INSTITUTO POLITÉCNICO DE BRAGANÇA
Escola Superior Agrária

Random Forest effectiveness for Bragança Region mapping: comparing indices, number of the decision trees, and generalization

Mariam AMEDJAR

Dissertation presented to the School of Agriculture of Bragança in partial fulfilment of the requirements for the degree of Master of Science in Forest Resources Management

Advisors

Prof. Dr. João Paulo Miranda de Castro

Prof. Dr. Marina Castro

School of Agriculture of Bragança (PORTUGAL)

Prof. Dr. Mustapha Naimi

Prof. Dr. Imane Sebari

Agronomic and Veterinary Institute Hassan II (MOROCCO)

Bragança

October 2020

Acknowledgements

First of all, I would like to thank deeply my thesis advisors Professor **João Paulo Miranda de Castro** and Professora **Marina Castro** of the of Escola Superior de Agrária at Instituto Politécnico de Bragança. They steered me in the right direction whenever they thought I needed it.

My special thanks to Professor **Mustapha Naimi** and Professora **Imane Sebari** from Agronomic and Veterinary Institute Hassan II (MOROCCO) for their advices toward the work.

My sincere thanks to Erasmus+ Project, namely to Professor **Paulo Cortez**, and to the International Relations Office (IRO), and Dr **Joana Aguiar** for their help and backness, to offer up this chance to do my Master Degree in Portugal. A big thank, to Professor **Chtaina Noureddine,** Professor **Bouhaloua Mhammed** and Mrs. **Sanaa Zebakh** from Agronomic and Veterinary Institute Hassan II in Morocco,

My special thanks to all the teachers of the units courses I received during the Academic Year of 2019/2020, related to the Master Degree on Management of Forest Resources in the School of Agiculture of Polytechnic Institute of Bragança in Portugal, for their big support and also their love and respect.

My special thanks go to my colleagues in EMEN class and all my friends, especially: **Achraf** ISMAILI ALAOUI, **Ghita** MANGAD, **Kaoutar** BENGUEDDOUR and **Oumayma** BASSAIRATE

Finally, I would like to express my endless gratefulness to my parents and all my sisters &brothers, for providing me with unfailing support and continuous encouragement throughout my years of study, this accomplishment would not have been possible without them. Thank you.

I Love you all.

Mariam AMEDJAR

Abstract

Remote sensing is a domain that tends to use satellite images for classification and Land Use/Cover (LULC) mapping. For this purpose, classification algorithms are used, which are numerous and diverse, and it is necessary to establish decision criteria when choosing the algorithm. Ultimately, the main decision criterion will be the accuracy obtained in classification because the accuracy of classification may differ from one algorithm to another, even within the same algorithm, according to its variables. But there are other equally important criteria: it depends on the nature of the task, the quantity and types of data available, the type of response expected, the time and computational resources available, the depth of our knowledge about the algorithms.

The methodology of each part of the work was described and the criteria for comparison were established. In this research, with the same training data, the same validation data, the same application context (7 classes), and the same image data (Sentinel-2), we tested 15 iterations with the Random Forest classification algorithm, with different tree number decision values, and 3 iterations with vegetation and soil indexes, for the production of the LULC map of the Bragança region (northeast Portugal). Finally, we evaluate the accuracy of the classification, before and after the post-classification tasks (generalization, fragmentation and removal of isolated pixels).

The results obtained show that a classification with an nb-trees = 1000, including vegetation and soil indices, and after post-classification tasks, provided excellent precision results (Coefficient Kappa = 0.93, Overall accuracy = 96%, and marginal errors of omission & commission below 4%).

Resumo

A teledetecção é um domínio que tende a utilizar imagens de satélite para classificação e mapeamento de Uso/Cobertura da Terra (LULC). Para este fim, são utilizados algoritmos de classificação, que são numerosos e diversos, sendo necessário estabelecer critérios de decisão ao escolher o algoritmo. Em última análise, o principal critério de decisão será a precisão obtida na classificação, porque a precisão da classificação pode diferir de um algoritmo para outro, mesmo dentro do mesmo algoritmo, de acordo com as suas variáveis. Mas existem outros critérios igualmente importantes: depende da natureza da tarefa, da quantidade e tipos de dados disponíveis, do tipo de resposta esperada, do tempo e dos recursos computacionais disponíveis, da profundidade dos nossos conhecimentos sobre os algoritmos.

A metodologia de cada parte do trabalho foi descrita e os critérios de comparação foram estabelecidos. Nesta investigação, com os mesmos dados de formação, os mesmos dados de validação, o mesmo contexto de aplicação (7 classes), e os mesmos dados de imagem (Sentinel-2), testámos 15 iterações com o algoritmo de classificação Random Forest, com diferentes valores de decisão de número de árvores, e 3 iterações com índices de vegetação e solo, para a produção do mapa LULC da região de Bragança (nordeste de Portugal). Finalmente, avaliámos a exactidão da classificação, antes e depois das tarefas de pós-classificação (generalização, fragmentação e remoção de pixels isolados).

Os resultados obtidos mostram que uma classificação com um nb-trees = 1000, incluindo índices de vegetação e solo, e após tarefas de pós-classificação, forneceu excelentes resultados de precisão (Coeficiente Kappa = 0.93, Precisão geral =96%, e erros marginais de omissão & comissão abaixo de 4%).

Table of contents

List of figures

List of tables

List of equations

List of Annexes

Symbols list

AC: Annual Crops

B: Blue

BI: Brightness Index

BOA: Bottom Of Atmosphere

Cab: Chlorophyll content in the leaf

CI: Colour Index

Conif: Coniferous

COS: Carta de Uso e Ocupação do Solo

DN : Digital number

ESA: European Space Agency

FAPAR: Fraction of Absorbed Photosynthetically Active Radiation

FCOVER: Fraction of vegetation cover

G: Green

G: Grasslands

GMES: Global Monitoring for Environment and Security

LAI: Leaf Area Index

LULC: Land Use/Cover

MD: Mahalanobis distance

NASA: National Aeronautics and Space Administration

NIR: near-infrared

OA: Overall Accuracy

OOB: out-of-bag

PA: Producer Accuracy

PAR: Photosynthetically active radiation

PC: Permanent Crops

R: Red

RF: Random Forest

RI Redness Index

ROI: Region of interest

RS: Remotre sensing

Shrubs: Shrublands

SPOT : Systeme Probatoire d'Observation de la Terre

S**VM**: Support Vector Machine

SWIR: shortwave infrared

TD: Training data

TOA: Top Of Atmosphere

TS: temporal resolution

UA: User Accuracy

VD: Validation data

VIS: visible

1 Context and objectives

1.1 Context

Remote sensing measures the Earth and its features deprived of making physical contact, the data can be gathered from entire continents over a long time period, so the possibility of tracking the Earth changing. By using specialized aircrafts and sophisticated satellites to gather DATA with remote sensing methods.

The modern era of Earth remote sensing from satellites began in 1978. Three very important new satellite systems were launched in space in that year (Cracknell, 2018), and since then the focus has shifted to the improvement achieved. For instance, (1) a small satellite design approach was followed by various companies and space agencies throughout the world (Kramer & Cracknell, 2008), (2) the small size satellites achievement has positively affected the cost and the time needed thus a short development times and cost efficiency (Xue et al., 2008), (3) the improvement of satellite output quality, that can be accurate for measurements that can be indicated with a precision only a few meters, and within centimetre's(Pelton & Camacho-Lara, 2013).

Land use & land cover are two distinct terminologies which are frequently used mutually (DIMYATI et al., 1996). Land cover indicates the physical characteristics of the land surface such as the distribution of vegetation, water, soil and other physical characteristics of the Earth, including those created exclusively by human activities, for instance settlements. On the other hand, land use refers to the approach in which land has been used by man, for his habitation, but usually with a focus on the functional role of land for his economic activities (Rawat & Kumar, 2015)

In terms of teledetection, Land Use and Land Cover (LULC) map which represent the spatial information, is an ending of classification performed on satellite images data (Sentinel for instance) using classifiers dedicated to create classes based on pixels. There are plenty of classifiers and each one should be selected with criteria because they can provide better or worse accuracy. Even each classifier (Algorithm), its accuracy depends on properties and indexes that can be added. In this research, we focused on the production of the LULC map of the Bragança region (northeast of Portugal) by Sentinel-2 satellite image classification, using the Random Forest (RF) classification algorithm. Seven classes were considered. Several

iterations were tested, with different tree number decision values, with or without vegetation indexes and soil indexes.

Remote sensing community focusing on image classification has been attracted by remote sensing owing to the accuracy of results acquired, its important role in detecting, mapping, understanding, and predicting changes in the environment (Rose et al., 2015). Early applications mainly assessed land use and cover change, such as estimation of forest cover losses (Tyukavina et al., 2015) and mapping carbon stocks in the amazon (Asner et al., 2010), to identify critical migrant bird breeding habitat (Goetz et al., 2010) and assess the influence of light pollution on seabirds (Rodrigues et al., 2011), detect abandoned cropland (Löw et al., 2015) More recently, remote sensing has been used, for example, to Assess the Ecosystem Service Potential of Urban Rivers (Beißler & Hack, 2019), in prediction of soil erosion (Jazouli et al., 2019) mapping the proliferation of aquaculture ponds (Al Sayah et al., 2020), moreover the assessment of groundwater use in irrigated agriculture (Nhamo et al., 2020), evaluating the ecological quality of a mining area for management goals (Zhu et al., 2020).

Thus, LULC mapping plays a key role in studying and analysing the overall changed scenario. Currently, data availability on such changes is crucial, for providing critical contribution to decision-making of environmental planning and ecological management for future (Zhou et al., 2008).

1.2 Objectives

Objective 1: To evaluate the criteria to be taken into account when choosing a satellite image for the production of LULC maps of the Bragança region using Random Forest classifier (machine learning).

Objective 2: finding the optimum number of trees in order to have a precise classification.

Objective 3: analysing the effect (accuracy) of post-classification using GIS tools (ArcGIS)

Objective 4: assess the contribution of integrating vegetation and soil indices on the classification accuracy

2 Bibliographical Review

2.1 Remote sensing & GIS overview

2.1.1 Introduction

Earth has only been observed as a blue sphere in the middle of Space by privileged people: astronauts who have travelled to the Moon. Apollo 8 was the initial crewed spacecraft to successfully orbit the Moon and return to Earth in late 1968. The Apollo 8 crew were the first to benefit of witnessing and photographing an Earthrise (Loff, 2018) (Figure 1) . Fortunately, we can all have this vision of Earth today thanks to remote sensing satellites. But then how can the term "Remote Sensing", "Teledection" be defined? Remote sensing can be broadly defined as the collection and interpretation of data about an object, an area or an event without being in physical contact along with the object, this set of techniques make it possible, through the acquisition of images, to acquire information about the Earth's surface (including the atmosphere and the oceans), without any direct contact with it. RS encompasses the entire process of capturing and recording to process and analyse the information, in order to apply this latter.

Figure 1: The view of the rising Earth by the Apollo 8 astronauts in December 1968 (Loff, 2018)

2.1.2 Definition & history

Remote sensing is a set of techniques which differ from each other by the type of vector (airplane, satellite or space shuttle, ground based sensors) (D Rudd et al., 2017), the acquisition mode (analogue or digital, active or passive), the spatial resolution, the spectral range used and the observed area. The remote sensing data correspond to measurements integrated on a spatial

scale (pixel) of certain characteristics of the surface brightness (Levin, 2017), apparent temperature (Venter et al., 2020), backscatter coefficient (Belenguer-Plomer et al., 2019). This is another type of data that should be calibrated and transformed into useful variables for models: Determination of the Ground Albedo (Wu et al., 2018), downscaling both sea-temperature (Minnett et al., 2019) and land-temperature (Y. Yang et al., 2017) , 3D estimation of surface biomass (Wallace et al., 2017), primary production (Gómez-Giráldez et al., 2019), etc.)

The photograph in the visible part of the electromagnetic wavelength was the primary form of remote sensing but technological progresses have made it possible to acquire information in further wavelengths, including near infrared, thermal infrared and microwaves (Merlin et al., 2010). Collection of information over a large number of wavelength bands is called multispectral (Landgrebe, 2005) or hyperspectral (Govender et al., 2007) data. The development and deployment of manned and unmanned aerial vehicle have improved remote sensing data collection and provides an inexpensive way to obtain information on large areas (Rango et al., 2009). However, the capacities offered by the observations provide development and the widespread use of this type of technology in a systematic way.

The field scope can be developed by examining its history, in order to trace the evolution of its concepts (Table 1).

Remote sensing, as it is understood today, made its debut in 1946-1950 period with the launch of the first V-2 rockets (Figure 2) acquired from Germany after second world war at high altitude from White Sands, New Mexico (United States). These V-2 rockets contained a small integrated cameras and sensors installed that offered a large-scale geographic view of the Earth showing an area of approximately 800.000 sq. Miles (Figure 3) (Edgington, 2012).

From that moment, various projects and missions with other rockets, ballistic missiles and satellites took photos of the Earth. Although the quality of the first photos of the space was not good, it revealed the real potential associated to this technique. Systematic Earth observation has started in 1960 with the launch of TIROS-I (Schnapf, 1985), the first meteorological satellite equipped with a low spatial resolution television, which allows meteorologists to distinguish between clouds, water, ice and snow (Fiolek, 2011) . The series of satellites TIROS, known since 1970 as NOAA (National Oceanic and Atmospheric Administration), are still active NOAA-20 called JPSS-1 is; for now; the last satellite placed in orbit on 2017, and the

next one in the series, JPSS-2, is scheduled to launch in the first quarter of 2022 (NASA's Scientific, 2020).

The encouraging future that awaited Remote Sensing is definitely confirmed with the first space programs of the 1960s (Mercury, Gemini and Apollo lunar mission)(Lowman, 1999). In the orbits described by Apollo 9 around the Earth and before landing on the Moon, the first controlled experiment was carried out of multi spectral photography to study the natural resources of the Earth's surface (Sharp, 2018).

Table 1: Milestones in the history of remote sensing

Year	Invention	Author
1839	Beginning of practice of photography	Daguerre, Talbot and Niepce
1940	From photo produce topographic map	French
1858	Photography from balloons	Gaspard Felix Tournachon
1909	Photography from airplane	Wilbur Wright
1910's	World War I: aerial reconnaissance	British Royal Air Force
1960-1970	Satellite era: First use of term remote sensing	
	TIROS: weather satellite	NASA
	Skylab remote sensing observation from space	CORONA
1972	Launch of Landsat 1	NASA
1970-1980	Rapid advances in digital image processing	-
1980-1990	Launch of new generation of Landsat sensor	NASA
1986	SPOT French Earth observation satellite	French centre National d'Etudes Spatial (CNES)
1980's	Development of hyperspectral sensors	-
1990's	Global remote sensing system, LIDAR	-
1990's	First Commercial satellite IKONOS	Space Imaging

Figure 2: The first V-2 rocket launched (Hollingham, n.d.)

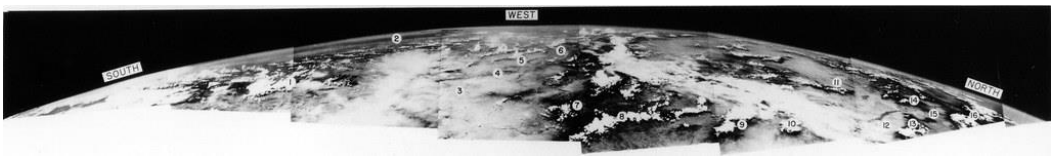


Figure 3: V-2 Rocket-eye view from 100 Km up, July 26 1948, Source: Photo Courtesy of the Applied Physics Laboratory at Johns Hopkins University.

2.1.3 Process of remote sensing

Figure 4 shows the requirements needed for remote sensing process. Details are given below:

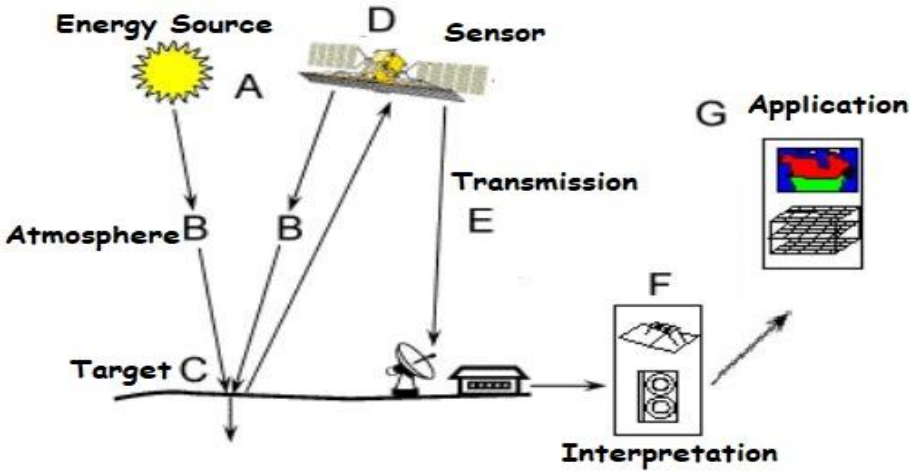


Figure 4: spectral signature of soil, water and vegetation at different wavelengths. (Khan et al., 2018)

A- Energy source or illumination: this is the first requirement in the field of remote sensing. The source (i.e. sun) provides electromagnetic energy to the object of interest (crops, forests, buildings ...)

B- The Atmosphere: from its source to the Earth's surface, the radiation is in contact and interaction with the atmosphere passing through it. This interaction can take place another time as energy returns to the sensor from the target.

C- Target: once the energy reaches the target through the atmosphere. The radiation interacts; based on several properties; with the targeted surfaces.

D- Sensor: the sensor on board a satellite collects and records EM radiation reflected from the Earth's surface and the atmosphere.

E- Transmission, Reception, & Processing: The radiation received by the sensor is transmitted in electronic form, to a host station in order to process the data in an image (paper / digital).

F- Interpretation & Analysis: at this level, the image received and processed is interpreted. The operation is done in visually / digitally or electronically way in order to extract information about the defined target.

G- Application: the last step in the remote sensing process is to apply data extracted from the imagery in order to better understand it. Extract other more detailed information that serves to solve certain problems in the study area.

2.1.4 Sensors in remote sensing

Remote sensors are mechanical devices that have the role of collecting information in the form of energy (EMR or other), usually in storable form. All remote sensors record in selected wavelength bands, variation in the amount of energy either reflected or emitted by various materials on the surface of the Earth. Remote sensing data is characterised in terms of spatial, spectral, temporal and radiometric resolutions.

The **spatial resolution**, also known as ground resolution, specifies the pixel size of images covering the Earth's surface (Figure 5). The detail that can be discerned in an image depends on the sensor's spatial resolution used. It is a function of the ability to detect the smallest element. A sensor's spatial resolution depends mainly on its Instantaneous Field of View (IFOV). As shown in Figure 5, the IFOV is defined as the visible cone of the sensor (A) that determines the area of the surface "visible" at a given altitude & a specific time (B). The size of this area is obtained by multiplying the IFOV by the distance from the surface to the sensor (C) (P.K. Mani, 2014)

However, the spatial resolution of a satellite image plays an essential role in the cartographic process. This property is decisive and varies according to the needs of the cartographer. Depending on the work to be carried out, the spatial resolution will be different. Indeed, the identification of urban structures, for example, requires images with very high spatial resolution, while a mapping of land use patterns can be satisfied with a spatial resolution of more than 20 meters.

To more illustrate, the Figure 6 shows the quality of the view by changing the resolution.

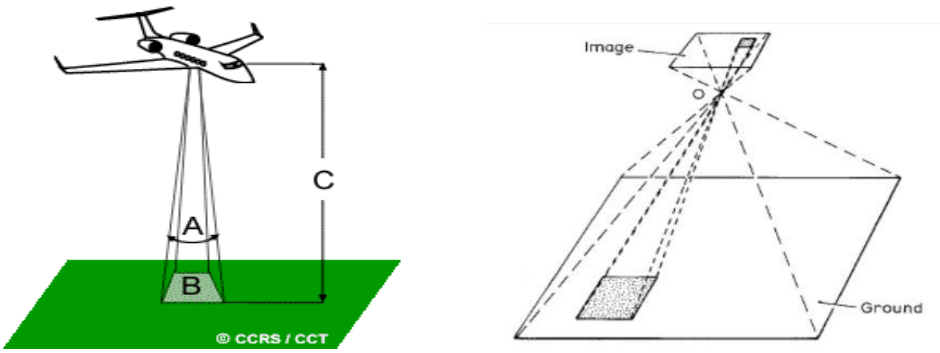


Figure 5: the explanatory elements of spatial resolution. Visibility of the sensor(A), area seen from a given altitude at a given time (B), altitude (C) (left) and the Illustration of the geometrical instantaneous field of view reconstructed by projection from a pixel in the image plane (right).

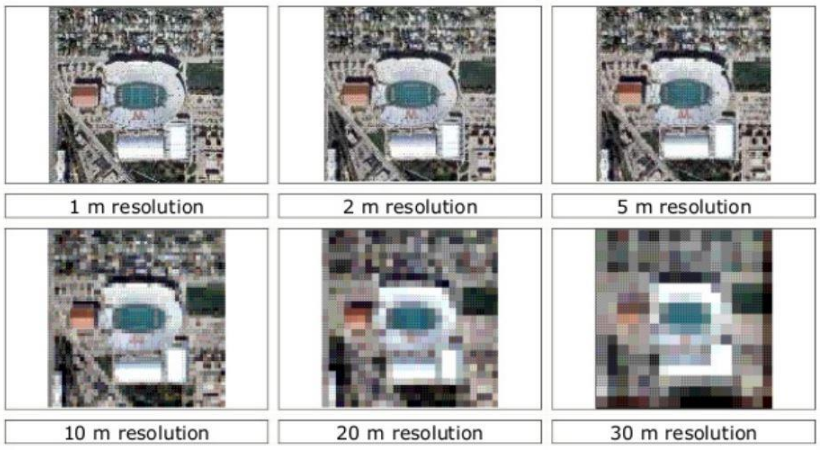


Figure 6: Camp Randal Stadium, University of Wisconsin, viewed at different resolutions (image: Institute for Environmental Studies, University of Wisconsin)

The **Spectral resolution** is the wavelength width of the bands recorded. A sensor's spectral resolution identifies the number and the position of spectral bands inside the electromagnetic spectrum that a sensor can gather reflected radiation. Hence, a sensor should at the same time be sensitive to a large portion of the electromagnetic spectrum and captures many bands within that portion. The higher the number of small bands captured, greater is the spectral resolution showing detailed distinction. That means each individual element is in state of reflecting or emitting electromagnetic energy in a unique way, this differentiates elements in the scene. Multispectral images record energy over several separate wavelength ranges at various spectral resolutions. allowing the extraction of extra information that a human eye fails to capture with its receptors for red, green and blue. Multispectral imaging was originally developed for space-based imaging applications.

The simplest form of spectral resolution is a sensor with one band only, which senses visible light. For instance, an image from that sensor, with only one band, is compared to an aircraft white & black image. In case the sensor has in the visible region of EM spectrum three spectral band, this indicates the possibility of capturing the same data captured by human eyes.

The **temporal resolution** is a measure of the frequency with which a sensor returns to the same part of the Earth's emplacement imaging the exact same area at the same view angle (Yao, 2009). However, some areas of the surface may be observed more frequently since there is an overlap between adjacent orbits and these areas of overlap become larger and larger as they approach the poles with increasing latitude. The frequency is going to differ from numerous times per day, for a characteristic weather satellite, to 8→20 times a year for a moderate ground resolution satellite.

The **radiometric resolution** describes the information contained in an image. Each time an image is token by a sensor, its sensitivity to the electromagnetic energy's intensity regulates the radiometric resolution (Muhammad Zulkarnain). The radiometric resolution of a remote sensing system describes its ability to recognize small differences in electromagnetic energy. It is a measure of how many grey levels are measured between pure black -no reflectance- and pure white. The finer is the radiometric resolution, more sensitive to small differences in the intensity of the energy received the sensor is. The radiometric resolution is measured in bits and is used to encode values in binary format: $1bit = 2^1 = 2$ grey levels; $2bit = 2^2 = 4$; ...; a sensor using 8 bits to record data will have $2^8 = 256$ available intensity levels (ranging from 0 to 255), or Digital Numbers (DN). The higher the number of bits the higher the radiometric resolution.

2.1.5 Examples of some Earth observation sensors

Landsat was the first satellite to monitor the Earth, launched in 1972. All Landsat satellites cross the equator in the morning to take advantage of optimal lighting conditions. Landsat1, 2 and 3 were placed at an altitude of 900 km with a repeatability of 18 days while the other satellites (Landsat 4 to Landsat 5) were placed at an altitude of 700 km with a repeatability of 16 days. From 1982, by creating the Landsat 4, there was an improvement, better spatial and radiometric resolution, a larger number of bands, spectral bands that are narrower. The Table 2 below summarizes the spectral domains of the different bands of the Landsat-8 with their resolution. Landsat 8 imagery captures data with a radiometric resolution of 12-bits.

Table 2: Sensitivity and Resolution Used on Landsat-8,

Bands	Spectral Sensitivity (µm)	Nominal Spectral Location	Ground Resolution(m)
Band 1	0.43 - 0.45 µm	Visible	30 x 30
Band 2	0.450 - 0.51 µm	Visible	30 x 30
Band 3	0.53 - 0.59 µm	Visible	30 x 30
Band 4	0.64 - 0.67	Red	30 x 30
Band 5	0.85 - 0.88	Near-Infrared	30 x 30
Band 6	1.57 - 1.65	SWIR 1	30 x 30
Band 7	2.11 - 2.29	SWIR 2	30 x 30
Band 8 (Panchromatic)	0.50 - 0.68	False colour	15 x 15
Band 9	1.36 - 1.38	Cirrus	30 x 30

The **Spot** (Systeme Probatoire d'Observation de la Terre) is a series of Earth observation satellites have been designed and launched by the French National Centre for Space Studies (CNES) in collaboration with Sweden and Belgium for the acquisition of remote sensing data for commercial purposes. All the SPOT satellites are placed at an altitude of 830 km with a repeatability of 26 days, equipped with sensors that develops with each launch, offering spatial resolution reaching 5m. The Table 3 below summarizes the spectral domains of the different bands of the SPOT-4 with their resolution. The pixel depth of SPOT- 6 /-7 is 12 bit, an improvement over previous versions (SPOT- 4 /-5) which was 8 bit.

