

POLITECNICO DI MILANO

Geoinformatics Engineering Master of Science

School of Civil Environmental and Land Management Engineering

School of Industrial and Information Engineering



Multi-scale analysis and comparison of Burned Area products derived from Earth Observation (EO) data over the African continent

Supervisor: Prof.ssa Giovanna Sona

Co-Supervisors: Dott.ssa Daniela Stroppiana

Dott. Sali Matteo

Thesis by:

Martina Pasturensi

Student ID: 946886

Academic Year 2020-2021

Multi-scale analysis and comparison of Burned Area products derived from Earth Observation (EO) data over the African continent

Table of contents

Table of contents	3
Index of Figures	5
Index of Tables.....	10
Abstract	11
Abstract (italian version)	12
Introduction.....	14
1 Remote Sensing of Fires	16
1.1 Fires and Climate Change	16
1.2 Remote Sensing	18
1.3 Remote Sensing for Fire Monitoring.....	23
2 Fire CCI Project	25
3 Objectives	27
4 Study areas	28
5 Input Dataset and Products.....	30
5.1 Planet	30
5.2 Sentinel-2 Multispectral Instrument	34
5.3 FireCCI51 BA product	48
5.4 Active Fires.....	49
6 Burned area mapping	55
6.1 Introduction to Google Earth Engine	55
6.2 Algorithm Modules.....	60
6.2.1 Selection of Sentinel-2 time series.....	61
6.2.2 Time Series and RGB Download	69
6.2.3 Classification (Random Forest Algorithm)	71
6.3 Classification outputs	83
7 Validation.....	88
7.1 Validation Methods	89
7.1.1 Data Preparation	92
7.1.2 Comparison and accuracy metrics computation	95
7.1.3 Computation of the Pareto boundary.....	100

Multi-scale analysis and comparison of Burned Area products derived from Earth Observation (EO) data over the African continent

7.2 Results and sources of error	104
7.3 Comparison between BA maps from S2 and the FireCCI product.....	124
7.4 Pareto Boundary	126
Discussion and Conclusion	129
Bibliography.....	132
Sitography.....	135

Index of Figures

Figure 1. 1 Fire Triangle	16
Figure 1. 2 Electromagnetic Spectrum	19
Figure 1. 3 Reflection, absorption and transmission of a radiation	21
Figure 1. 4 Spectral Signatures for different land cover types.....	22
Figure 4. 1 The location of the Areas Of Interest (AOIs) over Africa. In the zoom panels the area covered by the Planet mosaics (red polygons) and the S2 tile area that intersect.....	28
Figure 5. 1 Example Planet mosaic images over Site 1A, Senegal, displayed as RGB true colour composites for 14/01/2019 and 19/03/2019 first and last date of the site’s time series, respectively.	33
Figure 5. 2 PlanetScope output layers of the Planet Classification done by CNR-IRE for the short time unit.....	34
Figure 5. 3 The Copernicus Sentinel mission (https://www.esa.int/Applications/Observing_the_Earth/Copernicus/The_Sentinel_missions)	36
Figure 5. 4 Schematic View of the Deployed SENTINEL-2 Spacecraft (image credit: EADS Astrium)	37
Figure 5. 5The Twin-Satellite SENTINEL-2 Orbital Configuration	38
Figure 5. 6Coverage and revisit time foreseen for Sentinel-2 MSI acquisitions	39
Figure 5. 7 TOA Level-1C image data (left) and associated Level-2A BOA image data (right) generated using Sen2Cor processor.....	41
Figure 5. 8 Spectral response functions of Sentinel-2A MSI bands.....	43
Figure 5. 9 The S'' RGB and SCL layers for one example date for each AOI for the six tiles covering the regions: a)28PFA, b) 28PGV, c) 33MXQ-33MXR, d)36KWE, e)36KYE.....	46
Figure 5. 10 Acquisition dates for S2 (red) and Planet time series (blue).....	48

Figure 5. 11 Plots of FIRECCI51 for Sentinel study areas: a) 28PFA, b) 28PGV c) 33MXQ, d) 33MXR, e) 36KYE, f) 36KWD	49
Figure 5. 12 Ground observation MODIS (https://earthdata.nasa.gov/faq/firms-faq) .	53
Figure 5. 13 Ground observation VIIRS (https://earthdata.nasa.gov/faq/firms-faq)	53
Figure 5. 14 Example of active fire points for the AOIs 1A and 1B in Northern Africa. Colors show the date of detection of Afs between January 17 th and March 18 th , 2019. The blue square is the size of the S2 tiles covering the regions and the red rectangles is the a.....	54
Figure 5. 15 Seasonality of fires as observed in the Afs datasets for the five AOIs in Africa and the year 2019. Active fire points for each region are subset between the first and last date of the temporal series of satellite images.	54
Figure 6. 1 Screenshot of Explorer Interface on the GEE platform	56
Figure 6. 2 Screenshot of Code Editor Interface on GEE platform	58
Figure 6. 3 Flowchart of the Algorithm for burned areas mapping	61
Figure 6. 4 Screenshot of Task Manager Interface on GEE platform	64
Figure 6. 5 Example of time series of s2 images over tile 36KYE of the three consecutive S2 image	69
Figure 6. 6 Screenshot of Asset details for 36KYE	70
Figure 6. 7 Screenshot of the imported table for 36KYE.....	70
Figure 6. 8 Decision tree for classification problem	72
Figure 6. 9 Random Forest.....	72
Figure 6. 10 Difference between pre-fire image and post-fire image for the tile 36KYE	78
Figure 6. 11 Training polygons on difference layer for the tile 36KYE	79
Figure 6. 12 Screenshot of the two geometries imported as variables	80
Figure 6. 13 Example of an inaccurate first output	81
Figure 6. 14 Example of output of classification after manual correction	82
Figure 6. 15 Zoom on the top left part of the previous figure	82

