., BÖTTCHER, K., HAASTRUP, P. and LE BAS, C., 2010. Common criteria for the redefinition of Intermediate Less Favoured Areas in the European Union. *Environmental Science & Policy*, **13**, pp. 766–777.

EUROPEAN ENVIRONMENT AGENCY, 2017. Copernicus Land Monitoring Service - Reference Data: EU-DEM.

EUROPEAN ENVIRONMENT AGENCY, 2018. Copernicus Land Monitoring Services - High Resolution Layers (HRL) for 2012, 2015 and 2018. Datasets. Available: https://land.copernicus.eu/pan-european/high-resolution-layers [8/01, 2019].

EUROPEAN ENVIRONMENT AGENCY, 2019a. Copernicus Land Monitoring Services - Corine Land Cover 1990 to 2018. Datasets. Available: https://land.copernicus.eu/pan-european/corine-land-cover [10/01, 2020].

EUROPEAN ENVIRONMENT AGENCY, 2019b. Copernicus Land Monitoring Services - Corine Land Cover 1990 to 2018. Datasets.

EUROPEAN ENVIRONMENT AGENCY, 2019c. Nationally designated areas (CDDA). Datasets. Available: https://www.eea.europa.eu/data-and-maps/data/nationally-designated-areas-national-cdda-15 [10/01, 2020].

EUROPEAN ENVIRONMENT AGENCY, 2019d. Natura 2000 data - the European network of protected sites. Datasets. Available: https://www.eea.europa.eu/data-and-maps/data/natura-11 [10/01, 2020].

EUROSTAT, 2020. Agri-environmental indicator - soil erosion. Available: https://ec.europa.eu/eurostat/statistics-explained/index.php/Agri-environmental_indicator_-_soil_erosion#Data_sources [11/10, 2020].

GEORGIADIS, C., TASSOPOULOU, M., TORRALBA, J., VERDE, N., GÓMEZ-CONEJO, R., GROMMY, E., ALEKSANDROWICZ, S. and KRUPIŃSKI, M., 2021. Deliverable 2.3 Report on Methodology development. Identifying Marginal Lands in Europe and strengthening their contribution potentialities in a CO2 sequestration strategy (MAIL). H2020 MSCA RISE 2018.

GERWIN, W., REPMANN, F., GALATSIDAS, S., VLACHAKI, D., GOUNARIS, N., BAUMGARTEN, W., VOLKMANN, C., KERAMITZIS, D., KIOURTSIS, F. and FREESE, D., 2018. Assessment and quantification of marginal lands for biomass production in Europe using soil-quality indicators. *SOIL*, **4**, pp. 267–290.

GOPALAKRISHNAN, G., NEGRI, M. C. and SNYDER, S. W., 2011. A Novel Framework to Classify Marginal Land for Sustainable Biomass Feedstock Production. *Journal of Environment Quality*, **40**, 1593.

HANSEN, M. C., POTAPOV, P. V, MOORE, R., HANCHER, M., TURUBANOVA, S. A., TYUKAVINA, A., THAU, D., STEHMAN, S. V, GOETZ, S. J., LOVELAND, T. R., KOMMAREDDY, A., EGOROV, A., CHINI, L., JUSTICE, C. O. and TOWNSHEND, J. R. G., 2013. High-Resolution Global Maps of 21st-Century Forest Cover Change. *Science*, **342**, 850 LP – 853.

HIEDERER, R., 2013. Mapping Soil Properties for Europe - Spatial Representation of Soil Database Attributes. Luxembourg.

HOLLANDER, J. H., 1895. The Concept of Marginal Rent. *The Quarterly Journal of Economics*, **9**, 175.

ISRIC-WORLD SOIL INFORMATION., 2020. SoilGrids250m 2.0 - Cation exchange capacity at ph7. DOI: 10.17027/isric-soilgrids.713396fc-1687-11ea-a7c0-a0481ca9e724.

IVANINA, V., ROIK, M. and HANZHENKO, O., 2016. Report on MagL concepts, debate and indicators (D2.3). Ukraine.

KANG, S., POST, W., WANG, D., NICHOLS, J., BANDARU, V. and WEST, T., 2013a. Hierarchical marginal land assessment for land use planning. *Land Use Policy*, **30**, pp. 106–113.

KANG, S., POST, W. M., NICHOLS, J. A., WANG, D., WEST, T. O., BANDARU, V. and IZAURRALDE, R. C., 2013b. Marginal Lands: Concept, Assessment and Management. *Journal of Agricultural Science*, **5**.

LI, G., MESSINA, J. P., PETER, B. G. and SNAPP, S. S., 2017. Mapping Land Suitability for Agriculture in Malawi. *Land Degradation and Development*, **28**, 2001–2016.

MALINOWSKI, R., LEWIŃSKI, S., RYBICKI, M., GROMNY, E., JENEROWICZ, M., KRUPIŃSKI, M., NOWAKOWSKI, A., WOJTKOWSKI, C., KRUPIŃSKI, M., KRÄTZSCHMAR, E. and SCHAUER, P., 2020. Automated Production of a Land Cover/Use Map of Europe Based on Sentinel-2 Imagery. *Remote Sensing*, **12**, 3523.

MELLOR, P., LORD, R. A., JOÃO, E., THOMAS, R. and HURSTHOUSE, A., 2021. Identifying non-agricultural marginal lands as a route to sustainable bioenergy provision - A review and holistic definition. *Renewable and Sustainable Energy Reviews*, **135**, 110220.

PEKEL, J. F., COTTAM, A., GORELICK, N. and BELWARD, A. S., 2016. High-resolution mapping of global surface water and its long-term changes. *Nature*, **540**(7633), pp. 418-422.

PENMAN, H. L., 1948. Natural evaporation from open water, hare soil and grass. *Proceedings of the Royal Society of London*. Series A, Mathematical and physical sciences.

REGER, B., OTTE, A. and WALDHARDT, R., 2007. Identifying patterns of land-cover change and their physical attributes in a marginal European landscape. *Landscape and Urban Planning*, **81**, pp. 104–113.

SAUER, T. J., JAMES, D. E., CAMBARDELLA, C. A. and HERNANDEZ-RAMIREZ, G., 2012. Soil properties following reforestation or afforestation of marginal cropland. *Plant and Soil*, **360**, pp. 375–390.

STRIJKER, D., 2005. Marginal lands in Europe—causes of decline. *Basic and Applied Ecology*, **6**, pp. 99–106.

ZHANG, J., HE, C., CHEN, L. and CAO, S. 2018. Improving food security in China by taking advantage of marginal and degraded lands. *Journal of Cleaner Production*, **171**, pp.1020–1030.

ZOLEKAR, R. B. and 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**, pp. 300–321.