Table 3: Sensitivity and Resolution Used on SPOT-4-5-6-7

Sensor	Electromagnetic spectrum	Pixel size	Spectral bands
SPOT 6-7	Panchromatic	1.5m	0.45 - 0.75 µm
	B1 : blue	6 m	0.45 – 0.53 µm
	B2 : green	6 m	0.53 – 0.59 µm
	B3 : red	6 m	0.63 – 0.70µm
	B4 : near infrared	6 m	0.76 – 0.89µm
SPOT 5	Panchromatic	2.5 m or 5 m	0.48 - 0.71 µm
	B1 : green	10 m	0.50 - 0.59 µm
	B2 : red	10 m	0.61 - 0.68 µm
	B3 : near infrared	10 m	0.78 - 0.89 µm
	B4 : short wave infrared (SWIR)	20 m	1.58 - 1.75 µm
SPOT 4	Monospectral	10 m	0.61 - 0.68 µm
	B1 : green	20 m	0.50 - 0.59 µm
	B2 : red	20 m	0.61 - 0.68 µm
	B3 : near infrared	20 m	0.78 - 0.89 µm
	B4 : short wave infrared (SWIR)	20 m	1.58 - 1.75 µm

The **Sentinel** are a series of Earth observation missions developed by European Space Agency (ESA), on behalf of the joint ESA/European Commission initiative GMES (Global Monitoring for Environment and Security). The objective of Sentinel program is to substitute the existing elder Earth observation missions to avoid interruptions in progress studies. Every mission will concentrate on a varied aspect of Earth surveillance; Oceanic, Atmospheric, and Land monitoring (Missions - Sentinel Online.). Each Sentinel mission is based on a constellation of two satellites to fulfil and revisit the coverage requirements for each mission, providing robust datasets. All the Sentinel (except Sentinel-4) satellites are placed at an altitude between of 786-830 km which gives an orbit cycle range between 10-29 days, as already mentioned, each mission got two satellites. Thus, the repeatability is the half of orbit cycle period. Sentinel-4 is placed in about 35786km above the Earth equator, its orbital period is equal to the Earth's rotational period.

The Sentinel-2 instrument acquires measurements at 12 bits. These measurements are converted to reflectance and stored as 16 bit integers in the S2 product (*Radiometric - Resolutions - Sentinel-2 MSI - User Guides - Sentinel Online*). Sentinel-2 is considered as the newest mission for Earth observation by the ESA. It has obtained a great attention in research due to its open access and global coverage. A varied range of studies have been made with Sentinel-2 , such as soil moisture mapping (El Hajj et al., 2017), urban surface water bodies mapping (X. Yang et al., 2017), forest stress monitoring (Eitel et al., 2011), and ground biomass quantifying. The Table 4 below summarizes the spectral domains of the different bands.

Table 4: Spectral Sensitivity and Ground Resolution Used on Sentinel-2 (A and B)

Bands	**Spectral Sensitivity (µm)**		**Nominal Spectral Location**	**Ground Resolution (m)**
	Sentinel-2A	Sentinel-2B		
Band 1	0.432 - 0.453	0.432 - 0.453	Coastal aerosol	60 x 60
Band 2	0.459 - 0.525	0.459 - 0.525	Blue	10 x 10
Band 3	0.542 - 0.578	0.541 - 0.577	Green	10 x 10
Band 4	0.649 - 0.680	0.649 - 0.680	Red	10 x 10
Band 5	0.697 - 0.712	0.696 - 0.712	Vegetation red edge	20 x 20
Band 6	0.733 - 0.748	0.732 - 0.747	Vegetation red edge	20 x 20
Band 7	0.773 - 0.793	0.770 - 0.790	NIR	20 x 20
Band 8	0.780 - 0.886	0.780 - 0.886	Narrow NIR	10 x 10
Band 8A	0.854 - 0.875	0.853 - 0.875	Water vapour	20 x 20
Band 9	0.935 - 0.955	0.933 - 0.954	SWIR	60 x 60
Band 10	1.358 - 1.389	1.362 - 1.392	SWIR – Cirrus	60 x 60
Band 11	1.568 - 1.659	1.563 - 1.657	SWIR	20 x 20
Band 12	2.115 - 2.290	2.093 - 2.278	SWIR	20 x 20

2.1.6 Discussion on the choice of sensor

Before acquiring an image in order to perform a classification, the following question must be asked first "What are the criteria to be taken into account when choosing a satellite image?" knowing that there are a multitude of images available.

For our study we chose Sentinel-2 as the sensor which provides the images, the first motif is its spectral response, (here we make a comparison between Sentinel-2 (MSI) and Landsat 8 (OLI) bands). The MSI sensor (Table 4) captures images in the visible (VIS), near-infrared (NIR), and shortwave infrared (SWIR) spectral regions through 4 spectral bands of 10 m, 7 bands of 20 m, and 3 bands of 60 m, and concerning the Landsat sensor (Table 2), it captures images in the visible (VIS), near-infrared (NIR), and shortwave infrared (SWIR) spectral regions, through 9 spectral bands of 30 m spatial resolution and an additional panchromatic band of 15 m. from (Figure 7) we can see that, Both sensors are offering 4 matching bands (NIR, R, G and B) respectively 8, 4, 3 and 2 for MSI and 5, 4, 3 and 2 for OLI. but the sentinel offers 3 other bands more 5,6 and 7 which are dedicated to capture the vegetation ramp, which is extremely useful for vegetation studies.

Secondly, the choice of sensor is related to temporal resolution (TS), Landsat has a TS of 16 days, whereas, with the availability of both Sentinel-2A and Sentinel-2B, temporal resolution is increased to five days, increasing the chance of acquisition of cloud-free images.

Finally, because they are provided at no cost to the European user. In addition, the European Space Agency's (ESA) provides the SNAP software (and Sen2cor for Correction of Atmospheric Effects), also free of charge, which was used in the processing of the Sentinel-2 images.

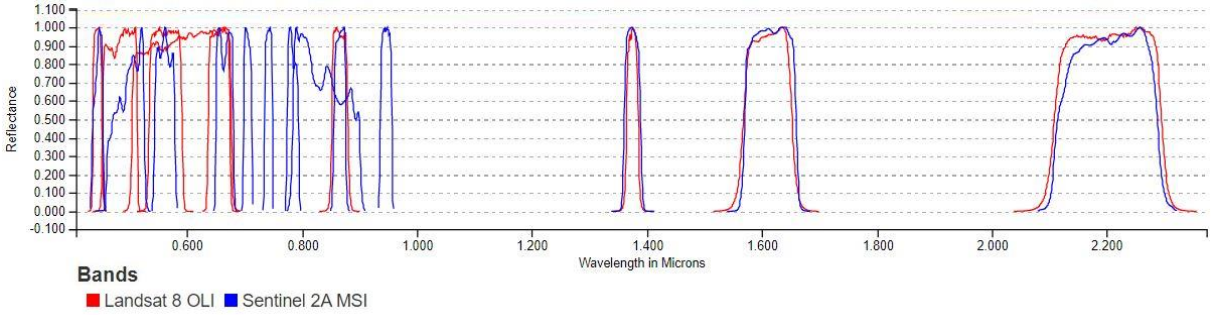


Figure 7:Spectral response of Sentinel-2 and Landsat 8 (*Spectral Characteristics Viewer | Landsat Missions*)

2.1.7 Classification algorithms

This section delineates the supervised image classification techniques that are used recently.

The **Maximum likelihood classifier** is one of the most popular supervised classification techniques used with remote sensing image data. This method is based on the probability value that a pixel is belonging to a specific group (Thakur & Maheshwari, 2017). The basic theory assumes that the input bands have normal distributions and that the probabilities are equivalent for entire classes. However, this technique needs large computational time, but it's highly efficient when it comes to classifying the satellite images, especially the multi-spectral images.

The maximum likelihood classification has a number of advantages including:

- Take several variables (bands and band transformations) into consideration.
- Take into account the variability of the classes using the covariance matrix.

This classification also has drawbacks, namely:

- An important equation which takes a long time to calculate. The latter increases with the number of input bands.
- Maximum likelihood strongly depends on a normal distribution of the data in each input band.
- Tends to outclass signatures with relatively large values in the covariance matrix. If there is a large distribution of pixels in a sample site, the covariance matrix for that signature contains large values.

The **Support Vector Machine** (SVM) is a supervised machine learning technique that can be used for both regression and classification problems. It creates a hyperplane in multidimensional space to split dataset into separated classes in the correct possible way (Figure 8). The objective of an SVM is to select the most optimal plan (hyperplane) which separates data into representative classes, the hyperplane for which the margin (the distance between the hyperplane and the nearest data point from either set) is maximum (Figure 8). The approach in which an SVM admits the optimal hyperplane is Calculating the Margin; gap between the support vectors and the plane; and choosing the plan which enlarges the gap between classes (*Support Vector Machines Tutorial - Learn to Implement SVM in Python*, 2017). SVMs are able to handle problems where classes are not linearly separable by transforming the data using a kernel function such as the radial basis function (RBF) kernel.

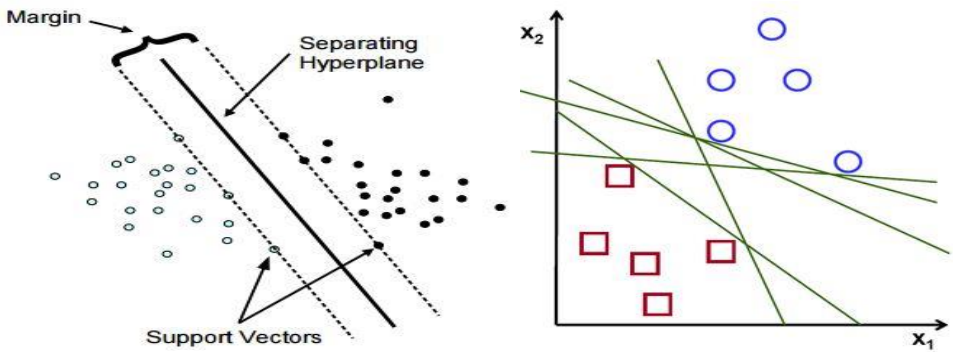


Figure 8: Hyperplane (decision plane that splits and classifies data), Support vectors (the nearest data points to the hyperplane), Margin (the distance between the hyperplane & the closest data point from either set)

The SVM algorithm is very effective in solving different problems. It has several advantages such as:

- The mathematical foundations on which it is based are very solid.
- The nonlinear classification problem is brought back to the linear one by projecting it into a new large space.
- SVM is more efficient in large spaces.
- The kernel trick is the real strength of SVM. With proper kernel function, we can solve any complex problem.
- It adapts relatively well to large data.

Disadvantages:

- SVM does not work very well, when the set of target classes overlap.
- In cases where the number of features for each data point exceeds the number of training data samples, the SVM will underperform.
- Choosing the best kernel function is not easy.
- Long training time for large data sets.

The **Mahalanobis distance** (MD) is a technique based on the distance between two points in multivariate space. In a regular Euclidean space, (Annex II) variables (e.g. x, y, z) are represented by axes drawn at right angles to each other; The distance between any two points can be measured with a ruler. For uncorrelated variables, the Euclidean distance equals the MD. However, if two or more variables are correlated, the axes are no longer at right angles, and the measurements become impossible with a ruler. The MD method takes into account the

statistical criteria (Figure 9), and tends to classify signatures with relatively large values in the covariance matrix. It's a slow technique in term of computing.

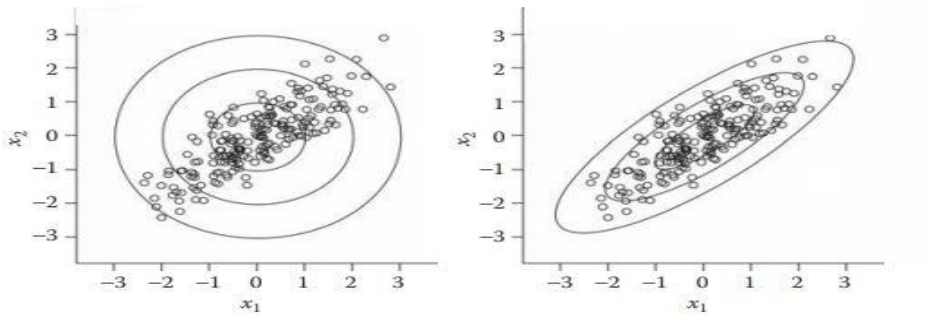


Figure 9: The difference between Euclidean distance (left) and Mahalanobis distance (right) is shown.

The three lines (circles and ellipses) (Figure 9, left) correspond to distances of 1, 2, and 3, from the origin, respectively. The Mahalanobis distance (on the right) also accounts for the covariance structure (correlation of the variables) of the data (Varmuza & Filzmoser, 2016).

The Mahalanobis distance classification has a number of advantages, including:

- takes into account the correlations of the different variables of the classes
- useful in cases where statistical criteria must be taken into account.
- treat class groups that have a non-hyperspherical shape

The Mahalanobis distance classification is not totally perfect, it has a number of drawbacks, namely:

- Tends to outclass signatures with relatively large values. Because when there is a large dispersion of pixels in a sample, the covariance matrix of that signature contains large values.
- takes tame to calculate.

The **Random Forest** (RF) classifier is an ensemble learning method used for classification and regression. Developed by (Breiman, 2001), the method construct a collection of decision trees (Figure 10) with controlled variation. Each decision tree in the ensemble is constructed using a sample with replacement from the training data.

Firstly, each tree is constructed using a random bootstrapped sample of the training data. Secondly, rather than testing all features for the best split, a random subset of variables is tested at each split in each tree. To have a less subjective outcome to any random fluctuation in the training data, the randomness in the construction of the trees should be introduced (Diesing & Stephens, 2015). The prediction is made for unobserved data by taking a majority vote of the individual trees. The samples not part of the bootstrapped sample for each tree, referred to as 'out-of-bag' (OOB) samples, are used to create a cross-validated prediction error for the forest. The random Forest package in R was used for the implementation (Liaw & Wiener, 2002).

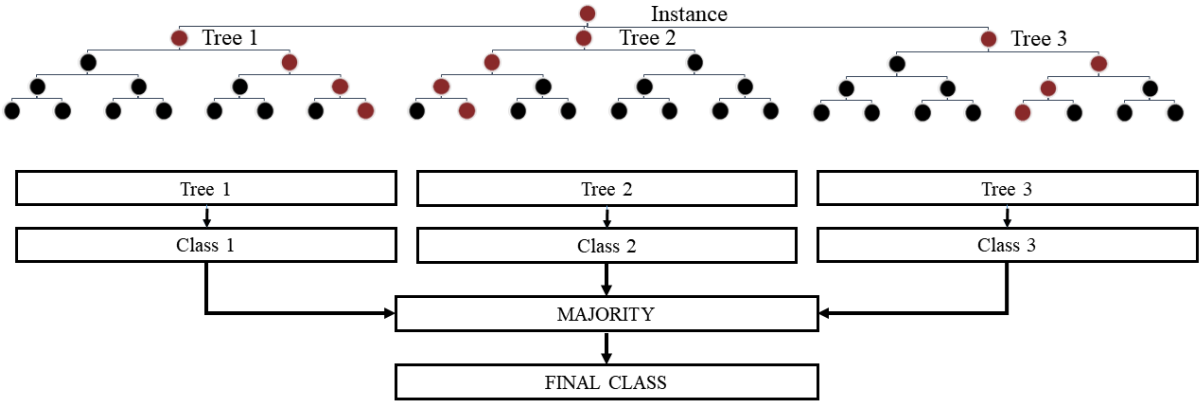


Figure 10: Random Forest decision tree.

Advantages:

- Random forest can solve both type of problems that is classification and regression and does a decent estimation at both fronts.
- One of benefits of Random Forest which exists me most is, the power of handle large data sets with higher dimensionality.
- It can automatically balance data sets when a class is more infrequent than other classes in the data.
- It has an effective method for estimating missing data and maintains accuracy when large proportion of the data are missing.
- It has methods for balancing errors in data sets where classes are imbalanced.
- The capability of the above can be extended to unlabelled data, leading to unsupervised clustering, data views and outlier detection.

The main limitations of random forest are:

- A large number of trees can make the algorithm too slow and ineffective for real-time predictions.
- Random forest is a predictive modelling tool and not a descriptive tool.
- Need to choose the number of trees

2.1.8 Discussion of the algorithm choice

In the following study we will use Random Forest Classifier (RFC) as a classification algorithm, which is commonly used because of its effectiveness. The choice is made according to several previous studies, which show the efficiency of this algorithm in terms of precision. For instance, (Sabat-Tomala et al., 2020) achieved better overall accuracy (OA) for the RFC (67.7%) than for the nearest neighbour method (62.3%). According to (Hastie et al., 2009), the chosen classifier achieves only 4.88% misclassification error. We also found that some articles admit that the RFC is a faster image classifier compared to other algorithms used in the same data set (Sabat-Tomala et al., 2020).

Random forest obtains a class prediction from each tree, and then ranks all predictions using the majority vote. The accuracy of the classification depends on the number of trees, with (Hastie et al., 2009) stating that the number 250 is the value from which the classification stabilises. However, it is suggested to verify the effect of the variation in the number of trees on the precision of the classification and, at the same time, the processing time.

2.1.9 Comparison of Random Forest performance

The comparison criteria are those that allow an evaluation of the performance of this machine learning algorithm from the results obtained. They allow to study the behaviour of RF according to number of trees, utilization of indices, processing a generalization and according the thematic classes of the image.

The quality of the results provided at each iteration is evaluated by estimating the accuracy of the resulting image. Indeed, the precision of the classifications results is the main criterion of comparison, because the essential objective is to know the parameters to be chosen beforehand using RF classifier for the improvement of the final results precision. It is calculated with respect to ground truth data.

The result of the comparison between the classification made by RF and ground truth data is generally presented in the form of a table called the confusion matrix.

There are several types of accuracy in terms of classification results evaluation depending on the interpretation made of the calculated confusion matrix

- Global indices:
  - **Kappa coefficient** (Equation 1): is used to judge the quality of the multi-class classification
  - **Overall Accuracy (OA)** (Equation 2): is the average of the percentages of good classification
- Marginal indices:
  - **Producer Accuracy (PA)** (Equation 3): is the proportion in percentage of the pixels of a class effectively classified related to classes made by true data.
  - **User Accuracy (UA)** (Equation 4): is the percentage proportion of the pixels effectively well classified related to classes made by classifier.
  - **Commission errors** (Equation 5): how many test pixels were incorrectly classified as a class.
  - **Omission errors** (Equation 6): incorrect classified pixels.

here is the calculation equation of each precision measures:

Equation 1: KAPPA Coefficient

$$\text{KAPPA Coefficient} = \frac{\text{Overall Accuracy} - \text{Estimated Chance Agreement}}{1 - \text{Estimated Chance Agreement}}$$

Equation 2: Overall Accuracy

$$\text{Overall Accuracy} = \frac{\text{Number of Correctly Classified Samples}}{\text{Number of Total Samples}}$$

Equation 3: Producer Accuracy

$$\text{Producer Accuracy} = \frac{\text{Number of Correctly Classified Samples in Class}}{\text{Number of Samples from Reference Data in Class}}$$

Equation 4: User Accuracy

$$(4) = \frac{\text{Number of Correctly Classified Samples in Class}}{\text{Number of Samples Classified to that Class}}$$

Equation 5: Commission error

$$(5) = 1 - (4)$$

Equation 6: Omission error

$$(6) = 1 - (3)$$

The Kappa coefficient (K) developed by (Cohen, 1960) is a measure of concordance (indicated by the diagonal elements of the confusion matrix) minus the possibility of concordance (indicated by the product of marginal rows and columns); the Kappa coefficient takes into account all the cells of the confusion matrix, not just the diagonals of the matrix. (Landis & Koch, 1977) explain the strength of agreement according to the values of kappa obtained (Table 5).

Table 5: Rating criteria of Kappa statistics.(Landis & Koch, 1977)

Strength of agreement	**Kappa statistics**
Poor	<0.00
Slight	0.00 - 0.20
Fair	0,80 0.21 - 0.40
Moderate	0.41 - 0.60
Substantial	0.61 - 0.80
Almost perfect	0.81 - 1.00

A comparative study was carried out by (Fung & LeDrew, 1988) on a set of methods for evaluating classifications accuracy. Several authors concluded that the Kappa coefficient was the most appropriate for a correct description of the accuracy of the classifications. (Rosenfield & Fitzpatrick-Lins, 1986) also recommend the Kappa coefficient (Equation 1) as a measure of accuracy for the precision of classification. The Kappa coefficient will therefore be used in this study to calculate the precision of our classifications.

2.1.10 Geographic information systems GIS

Geographic Information Systems (GIS) is a multidisciplinary and spatial analysis environment. They offer tools for entering and managing information in digital form, tools for analysis, modelling and cartographic representation.

Thematic mapping is one of the scientific fields at the origin of GIS, which has always favoured the description and analysis of the spatial dimension. The operation of a GIS is based on the storage of information in the form of thematic layers that can be linked to each other by their geographical location. It has several fields of application, in particular environment, demography, public health, as well as the study of environmental impacts, etc.

2.2 Role of Remote sensing in supporting resource management

2.2.1 Importance of remote sensing

Remote sensing is a modern monitoring tool, capable of providing managers with reliable data on even the most remote and inaccessible regions, with great flexibility at the desirable date and time. Given the facilities it offers managers, remote sensing is increasingly replacing traditional management methods, which are time and resource demanding, such as a satellite image, which can replace a hundred of aerial photographs. Remote sensing finds its applications in various fields: geology, hydrology, oceanography, forestry, etc.

2.2.2 Land use/ Land cover (LULC) mapping

The mapping of LULC serves as a basic inventory of soil resources. Remote sensing offers an efficient means of acquiring and presenting LULC data both regionally and locally. The land use /land cover mapping serves to obtain an inventory of soil resources for monitoring and management. Remote sensing (RS) is a practical and economical tool to obtain a perspective of LULC. RS data permits to observe changes in the phenology (growth) of plants during the growing seasons. The multisource image which allows the integration of two data sources has the advantage of increasing the information content (J. Zhang, 2010).

2.2.3 Forest

In forestry, remote sensing plays a very important role in monitoring of forest health disturbances. Indeed, using sensors optics capable of measuring not only the amount of radiation reflected in the visible but also in the near infrared which is very sensitive to plant health, we can easily detect stress in forests. It also offers managers the possibility of identifying and / or delimiting the various types of stands, an operation that would be painful, long and expensive with the classical methods of land surveys. Depending on the detail sought, we set the resolution and the appropriate spectral characteristics. Alongside these applications, remote sensing plays a key role in monitoring changes in wooded areas and illegal logging, especially in the most remote and inaccessible areas and far from forester surveillance. It would be enough

to acquire previous and recent images of the region in question, process them and analyse them in order to compare and deduce the differences in the extent of logging and forest losses.

2.2.4 Agriculture

Economic profitability with low ecological impacts is the goal of all agricultural producers, large scale managers and regional agricultural organisations. Identifying and mapping crops is important because the food production is fundamental. These maps are used to forecast grain supplies (production prediction), collect production statistics, maintain crop rotation records, map soil production, identify factors that influence crop health, assess storm & drought damage and monitor agricultural activities.

The spectral reflectance of a field varies according to the phenological (growth) stage, the type of plants and their state of health. Microwaves are sensitive to the alignment, structure and amount of water present in plants and soil. Thus, achieving a more accurate classification.

Results of the interpretation of remote sensing data can be integrated into a GIS and crop management system, and can also be combined with ancillary data to provide information on management practices.

3 Methodology and Materials

3.1 Study Area Presentation

3.1.1 Geography

The District of Bragança is an administrative division of Portugal, located in the northeast corner of the country (Figure 11), covering 6,600 km^2 which represent 7.4% of the national continental land mass.

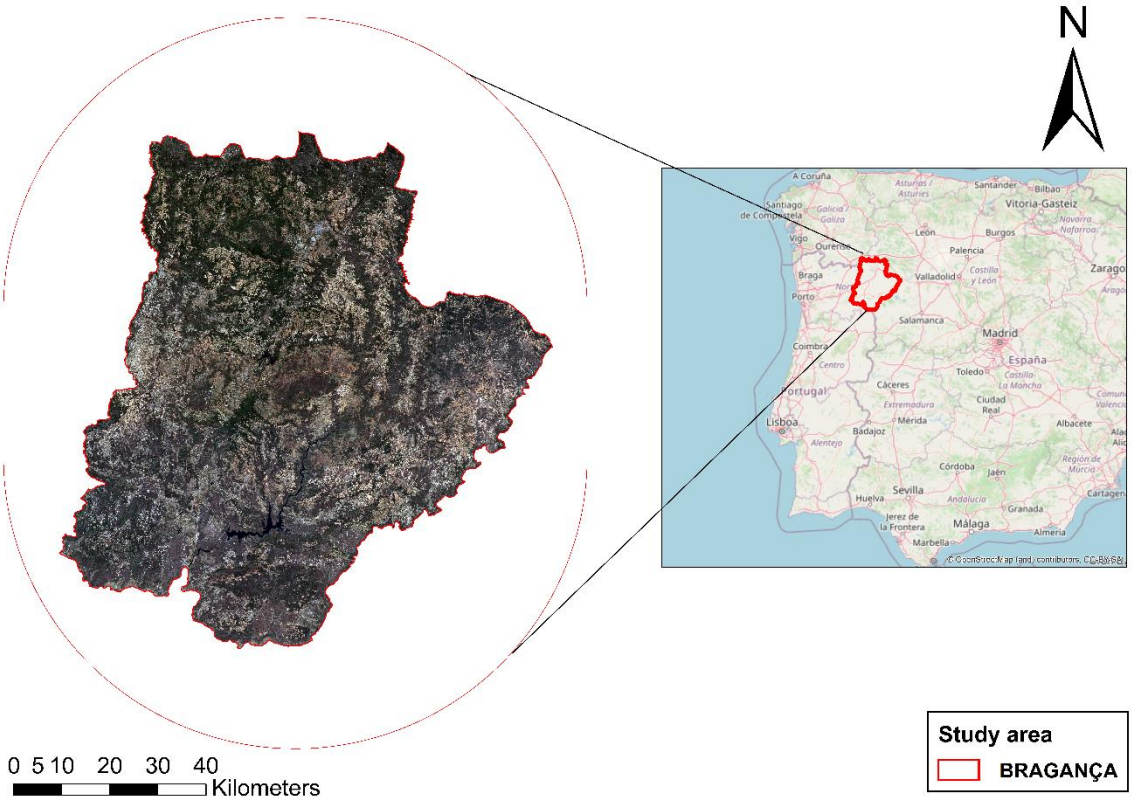


Figure 11: Location of study area "Bragança"

3.1.2 Climate characteristics

The climate of Bragança is classified as Csb according to the Köppen classification, temperate with warm dry summer. In the period 1981-2010, the average maximum temperature in January for Bragança was around 8.5 °C while in July it was around 28.5 °C (*IPMA*, 2020).