Figure 6. 16 a)Time series of S2 images over a sample tile (35LMD, Zambia, Tropical savanna High); b) fire polygon derived with RF classification for each S2 image pair; c) synthetic fire reference perimeter products over the series length.....	83
Figure 6. 17 Cloud Mask layer for the tile 28PVG 17 Jan 2019 - 27 Jan 2019.....	84
Figure 6. 18 Cloud Mask Button on Console Interface.....	85
Figure 6. 19 Training Polygons vector layer and region for 28PGV 17 Jan 2019 - 27 Jan 2019	87
Figure 7. 1 Flowchart of Validation phase	91
Figure 7. 2 Attribute table for Planet BA ng unit product, site 2C as example	92
Figure 7. 3 Planet BA product after Data Preparation phase.....	93
Figure 7. 4 Attribute table for Sentinel BA product long unit, 2C	94
Figure 7. 5Data preparation results from short unit to clipped long unit for the site 1A	95
Figure 7. 6 Plotted chart on GEE Code Editor Console. Reflectance trends as function of dates for each band for 28PGV tile, agreement sample area	99
Figure 7. 7 Spectral Signature for 28PGV derived for a region of Agreement (both S2 and Planet identified the area as burned). Spectral signature as a function of the wavelength of S2 bands (x-axis) is displayed for different dates (colors).....	100
Figure 7. 8 Step for the Pareto Boundary Computing. a)Sentinel Grid and overlapping Planet BA product as input; b) intersection between cells and BA in terms of percentage; c) Evaluating omission and commission for different threshold; d) Pareto Boundary Plot.....	102
Figure 7. 9 Example of the first 18 rows of DataFrame output of the intersection.....	104
Figure 7. 10 Map of agreement over site 1A showing the spatial distribution of correctly classified burned areas (green) and omission (red) and commission (blue) errors.	106
Figure 7. 11 Zoom on Omission Error due to difference in date of acquisition of the first image in the time series, site 1A (left) and Active Fires (right).	107

Figure 7. 12 Zoom on Omission Error, site 1A (left) and the active fires detected over the same area (right)	108
Figure 7. 13 Zoom on the boundary on classification polygons, site 1A.....	109
Figure 7. 14 Map of agreement over site 1B showing the spatial distribution of correctly classified burned areas (green) and omission (red) and commission (blue) errors.	110
Figure 7. 15 Zoom on Commission Error, site 1B.....	111
Figure 7. 16 Spectral Signature for the Commission Error zone, site 1B in the s2 wavebands.....	112
Figure 7. 17 site 1B a) region of commission error (blue) over the RGB false color composite of the Planet images acquired on January 17 th ; b) RGB false color composite of the same image showing on going fire; c) RGB false color composite of the S2 images on the sam.....	113
Figure 7. 18 Map of agreement over site 2C showing the spatial distribution of correctly classified burned areas (green) and omission (red) and commission (blue) errors.	114
Figure 7. 19 The three sample regions in site 2C selected for the extraction of the spectral signature over unburned (black line), burned (blue) and error (yellow) regions	115
Figure 7. 20 Spectral Signature for the three sampled regions shown in the figure above for Omission Error, site 1C 27/07/2019.....	115
Figure 7. 21 Map of agreement over site 3A showing the spatial distribution of correctly classified burned areas (green) and omission (red) and commission (blue) errors.	117
Figure 7. 22 Zoom on omission error, site 3A for Planet RGB (left) and Sentinel RGB (right)	118
Figure 7. 23 Zoom on commission error, site 3A for Planet RGB (left) and Sentinel RGB (right)	119
Figure 7. 24 Spectral Signature for Omission Error, site 3A 29/08/2019	120
Figure 7. 25 Zoom on thematic map, site 3A	120

Figure 7. 26 Comparison's errors, site 3B.....	122
Figure 7. 27 Zoom on omission error area, site 3B	123
Figure 7. 28 a) RGB on 27/08/2019; b) RGB on 11/09/2019, both for a zoom area of the site 3B	124
Figure 7. 29 Histogram for BA products (S2 products are blue and FireCCI51 products are orange)	125
Figure 7. 30 The results of the coparison between BA maps from S2 and FireCCI51 for the six S2 tiles of the AOIs:28PFA (a), 28PGV (b), 33MXQ (c), 33MXE (d), 36KWE (e) and 36KYE (f). Colors represent agreement and disagreement areas.	126
Figure 7. 31 Pareto Boundary for all the five sites	127
Figure 7. 32 Pareto Boundary with MODIS grid as input	128

Index of Tables

Table 4. 1 Site Biome Location Coordinates (Lat-Lon) of the Central Point S2 tile.....	29
Table 5. 1 Planet Products characteristics.....	31
Table 5. 2 Information about the five sites chosen for Planescope dataset.....	33
Table 5. 3 Sentinel-2 product types.....	40
Table 5. 4 Sentinel-2 L2A Characteristics	43
Table 5. 5 Bands in Sentinel-2 MSI L2A	45
Table 5. 6 Scene Classification Classes	45
Table 5. 7 Sentinel-2 tiles used for this thesis	48
Table 6. 1 JavaScript code for selection of satellites data.....	62
Table 6. 2 Example of two first attributes of .csv output related to tile 36KYE where 'system:index' is the image ID and 'CLOUDY_PIXEL_PERCENTAGE' is the percentage of image covered by clouds according to metadata information.	66
Table 6. 3 Differences between Decision Tree and Random Forest Algorithm	73
Table 7. 1 Areas' values and Percentage of errors and similarity, site 1A	105
Table 7. 2 Areas' values and Percentage of errors and similarity, site 1B	109
Table 7. 3 Areas' values and Percentage of errors and similarity, site 2C	113
Table 7. 4 Areas' values and Percentage of errors and similarity, site 3A	116
Table 7. 5 Areas' values and Percentage of errors and similarity, site 3B	121
Table 7. 6 Metrics about comparison between FIRECCI51 and Sentinel-2 BA products	125
Table 7. 7 Results of Validation phase for all five sites	Errore. Il segnalibro non è definito.

Abstract

This work focuses on the comparison of multi-source and multi-scale Burned Area (BA) products derived from Earth Observation (EO) data and, in particular, from satellite images provided by the Copernicus missions. The analyses and comparisons between different BA products have been carried out over Area Of Interest (AOIs) for the African continent and the year 2019.

A supervised algorithm for Burned Area (BA) classification based on Random Forest (RF) was implemented and applied to the Sentinel-2 data; the aim is to map burned areas from the time series of each AOIs. The result of this operation, carried out in the Google Earth Engine working environment, are layers containing the perimeters of the burned areas (BA) over the S2 tiles (100 km x 100 km) overlapping the AOIs.

Then, BA products are validated in QGIS: BA maps from Sentinel-2, are compared with the burned area perimeters derived from Planet data (which have a better resolution and are considered as ground truth). Omission and commission error metrics and the Dice coefficient were calculated. Error sources were investigated and analysed.