The January minimum temperature hovers around the freezing point (0.3 °C). And winter temperatures can fall to as low as -11.6 °C. The annual mean temperature is around 13 °C. The average of total rainfall was around 800 mm per year in the period 1981-2010 (*IPMA*, 2020). The year of 2005 was particularly dry in Portugal and Bragança suffered from water shortages and devastating forest fires in the rural areas (*International Forest Fire News (IFFN)*, 2006).

We present here the ombrothermic diagram of Bragança from the period 1981-2010 (Figure 12) using temperature values from the meteorological station of Bragança (IPMA, 2020) and precipitation values from meteorological station of Vinhais in Bragança (*Portuguese Environment Agency*, n.d.).

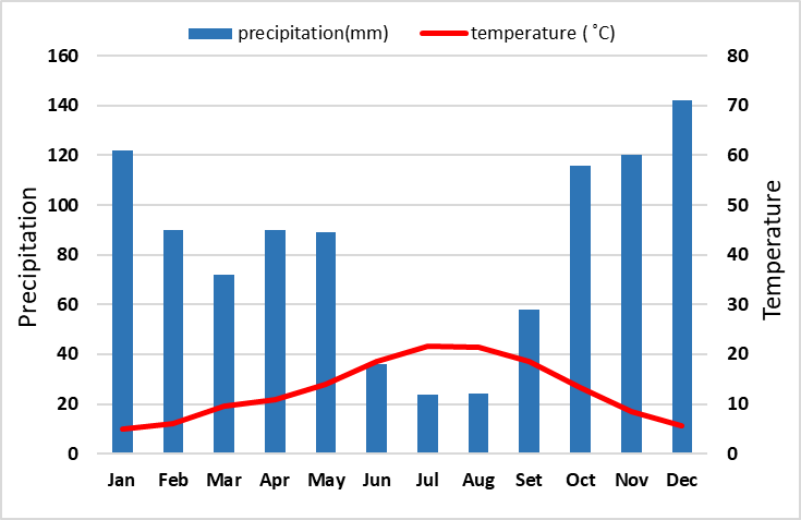


Figure 12: Ombrothermic diagram of Bragança of the period 1981-2010

To conclude, the inland north-eastern Portugal has a hot summers and cold long winters, hence continental features, and the diurnal temperature variation can reach 20 °C (according to the Figure 12 above).

3.1.3 Landscape characteristics

The study area of the district of Bragança is characterized by a high climatic and landscape diversity (Castro & Fernandez-Nuñez, 2014). The region is consisted of a heterogeneous patchwork of land covers and land uses that are juxtaposed. We found: **Shrublands**, **Agriculture** (annual crops; permanent crops), **Woodland** (broadleaves; coniferous), **Grasslands**, **Urban area** and **Water bodies**. It should be noted that from 1958 to 2005 the study landscape went through relevant modifications increasing the patch size for the most combustible land classes (shrublands and forests) (Azevedo, César, Castro 2011),

Shrublands: Shrubs is one of the categories existing in terrestrial land of study area even it is the most dominant type. These are plants not able to reach more than 5m height. The dominance of shrubs in this area is due in first place to the management plan of lands, also due to the mountainous relief of the region, and even with a slight contribution we can add agriculture abandonment effect (Azevedo et al., 2011).

Agriculture: according to Azevedo et al (2011), this cover class is considered in the second rank in terms of dominance. This is areas dedicated to agriculture, of which we particularly specify two main types of soil exploration in the purpose of plant production, the annual crop (e.g. cereal, horticulture), and the permanent crops (e.g. vineyards, olive groves, almond groves, orchards).

Woodland: the forest is an important class in the region of Bragança the reasons for the dominance of this class are the same as shrublands class. This class is defined as "Areas covered by trees with a canopy of at least 10%" (*Land Cover and Land Use, Landscape (LUCAS) (Lan)*).

A forest category is defined by its composition (i.e. type of trees that forms this forest)

Broadleaved Woodland: these are areas with broadleaf species, mainly oak trees, potentially grazed by small ruminants, so this class is called grazing forest.

Coniferous woodland: this class refers to areas with conifers (mainly pine). It is an un-grazed land, which means that ruminant animals show no preference for this land cover category.

Grasslands: this class is a terrestrial ecosystem predominantly covered by communities of grassland, grass-like plants and forbs. Grassland zone is deemed as natural permanent pasture for cattle, or origin of cutting hay or silage.

Artificial Land (urban): areas characterized by an artificial cover which comprises the entire construction made by anthropological actions. The residences, sport, leisure facilities and industrial or commercial sites are the main manifestations of these surfaces.

Water areas: patches that constitute inland areas with water accumulation.

3.2 Digital processing methodology

3.2.1 Digital processing flowchart

This section presents the methodology adopted for digital processing. We summarise in the following flowchart the different steps to be taken to meet the objectives set.

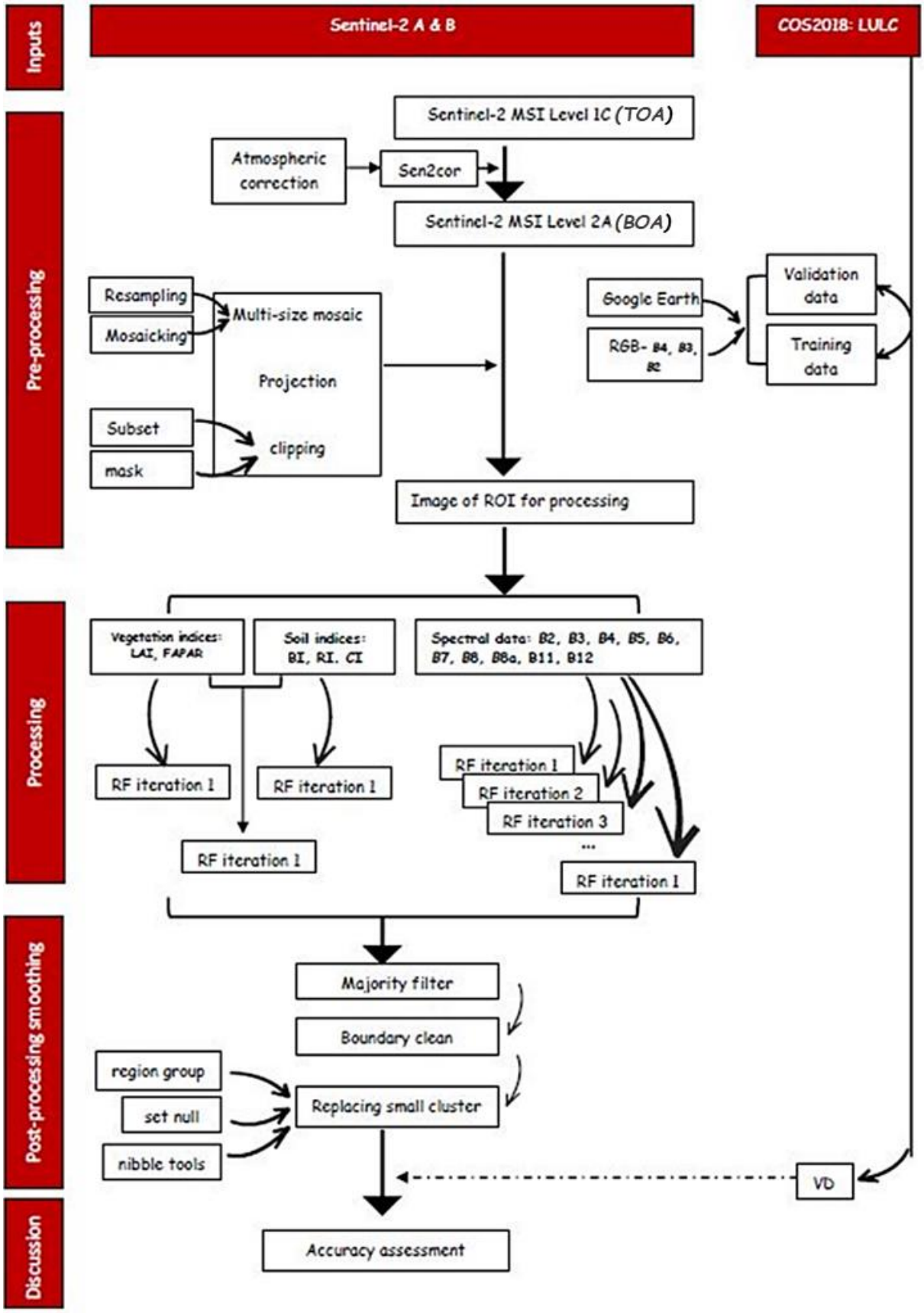


Figure 13: Processing flowchart

3.2.2 Material and data

This chapter is devoted to the presentation of the data used, to the application of the RF algorithm with several value of tree numbers and the integration of indices, to data obtained and the criteria used in the comparison of iterations. The satellite images used are those of the European Space Agency (Sentinel-2) Level-C (Table 6). The products consist of image data, delivered as ZIP-compressed, and the associated metadata, all capsuled within a "SAFE" file container. The multispectral Sentinel-2 Level-C bands used were those listed in the Table 7.

Table 6:Satellite data with the date of acquisition and cloud cover percentage

Satellite	Zone path	Acquisition date	Cloud cover (%)
Sentinel-2B	30TTL	20190805	0.0
Sentinel-2A	29TPG	20190803	0.0583
Sentinel-2B	29TQG	20190719	0.0016
Sentinel-2B	29TPF	20190719	0.2546

Table 7: Multispectral Sentinel

Multispectral band	Central wavelength	Spatial resolution (m)
B2	Blue, 490 nm	10
B3	Green, 560 nm	10
B4	Red edge, 705 nm	10
B5	Red edge, 749 nm	20
B6	Red edge, 783 nm	20
B7	Near-Infrared, 842 nm	20
B8	Near-Infrared, 842 nm	10
B8b	Near-Infrared, 865 nm	20
B11	Short-wave IR, 1610 nm	20
B12	Short-wave IR, 2190 nm	20

We applied the Random forest algorithm, whose performance varies according to the number of trees used, as well as the incorporation of indices (vegetation, soils). Our comparative study is carried out in a context of type of cover. For **training data** (TD) and **validation data** (VD) we choose 7 clusters of homogeneous pixels based on the Sentinel-2 RGB colour composition, aided by the high-resolution images of Google Earth and the ShapeFile of the official Land Use and Land Cover Map of 2018 (COS2018) provided by Direção-Geral do Território in Portugal (Dgterritorio). The TD has been selected so that it is the basis on which the algorithm classifies the data. Its size is 69044.2 pixels distributed as follows: 40470, 1697, 1693, 705, 386.4, 492.8

and 23600 respectively for water, Annual crops, Broadleaves, Conifers, Grasslands, Shrublands and Permanent crops. The VD was chosen for the entire study area to validate the classification. Its size is 89679 pixels distributed as follows: 60240, 1294, 6628, 926, 407, 19340 and 844 respectively for water, Annual crops, Broadleaves, Conifers, Grasslands, Shrublands and Permanent crops.

The software used was the European Space Agency's (ESA) open-source SNAP and Sen2cor for Correction of Atmospheric Effects, both designed for the exploitation of Earth observation data. ArcGIS software was also used, for processing vector data and geodatabases. In this study, this software is used for the vectorization of the map and for the analysis of raster data, post classification, obtaining confusion matrices, etc. We used Microsoft Office tools for result analysis, data processing, chart and report production and word processing.

3.2.3 The pre-processing methodology

"Sen2cor SNAP" is a Level-2A processor that has as main objective to correct Sentinel-2 Level-1C Top of the Atmosphere (TOA) products in Level-2A Bottom-Of-Atmosphere (BOA) reflectance products, eliminating, or at least reducing, the effects of light dispersion caused by components of the atmosphere (water vapour, aerosols, among others). This process of **atmospheric correction** is part of the **pre-processing**, which precedes the classification of images.

Image **resampling** is necessary to equalise the spatial resolution of the 13 bands (see Table 7). The multi-size mosaic tool is applied, resulting in a series of 10x10 pixel size images.

Mosaicking allows the integration of several images of the same scene, which partially overlap, to obtain a full representation of the study area; in our case 4 images are brought together to cover the entire study area. The mosaic process also allows the transformation of the original WGS 84 geographical coordinates into projected coordinates, in this case for the UTM Zone 29 WGS84 coordinate system.

A subset of the image is created based on the **Region of Interest** (often abbreviated ROI), which corresponds to the official administrative boundaries of the Bragança district (ShapeFile polygon) provided by government authorities (Direção-Geral do Território).

3.2.4 The classification methodology

The classification of satellite images is a complex process which can be affected by several factors, in our case we adopt the random forest algorithm approach, so the parameters that can be modified are RF decision tree and the insertion vegetation & soil indices. There are three steps to the classification process:

- Establishment of training data.
- The classification of each of the pixels according to the TD.
- Verification of results with validation data.

Class determination

Visually analysing the satellite image and with the help of COS2018 ShapeFile, we identified 7 thematic classes with interest, according to the objectives described above (Table 8).

Table 8: LULC classes (Dgterritorio)

classes		**constituents**
Shrublands		Matos
Agriculture	AC	Annual crops (cereals, horticulture, …)
	PC	Permanent crops (olive groves, orchards, vineyards, …)
Woodland:	Broadleaves	cork oak, holm oak, chestnut, eucalyptus, …
	Coniferous	maritime pine, stone pine, …
Grasslands		Spontaneous pastures, improved pastures
Artificial Land		Industry, Trade, Infrastructure, Transport, …
Water Areas		water courses, lakes and reservoirs, …

Choice of training data

The careful choice of training data is very important in order to achieve a successful classification. A sufficient number of training samples and their representativeness are essential criteria for the image classification, for this purpose we have chosen samples throughout our study area to minimise the slopes effect on the classification accuracy, since the area is mountainous. The training areas will be used to define the signatures of each of the thematic classes.

Application of the Random Forest algorithm

The classification has the following requirements:

- The same thematic classes are adopted in the different iterations.

- Similar training data are maintained for all iterations.
- The same validation data is used for the different iteration. We specify here that this validation data is independent of the training one.
- The colours assigned to each of the thematic classes are the same for the different classified images, to ensure reliable visual analysis**.**

Steps of application: In the main SNAP toolbox, we choose Random Forest Classifier, by clicking on RASTER then Classification then supervised classification:

- ProductSet- Reader operator

At the ProductSet- Reader operator level, we add the satellite image:

- Random Forest Classifier operator

In the random forest classifier operator section, we select the training vectors, and the bands to be used (Table 7), according to the purpose of the classification.

At this section we also decide the number of trees on which we base the classification:

- write operator

In write operator we define the name of the output and its directory.

Application of RF algorithm for number of decision trees purpose:

Considering the 7 classes of training data already selected, the algorithm separates the data into 7 proposed classes. Throughout this part of work, we did 15 iterations. The numbers used in each iteration are: 10, 50, 100, 150, 200, 250, 300, 350, 400, 500, 600, 700, 800, 900 and 1000, it should be noted that the maximum nb-trees we have been able to perform is 1000, due to time constraint, so we can admit the following assumption "with nb-trees not used in our study (>1000), the results can be superior to those obtained". The decision on the optimal number of trees to be included in the classification was based on the assessment of the quality of interpretation through the analysis of confusion matrices.

Application of RF algorithm for Indices purpose:

We make 3 iterations: 1st iteration: with Soil Indices, 2nd Iteration: Vegetation indices and 3rd iteration: with Soil & Vegetation indices

1st iteration (with Soil Indices): on the menu bar we click on Raster then band math, filling in according to the corresponding equations for the 3 soil indices desired to be integrated into the

classification, the Brightness Index (BI) algorithm, representing the average of the brightness of a satellite image (Equation 7), the Redness Index (RI) algorithm, developed to identify soil colour variations (Equation 8) - Pouget et al.(1990) and the Colour Index (CI) algorithm, developed to differentiate soils in the field (Equation 9) - Pouget et al.(1990).

Equation 7: Brightness index

$$BI = \sqrt{\frac{B_4 \cdot B_4 + B_3 \cdot B_3}{2}}$$

Equation 8: Redness index

$$RI = \frac{B_4 \cdot B_4}{B_3 \cdot B_3 \cdot B_3}$$

Equation 9: Colouration index

$$CI = \frac{B_4 - B_3}{B_4 + B_3}$$

2nd Iteration: Vegetation indices: For this iteration we go to Optical section on the main menu, then we choose Thematic land processing, after we click on biophysical processor, there the calculation is carried out automatically of the 4 selected vegetation indices, automatically calculated, the Fraction of Absorbed Photosynthetically Active Radiation (FAPAR) (Equation 10), the Leaf Area Index (LAI) (Equation 11), the Fraction of vegetation cover (FCOVER), and the Chlorophyll content in the leaf (Cab).

Equation 10: Fraction of Absorbed Photosynthetically Active Radiation (Begue et al., 1991; Fensholt et al., 2004)

$$FAPAR = \frac{((\mathrm{PAR_i} - \mathrm{PAR_{cr}}) - \left((\mathrm{PAR_{tr}}(1 - \alpha_s))\mathrm{PAR_i}\right)}{\mathrm{PAR_i}}$$

where PARi is incoming Photosynthetically Active Radiation (PAR), PARcr is reflected PAR from the canopy and αs is the soil albedo (derived over bare soil). PARtr is the transmittance of PAR through the canopy

Equation 11: Leaf Area Index (Saito et al., 2001)

$$LAI = 0.57e^{2.33\frac{B_8 - B_4}{B_8 + B_4}}$$

Equation 12: Fraction of vegetation cover

$$FCOVER = \frac{\text{NDVI} - \text{NDVIs}}{\text{NDVIv} - \text{NDVIs}}$$

Where NDVIv and NDVIs are difficult to obtain, but the estimation can be summarized as the maximum and minimum NDVI in a study area (Gutman & Ignatov, 1998)

3rd Iteration: integrate vegetation and soil indices into the classification process.

3.3 Post classification:

Post-classification generalization was applied to the resulting image using a **Majority filter** where each pixel is replaced by the most frequent value of the surrounding pixels. Therefore, we used the **Boundary clean** tool to smooth out boundaries between zones without changing the integrity of classes. Finally, we used the **Nibble** tool to replace certain small groups of pixels after applying a mask which assigns each pixel a unique identification.

3.4 Accuracy assessment

By comparing the classification results with the ground truth already set by validation data, we can assess the proportion of pixels that are actually well classified, as well as those that are misclassified. The proportion of well-classified pixels gives an idea of the accuracy of the classification and that of misclassified pixels an idea of the classification error.

The analysis of these Accuracies (Equation 1,Equation 2,Equation 3,Equation 4,Equation 5 and Equation 6) is enough to make a good evaluation of RF Classification results quality. The results exactitude is the fundamental criterion in a study aiming to evaluate and compare several iterations, because the accuracies permit to make:

- Comparison between number of trees and deduce the most appropriate value to use in the classification,
- Assessment of indices effect, either vegetation or soil one, on the classification quality,
- Evaluation of the effect of post-classification generalization on the final classification.

The execution time of the classification, in our case, is only the time to make the classification, without taking into account the preparation time of the entries or the time for processing the results. The assessment of the overall duration of a classification operation can sometimes be subjective depending on the experience of the user.

The following study is done on Intel(R) Core(TM) i7 CPU M 640 @ 2.80GHz 2.8 GHz.

4 Results and discussion

The comparison of the performance of the iterations made by the Random forest classification algorithm requires an evaluation of the behaviour of each of them with respect to the main influencing parameters. Depending on the different applications carried out, we retain the following essential criteria to conduct this practical comparison. It's about:

- Global precision in the discrimination of each type of class;
- Behaviour in relation to the statistical parameters (means and variances) of the classes;
- Calculation time taken by each iteration to classify the image.

The first 15 iterations will focus on the comparison of number of trees and their effect on the classification accuracy. Concerning the 2nd group of iterations (3 iterations) it is about comparing the effects of the soil and vegetation indices on the classification.

4.1 Comparative analysis of classification accuracy for each number of trees iteration

4.1.1 Global Indices:

Nb trees	Kappa
10	0.82
50	0.84
100	0.84
150	0.85
200	0.85
250	0.86
300	0.87
350	0.87
400	0.86
500	0.87
600	0.86
700	0.87
800	0.87
900	0.87
1000	0.88

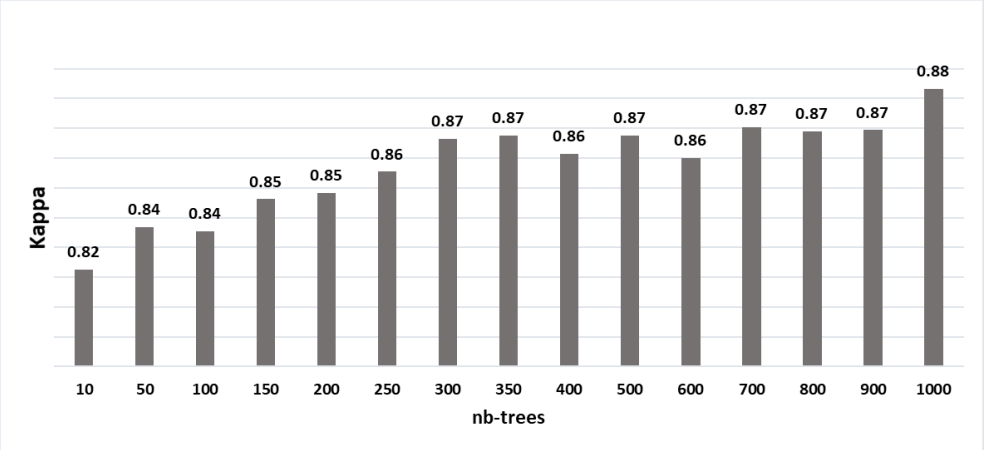


Figure 14: kappa index of agreement (nb-trees analyse)

Nb trees	OA
10	90.91
50	91.73
100	91.63
150	92.28
200	92.33
250	92.70
300	93.32
350	93.32
400	93.03
500	93.34
600	92.89
700	93.45
800	93.40
900	93.44
1000	94.18

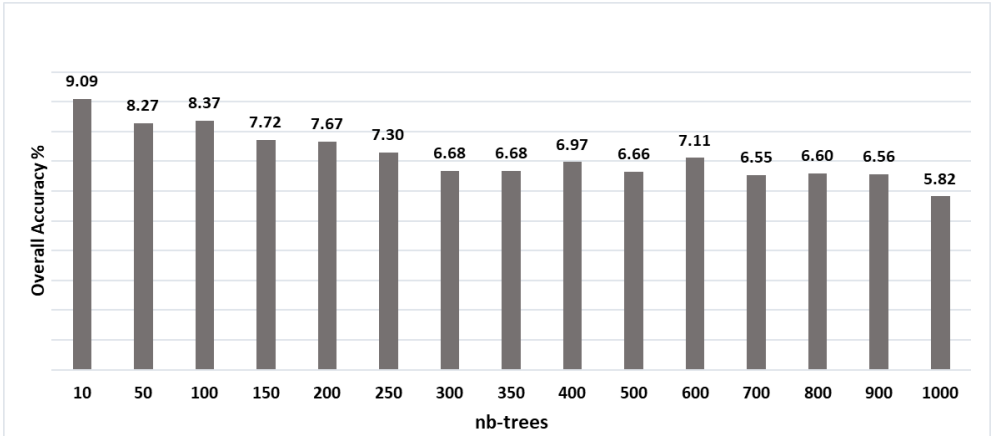


Figure 15: Overall Accuracy (nb-trees analyse)

According to the two graphs (Figure 14 and Figure 15) which represent respectively the kappa index and the overall accuracy of each classification, it can be seen that the iteration with a number of trees of 1000 is considered as the best and most accurate one. It is also clear that precision tends to increase with the increase in the number of decision trees.

4.1.2 Marginal indices

In order to establish a comparative analysis, the confusion matrices are used to record the classification accuracy of each class according to each number of trees used in classification. The Figure 16 represents the preliminary data "precision of the producer" of this analysis.