Furthermore, an analysis of the effect of the resolution of the input images on the accuracy of BA maps was carried out through the analysis of the Pareto Boundary that was carried out in R environment.

Sentinel-2 fire perimeters were also compared with BA maps from the FireCCI51 products, to complete the multiscale comparison. The FireCCI51 has lower spatial resolution (250m) but it delivers multi annual global BA maps. Lower accuracy mapping is therefore expected compared to higher resolution S2 fire perimeters.

Results show that a satisfactory accuracy is achieved at each site: Sentinel-2 BA maps have omission error in the range [1,60-20,97 %] and commission error in the range [5,93-19,85 %]. Greater accuracy is obtained over larger burned areas patches and with regular boundaries; in these cases, classification errors due to different resolution of the input image data are less influential. The supervised RF algorithm however provides more accurate than an automatic algorithm since it is trained over each site.

Abstract (italian version)

Il lavoro di questa tesi si concentra sul confronto di prodotti multi-sorgente e multi-scala di aree bruciate (BA) derivati da dati di Osservazione della Terra (EO) e, in particolare, da immagini satellitari fornite dalle missioni Copernicus. Le analisi e i confronti tra i diversi prodotti BA sono stati effettuati su aree di interesse (Areas Of Interest, AOIs) in Africa e per l'anno 2019.

Un algoritmo supervisionato per la classificazione delle aree bruciate (BA) basato su Random Forest (RF) è stato implementato e applicato ai dati Sentinel-2 per mappare le aree bruciate delle serie temporali di ogni AOIs. Il risultato della classificazione, effettuata nell'ambiente di lavoro Google Earth Engine, sono i perimetri delle aree bruciate (BA) sui tile (100kmx100km) S2 che si sovrappongono alle AOI.

Successivamente, i prodotti BA sono validati in QGIS: I risultati della classificazione di Sentinel-2 sono stati confrontati con i perimetri delle aree bruciate derivati dai dati Planet (che hanno una migliore risoluzione e sono considerati come verità a terra). Sono state calcolate le metriche degli errori di omissione e commissione e il coefficiente di Dice, studiando e analizzando le fonti di errore.

Inoltre, è stata effettuata un'analisi dell'effetto della risoluzione delle immagini di input sull'accuratezza delle mappe BA attraverso l'analisi ed il calcolo della frontiera di Pareto (Pareto Boundary) in ambiente R Studio.

I perimetri degli incendi Sentinel-2 sono poi stati confrontati con le mappe BA dei prodotti FireCCI51, per completare il confronto multi-scala. Il dato FireCCI51 fornisce mappe BA globali e pluriennali (2001-2020) con risoluzione di 250 m. Ci si aspetta, quindi, una mappatura di minore accuratezza rispetto ai perimetri di incendio S2 che ha una più alta risoluzione (10/20 m).

I risultati mostrano che è stata raggiunta un'accuratezza sufficiente in ogni sito: le mappe BA ottenute da Sentinel-2 hanno errori di omissione e commissione, rispettivamente, nel range [1,60-20,97 %] e [5,93-19,85 %]. L'accuratezza maggiore è stata osservata per aree bruciate più grandi e con confini regolari; in questi casi, l'errore di classificazione dovuto alla diversa risoluzione dei dati dell'immagine di input

è meno influente. L'algorithmo supervisionato RF fornisce comunque un'accuratezza maggiore di un algorithmo automatico, poiché è stato addestrato su ogni sito.

Introduction

The physical phenomenon studied in this thesis is wildfires.

It is of extreme importance to analyse study and look for a solution to deal with these phenomena because they have a great impact on the environment and the ecosystems. The major consequences of fires are in terms of increased concentrations of combustion gases in the atmosphere, consequences in the evolution and biodiversity of territories especially for flora and fauna both on a small and large scale.

Biomass burning is widely recognized as one of the critical factors affecting atmospheric chemistry, as a significant share of aerosols and greenhouse gas emissions are produced from burning (Knorr et al., 2016; van der Werf et al., 2010).

The causes of forest fires in most cases are of anthropic reasons, but the temporal increase in the number of fires and their severity is urged by climate change. Climate change has altered and greatly modifies the conditions of air temperature and drought, altering the natural physical-chemical balances of the territories.

Five study areas are considered for this thesis, all on the African continent. NASA's Earth Observatory (<https://earthobservatory.nasa.gov/>) defines Africa as the "continent of fires" (the instrument, with which the satellites of the Aqua and Terra agency are equipped, detects, on an average day in August, 100,000 active fires throughout the planet; of these, 70% are in Africa). Unlike other places in the world where natural phenomena and deforestation have the greatest influence on the fire season, this is not the case in Africa. In addition, fire data on the African continent is consistent from one year to the next. In fact, according to Global Forest Watch Fire data, the continent has been affected by 250,000 fires in the past two years.

When analysing the African continent, fires seem to affect mainly Savannah (therefore a predominantly grassy biome) and fields cultivated by small farmers. In sub-Saharan Africa, farmers and breeders use to light fires before the rainy season because regular fires allow for renewal of vegetation. The ash deposited then provides a nutrient-rich layer for subsequent crops and animal grazing.

Despite the positive effects of fires on the ecosystem, it should be noted that smoke from fires contributes to air pollution. Fires are studied through the technique of

Remote Sensing, which allows us to observe the Earth from satellite images to extrapolate useful information. The resolution of current satellites has made it possible to record even small fires that could not be seen with previous instruments (Roteta et al., 2019), and whose impact in terms of emissions must be taken into account in climate models.

The purpose of this thesis is to compare burned areas derived from the classification of satellite imagery at different scales and with a difference in spatial resolution.

The innovative contribution of this work is:

- The use of new generation of Copernicus satellite sensors for EO providing high spatial, spectral, and temporal resolution; these satellites are useful to support the mapping of burned areas at detailed scales, such as the Multispectral Instrument (MSI) sensor aboard Sentinel-2 (S2) satellites, which is also equipped with spectral bands in the red-edge domain (Smiraglia, D.; Filipponi, F.; Mandrone, S.; Tornato, A.; Taramelli, A. 2020);
- The application of machine learning algorithm to Earth Observation: a supervised classification algorithm (Random Forest) was implemented for the classification of burned areas, working in a cloud environment on large amounts of data;
- Critical analysis using different tools not only to quantify the error in the mapping of Burned Area, but also to study the source of errors;
- Study of product accuracy in relation to spatial resolution (Pareto Boundary);
- Comparison with BA maps delivered by the operation FireCCI51 global product.