Nb trees	Producers accuracy
10	90.91
50	91.73
100	91.63
150	92.28
200	92.33
250	92.70
300	93.32
350	93.32
400	93.03
500	93.34
600	92.89
700	93.45
800	93.40
900	93.44
1000	94.18

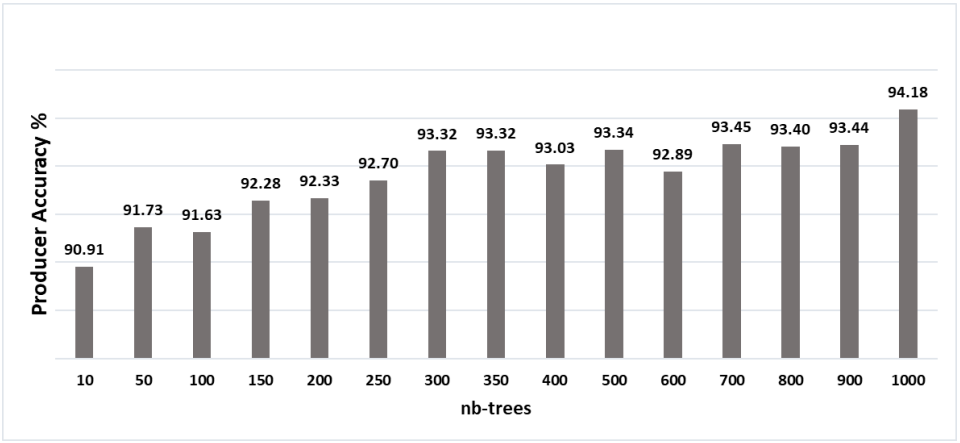


Figure 16: Producer Accuracy (nb-trees analyse)

Nb-trees	O &C errors
10	9.09
50	8.27
100	8.37
150	7.72
200	7.67
250	7.30
300	6.68
350	6.68
400	6.97
500	6.66
600	7.11
700	6.55
800	6.60
900	6.56
1000	5.82

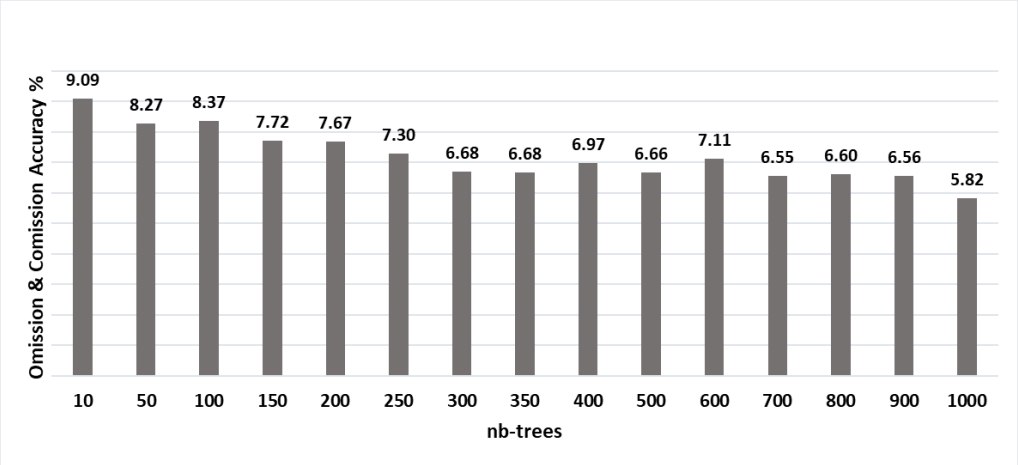


Figure 17: Omission & commission error (nb-trees analyse)

From the graphs of the Figure 16 and Figure 17, we can see that with a number of trees of 1000, the RF Classifier assigns over 94% of the image pixels to their true LULC class.

As shown in the omission & commission index graph, if the number of trees in the analysis increases, the percentage of image pixels assigned to incorrect classes is lower. For a number of trees of 1000, the classification makes the minimum error (less than 6%).

Accuracy According to each Class

Table 9: Producer Accuracy of Random Forest

classes	**10**	**50**	**100**	**150**	**200**	**250**	**300**	**350**	**400**	**500**	**600**	**700**	**800**	**900**	**1000**
Annual Crops	96	97	98	98	98	98	98	99	98	98	99	98	98	98	99
Broadleaves	96	97	97	96	97	97	98	96	97	97	97	96	97	97	98
Conifers	76	76	76	76	76	76	76	78	76	76	76	76	76	76	75
Grassland	88	89	89	89	89	89	90	90	89	88	88	89	89	89	88
Perm. Crops	37	49	46	50	48	47	47	47	47	47	44	47	48	45	48
Shrublands	64	66	66	69	69	71	74	74	73	74	72	75	69	75	78
Water	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100

Figure 18: the dispersion of the classification precision for each class (nb-trees analyse)

The graph in the Figure 18 present the classification accuracy dispersion in terms of number of trees in each ROI image class. We cannot conclude the number that affect the classification the most, but by zooming we can note that the 1000 approach has a tendency to well affect the classification for each class.

Analysing the Table 9, we can notice that we obtain a satisfactory accuracies for some classes before reaching nb-trees=1000 (perm. crops at nb-trees=50, grassland at nb-trees=10), but if we consider the whole column we can state that 1000 has gathered the highest accuracies for all classes.

The Permanent Crops, Shrublands, Conifers and Grassland classes, obtained less agreement for classification. This misclassification is due to incorrect assignment of pixels belonging to one class to another class.

The class with the lowest concordance was the Permanent Crops class (<55%), as a result of the diversity of this class, which includes vineyards, chestnut and almond trees, olive groves, which are very different in the remote sensing images. But in this case, this lack of agreement is due to the soil factor, which is predominant, especially in young plantations of fruit orchards and olive groves, and even in vineyards. In such cases temporary crops can be confused with permanent crops or even with pastures, especially in summer when the soil is also devoid of green vegetation.

We can also relate this confusion to the species themselves. For instance, the *Juniperus oxycedrus*, is is considerably present in our study area (Costa et al., 1993) and this specie is

considered to be both a conifer (H. W. Zhang & Chan, 2015) and a shrubland (*Juniperus Oxycedrus - Trees and Shrubs*)

4.1.3 Descriptive statistics:

For each iteration accuracy, the parameters calculated are presented in (Table 10). By observing the box and whisker figure (Figure 19), we can get an idea of the central tendency of the precision values of each classification iteration, the box length gives an idea on the variability for each number of trees value. We notice that the length of 1000, 900,700, 500 and 350 boxes are small, which explains a small variation of the values, but in term of symmetry we can say that 1000 has a less skewed data. Hence, the latter presents high precision values compared to the other box and whiskers.

Table 10: Descriptive statistics for all the precision observations of each iteration (nb-trees analyse)

nb tree	N	Maximum	Minimum	Mean	standard deviation	Variance
10	7	99.972	37.055	79.591	22.779	518.904
50	7	99.983	49.169	82.073	19.128	365.873
100	7	99.983	48.456	81.718	19.885	395.409
150	7	99.977	43.943	82.654	18.616	346.556
200	7	99.980	46.200	82.333	19.229	369.773
250	7	99.987	49.525	82.638	19.125	365.753
300	7	99.977	46.675	83.079	19.359	374.790
350	7	99.980	46.675	83.391	19.044	362.685
400	7	99.982	47.268	82.815	18.991	360.645
500	7	99.982	45.368	82.878	19.017	361.629
600	7	99.982	47.268	82.284	20.133	405.353
700	7	99.978	46.437	82.984	18.942	358.789
800	7	99.982	47.506	82.333	19.229	369.773
900	7	99.985	46.793	82.698	19.390	375.984
1000	7	99.985	46.556	83.639	18.477	341.391

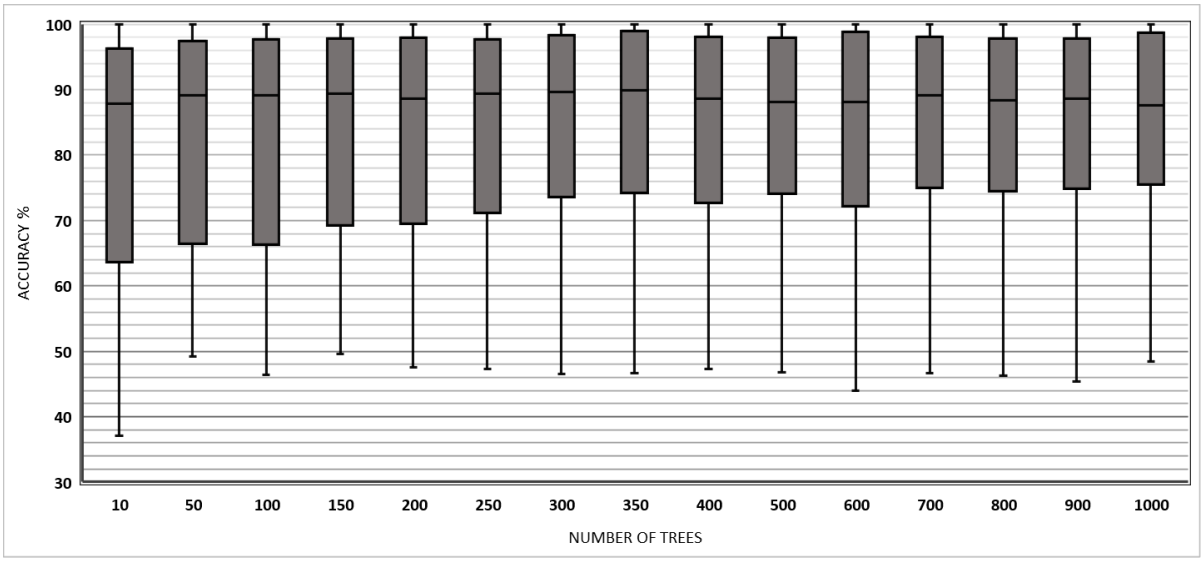


Figure 19: the box and whisker of different observations of producer precision (nb-trees analyse)

4.1.4 Analysis of variance:

The studied variable "Producer Accuracy" is given as a percentage which requires an arcsine transformation (Equation 13) (*Data Analysis in the Geosciences*), then we perform **one-way analysis of variance** (ANOVA) so we get the Table 11

Equation 13: arcsine transformation for one-way analysis of variance

$$ARC\sin(\sqrt{PA})$$

We notice that the value of the significance for the of the LULC classes and the number of trees factors are lower than the 5% threshold, so we reject H0, and we conclude that the difference is significant whether between the LULC classes or between classifications with different number of trees.

Table 11: Significant test of inter & intra factor

Source of Variation	**Sum of Squares**	**df**	**Mean Sum of Squares (MSS)**	**F**	**Significance**
Classes	7.132886484	6	1.188814414	2716.246493	6.14E-94
Nb-trees	0.015444248	14	0.001103161	2.520541439	0.004758171

4.1.5 Execution velocity

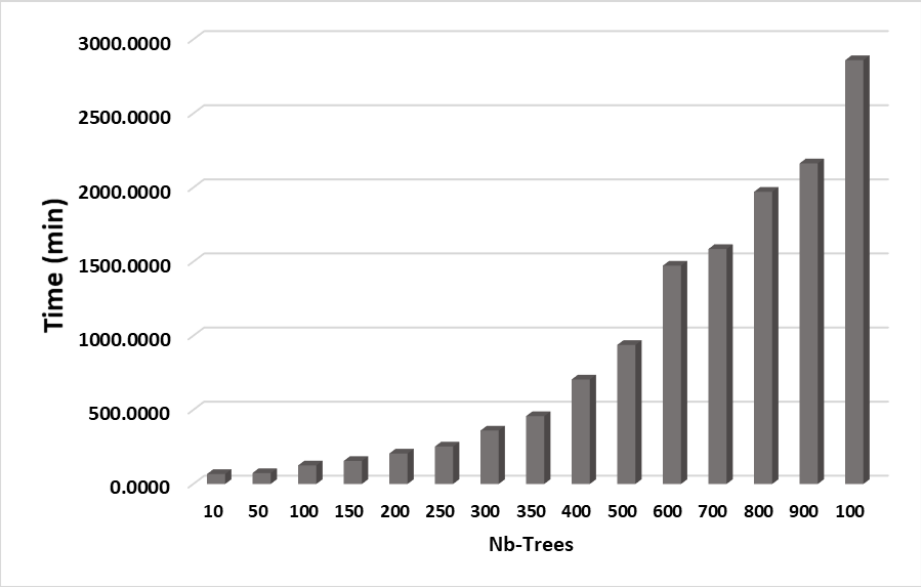


Figure 20: Execution time of different iterations (nb-trees analyse)

The graph of Figure 20 shows that the execution time is related to the number of trees chosen. As the number is large the execution time is longer.

4.1.6 Conclusion

By making an overview of the results obtained from the 15 iteration, the classification with a nb-trees=1000 is considered as the best in terms of precision, either for global Indices (Kappa index, Overall accuracy) or Marginal Indices (Producer accuracy, User Accuracy, Omission & commission errors). We present in Figure 21 the LULC map for the district of Bragança, obtained through the classifier Random forest with nb-trees=1000.

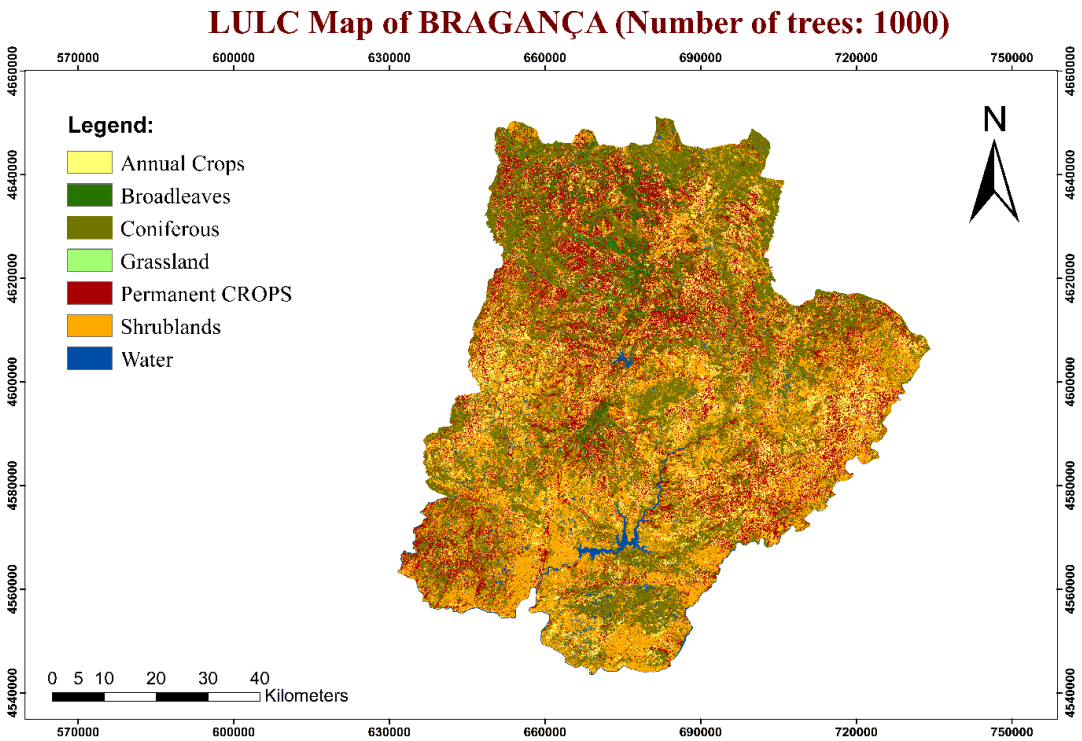


Figure 21: LULC classification of Bragança with nb-trees=1000

4.2 Comparative analysis of classification data for indices effect:

Once we had researched the effect of the number of decision trees on the classification process, we decided to see if we could further improve the classification result by integrating soil and vegetation indices, with constant decision trees (nb-trees=1000)

4.2.1 Global Indices

iteration	Kappa %
without indices	0.883
soil indices	0.908
Vegetation indices	0.885
Both Indices	0.899

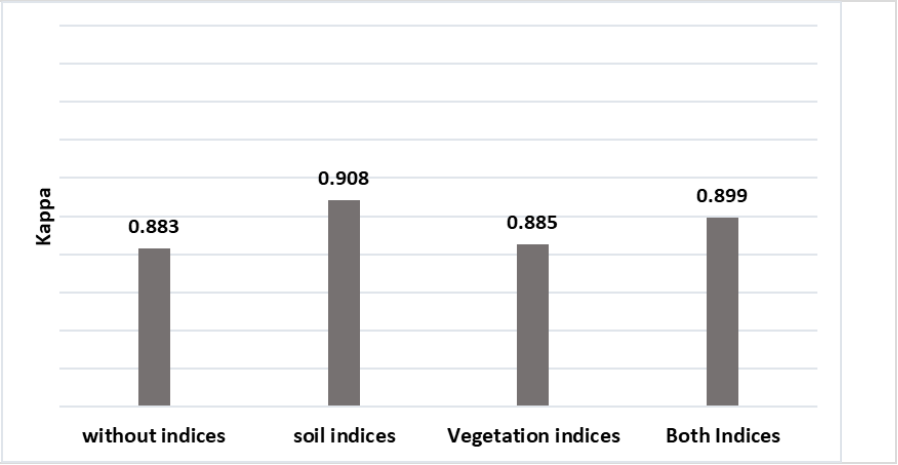


Figure 22: kappa index of agreement (Indices analyse)

iteration	Overall accuracy
without indices	94.178
soil indices	95.421
Vegetation indices	94.397
Both Indices	95.014

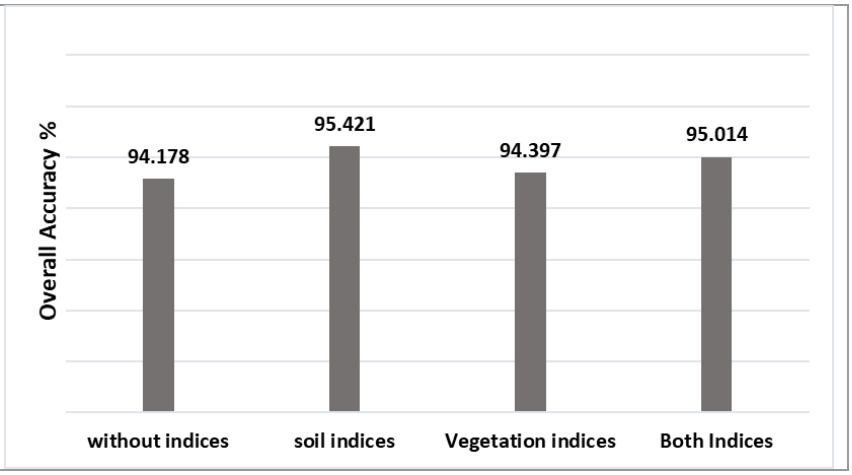


Figure 23: Overall Accuracy (Indices analyse)

According to the Figure 22 which presents the kappa index of each iteration made, we notice that the classification with the soil indices is the most precise (more than 0.9) compared to the other iterations. And also the overall accuracy, represented in the Figure 23, shows that it is the most precise iteration (more than 95%).

The 2 figures above reveal that the use of vegetation and soil indices improves the quality of the classification, however, the soil indices are more efficient.

according to (Macintyre et al., 2020) the classification is related to the season of the satellite images, therefore we can strongly link the clear response of the soil indices to the season, especially in summer the soil is deprived of the green layer which promotes soil indices role.

4.2.2 Marginal indices:

iteration	Producer
without indices	94.178
soil indices	95.421
Vegetation indices	94.397
Both Indices	95.014

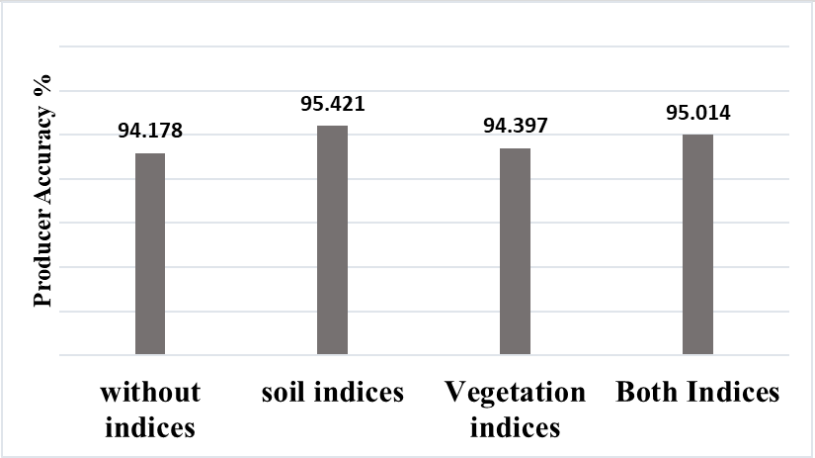


Figure 24: Producer Accuracy (Indices analyse)

iteration	O&C error
without indices	5.822
soil indices	4.579
Vegetation indices	5.603
Both Indices	4.986

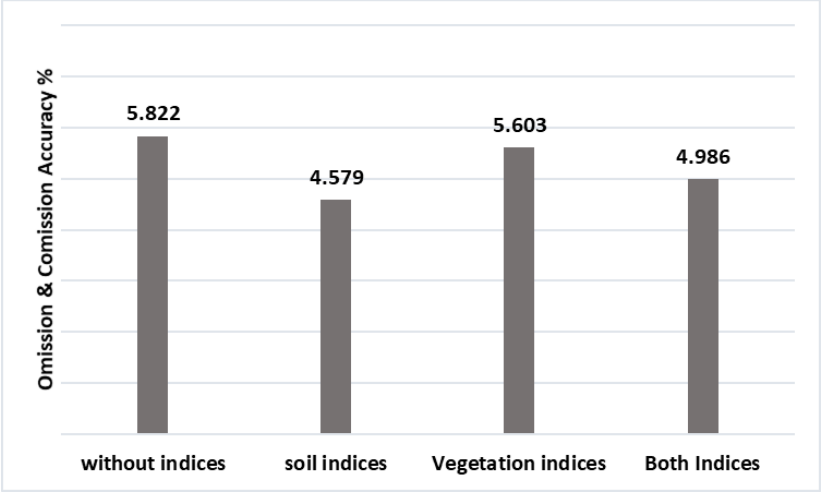


Figure 25: Omission & Commission error (Indices analyse)

From the producer's perspective (Figure 24), it should be noted that with the inclusion of soil indices the RF Classifier allocates more than 95% of the pixels to its correct class. In general, the integration of indices, both vegetation and soil, helps the classifier to assign pixels to its true class.

According to Figure 25, the iteration with soil indices is still the best, because only 4.5% of the pixels that are incorrectly classified.

Accuracy according to each class

Table 12: preliminary data for the classification results (Indices analyse)

Classes	Without Indices	With Soil Indices	With Vegetation Indices	With Both Indices
Annual Crops	99	98	99	99
Broadleaves	98	97	95	95
Conifers	75	75	73	73
Grassland	88	90	80	87
Permanent Crops	48	51	52	49
Shrublands	78	83	79	82
Water	100	100	99	99

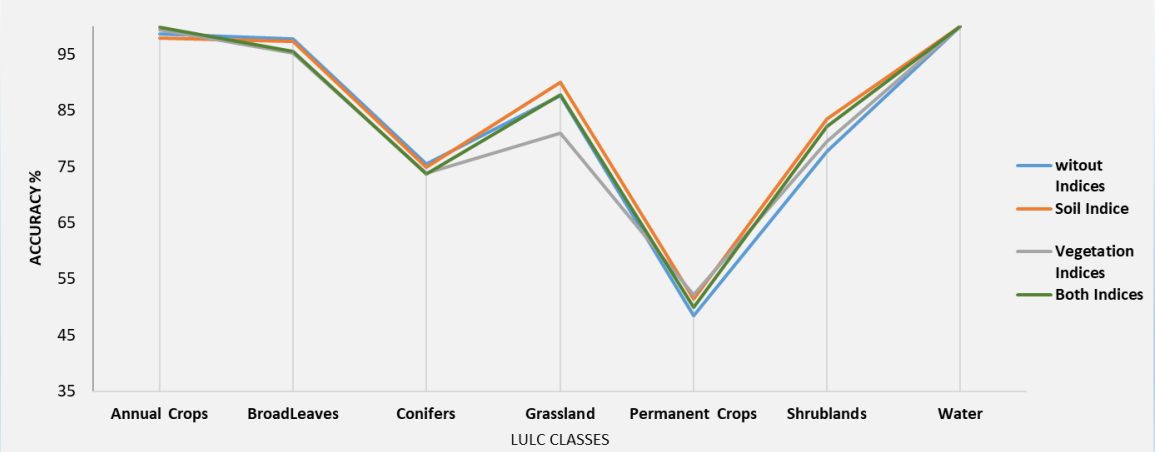


Figure 26: dispersion of accuracy of classification by land use classes (Index inclusion analysis)

As presented (Table 13, Figure 27) the indices have, in one way or another, an effect on the classification. For the permanent crop, the precision is improved with 4% by integrating the vegetation indices, on the other side, the classification of the Grassland and Shrublands classes has become more accurate by using the soil indices with 2% and 3% respectively. On the other hand, there is a deterioration for the Grasslands Accuracy using the vegetation indices, which is supported by (Baret et al., 2006), indicating an under-representation of grass and broadleaf surfaces where LAI / FAPAR measurements have been performed.

4.2.3 Descriptive statistics

Table 13:Descriptive statistics for all the precision observations of each iteration (Indices analyse)

use of indices	N	Maximum	Minimum	Mean	Standard deviation	Variance
Without indices	7	99.9834	48.4561	83.6387	18.4768	341.3908
Soil Indices	7	99.9801	51.4252	84.9860	17.3043	299.4398
Vegetation Indices	7	99.9884	52.1378	82.9727	17.1052	292.5884
Both Indices	7	99.9834	49.8812	84.1424	17.9267	321.3657

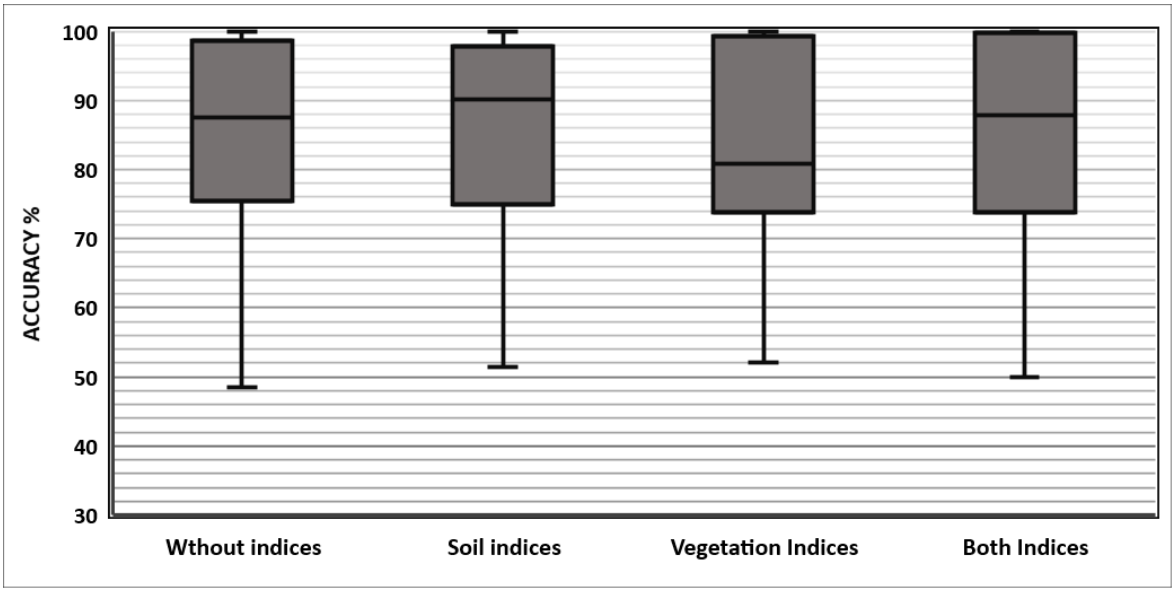


Figure 27: the box and whisker of different observations of producer precision (Indices analyse)

As shown in (Figure 27), the 4 boxes have the same length, which means the resemblance of the 4 iterations in terms of values variation. Also we can see the fact that half of the calculated AP have an accuracy greater than 88% (this is the case of without indices, soil indices and both indices)

4.2.4 Execution velocity

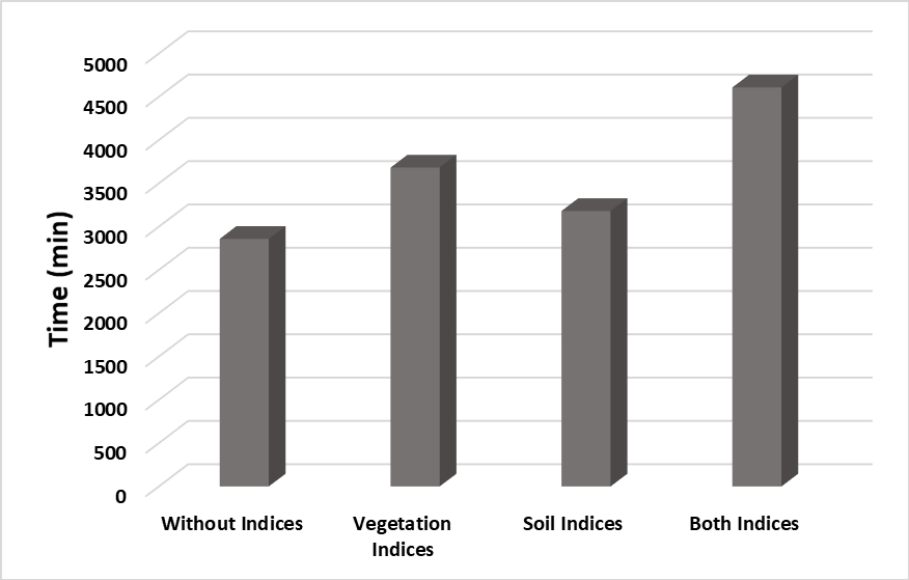


Figure 28: Execution time of different iterations (Indices analyse)

The Figure 28 shows that the time to perform the classification with the indices is longer, especially if we combine the 2 indices.

4.2.5 Conclusion

After having finalized this part of work we can conclude the effect of including all the Indices. It was found that the classification using LAI, FAPAR and FCOVER brought less improvement in terms of accuracy compared to soil indices (BI, CI and RI). This can be explained by the sensitivity of vegetation indices to soils (Viña et al., 2011). According to these authors, vegetation indices can be affected by soil background reflectance, which reduces their contribution to the LULC classification.

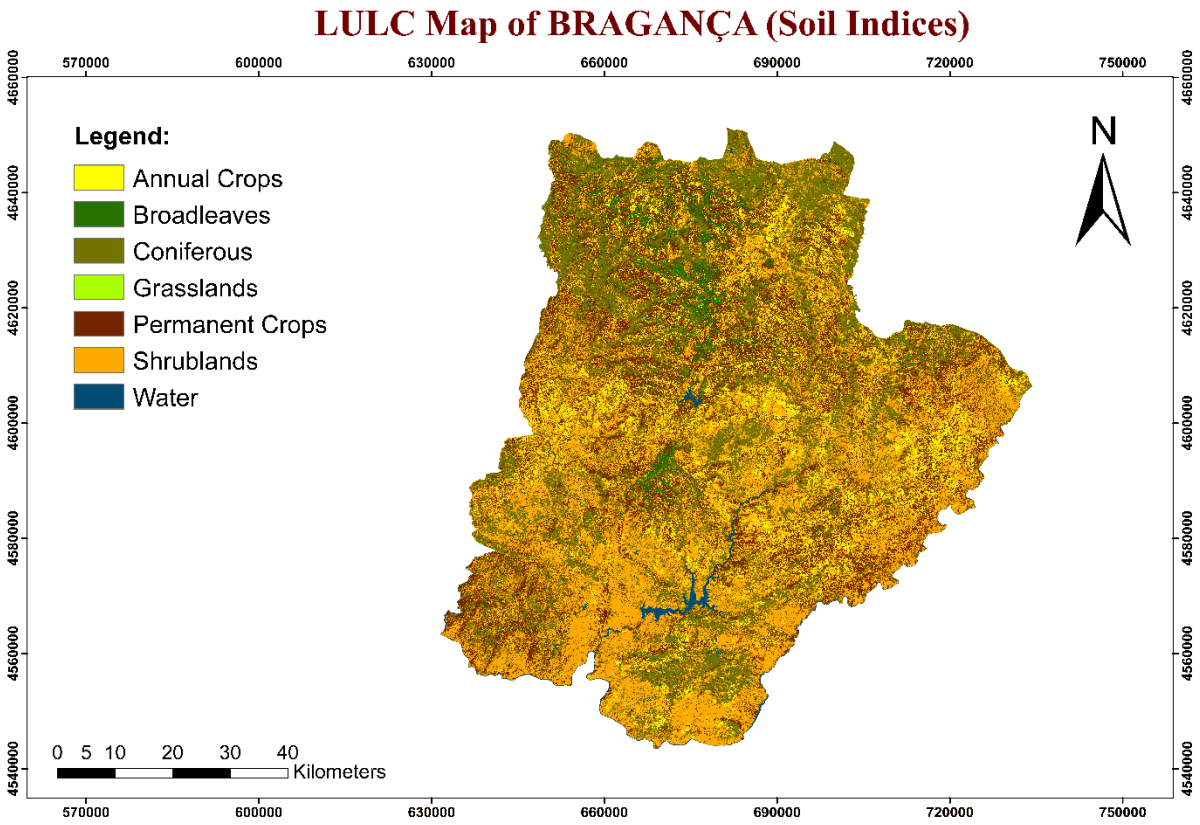


Figure 29: LULC classification of Bragança district with Soil Indices

4.3 Post-Classification

The best classification result obtained was used to evaluate possible losses of precision caused by post-classification tasks.

4.3.1 Global Indices:

iteration	Kappa
before post-classification	0.908
after post-classification	0.927

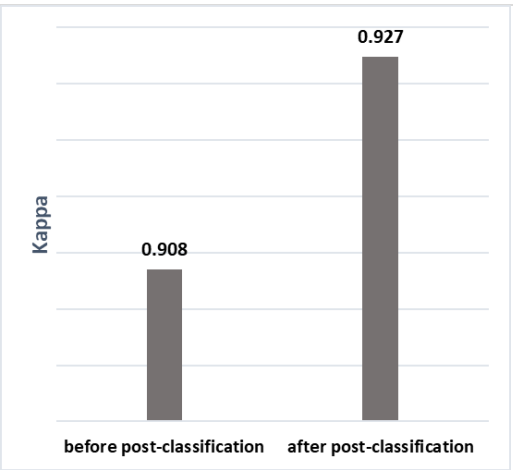


Figure 30: kappa index of agreement (post-classification analyse)

iteration	Overall accuracy
before post-classification	95.421
after post-classification	96.375

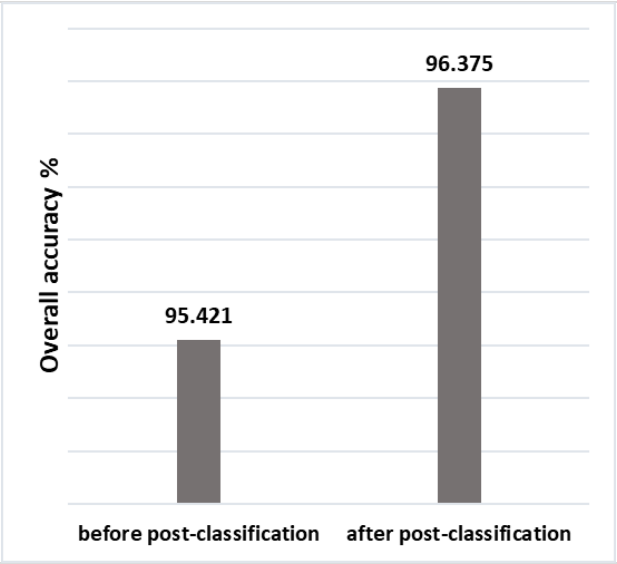


Figure 31: Overall Accuracy (post-classification analyse)

From the figures (Figure 30 and Figure 31) below we observe an improvement in the Kappa Index is from 0.908 to 0.927 and in the overall accuracy, from 95.421 to 96.375. Consequently, in this case, post-classification tasks contributed positively to the accuracy of the classification obtained.

4.3.2 Marginal Indices:

iteration	Producer
before post-classification	95.421
after post-classification	96.375

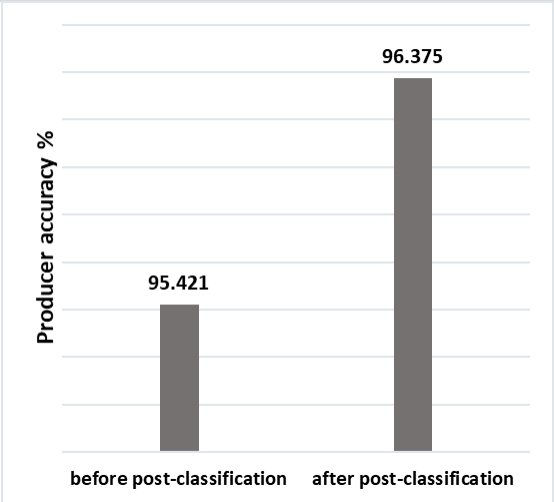


Figure 32: Producer Accuracy (post-classification analyse)

iteration	O&C error
before post-classification	4.579
after post-classification	3.625

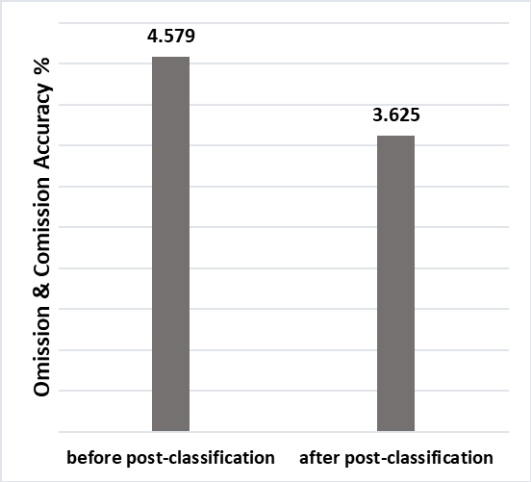


Figure 33: Omission & Commission error (post-classification analyse)

According to Figure 32 and Figure 33, almost 1% of the pixels not well classified before, became well placed in their correct classes by performing the post-classification tasks.

Accuracy according to each class

Table 14: preliminary data for data analysis (post-classification analyse)

classes	Before post-classification	After post-classification
Annual Crops	97.824	99.767
BroadLeaves	97.255	97.617
Conifers	74.893	73.712
Grassland	90.074	89.826
Permanent Crops	51.425	55.344
Shrublands	83.450	87.477
Water	99.980	99.992

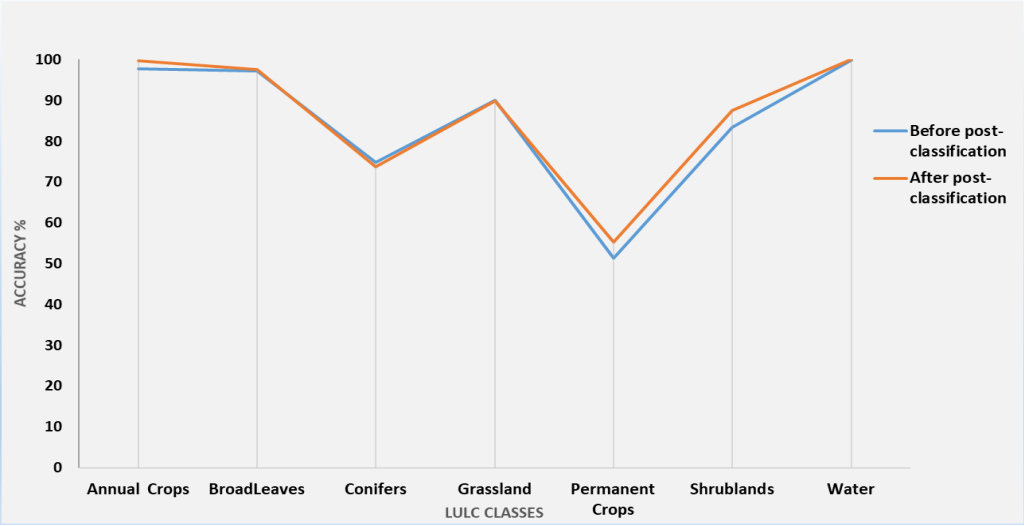


Figure 34: the dispersion of the classification precision for each class (post-classification analyse)

The effect of the post-classification on the classes is mainly existing at the level of the following classes: AC, PC, Shrubs with an improvement of 2%, 4% and 4% respectively. On the other hand, the precision has decreased for Grassland with 0.3%.

This increase or decrease in terms of precision is generally explained by the size of each cluster. for instance, the grassland class is divided into small cluster size which explains its elimination at the time of generalizations, and the same for the class AC, PC and Shrubs. The generalization has eliminated all type of noise in the middle of the clusters dedicated to these 3 classes.

4.3.3 Descriptive statistics:

Table 15: Descriptive statistics for all the precision observations of each iteration (post-classification analyse)

POST-CLASSIFICATION	N	Maximum	Minimum	Mean	Standard deviation	Variance
Before post-classification	7	99.980	51.425	84.986	17.304	299.440
After post-classification	7	99.992	55.344	86.248	16.469	271.233

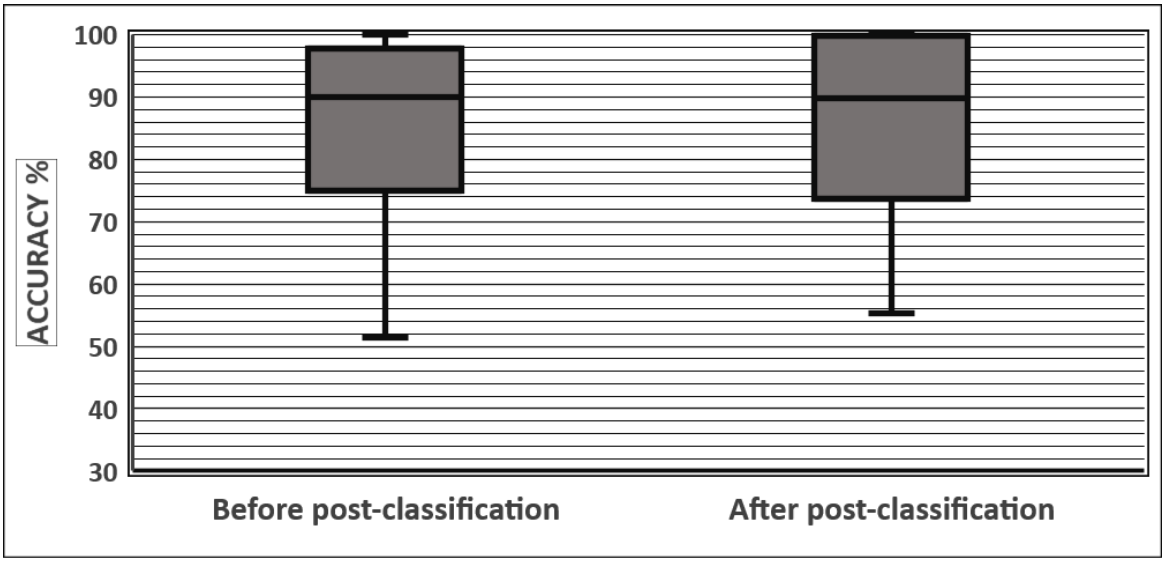


Figure 35: the box and whisker of different observations of producer precision (post-classification analyse)

By observing the box and whisker (Figure 35), we can admit that the post classification has an effect on the classification in terms of precision. For instance, the values are less skewed, and the variance is lower.

4.3.4 Conclusion

After performing a generalization to the classification, we got an improvement in terms of accuracy, the map below shows the classification pf Bragança, with (nb-trees=1000, soil indices and generalization process) (Figure 36).

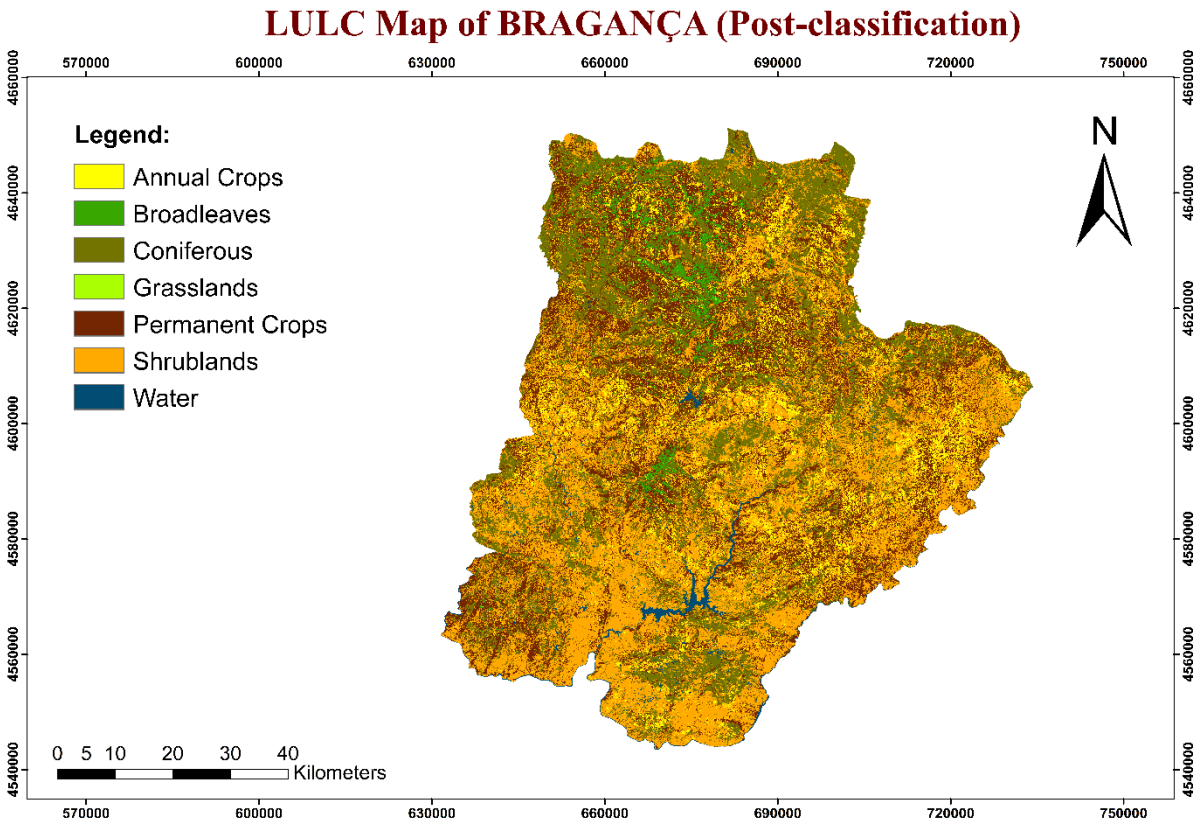


Figure 36: LULC classification of Bragança district with post-classification processing

5 General conclusion:

This project allowed us to approach the world of remote sensing, to understand the use of SNAP, and also to understand how LULC maps are developed from Sentinel-2 satellite images.

At the end of this study on the evaluation and the comparison of the performances and the behaviour of Random Forest Classifier related to the variations of some important parameters in the classification of images especially (nb-trees, indices and generalization).

Furthermore, the comparison of the results of the study of iterations and their application made it possible to identify the most efficient methods.

the evaluation of the global indices: "kappa coefficient" and "Overall accuracy". according to the three application contexts (nb-trees, indices and generalization), revealed that the iteration with a nb-trees = 1000 is more suited to the classification of Sentinel-2 image data, this classification is considered as the most robust in terms of precision giving good results compared to other iterations with other nb-trees. And for the fact of integrating indices in the classification, it makes the classification precise especially for the soil indices. Concerning the generalization, it improved the classification to be more accurate than a classification with 1000 as nb-trees or a classification with indices.

Here is the final map of Bragança region, by adding the urban class (*Dgterritorio*) which is predefined since the demographic growth is very low.(*Bragança · Population*).

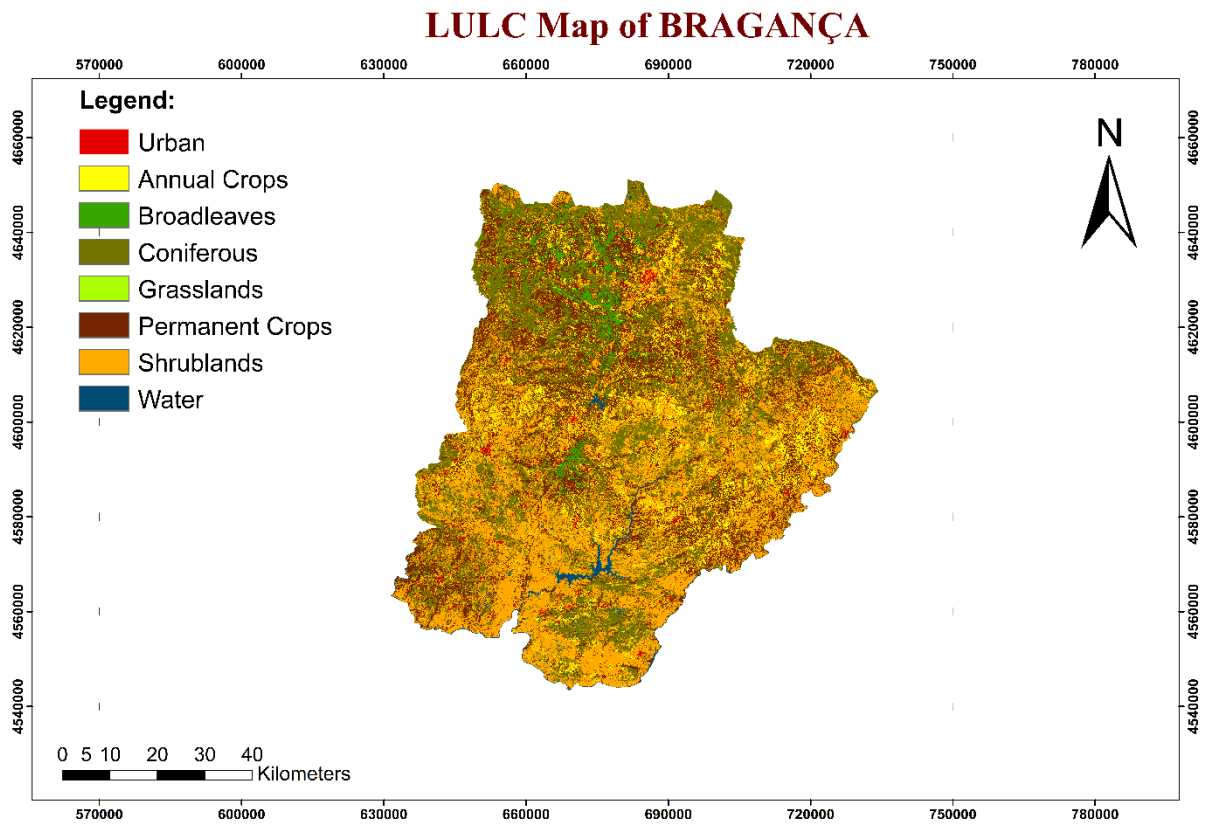


Figure 37: Final LULC map of Bragança district

We conclude that the classification of Sentinel-2 images with Random Forest Classifier on SNAP software will be more robust by:

- increasing the nb-trees
- integrating vegetation / soil indices
- applying generalization to eliminate intruder pixels.

The main limitations during this project are as follows:

- although SNAP is very easy to use, but occasionally it bugs before starting iteration.
- Processing times are relatively long
- The study area is very large which makes it difficult to assess the classifications made

Finally, for a possible similar study, we recommend:

- Test other satellite images covering other areas to further compare the results
- Use other classification algorithms in the comparison.
- Test other satellite images season in the comparison.
- Test RF with a number of trees Superior to 1000

References:

Al Sayah, M. J., Nedjai, R., Abdallah, C., & Khouri, M. (2020). On the use of remote sensing to map the proliferation of aquaculture ponds and to investigate their effect on local climate, perspectives from the Claise watershed, France. *Environmental Monitoring and Assessment*, *192*(5), 301. https://doi.org/10.1007/s10661-020-08250-0

Asner, G. P., Powell, G. V. N., Mascaro, J., Knapp, D. E., Clark, J. K., Jacobson, J., Kennedy-Bowdoin, T., Balaji, A., Paez-Acosta, G., Victoria, E., Secada, L., Valqui, M., & Hughes, R. F. (2010). High-resolution forest carbon stocks and emissions in the Amazon. *Proceedings of the National Academy of Sciences*, *107*(38), 16738–16742. https://doi.org/10.1073/pnas.1004875107

Azevedo, J. C., Moreira, C., Castro, J. P., & Loureiro, C. (2011). Agriculture Abandonment, Land-use Change and Fire Hazard in Mountain Landscapes in Northeastern Portugal. In C. Li, R. Lafortezza, & J. Chen (Eds.), *Landscape Ecology in Forest Management and Conservation: Challenges and Solutions for Global Change* (pp. 329–351). Springer. https://doi.org/10.1007/978-3-642-12754-0_14

Baret, F., Morissette, J. T., Fernandes, R. A., Champeaux, J. L., Myneni, R. B., Chen, J., Plummer, S., Weiss, M., Bacour, C., Garrigues, S., & Nickeson, J. E. (2006). Evaluation of the representativeness of networks of sites for the global validation and intercomparison of land biophysical products: Proposition of the CEOS-BELMANIP. *IEEE Transactions on Geoscience and Remote Sensing*, *44*(7), 1794–1803. https://doi.org/10.1109/TGRS.2006.876030

Begue, A., Desprat, J. F., Imbernon, J., & Baret, F. (1991). Radiation use efficiency of pearl millet in the Sahelian zone. *Agricultural and Forest Meteorology*, *56*(1), 93–110. https://doi.org/10.1016/0168-1923(91)90106-Z

Beißler, M. R., & Hack, J. (2019). A Combined Field and Remote-Sensing Based Methodology to Assess the Ecosystem Service Potential of Urban Rivers in Developing Countries. *Remote Sensing*, *11*(14), 1697. https://doi.org/10.3390/rs11141697

Belenguer-Plomer, M. A., Tanase, M. A., Fernandez-Carrillo, A., & Chuvieco, E. (2019). Burned area detection and mapping using Sentinel-1 backscatter coefficient and thermal anomalies. *Remote Sensing of Environment*, *233*, 111345. https://doi.org/10.1016/j.rse.2019.111345

Bragança · Population. (n.d.). Retrieved October 25, 2020, from http://population.city/portugal/braganca/

Breiman, L. (2001). Random Forests. *Machine Learning*, *45*(1), 5–32. https://doi.org/10.1023/A:1010933404324

Castro, M., & Fernandez-Nuñez, E. (2014). Soil properties and understory herbaceous biomass in forests of three species of Quercus in Northeast Portugal. *Forest Systems*, *23*(3), 425. https://doi.org/10.5424/fs/2014233-05086

Cohen, J. (1960). *A Coefficient of Agreement for Nominal Scales—Jacob Cohen, 1960*. https://journals.sagepub.com/doi/abs/10.1177/001316446002000104?journalCode=epma

Costa, J., Capelo, J., Lousã, M., & Aguiar, C. (1993). Communautés de Juniperus au Portugal. *Colloq. Phytosoc.*, *22*, 499–526.