1 Remote Sensing of Fires

1.1 Fires and Climate Change

Fire is an uncontrolled oxidative reaction that develops without limitation in space and time. A fire can be triggered by natural or man-made causes and occurs if three basic elements are present: flammable material, the oxidiser (oxygen) and heat. These three elements are combined in the so-called “Fire Triangle” [Figure 1. 1].

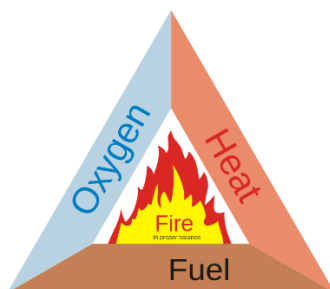


Figure 1. 1 Fire Triangle

Once a fire is started there are other factors that influence its intensity and determine how it spreads:

- *Weather conditions: temperature, wind and drought.*

The temperature affects the intensity of the fire because heat is one of the three pillars for starting it, therefore, a high temperature will be a more favourable condition for the spread of the fire.

Wind and drought are elements that favour the spread of the fire, because the wind provides a lot of oxygen (which like heat is one of the three elements necessary for a fire to start) and feeds it in an unpredictable and uncontrollable way. A study by Dr. Terry Clark, a scientist at the National Center for Atmospheric Research (NCAR), states that not only do winds feed the fire, but the fire itself generates its own winds, which are up to 10 times faster than the ambient wind (Bonsor, 2001).

- *Quantity of vegetation and humidity conditions.*

The risk of fire varies with the fuel dimension and humidity level. When there is a lot of flammable vegetation, but with a high-water content, the fire will be of low intensity, vice versa if I have little water content, the fire intensity will be higher.

The maximum risk condition is when there is a high amount of fuel with little water content.

- *Topography*

Terrain characteristics play an important role, as fires travel much faster uphill than downhill.

The causes of fires can be of various kinds, anthropological or otherwise, but the consequences on the territory could have disastrous impacts. In addition, the evolution of fires is often not controllable due to the strong uncertainty of the influencing factors listed above, so the risk factor increases. The effects can affect people, infrastructure and the surrounding environment; in addition to the damage that fires cause as they burned, other effects may occur many months after the fire is out, such as the impact on air quality and climate change.

One of the main effects is the gases that biomass combustion emits into the atmosphere; it includes human-induced fires due to deforestation and land-use change, as well as natural fires induced by lightning. Scientists estimate that humans are responsible for about 90 per cent of biomass fires.

The burning of vegetation releases large quantities of volatile organic compounds (VOCs)¹ and gases, including greenhouse gases such as CO₂, CH₄ and N₂O and chemically active gases such as NO, CO. Greenhouse gases can lead to increased global warming or man-made global climate change. Vegetation acts as a natural storehouse of carbon dioxide, storing it over time through the process of photosynthesis. When combustion occurs, it can release hundreds of years of stored carbon dioxide into the atmosphere in a matter of hours.

In addition, particulate matter produced by burning biomass has an impact on the climate and can also affect human health when inhaled, causing respiratory problems.

Biomass burning is estimated to be the world's leading source of some air pollutants such as black carbon (BC) and primary organic aerosols (POA).

¹ “Volatile organic compounds (VOC) means any compound of carbon, excluding carbon monoxide, carbon dioxide, carbonic acid, metallic carbides or carbonates and ammonium carbonate, which participates in atmospheric photochemical reactions” definition by <https://www.epa.gov/>

About 59% of global BC emissions and 85% of POA emissions are related to biomass burning. Forest fires alone account for about one third to one half of total global carbon monoxide (CO) emissions and 20% of nitrogen oxides (NO_x) emissions.

Simulation models have also shown that in the absence of forest fires, atmospheric concentrations of CO₂ would be around 40 ppm lower, underlining the importance of forest fire emissions in the atmospheric carbon budget (Meinrat, 2019).

So far only the negative effects of fires have been listed, but there are also some positive ones. For some types of trees, fires are useful for getting rid of infesting species, for others they are even necessary for reproduction (this is the case, for example, of *Pinus contorta*, tree species that grows in the western United States: the scales of its cones are practically sealed by a large quantity of resin that only temperatures around 65°C can liquefy, freeing the seeds). Another example is the fires in the African Savannas, an area analysed for this thesis: fires are frequent and part of the plant growth cycle. Since the plants in question are herbaceous, they burn quickly, so the flames spread quickly without burning the soil or the roots; once the fires have passed, the plants grow back very quickly.

Therefore, it can be summed up to say that fires have a great ecosystem complexity, which is why there are many projects and research all over the world to monitor and study them. In this frame, this thesis is developed on the classification of burned areas and subsequent validation of the results using Remote Sensing.

1.2 Remote Sensing

Remote sensing, or the observation of the earth from space, is the methodology of obtaining information about an object through the analysis of data acquired by an instrument (in this case, a sensor mounted on a carrier) that is not in direct contact with the object under investigation. The sensors record the different ways in which electromagnetic energy is reflected by objects on the surface (Lillesand and Kiefer, 1999).

Physical introduction to Remote Sensing

One of the different forms in which electromagnetic energy occurs is light.

It is described as a form of energy travelling sinusoidally at the speed of light in the form of a wave. Every electromagnetic wave has a defined frequency and a wavelength associated with this frequency. Thanks to these characteristics waves can be ordered in a graph known as the electromagnetic spectrum [Figure 1. 2].