Cracknell, A. P. (2018). The development of remote sensing in the last 40 years. *International Journal of Remote Sensing*, *39*(23), 8387–8427. https://doi.org/10.1080/01431161.2018.1550919

D Rudd, J., Roberson, G., & J. Classen, J. (2017, January). *Application of satellite, unmanned aircraft system, and ground-based sensor data for precision agriculture: A review | Request PDF*. ResearchGate. https://doi.org/10.13031/aim.201700272

Data Analysis in the Geosciences. (n.d.). Retrieved October 20, 2020, from http://strata.uga.edu/8370/rtips/proportions.html

Dgterritorio. (n.d.). Retrieved August 31, 2020, from https://www.dgterritorio.gov.pt/

Diesing, M., & Stephens, D. (2015). A multi-model ensemble approach to seabed mapping. *Journal of Sea Research*, *100*, 62–69. https://doi.org/10.1016/j.seares.2014.10.013

DIMYATI, M., MIZUNO, K., KOBAYASHI, S., & KITAMURA, T. (1996). An analysis of land use/cover change in Indonesia. *International Journal of Remote Sensing*, *17*(5), 931–944. https://doi.org/10.1080/01431169608949056

Edgington, R. (2012). An 'all-seeing flying eye': V-2 rockets and the promises of Earth photography. *History and Technology*, *28*(3), 363–371. https://doi.org/10.1080/07341512.2012.722796

Eitel, J. U. H., Vierling, L. A., Litvak, M. E., Long, D. S., Schulthess, U., Ager, A. A., Krofcheck, D. J., & Stoscheck, L. (2011). Broadband, red-edge information from satellites improves early stress detection in a New Mexico conifer woodland. *Remote Sensing of Environment*, *115*(12), 3640–3646. https://doi.org/10.1016/j.rse.2011.09.002

El Hajj, M., Baghdadi, N., Zribi, M., & Bazzi, H. (2017). Synergic Use of Sentinel-1 and Sentinel-2 Images for Operational Soil Moisture Mapping at High Spatial Resolution over Agricultural Areas. *Remote Sensing*, *9*(12), 1292. https://doi.org/10.3390/rs9121292

Fensholt, R., Sandholt, I., & Rasmussen, M. S. (2004). Evaluation of MODIS LAI, fAPAR and the relation between fAPAR and NDVI in a semi-arid environment using in situ measurements. *Remote Sensing of Environment*, *91*(3), 490–507. https://doi.org/10.1016/j.rse.2004.04.009

Fiolek, A. (2011). *Selected Publications on TIROS Satellites and Satellite Meteorology Available from the NOAA Central Library Network*.

Fung, T., & LeDrew, E. (1988). The Determination of Optimal Threshold Levels for Change Detection Using Various Accuracy Indices. *PHOTOGRAMMETRIC ENGINEERING*, 6.

Goetz, S. J., Steinberg, D., Betts, M. G., Holmes, R. T., Doran, P. J., Dubayah, R., & Hofton, M. (2010). Lidar remote sensing variables predict breeding habitat of a Neotropical migrant bird. *Ecology*, *91*(6), 1569–1576. https://doi.org/10.1890/09-1670.1

Gómez-Giráldez, P. J., Aguilar, C., Caño, A. B., García-Moreno, A., & González-Dugo, M. P. (2019). Remote sensing estimation of net primary production as monitoring indicator of holm oak savanna management. *Ecological Indicators*, *106*, 105526. https://doi.org/10.1016/j.ecolind.2019.105526

Govender, M., Chetty, K., & Bulcock, H. (2007). A review of hyperspectral remote sensing and its application in vegetation and water resource studies. *Water SA*, *33*(2), Article 2. https://doi.org/10.4314/wsa.v33i2.49049

Gutman, G., & Ignatov, A. (1998). The derivation of the green vegetation fraction from NOAA/AVHRR data for use in numerical weather prediction models. *International Journal of Remote Sensing*, *19*(8), 1533–1543. https://doi.org/10.1080/014311698215333

Hastie, T., Tibshirani, R., & Friedman, J. (2009). Random Forests. In T. Hastie, R. Tibshirani, & J. Friedman (Eds.), *The Elements of Statistical Learning: Data Mining, Inference, and Prediction* (pp. 587–604). Springer. https://doi.org/10.1007/978-0-387-84858-7_15

Hollingham, R. (n.d.). *V2: The Nazi rocket that launched the space age*. Retrieved August 4, 2020, from https://www.bbc.com/future/article/20140905-the-nazis-space-age-rocket

International Forest Fire News (IFFN). (2006). https://gfmc.online/iffn/iffn-34-content34.html

IPMA. (2020). [Instituto Português do Mar e da Atmosfera]. Normais Climatológicas - 1981-2010 (Provisórias)- Bragança. http://www.ipma.pt/en/oclima/normais.clima/1981-2010/003/

Jazouli, A. E., Barakat, A., Khellouk, R., Rais, J., & Baghdadi, M. E. (2019). Remote sensing and GIS techniques for prediction of land use land cover change effects on soil erosion in the high basin of the Oum Er Rbia River (Morocco). *Remote Sensing Applications: Society and Environment*, *13*, 361–374. https://doi.org/10.1016/j.rsase.2018.12.004

Juniperus oxycedrus—Trees and Shrubs Online. (n.d.). Retrieved October 17, 2020, from https://treesandshrubsonline.org/articles/juniperus/juniperus-oxycedrus/

Kramer, H. J., & Cracknell, A. P. (2008). An overview of small satellites in remote sensing. *International Journal of Remote Sensing*, *29*(15), 4285–4337. https://doi.org/10.1080/01431160801914952

Land cover and land use, landscape (LUCAS) (lan). (n.d.). Retrieved September 5, 2020, from https://ec.europa.eu/eurostat/cache/metadata/en/lan_esms.htm

Landgrebe, D. A. (2005). *Signal Theory Methods in Multispectral Remote Sensing*. John Wiley & Sons.

Landis, J. R., & Koch, G. G. (1977). The Measurement of Observer Agreement for Categorical Data. *Biometrics*, *33*(1), 159–174. https://doi.org/10.2307/2529310

Levin, N. (2017). The impact of seasonal changes on observed nighttime brightness from 2014 to 2015 monthly VIIRS DNB composites. *Remote Sensing of Environment*, *193*, 150–164. https://doi.org/10.1016/j.rse.2017.03.003

Liaw, A., & Wiener, M. (2002). *Classification and Regression by randomForest*. *2*, 6.

Loff, S. (2018, December 10). *Apollo 8* [Text]. NASA. http://www.nasa.gov/mission_pages/apollo/apollo-8.html

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

Lowman, P. D. (1999). Landsat and Apollo: The Forgotten Legacy. *PHOTOGRAMMETRIC ENGINEERING*, 5.

Macintyre, P., van Niekerk, A., & Mucina, L. (2020). Efficacy of multi-season Sentinel-2 imagery for compositional vegetation classification. *International Journal of Applied Earth Observation and Geoinformation*, *85*, 101980. https://doi.org/10.1016/j.jag.2019.101980

Merlin, O., Al Bitar, A., Walker, J. P., & Kerr, Y. (2010). An improved algorithm for disaggregating microwave-derived soil moisture based on red, near-infrared and thermal-infrared data. *Remote Sensing of Environment*, *114*(10), 2305–2316. https://doi.org/10.1016/j.rse.2010.05.007

Minnett, P. J., Alvera-Azcárate, A., Chin, T. M., Corlett, G. K., Gentemann, C. L., Karagali, I., Li, X., Marsouin, A., Marullo, S., Maturi, E., Santoleri, R., Saux Picart, S., Steele, M., & Vazquez-Cuervo, J. (2019). Half a century of satellite remote sensing of sea-surface temperature. *Remote Sensing of Environment*, *233*, 111366. https://doi.org/10.1016/j.rse.2019.111366

Missions—Sentinel Online. (n.d.). Retrieved August 27, 2020, from https://sentinel.esa.int/web/sentinel/missions

Muhammad Zulkarnain, A. R. (n.d.). *Characteristics-of-remote-sensing-data-Pre-processing.pdf*. Retrieved August 27, 2020, from https://people.utm.my/nurulhazrina/files/2015/04/L09-Characteristics-of-remote-sensing-data-Pre-processing.pdf

NASA's Scientific. (2020, May 8). *SVS: NOAA-20 satellite orbit with Suomi NPP and JPSS-2*. https://svs.gsfc.nasa.gov/4820

Nhamo, L., Ebrahim, G. Y., Mabhaudhi, T., Mpandeli, S., Magombeyi, M., Chitakira, M., Magidi, J., & Sibanda, M. (2020). An assessment of groundwater use in irrigated agriculture using multi-spectral remote sensing. *Physics and Chemistry of the Earth, Parts A/B/C*, *115*, 102810. https://doi.org/10.1016/j.pce.2019.102810

Pelton, J. N., & Camacho-Lara, S. (2013). Introduction to Satellite Navigation Systems. In J. N. Pelton, S. Madry, & S. Camacho-Lara (Eds.), *Handbook of Satellite Applications* (pp. 561–571). Springer. https://doi.org/10.1007/978-1-4419-7671-0_10

P.K. Mani. (2014, January 17). *Scanners, image resolution, orbit in remote sensing, pk mani* [Education]. https://www.slideshare.net/pabitramani/scanners-image-resolution-orbit-in-remote-sensing-pk-mani

Portuguese Environment Agency. (n.d.). Retrieved September 5, 2020, from https://apambiente.pt/index.php

Radiometric—Resolutions—Sentinel-2 MSI - User Guides—Sentinel Online. (n.d.). Retrieved October 23, 2020, from https://sentinel.esa.int/web/sentinel/user-guides/sentinel-2-msi/resolutions/radiometric

Rango, A., Laliberte, A., Herrick, J. E., Winters, C., Havstad, K., Steele, C., & Browning, D. (2009). Unmanned aerial vehicle-based remote sensing for rangeland assessment, monitoring, and management. *Journal of Applied Remote Sensing*, *3*(1), 033542. https://doi.org/10.1117/1.3216822

Rawat, J. S., & Kumar, M. (2015). Monitoring land use/cover change using remote sensing and GIS techniques: A case study of Hawalbagh block, district Almora, Uttarakhand, India. *The Egyptian Journal of Remote Sensing and Space Science*, *18*(1), 77–84. https://doi.org/10.1016/j.ejrs.2015.02.002

Rodrigues, P., Aubrecht, C., Gil, A., Longcore, T., & Elvidge, C. (2011). Remote sensing to map influence of light pollution on Cory's Shearwater in São Miguel Island, Azores Archipelago. *European Journal of Wildlife Research*, *58*, 147–155. https://doi.org/10.1007/s10344-011-0555-5

Rose, R. A., Byler, D., Eastman, J. R., Fleishman, E., Geller, G., Goetz, S., Guild, L., Hamilton, H., Hansen, M., Headley, R., Hewson, J., Horning, N., Kaplin, B. A., Laporte, N., Leidner, A., Leimgruber, P., Morisette, J., Musinsky, J., Pintea, L., … Wilson, C. (2015). Ten ways remote sensing can contribute to conservation. *Conservation Biology*, *29*(2), 350–359. https://doi.org/10.1111/cobi.12397

Rosenfield, G. H., & Fitzpatrick-Lins, K. (1986). A coefficient of agreement as a measure of thematic classification accuracy. In *Photogrammetric Engineering and Remote Sensing* (Vol. 52, Issue 2, p. 5).

Sabat-Tomala, A., Raczko, E., & Zagajewski, B. (2020). Comparison of Support Vector Machine and Random Forest Algorithms for Invasive and Expansive Species Classification Using Airborne Hyperspectral Data. *Remote Sensing*, *12*(3), 516. https://doi.org/10.3390/rs12030516

Saito, K., Ogawa, S., Aihara, M., & Otowa, K. (2001). *ESTIMATES OF LAI FOR FOREST MANAGEMENT IN OKUTAMA*. 6.

Schnapf, A. (1985). *Monitoring Earth's Ocean, Land, and Atmosphere from Space-Sensors, Systems, and Applications* (A. Schnapf, Ed.). American Institute of Aeronautics and Astronautics. https://doi.org/10.2514/4.865725

Sharp, T. (2018). *Saturn V Rockets & Apollo Spacecraft*. Space.Com. https://www.space.com/16698-apollo-spacecraft.html

Spectral Characteristics Viewer | Landsat Missions. (n.d.). Retrieved October 18, 2020, from https://landsat.usgs.gov/spectral-characteristics-viewer

Support Vector Machines Tutorial—Learn to implement SVM in Python. (2017, August 4). DataFlair. https://data-flair.training/blogs/svm-support-vector-machine-tutorial/

Thakur, N., & Maheshwari, D. (2017). *A REVIEW OF IMAGE CLASSIFICATION TECHNIQUES*. /paper/A-REVIEW-OF-IMAGE-CLASSIFICATION-TECHNIQUES-Thakur-Maheshwari/7899631cc2c59d488b621d27ff46d2260cf7360e

Tyukavina, A., Baccini, A., Hansen, M. C., Potapov, P. V., Stehman, S. V., Houghton, R. A., Krylov, A. M., Turubanova, S., & Goetz, S. J. (2015). Aboveground carbon loss in natural and managed tropical forests from 2000 to 2012. *Environmental Research Letters*, *10*(7), 074002. https://doi.org/10.1088/1748-9326/10/7/074002

Varmuza, K., & Filzmoser, P. (2016). *Introduction to Multivariate Statistical Analysis in Chemometrics*. CRC Press.

Venter, Z. S., Brousse, O., Esau, I., & Meier, F. (2020). Hyperlocal mapping of urban air temperature using remote sensing and crowdsourced weather data. *Remote Sensing of Environment*, *242*, 111791. https://doi.org/10.1016/j.rse.2020.111791

Viña, A., Gitelson, A. A., Nguy-Robertson, A. L., & Peng, Y. (2011). Comparison of different vegetation indices for the remote assessment of green leaf area index of crops. *Remote Sensing of Environment*, *115*(12), 3468–3478. https://doi.org/10.1016/j.rse.2011.08.010

Wallace, L., Hillman, S., Reinke, K., & Hally, B. (2017). Non-destructive estimation of above-ground surface and near-surface biomass using 3D terrestrial remote sensing techniques. *Methods in Ecology and Evolution*, *8*(11), 1607–1616. https://doi.org/10.1111/2041-210X.12759

Wu, S., Wen, J., You, D., Hao, D., Lin, X., Xiao, Q., Liu, Q., & Gastellu-Etchegorry, J.-P. (2018). Characterization of Remote Sensing Albedo Over Sloped Surfaces Based on DART Simulations and In Situ Observations. *Journal of Geophysical Research: Atmospheres*, *123*(16), 8599–8622. https://doi.org/10.1029/2018JD028283

Xue, Y., Li, Y., Guang, J., Zhang, X., & Guo, J. (2008). Small satellite remote sensing and applications – history, current and future. *International Journal of Remote Sensing*, *29*(15), 4339–4372. https://doi.org/10.1080/01431160801914945

Yang, X., Zhao, S., Qin, X., Zhao, N., & Liang, L. (2017). Mapping of Urban Surface Water Bodies from Sentinel-2 MSI Imagery at 10 m Resolution via NDWI-Based Image Sharpening. *Remote Sensing*, *9*(6), 596. https://doi.org/10.3390/rs9060596

Yang, Y., Cao, C., Pan, X., Li, X., & Zhu, X. (2017). Downscaling Land Surface Temperature in an Arid Area by Using Multiple Remote Sensing Indices with Random Forest Regression. *Remote Sensing*, *9*(8), 789. https://doi.org/10.3390/rs9080789

Yao, X. (2009). *Spatial Resolution—An overview | ScienceDirect Topics*. https://www.sciencedirect.com/topics/earth-and-planetary-sciences/spatial-resolution

Zhang, H. W., & Chan, K. (2015). *Juniperus Oxycedrus—An overview (pdf) | ScienceDirect Topics*. https://www.sciencedirect.com/topics/agricultural-and-biological-sciences/juniperus-oxycedrus/pdf

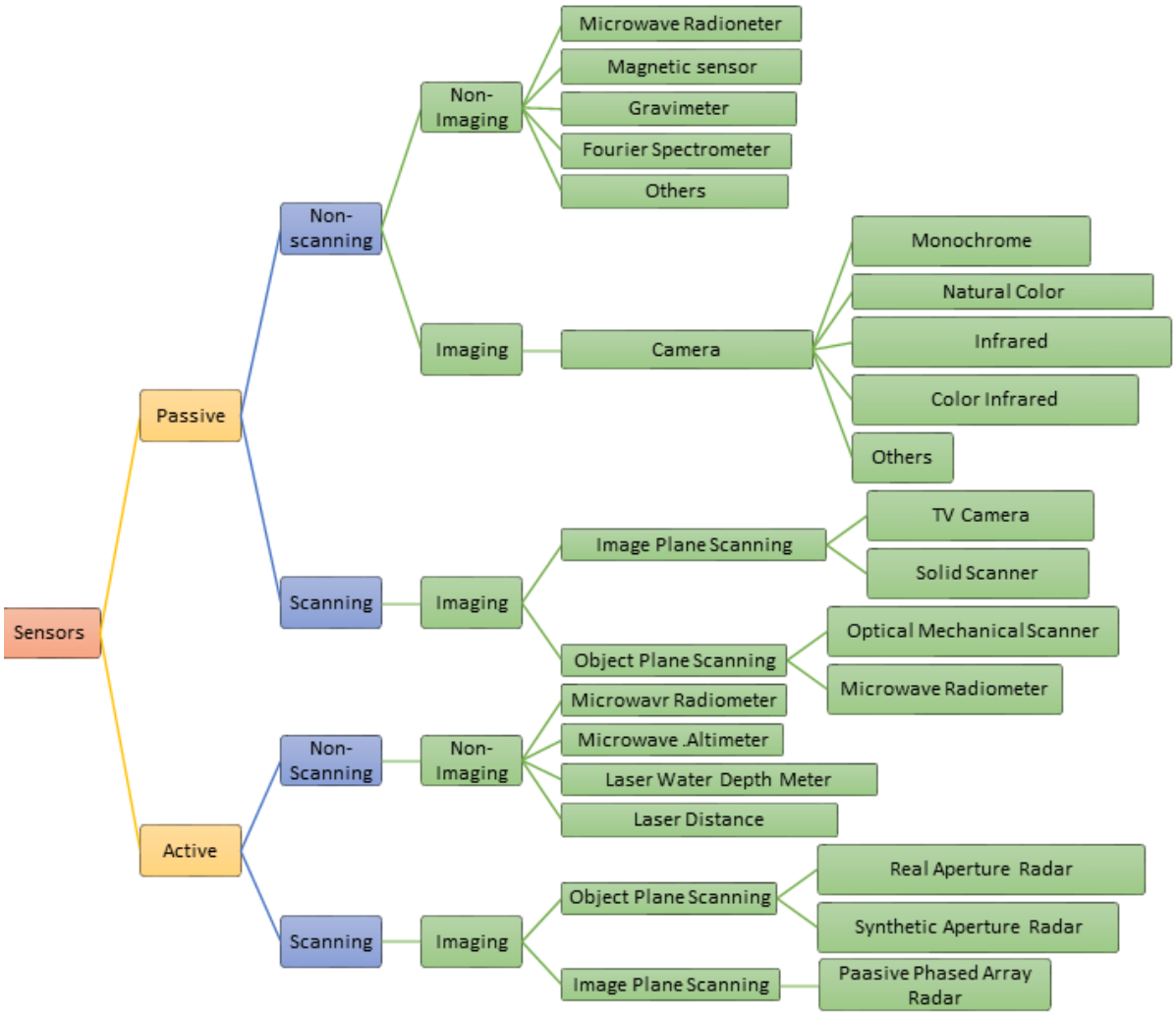
Zhang, J. (2010). Multi-source remote sensing data fusion: Status and trends. *International Journal of Image and Data Fusion*, *1*(1), 5–24. https://doi.org/10.1080/19479830903561035

Zhou, W., Troy, A., & Grove, M. (2008). Object-based Land Cover Classification and Change Analysis in the Baltimore Metropolitan Area Using Multitemporal High Resolution Remote Sensing Data. *Sensors*, *8*(3), 1613–1636. https://doi.org/10.3390/s8031613

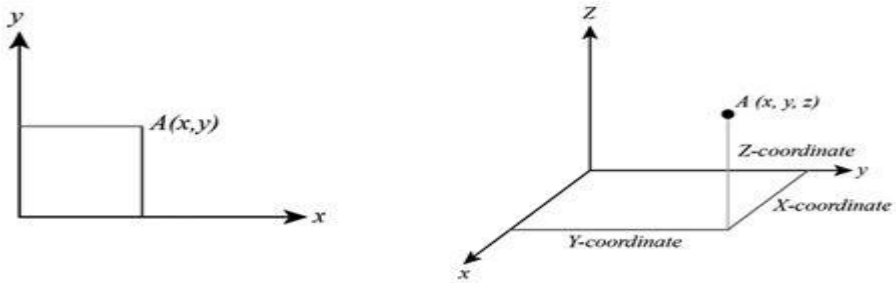
Zhu, D., Chen, T., Zhen, N., & Niu, R. (2020). Monitoring the effects of open-pit mining on the eco-environment using a moving window-based remote sensing ecological index. *Environmental Science and Pollution Research*, *27*(13), 15716–15728. https://doi.org/10.1007/s11356-020-08054-2

Annexes

Annex I: types of remote sensors



Annex II: Geometrical view of Euclidean plane and Euclidean space, left(two dimensional Euclidean plane), right (three dimensional Euclidean space)



Annex III: confusion matrix of Bragança Classification with Sentinel-2 Image, nb=1000

prdt_1000	True0	True1	True2	True3	True4	True5	True6	Ground truth	UA	CA
AC	1270	0	0	0	261	140	0	1671	76	24
BL	0	6479	226	26	0	0	0	6731	96	4
CONIF	0	56	703	0	0	2941	10	3710	19	81
G	0	8	0	353	37	0	0	398	89	11
PC	8	88	0	24	408	218	0	746	55	45
SHRUB	9	0	0	0	136	15013	0	15158	99	1
WATER	0	0	3	0	0	1029	60218	61250	98	2
Total	1287	6631	932	403	842	19341	60228	89664		
PA	99	98	75	88	48	78	100	94		
OA	1	2	25	12	52	22	0			
Overalll accuracy	94	index Kappa	0.88							

Annex IV: confusion matrix of Bragança Classification with Sentinel-2 Image, nb=900

pred_900	True0	True1	True2	True3	True4	True5	True6		UA	CA
AC	1259	0	0	0	290	141	0	1690	74	26
BL	0	6399	222	25	0	0	0	6646	96	4
CONIF	0	40	707	0	0	3810	9	4566	15	85
G	0	7	0	357	44	0	0	408	88	13
PC	23	185	0	21	382	252	0	863	44	56
SHRUB	5	0	0	0	126	14461	2	14594	99	1
WATER	0	0	3	0	0	677	60217	60897	99	1
Total	1287	6631	932	403	842	19341	60228	89664		
PA	98	97	76	89	45	75	100			
OA	2	3	24	11	55	25	0			
Overalll accuracy	93	index Kappa	0.87							

Annex V: confusion matrix of Bragança Classification with Sentinel-2 Image, nb=800

predict800	True0	True1	True2	True3	True4	True5	True6		UA	CA
AC	1259	0	0	0	280	135	0	1674	75	25
BL	0	6418	221	25	0	0	0	6664	96	4
CONIF	0	39	709	0	0	4047	12	4807	15	85
G	0	10	0	356	38	0	0	404	88	12
PC	23	164	0	22	389	247	0	845	46	54
SHRUB	5	0	0	0	135	14401	0	14541	99	1
WATER	0	0	2	0	0	511	60216	60729	99	1
Total	1287	6631	932	403	842	19341	60228	89664		
PA	98	97	76	88	46	74	100			
OA	2	3	24	12	54	26	0			
Overalll accuracy	93	index Kappa	0.87							

Annex VI: confusion matrix of Bragança Classification with Sentinel-2 Image, nb=700

predict700	True0	True1	True2	True3	True4	True5	True6		UA	CA
AC	1261	0	0	0	287	149	0	1697	74	26
BL	0	6372	219	22	0	0	0	6613	96	4
CONIF	0	39	710	0	0	4079	14	4842	15	85
G	0	6	0	359	38	0	0	403	89	11
PC	21	214	0	22	393	265	0	915	43	57
SHRUB	5	0	0	0	124	14486	0	14615	99	1
WATER	0	0	3	0	0	362	60214	60579	99	1
Total	1287	6631	932	403	842	19341	60228	89664		
PA	98	96	76	89	47	75	100			
OA	2	4	24	11	53	25	0			
Overalll accuracy	93	index Kappa	0.87							

Annex VII: confusion matrix of Bragança Classification with Sentinel-2 Image, nb=600

pre_600	True0	True1	True2	True3	True4	True5	True6		UA	CA
AC	1271	0	0	0	300	144	0	1715	74	26
BL	0	6410	217	24	2	0	0	6653	96	4
CONIF	0	49	712	0	0	5021	14	5796	12	88
G	0	7	0	355	32	0	0	394	90	10
PC	12	165	0	24	370	202	0	773	48	52
SHRUB	4	0	0	0	138	13957	0	14099	99	1
WATER	0	0	3	0	0	17	60214	60234	100	0
Total	1287	6631	932	403	842	19341	60228	89664		
PA	99	97	76	88	44	72	100			
OA	1	3	24	12	56	28	0			
Overalll accuracy	93	index Kappa	0.86							

Annex VIII: confusion matrix of Bragança Classification with Sentinel-2 Image, nb=500

predict500	True0	True1	True2	True3	True4	True5	True6	Ground truth	UA	CA
AC	1260	0	0	0	279	143	0	1682	75	25
BL	0	6435	219	26	0	0	0	6680	96	4
CONIF	0	56	711	0	0	4056	9	4832	15	85
G	0	11	0	355	34	0	0	400	89	11
PC	21	129	0	22	394	178	0	744	53	47
SHRUB	6	0	0	0	135	14321	0	14462	99	1
WATER	0	0	2	0	0	643	60219	60864	99	1
Total	1287	6631	932	403	842	19341	60228	89664		
PA	98	97	76	88	47	74	100	89664		0
OA	2	3	24	12	53	26	0			
Overalll accuracy	93	index Kappa	0.87							

Annex IX: confusion matrix of Bragança Classification with Sentinel-2 Image, nb=400

Pret_400	True0	True1	True2	True3	True4	True5	True6	G-truth	UA	CA
AC	1261	0	0	0	282	145	0	1688	75	25
BL	0	6423	220	24	0	0	0	6667	96	4
CONIF	0	42	712	0	0	4133	11	4898	15	85
G	0	8	0	357	42	0	0	407	88	12
PC	23	158	0	22	398	241	0	842	47	53
SHRUB	3	0	0	0	120	14047	0	14170	99	1
WATER	0	0	0	0	0	775	60217	60992	99	1
Total	1287	6631	932	403	842	19341	60228	89664		
PA	98	97	76	89	47	73	100	89664		0
OA	2	3	24	11	53	27	0			
Overalll accuracy	93	index Kappa	0.86							

Annex X: confusion matrix of Bragança Classification with Sentinel-2 Image, nb=350

Predict_350	True0	True1	True2	True3	True4	True5	True6	G-truth	UA	CA
AC	1273	0	0	0	276	136	0	1685	76	24
BL	0	6364	203	20	0	0	0	6587	97	3
CONIF	0	50	729	0	0	4128	12	4919	15	85
G	0	5	0	362	42	0	0	409	89	11
PC	8	212	0	21	393	247	0	881	45	55
SHRUB	6	0	0	0	131	14341	0	14478	99	1
WATER	0	0	0	0	0	489	60216	60705	99	1
Total	1287	6631	932	403	842	19341	60228	89664		
PA	99	96	78	90	47	74	100	89664		0
OA	1	4	22	10	53	26	0			
Overalll accuracy	93	index Kappa	0.87							

Annex XI: confusion matrix of Bragança Classification with Sentinel-2 Image, nb=300

prdict_300	True0	True1	True2	True3	True4	True5	True6	G-truth	UA	CA
AC	1257	0	0	0	276	141	0	1674	75	25
BL	0	6518	221	21	0	0	0	6760	96	4
CONIF	0	54	708	0	0	3769	9	4540	16	84
G	0	7	0	361	44	0	0	412	88	12
PC	24	52	0	21	392	192	0	681	58	42
SHRUB	6	0	0	0	130	14216	0	14352	99	1
WATER	0	0	3	0	0	1023	60219	61245	98	2
Total	1287	6631	932	403	842	19341	60228	89664		
PA	98	98	76	90	47	74	100			0
OA	2	2	24	10	53	26	0			
Overalll accuracy	93	index Kappa	0.87							

Annex XII: confusion matrix of Bragança Classification with Sentinel-2 Image, nb=250

prdict_250	True0	True1	True2	True3	True4	True5	True6	G-truth	UA	CA
AC	1257	0	0	0	277	139	0	1673	25	75
BL	0	6417	219	21	0	0	0	6657	4	96
CONIF	0	44	711	0	0	4628	11	5394	87	13
G	0	10	0	360	41	0	0	411	12	88
PC	25	160	0	22	398	241	0	846	53	47
SHRUB	5	0	0	0	126	13762	0	13893	1	99
WATER	0	0	2	0	0	571	60217	60790	1	99
Total	1287	6631	932	403	842	19341	60228	89664		
PA	98	97	76	89	47	71	100	89664		0
OA	2	3	24	11	53	29	0			
Overalll accuracy	93	index Kappa	0.86							

Annex XIII: confusion matrix of Bragança Classification with Sentinel-2 Image, nb=200

prdict_200	True0	True1	True2	True3	True4	True5	True6	G-truth	UA	CA
AC	1260	0	0	0	287	145	0	1692	74	26
BL	0	6416	218	23	0	0	0	6657	96	4
CONIF	0	69	710	0	0	5005	11	5795	12	88
G	0	5	0	357	35	0	0	397	90	10
PC	22	141	0	23	400	175	0	761	53	47
SHRUB	5	0	0	0	120	13426	0	13551	99	1
WATER	0	0	4	0	0	590	60217	60811	99	1
Total	1287	6631	932	403	842	19341	60228	89664		
PA	98	97	76	89	48	69	100	89664		0
OA	2	3	24	11	52	31	0			
Overalll accuracy	92	index Kappa	0.85							

Annex XIV: confusion matrix of Bragança Classification with Sentinel-2 Image, nb=150

predict150	True0	True1	True2	True3	True4	True5	True6	G-truth	UA	CA
AC	1258	0	0	0	306	151	0	1715	73	27
BL	0	6392	217	20	0	0	0	6629	96	4
CONIF	0	73	712	0	0	4546	8	5339	13	87
G	0	10	0	360	39	0	0	409	88	12
PC	24	156	0	23	417	176	0	796	52	48
SHRUB	5	0	0	0	80	13384	0	13469	99	1
WATER	0	0	3	0	0	1084	60220	61307	98	2
Total	1287	6631	932	403	842	19341	60228	89664		
PA	98	96	76	89	50	69	100	89664		0
OA	2	4	24	11	50	31	0			
Overalll accuracy	92	index Kappa	0.85							

Annex XV: confusion matrix of Bragança Classification with Sentinel-2 Image, nb=100

prdict_100	True0	True1	True2	True3	True4	True5	True6	G-truth	UA	CA
AC	1256	0	0	0	290	138	0	1684	75	25
BL	0	6401	223	22	0	0	0	6646	96	4
CONIF	0	57	709	0	0	5879	13	6658	11	89
G	0	9	0	359	39	0	0	407	88	12
PC	26	164	0	22	391	169	0	772	51	49
SHRUB	5	0	0	0	122	12829	0	12956	99	1
WATER	0	0	0	0	0	326	60215	60541	99	1
Total	1287	6631	932	403	842	19341	60228	89664		
PA	98	97	76	89	46	66	100	89664		0
OA	2	3	24	11	54	34	0			
Overalll accuracy	92	index Kappa	0.84							

Annex XVI: confusion matrix of Bragança Classification with Sentinel-2 Image, nb=50

prdict_50	True0	True1	True2	True3	True4	True5	True6	G-truth	UA	CA
AC	1247	0	0	0	262	139	0	1648	76	24
BL	0	6460	228	22	0	0	0	6710	96	4
CONIF	0	61	704	0	0	5644	10	6419	11	89
G	0	8	0	359	40	0	0	407	88	12
PC	36	102	0	22	414	141	0	715	58	42
SHRUB	4	0	0	0	126	12848	0	12978	99	1
WATER	0	0	0	0	0	569	60218	60787	99	1
Total	1287	6631	932	403	842	19341	60228	89664		
PA	97	97	76	89	49	66	100	89664		
OA	3	3	24	11	51	34	0			
Overalll accuracy	92	index Kappa	0.84							

Annex XVII: confusion matrix of Bragança Classification with Sentinel-2 Image, nb=10

prdict_10	True0	True1	True2	True3	True4	True5	True6	G-truth	UA	CA
AC	1238	0	0	0	309	149	0	1696	73	27
BL	0	6384	220	26	2	0	0	6632	96	4
CONIF	0	44	710	0	0	6699	7	7460	10	90
G	0	5	0	354	23	0	0	382	93	7
PC	39	198	0	23	312	186	0	758	41	59
SHRUB	10	0	0	0	196	12305	10	12521	98	2
WATER	0	0	2	0	0	2	60211	60215	100	0
Total	1287	6631	932	403	842	19341	60228	89664		
PA	96	96	76	88	37	64	100	89664		
OA	4	4	24	12	63	36	0			
Overalll accuracy	91	index Kappa	0.82							

Annex XVIII: confusion matrix of Bragança with Sentinel-2 Image, (Soil Indices)

pred_soil	True0	True1	True2	True3	True4	True5	True6	G-truth	UA	CA
AC	1259	0	0	0	155	139	0	1553	80.88	18.93
BL	0	6449	234	18	0	0	0	6701	95.87	3.76
CONIF	0	50	698	0	0	2060	7	2815	24.76	75.20
G	0	10	0	363	35	0	0	408	90.20	11.03
PC	22	122	0	22	433	628	0	1227	39.28	64.71
SHRUB	6	0	0	0	219	16140	5	16370	114.59	1.41
WATER	0	0	0	0	0	374	60216	60590	99.38	0.62
Total	1287	6631	932	403	842	19341	60228	89664	98.37	4.58
PA	97.82	97.26	74.89	90.07	51.43	83.45	99.98	95.42		
OA	2.18	2.74	25.11	9.93	48.57	16.55	0.02	4.58		
Overalll accuracy	95	index Kappa	0.91							

Annex XIX: confusion matrix of Bragança with Sentinel-2 Image (Vegetation Indices)

pred-veg	True0	True1	True2	True3	True4	True5	True6	Gtruth	UA	CA
AC	1278	0	0	0	221	123	0	1622	77.44	21.21
BL	0	6310	244	53	2	0	0	6609	97.20	4.52
CONIF	0	29	688	0	0	496	1	1214	57.41	43.33
G	0	3	0	326	32	0	0	361	101.94	9.70
PC	0	289	0	24	439	1140	3	1895	25.44	76.83
SHRUB	9	0	0	0	148	15378	3	15538	120.72	1.03
WATER	0	0	0	0	0	2204	60221	62425	96.46	3.53
Total	1287	6631	932	403	842	19341	60228	89664		5.60
PA	99.30	95.16	73.82	80.89	52.14	79.51	99.99			
OA	0.70	4.84	26.18	19.11	47.86	20.49	0.01			
Overalll accuracy	94	index Kappa	0.89							

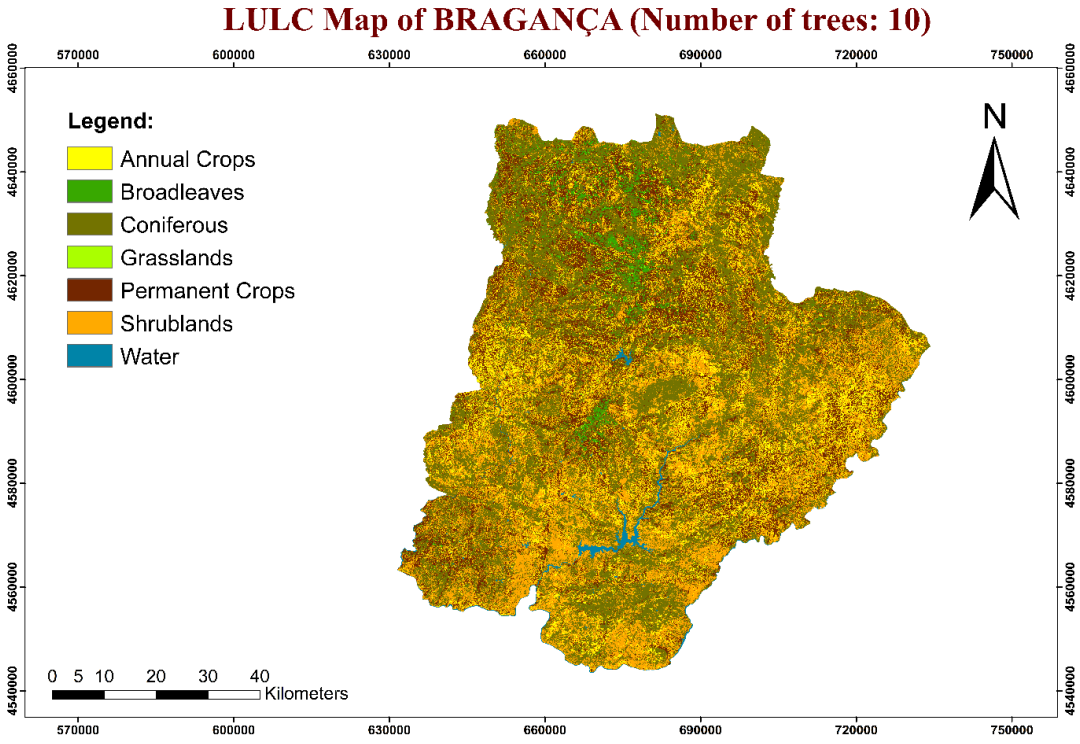
Annex XX: confusion matrix of Bragança with Sentinel-2 Image, both indices

pred-both	True0	True1	True2	True3	True4	True5	True6	Ground truth	UA	CA
AC	1285	0	0	0	114	131	0	1530	83.99	16.01
BL	0	6337	245	23	4	0	0	6609	95.88	4.12
CONIF	0	42	687	0	0	1168	3	1900	36.16	63.84
G	0	17	0	354	34	0	0	405	87.41	12.59
PC	0	235	0	26	420	867	1	1549	27.11	72.89
SHRUB	2	0	0	0	270	15892	6	16170	98.28	1.72
WATER	0	0	0	0	0	1283	60218	61501	97.91	2.09
Total	1287	6631	932	403	842	19341	60228	89664		
PA	99.84	95.57	73.71	87.84	49.88	82.17	99.98			
OA	0.16	4.43	26.29	12.16	50.12	17.83	0.02			
Overalll accuracy	95	index Kappa	0.90							

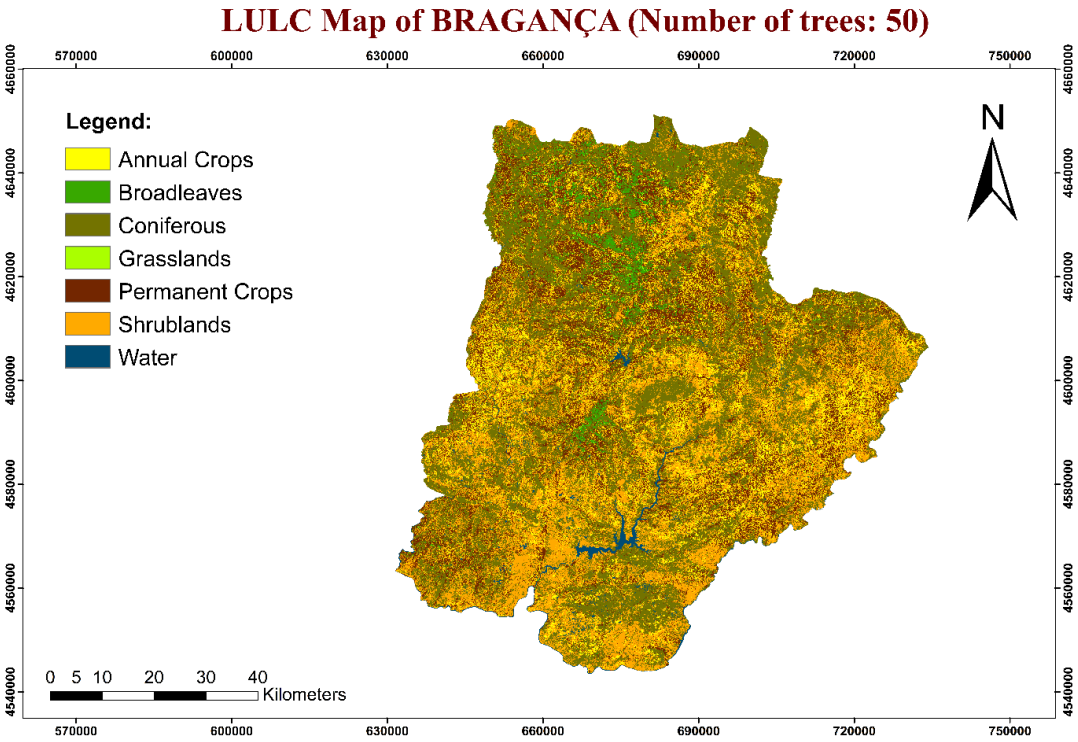
Annex XXI: confusion matrix of Bragança with Sentinel-2 Image, post-classification

pred-post	True0	True1	True2	True3	True4	True5	True6		UA	CA
AC	1284	0	0	0	103	122	0	1509	85.09	14.91
BL	0	6473	245	18	0	0	0	6736	96.10	3.90
CONIF	0	26	687	0	0	1544	0	2257	30.44	69.56
G	0	0	0	362	17	0	0	379	95.51	4.49
PC	0	132	0	23	466	441	0	1062	43.88	56.12
SHRUB	3	0	0	0	256	16919	5	17183	98.46	1.54
WATER	0	0	0	0	0	315	60223	60538	99.48	0.52
Total	1287	6631	932	403	842	19341	60228	89664		
PA	99.77	97.62	73.71	89.83	55.34	87.48	99.99			
OA	0.23	2.38	26.29	10.17	44.66	12.52	0.01			
Overalll accuracy	96	index Kappa	0.93							

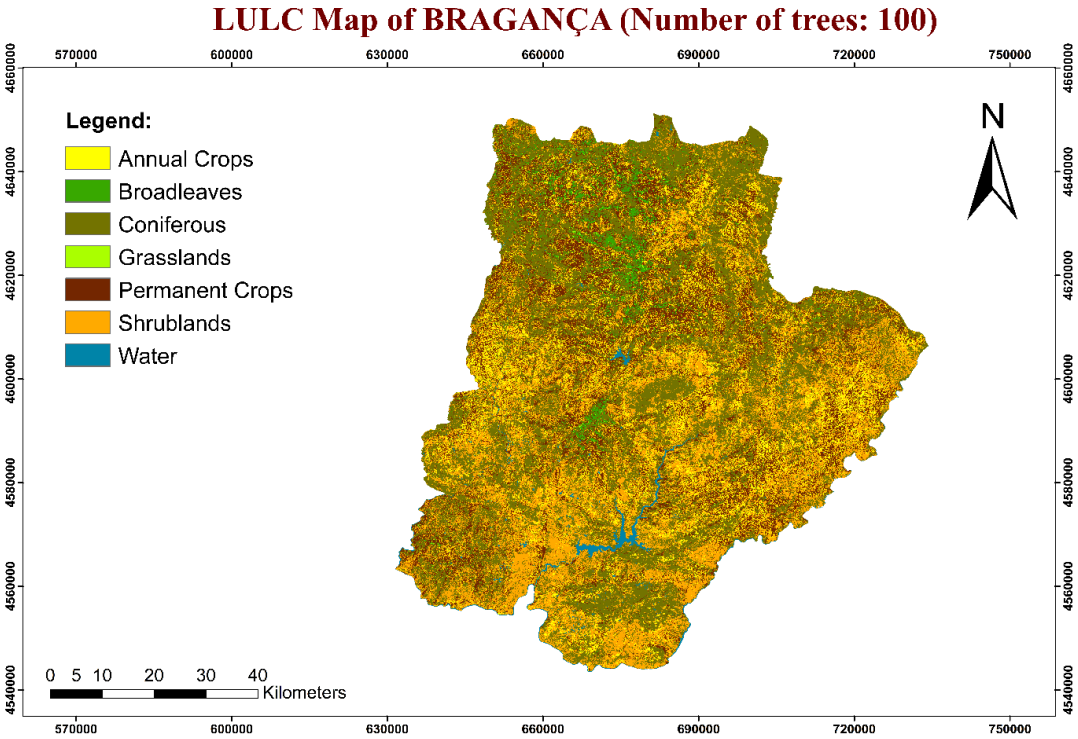
Annex XXII: LULC classification of Bragança with nb-trees=10



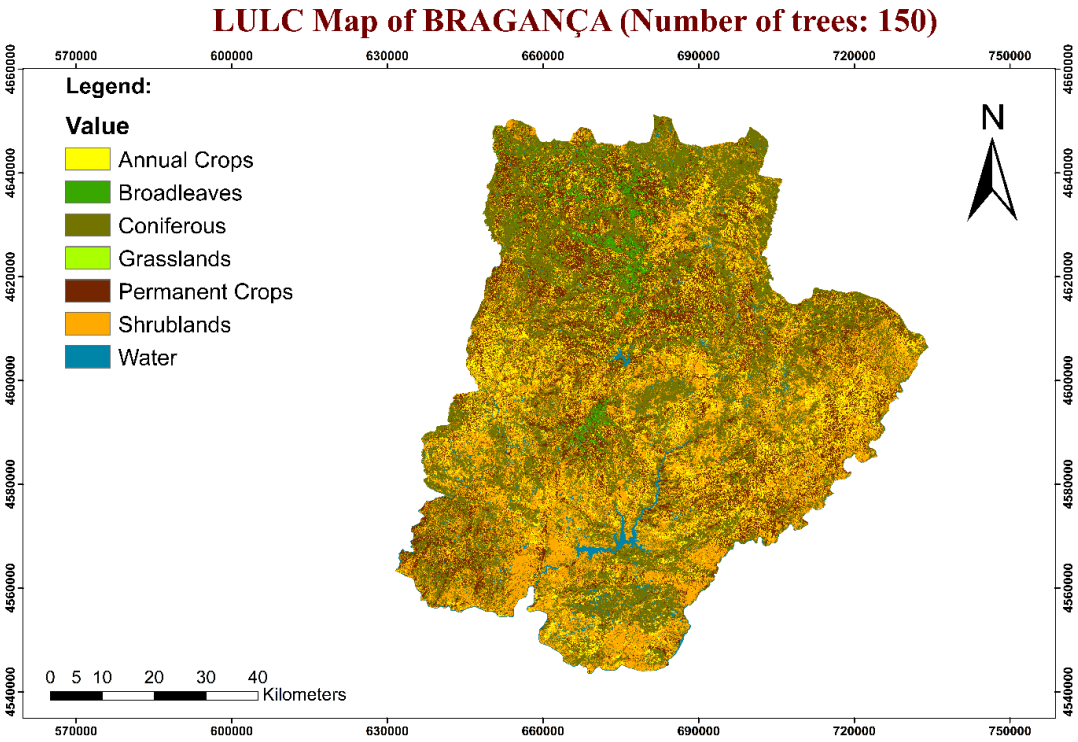
Annex XXIII: LULC classification of Bragança with nb-trees=50



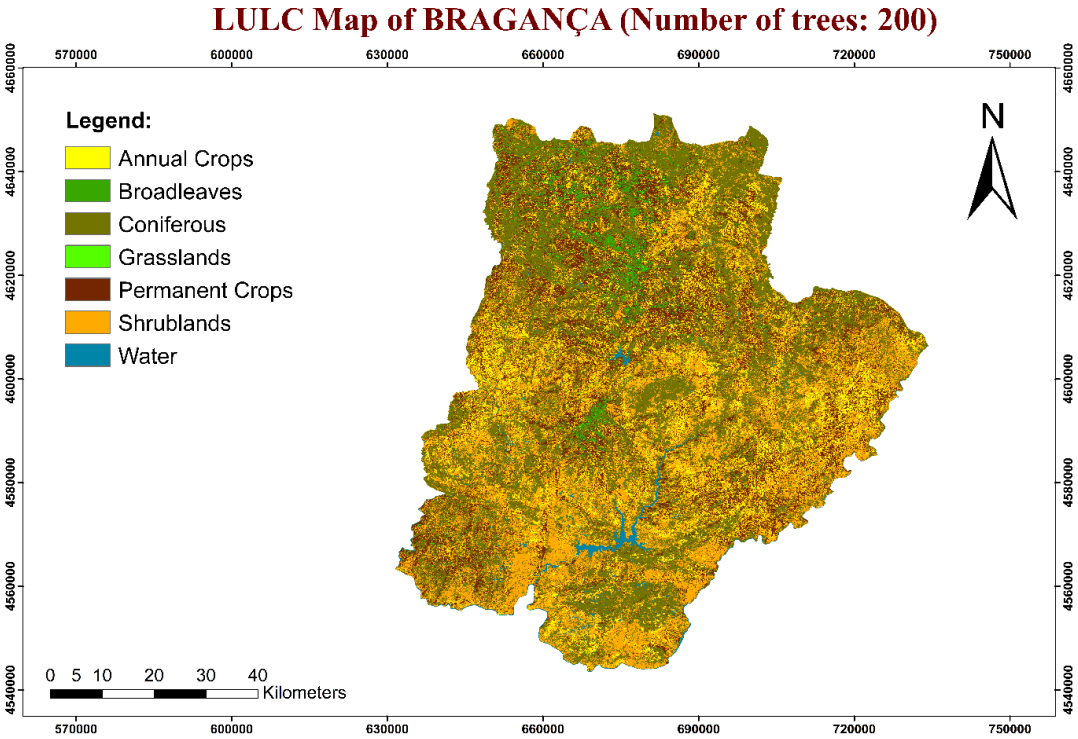
Annex XXIV: LULC classification of Bragança with nb-trees=100