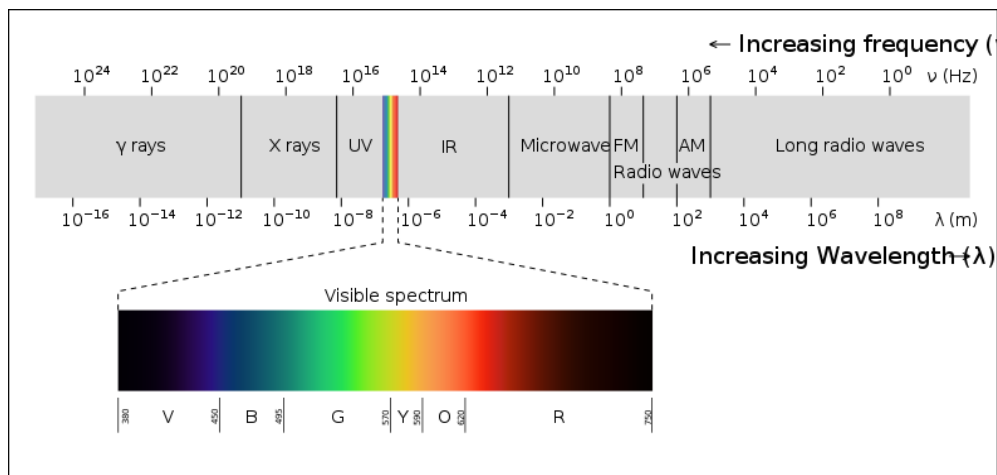


Figure 1. 2 Electromagnetic Spectrum

The electromagnetic spectrum has conventionally been divided into the following regions:

- Gamma rays (<0.03 nm) and X-rays (0.03 -300 nm): short-wave radiation with a high energy content.
- Ultraviolet radiation (0.28-0.40 μm): this region is subdivided into the near ultraviolet (UV-A, 0.32-0.40 μm), the far infrared (UV-B, 0.28-0.32 μm) and extreme ultraviolet (UV-C, <0.28 μm).
- The visible region occupies the range between 0.4 and 0.7 μm. Visible light can be divided into three segments, the so-called primary colours: from 0.4 to 0.5 μm light reflects blue, 0.5 to 0.6 μm green and 0.6 to 0.7 μm red.
- The infrared (IR) region, between 0.7 and 100 μm, is divided into three sub-intervals: near IR (0.7-1.3 μm), mid IR (1.3-3 μm), far IR or thermal IR (3-100 μm).
- The microwave region ranges from 0.1 to 100 cm.
- The region of longer wavelengths above 100 cm corresponds to the radio band.

All electromagnetic radiation used for remote sensing must pass through the atmosphere. Particles and gases in the atmosphere can attenuate incoming light. This is mainly determined by atmospheric scattering and absorption.

Atmospheric scattering is an unpredictable diffusion of radiation through atmospheric particles (Lillesand and Kiefer, 1999); radiation can interact with atmospheric molecules that are much smaller than the wavelength of the incident sunlight. The result is a decrease in the spectral contrast of the remote sensing image.

Atmospheric absorption reduces the energy of the radiation by a greater amount than dispersion. The most important trappers of solar radiation are water vapour, carbon dioxide and ozone, which can vary considerably in urban areas. Consequently, humidity, altitude, urbanisation and anthropogenic factors in general are variables that affect atmospheric absorption and its effects on the image.

These gases absorb solar energy in specific wavelength bands: solar radiation may or may not pass through the atmospheric layer depending on the wavelength.

This parameter thus defines the relative opacity of the atmosphere to the wave, identifying transparency intervals known as atmospheric windows. In correspondence with these wavelengths the absorption of radiation is reduced or null. Obviously, all the sensors mounted on satellite vectors have been calibrated to record images at these wavelengths.

When an electromagnetic wave (total incident energy) hits any object or surface, three fundamental interactions are possible [Figure 1. 3]:

- *Reflection*: part of the radiation is actually reflected by the target at different angles (depending on both the incident angle and the surface roughness.
- *Absorption*: part of the radiation is absorbed through molecular or electronic interactions; it will be partially re-emitted (emittance), especially at longer wavelengths, so solar radiation contributes to the heating of bodies that provide a thermal response.

- *Transmission*: part of the radiation penetrates certain media, for example water.

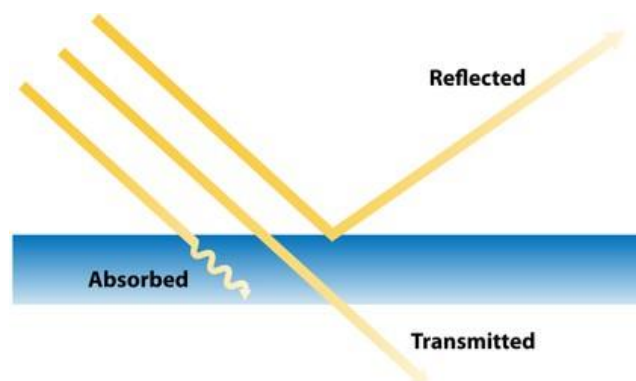


Figure 1. 3 Reflection, absorption and transmission of a radiation

This theory follows the principle of conservation of energy, that states:

$$E_r(\lambda) = E_i(\lambda) - E_a(\lambda) - E_t(\lambda)$$

Where:

E_r = Reflected energy (reflectance)

E_a = Absorbed energy

E_i = Total incident energy

E_t = Transmitted energy

(λ) = as a function of wavelength

The proportions of reflected, absorbed and transmitted energy vary depending on the conditions and the type of material affected. This allows the identification and separation of different substances in multispectral images, through the analysis of the spectrum of the reflected portion, that is described with the spectral signature. The spectral signature is a graph in which the reflectance of a surface is expressed as a function of wavelength; the conformation of spectral reflectance curves provides information on the spectral characteristics of a given object/surface.

Therefore, as a function of the spectral response one can distinguish in optical imagery numerous classes of features occupying an area, from pine forests to meadows, but also from clouds to burned areas. The figure below shows an example of spectral signatures for four typical surfaces [Figure 1. 4].

Multi-scale analysis and comparison of Burned Area products derived from Earth Observation (EO) data over the African continent

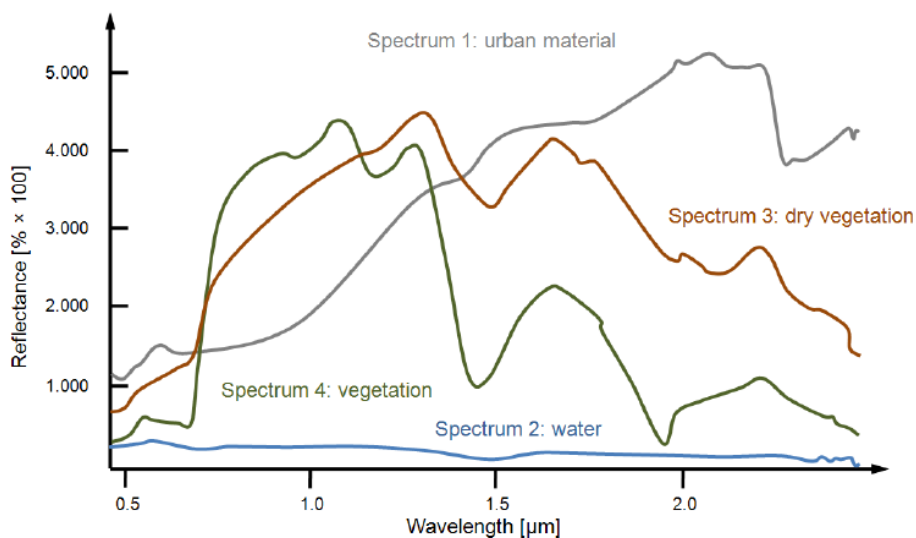


Figure 1. 4 Spectral Signatures for different land cover types

Instruments for the Earth Observation are called “Sensors”. Sensors can be on board different platforms (unmanned aerial vehicles-UAVs- aircrafts or satellites) to measure electromagnetic radiation at specific wavelength intervals (usually called bands). As a result, the measurements are quantized and converted into a digital image, where each image element (pixel) has a discrete value in Digital Number (DN) units (NASA, 2013). The resulting images have different characteristics (resolutions) depending on the sensor and platform:

1. The **temporal resolution** of a satellite in orbit is the revisit frequency of the satellite to a particular location with the same geometry of acquisition.
2. The **spatial resolution** of an instrument is the smallest area on the ground that can be imaged and it is mainly controlled by distance to target and the Instantaneous Field of View (IFOV).
3. The **radiometric resolution** of an instrument is a determination of the incremental level of intensity or reflectance that can be represented or distinguished by the system. The higher the radiometric resolution, the more capable the device will be of detecting differences in intensity or reflectance.

1.3 Remote Sensing for Fire Monitoring

Remote sensing is the only technology capable of monitoring fires on a global scale. These systems provide measurements of ground conditions before and after a fire. These measurements have been used to assist in fire risk mapping, fuel mapping, active fire detection, burned area estimation (BA mapping), fire severity assessment and vegetation recovery monitoring. (D. M. Szpakowski, J.R. Jensen, 2019).

The application of satellite imagery to BA mapping has a long history in remote sensing studies beginning in the early 1970s and 1980s. During this period of more than four decades, a wide range of techniques and algorithms have been developed and applied in BA mapping. Early BA products derived from satellite data were based on medium-resolution sensors, primarily the Landsat multispectral scanner (MSS) and, after 1982, TM imagery². Data from Landsat missions up to the latest Landsat-8 and the recently launched Landsat-9 constitute the longest medium-resolution archive for BA mapping and monitoring. Other medium-resolution sensors used in BA mapping have been the Haute Résolution Visible (HRV) sensor, aboard the Systeme Probatoire d'Observation de la Terre (SPOT) satellite since 1986 (Eastwood et al., 1998), the wide-field (WIFS) - linear imaging and self-scanning sensor (LISS), aboard the Indian Remote Sensing (IRS) satellite (Vázquez et al., 2001).

The most used remote sensing systems for fire research are space-based multispectral sensors, which divide the Earth at regular intervals, providing data on the amount of energy reflected in multiple regions of the electromagnetic spectrum (as seen in the previous section). In this thesis Sentinel-2 multispectral instruments (MSI) are used as input. The method used in this project is based on comparing images by differentiating post-fire and pre-fire images with spectral indices, observing the post-fire signal as well as its changes from the pre-fire reflectance. Spectral changes in Near Infrared (NIR) and Shortwave Infrared (SWIR) spectral regions are the foundation for most burned area mapping algorithms (Giglio et al., 2018, Van Der Werf et al., 2017, Zidane et al., 2019). Changes in canopy cover and brightness (due to scorching of leaves) are

² Landsat Thematic Mapper (TM) sensor was carried on Landsat 4 and Landsat 5, and created images consisting of six spectral bands with a spatial resolution of 30 meters for Bands 1-5 and 7, and one thermal band (Band 6) (<https://eos.com/find-satellite/landsat-5-tm>)

primarily observed in the NIR spectral region (Silva et al., 2004), whereas changes in landscape dryness is mostly observed in the SWIR spectral region (Trigg and Flasse, 2001). Other important spectral regions are the Red and Red-edge, related to strong absorption of the chlorophyll content in plants (Blackburn, 2007, Gitelson et al., 1996, Sims and Gamon, 2002).

The increasing availability of data combined with advanced analysis provided by Artificial Intelligence (AI) and Machine Learning (ML), together with the exceptional processing power of cloud computing, gives the possibility to map fires in fast, accessible and reliable way. Computer science, therefore, plays a fundamental role in these tasks.

A project that combines GIS, machine learning and remote sensing is the ESA FIRE_CCI in which this thesis work is embedded, as will be explained in detail in the following chapters.

2 Fire CCI Project

ESA's Climate Challenge Initiative (CCI) (<https://climate.esa.int/en/>) is a research programme, launched in 2008, that aims to combat global warming and reduce greenhouse gas emissions around the world using Earth Observation data.

Climate change is undoubtedly the greatest environmental challenge of the century. Increasing concentrations of carbon dioxide and other greenhouse gases in the atmosphere led to climate warming that reached approximately 1°C (above pre-industrial levels) in 2017, and it is increasing at around 0.2°C per decade (SPECIAL REPORT Global Warming of 1.5 °C, IPCC).

The consequences of global warming are evident and the United Nations Framework Convention on Climate Change (UNFCCC) is leading international efforts to combat climate change and limit global warming to 2 degrees Celsius above pre-industrial levels, as set out in the Paris Agreement. To achieve this goal and to make decisions on climate change mitigation and adaptation, the UNFCCC requires systematic observations of the climate system.

In response to this need for climate data from the UNFCCC and GCOS (Global Climate Observing System), the European Space Agency (ESA) has undertaken the CCI programme. The aim of the CCI is to realise the full potential of the long-term global EO archives that ESA, together with its Member States, has been establishing for the past 30 years as a significant and timely contribution to the databases required by the UNFCCC.