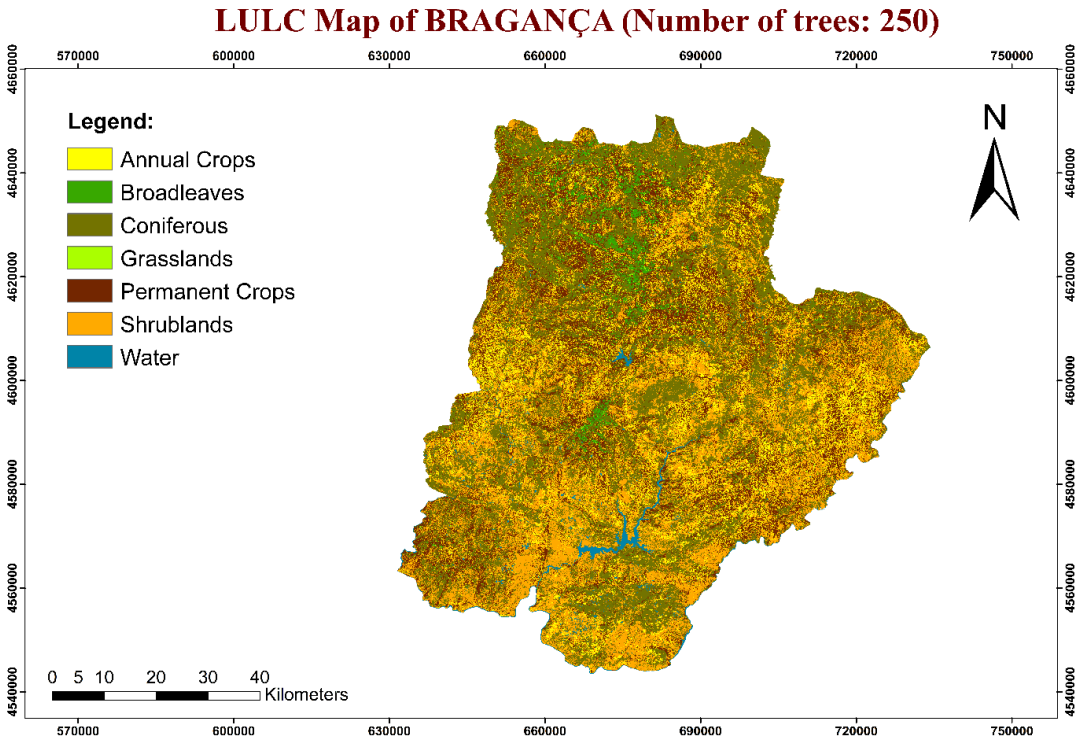
Annex XXV: LULC classification of Bragança with nb-trees=150



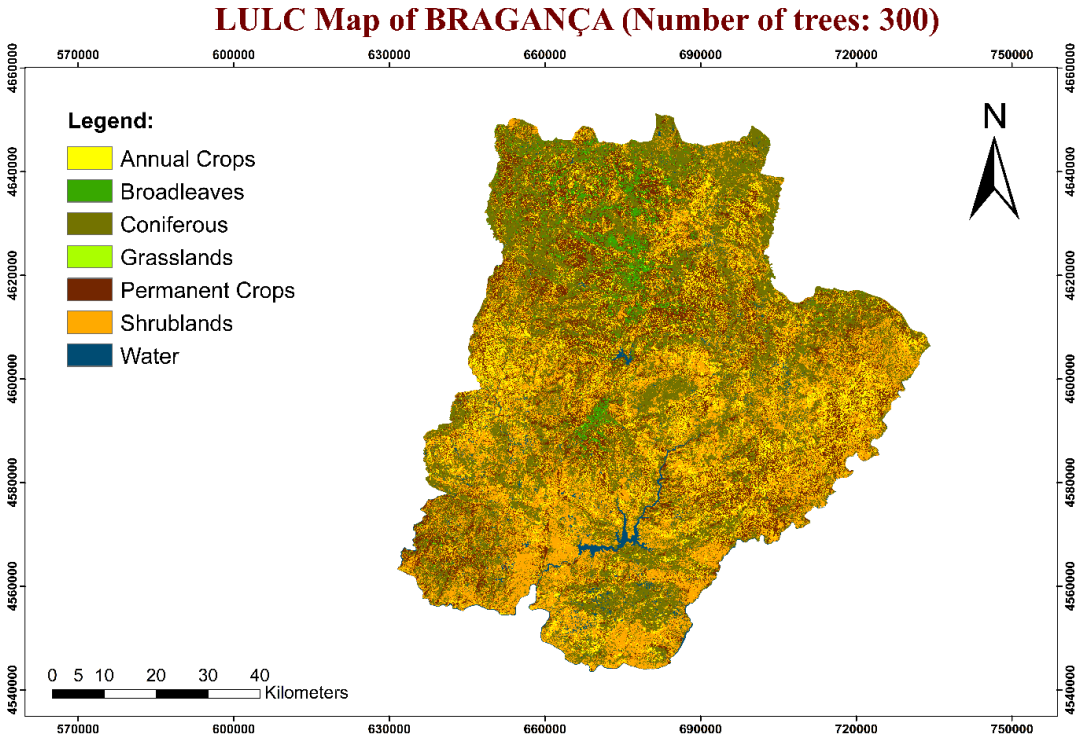
Annex XXVI: LULC classification of Bragança with nb-trees=200



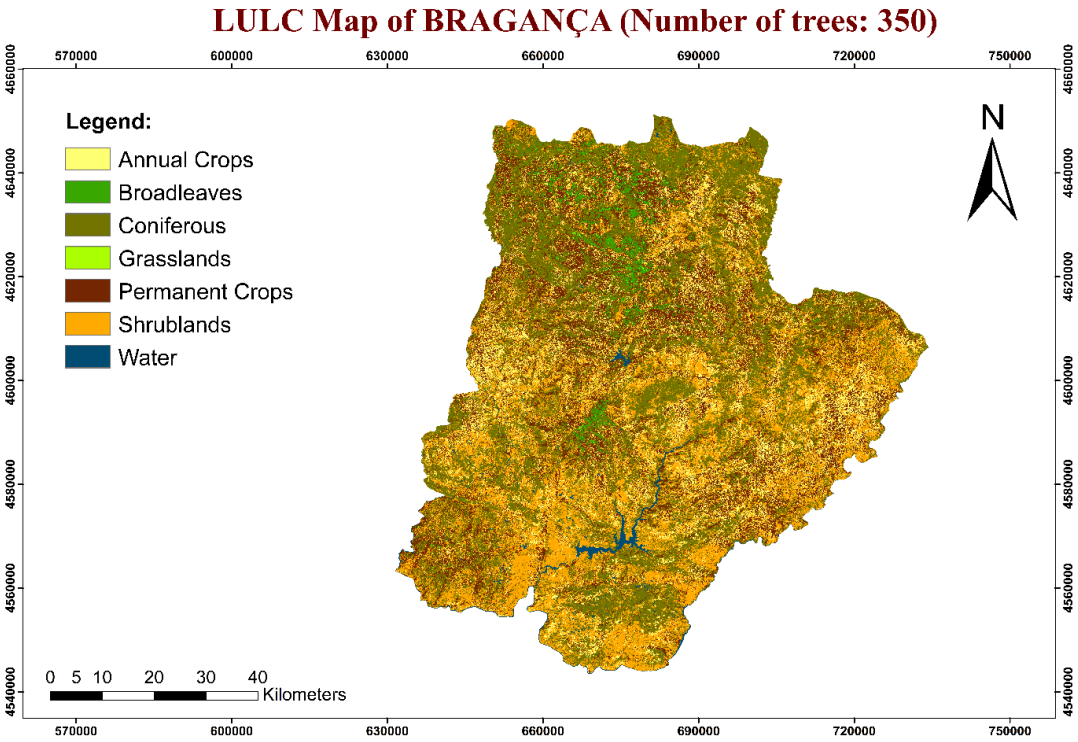
Annex XXVII: LULC classification of Bragança with nb-trees=250



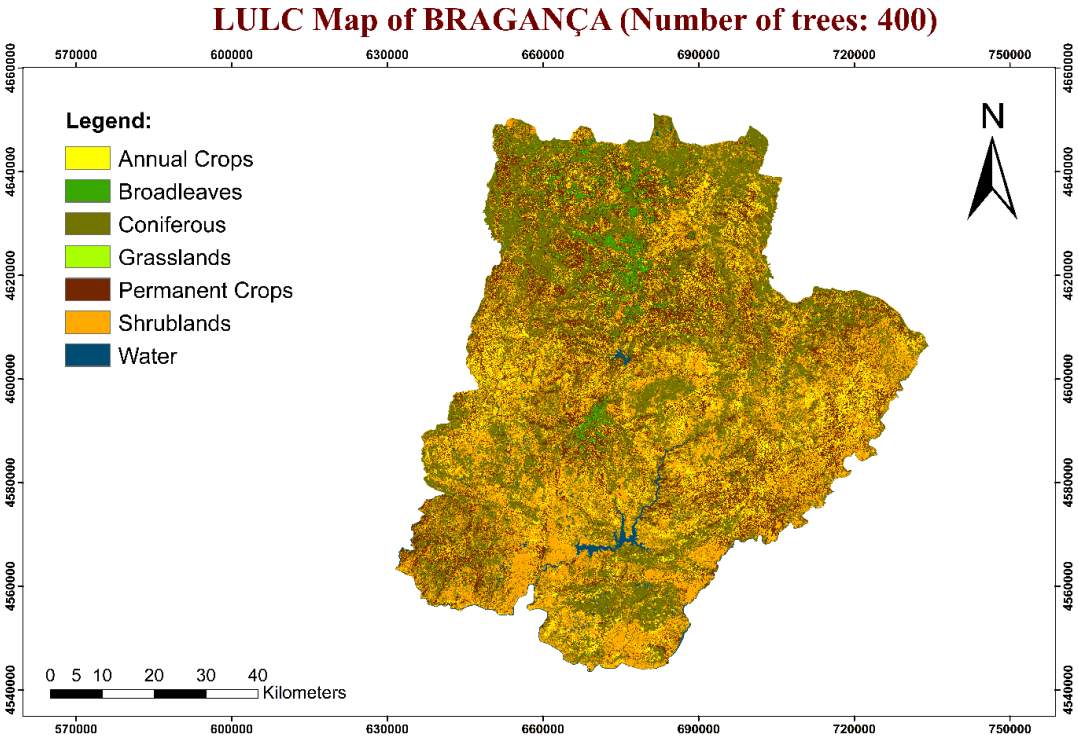
Annex XXVIII: LULC classification of Bragança with nb-trees=300



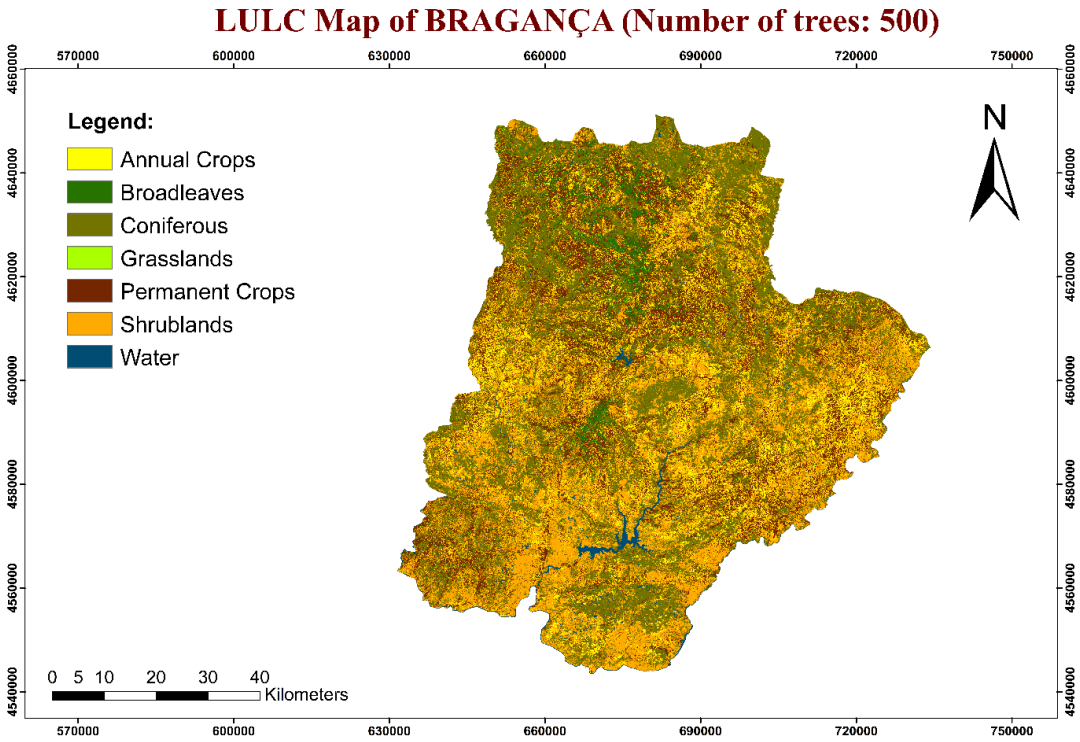
Annex XXIX:LULC classification of Bragança with nb-trees=350



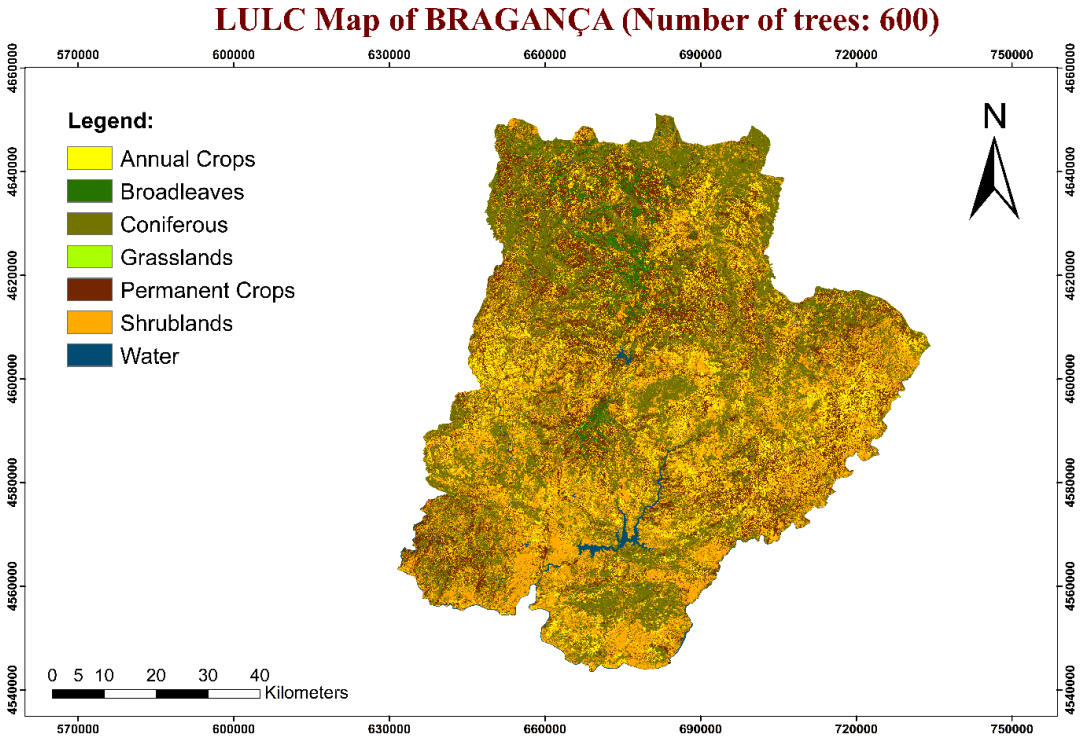
Annex XXX:LULC classification of Bragança with nb-trees=400



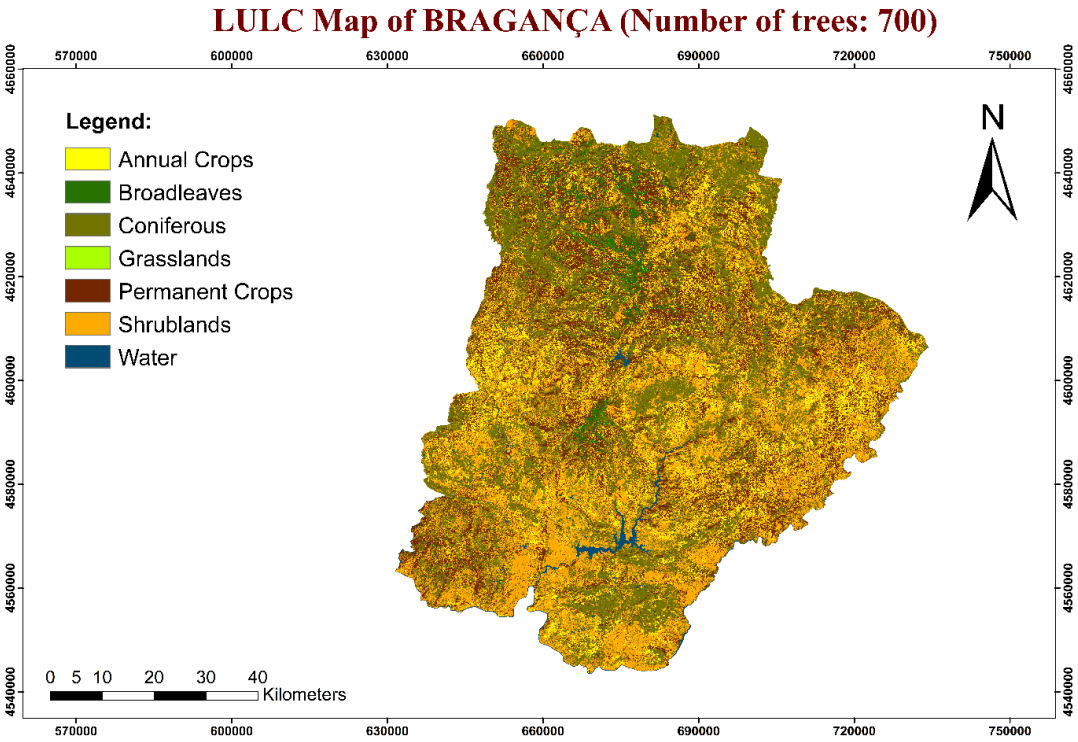
Annex XXXI:LULC classification of Bragança with nb-trees=500



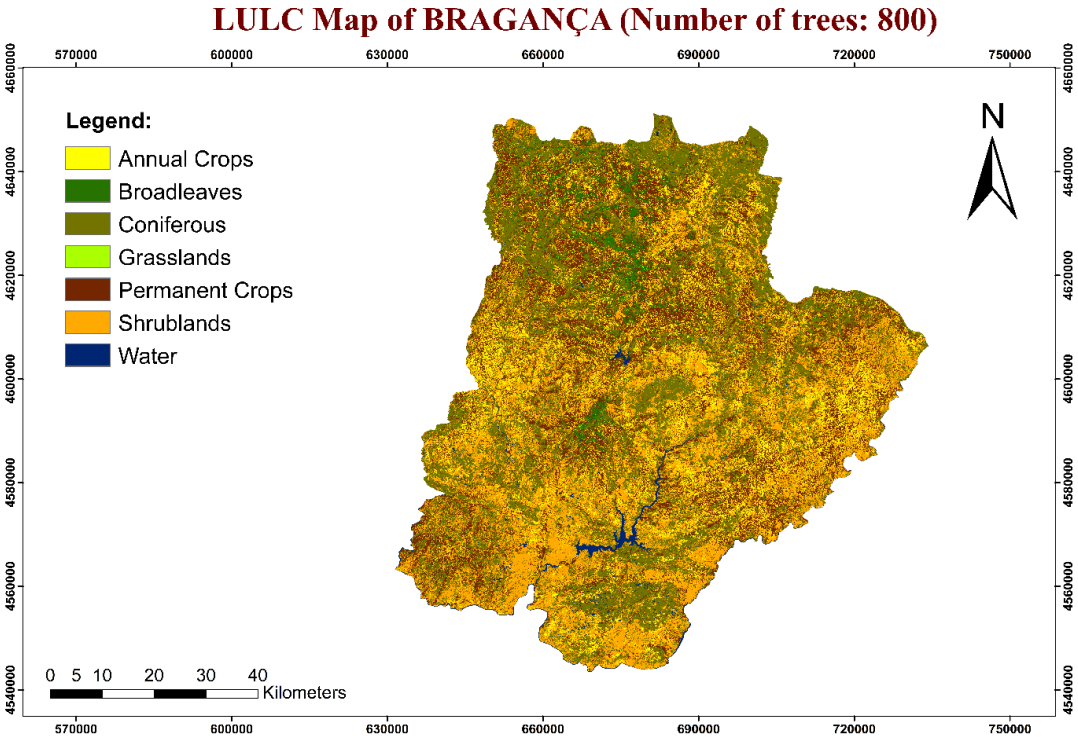
Annex XXXII:LULC classification of Bragança with nb-trees=600



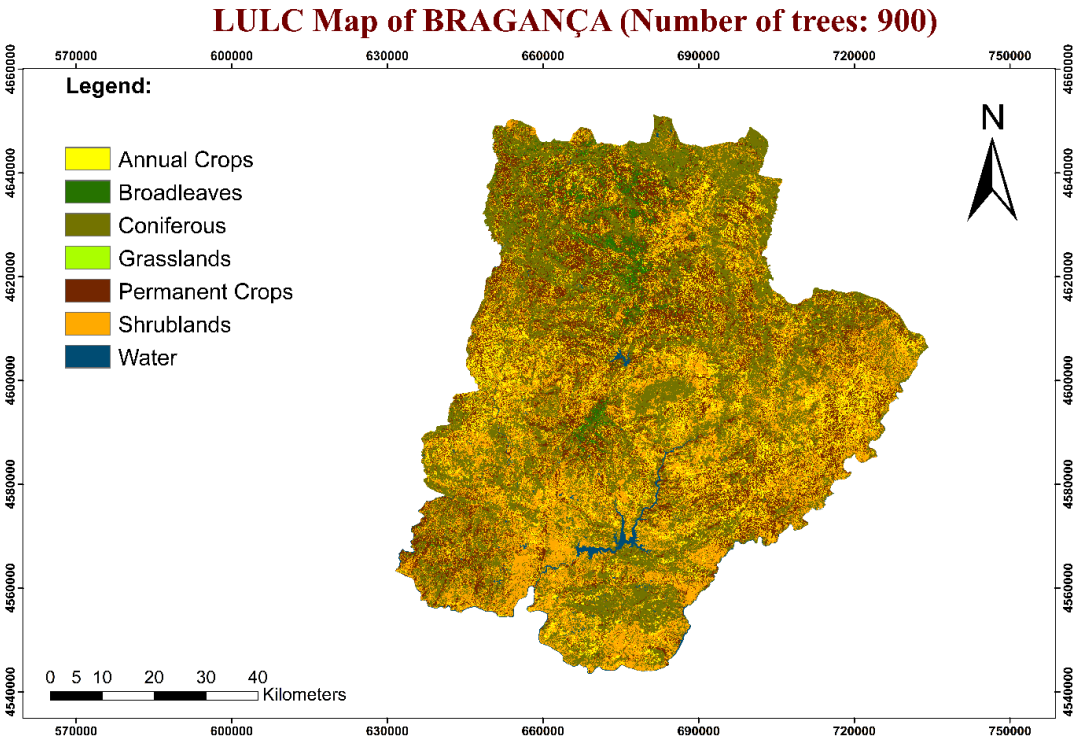
Annex XXXIII:LULC classification of Bragança with nb-trees=700



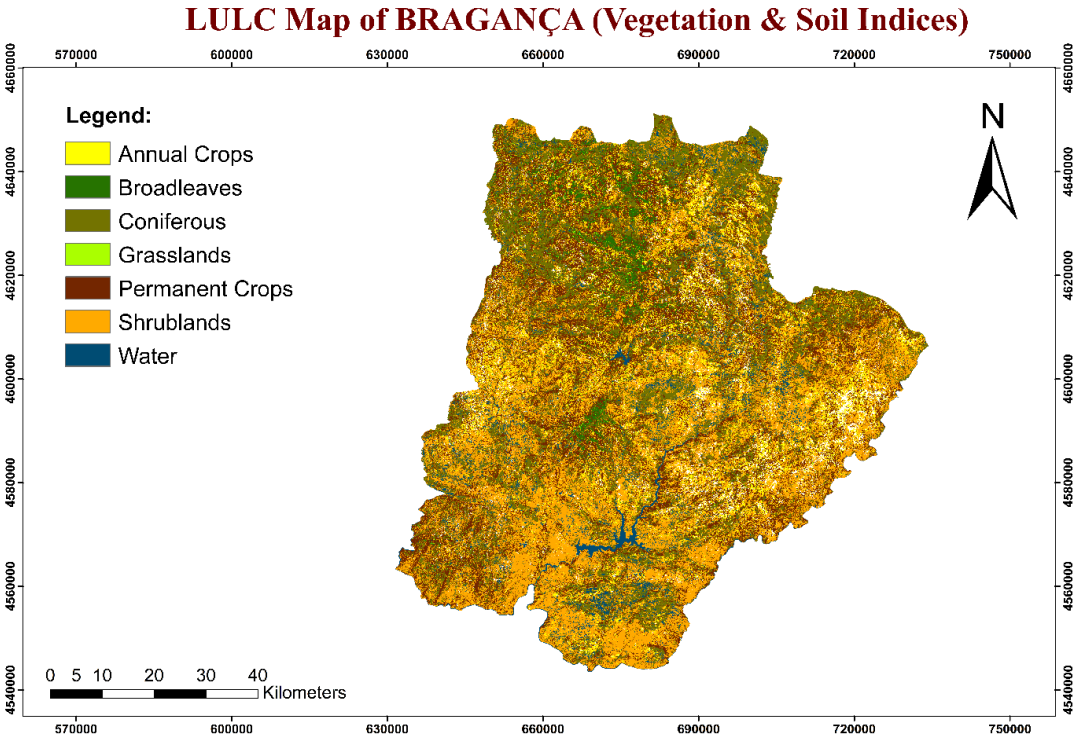
Annex XXXIV:LULC classification of Bragança with nb-trees=800



Annex XXXV:LULC classification of Bragança with nb-trees=900



Annex XXXVI: LULC classification of Bragança with vegetation& Soil Indices



Annex XXXVII: LULC classification of Bragança with Vegetation Indices