Within the CCI research programme of the European Space Agency, there are 26 Essential Climate Variable (ECV) projects: Aerosol, Biomass, Climate Modelling User Group (CMUG), Cloud, Fire, Greenhouse Gases (GHGs), Glaciers, High resolution land cover, Ice sheets (Antarctic), Ice sheets (green Land), Lakes, Land Cover, Land Surface Temperature, Ocean Colour, Ozone, Permafrost, RECCAP-2, Sea Ice, Sea Level, Sea level Budget Closure, Sea state, Sea Surface Salinity, Sea Surface Temperature, Snow, Soil Moisture and Water Vapour.

The FireCCI project aims at characterising spatial and temporal distribution of wildfires globally and regionally by processing long term archives of EO data with improved algorithms to deliver high quality products. Information on burned areas (BA) is important because approximately 25%-35% of greenhouse gases (GHG) are derived from the combustion of biomass and are therefore considered an important factor in climate change.

The Fire_CCI project developed algorithms for both pre-processing and detection of burned areas (BA), incorporating error characterization. The project team, led by the University of Alcalá (Spain), focuses on the following issues related to fire disturbance:

- Analysis and specification of climate-related scientific requirements
- Development and improvement of pre-processing and burned area algorithms
- Intercomparison and selection of burned area algorithms
- System prototyping and production of burned area datasets
- Product validation and evaluation

There are many partners who are set up to deal with specific tasks:

- Scientific Group for Earth Observation: University of Alcalá (UAH, Spain) – that is the project’s science leader-, Universidad Politécnica de Madrid (UPM, Spain), National Research Council of Italy - Institute for Electromagnetic Sensing of Environment (CNR - IREA, Italy);
- System engineers: Brockmann Consult GmbH (BC, Germany);
- Climate research group: Max Planck Institute for Meteorology (MPIM, Germany), National Centre for Scientific Research (CNRS, France);

and many other partners for different phases of the project.

A special focus should be made on the role of the National Research Council of Italy - Institute for Electromagnetic Sensing of Environment (CNR - IREA, Italy), where this thesis was carried out. Within the FireCCI project, CNR IREA is in charge of validation of BA products.

3 Objectives

The major objective of this thesis is to compare burned areas derived from Earth Observation (EO) data at different scales and characterized by different spatial resolution: Planet high resolution, Sentinel-2 medium resolution and MODIS low resolution data. BA maps from Planet and S2 are derived from data time series of consecutive image scenes that are classified with a supervised classification algorithm. In particular, Random Forest (RF) algorithms are trained with burned and unburned area polygons to identify areas that burned between two consecutive dates. Classification of image pairs are then combined in a synthetic Burned Area (BA) map by retaining the first date of detection. BA maps from S2 were produced during this thesis by implementing a RF algorithm in Google Earth Engine (GEE). BA maps from Planet were available at CNR IREA and derived by implementing the same type of algorithm in R open source software in the framework of the activities carried out in the ESA FireCCI project. However, Planet BA maps were derived in a similar way by applying a RF algorithm. In this comparison, BA derived from Planet images are assumed as ground truth, being the dataset derived from very high resolution imagery, and BA maps derived from S2 are compared to derive accuracy metrics. This validation provides accuracy of S2 BA maps produced within this thesis work. Moreover, S2 BA were compared to BA maps from the FireCCI product. The Pareto boundary has also been computed to investigate the impact of spatial resolution on the accuracy of the BA maps derived from S2 and MODIS (FireCCI51).

The comparison between BA maps was carried out in this thesis in QGIS and the computation of the Pareto boundary was done in R open software.

These activities were carried out for five selected sites distributed over the African continent for the year 2019. Sites were selected in different biomes, to study the behaviour of each individual ecoregion over time. The sites are described in the table [Table 4.1]. Results were analysed based on the characteristics of the sites in terms of fire regimes and land cover and the spectral characteristics of the burned areas as observed in the S2 wavebands. Ancillary geo-information was exploited to better depict fire seasonality (e.g., active fires).

4 Study areas

The analyses were carried out for five sites (Areas Of Interest, AOIs) selected over Africa as shown in figure below [Figure 4. 1]. The table [Table 4. 1] lists the characteristics of the sites, the location and the S2 tile(s) that cover the area of the Planet images. Two sites (1A, 1B) were chosen in Northern Africa in the semi-arid Sahelian savanna region (Senegal). In this region, a transitional belt between the humid savanna and the arid Sahara Desert, fires occur during the dry season (October to March); people set fires to manage farming and grazing lands. The second AOI (2C) is located in tropical central Africa (Democratic Republic of Congo, DRC); in this region the fire season typically occurs between May and August and they can affect humid tropical forests. In this region, persistent cloud cover can be an issue for remote sensing of fires since clouds can prevent from observing the surface conditions. The last two AOIs (3A, 3B) are located in southern Africa where fires generally affect savanna vegetation in the period June to September.

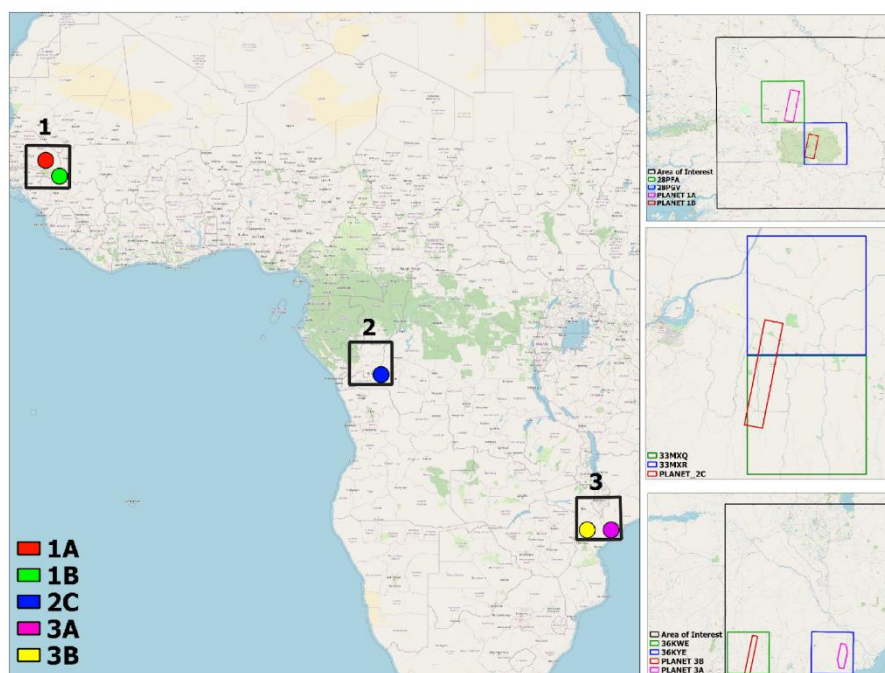


Figure 4. 1 The location of the Areas Of Interest (AOIs) over Africa. In the zoom panels the area covered by the Planet mosaics (red polygons) and the S2 tile area that intersect.

Multi-scale analysis and comparison of Burned Area products derived from Earth Observation (EO) data over the African continent

Site	Biome	Location	Coordinates (Lat-Lon) of the Central Point	S2 tile
1A	Tropical savanna	Tambacounda (Senegal)	13,8822 N 13,3713 E	28PFA
1B	Tropical savanna	Niokolo-Koba (Senegal)	13,0017 N 12,9552 E	28PGV
2C	Tropical savanna	Kinshasa (DRC)	-4,7076 N 16,0736 E	33MXQ 33MXR
3A	Temperate grassland	Capenga (Mozambique)	-18,6329 N 35,6433 E	36KYE
3B	Tropical savanna	Chimoio (Mozambique)	-18,6558 N 33,5820 E	36KWE 36KWD

Table 4. 1 Site Biome Location Coordinates (Lat-Lon) of the Central Point S2 tile

5 Input Dataset and Products

This chapter introduces the datasets and products used to carry out the classification and validation analyses. Input data for the analyses carried out in this thesis are satellite images acquired by the Sentinel-2 and Planet satellite missions, active fire products from the Moderate Resolution Imaging Spectroradiometer (MODIS) (MCD14DL) and Visible Infrared Imaging Radiometer Suite (VIIRS) 375 m (VNP14IMGTDL_NRT and VJ114IMGTDL_NRT) produced by NASA.

5.1 Planet

Planet Labs PBC is an American Public Earth imaging company based in San Francisco, California (Planet, 2017). The company designs and manufactures miniature Triple-CubeSat³ satellites called Doves that are delivered into orbit as a secondary payload on other rocket launch missions; each Dove is equipped with a high-powered telescope and a camera programmed to capture different swaths of the Earth.

Planet operates more than 200 satellites. The Planet constellation of satellites orbits the poles every 90 minutes, capturing more than 250 million square kilometres of satellite imagery every day with global coverage. Their goal is to visualise the entire Earth daily to monitor changes and discover relevant trends. Planet processes and delivers images quickly and efficiently to support decision making. The data pipeline ensures easy web and API access to Planet's images and archive. Images are collected and processed in a variety of formats to serve different use cases, whether it is mapping, deep learning, disaster response, precision agriculture or simple temporal analysis of images to create rich information products.

Planet products are multispectral images with colour correction and sensor calibration. Planet's multispectral products are PlanetScope (PS), RapidEye (RE), SkySat (SS). (These products are released under licence. <https://www.planet.com/products/>)

³ *CubeSats are a class of research spacecraft called nanosatellites. CubeSats are built to standard dimensions (Units or "U") of 10 cm x 10 cm x 10 cm (https://www.nasa.gov/mission_pages/cubesats/overview)*

The table below shows some information about them [Table 5. 1]

	PLANETSCOPE	RAPIDEYE	SKYSAT
Bands	4 (RGB, NIR)	5 (RGB, red edge NIR)	5 (RGB, NIR, pan)
Products	Color enhanced Visual Analytic	Color enhanced Visual Analytic	Visual Panchromatic Pansharpened multispectral Analytic
Pixel Resampled	3 m	5 m	Visual, panchromatic, pansharpened multispectral: 0.5m
Radiometric Resolution	Visual: 8 bit Analytic: 16 bit		Visual: 8 bit Analytic, panchromatic and pansharpened multispectral: 16 bit
Positional Accuracy	<10 m RMSE		
File Format	GeoTIF		

Table 5. 1 Planet Products characteristics

The Planet product that is used in this work is the PlanetScope. Unlike the Sentinel data, which were classified using a supervised classification algorithm on GEE, the Planet data were classified by CNR-IREA; the BA products were then used as references. The reference data will be used to validate the result of the classification by estimating errors of omission and commission (Chapter 7 of this thesis).

PlanetScope satellite constellation consists of more than 130 small satellites called Doves; each Dove satellite is a CubeSat made of three cubic units and thus measures only 10 cm x 10 cm x 30 cm.

PlanetScope satellite images are captured as a continuous strip of single-frame images, known as 'scenes' along a swath of about 25 Km; scenes can be composed in time-series so that changes over time can be seen [Figure 5. 1]. Scenes can be captured as a

single RGB (red, green, blue) frame or a split frame with one half RGB and one half NIR (near-infrared) depending on the satellite's capability. The data have been available since 2016; the constellation is constantly active and does not require an acquisition schedule.

The multi-spectral sensor mounted on this platform is characterized by four bands: three in the visible wavelengths (b1: 455-515 nm; b2: 500-590 nm; b3: 590-670 nm) and one in the NIR wavelengths (b4: 780-860 nm). The most recent third generation of Planet sensors (known as SuperDove or PSB.SD) is currently in orbit carrying a five spectral band multispectral sensor (Visible+NIR+Red Edge). The spatial resolution of Planet images is 3 m for all bands, which is particularly important to have an higher spatial resolution for the analysis of small burned areas (Lemajic et al., 2018).

Five sites were selected to be investigated with PlaneScope data; these sites are all on the African continent, but in different biomes, to study the behaviour of each individual burned areas over time. The sites are described in the table below [Table 5. 2].

Site	Area [km ²]	Acquisition date	N days
1A	1588.2	14/01/2019 21/01/2019 28/01/2019 25/02/2019 09/03/2019	55 days
1B	1245.5	17/01/2019 30/01/2019 06/03/2019 20/03/2019	63 days
2C	1367.3	13/06/2019 21/07/2019 27/08/2019	76 days

Multi-scale analysis and comparison of Burned Area products derived from Earth Observation (EO) data over the African continent

3A	974.7	05/09/2019 11/09/2019 29/09/2019	24 days
3B	1229.5	23/07/2019 08/08/2019 14/09/2019	54 days

Table 5. 2 Information about the five sites chosen for Planescope dataset

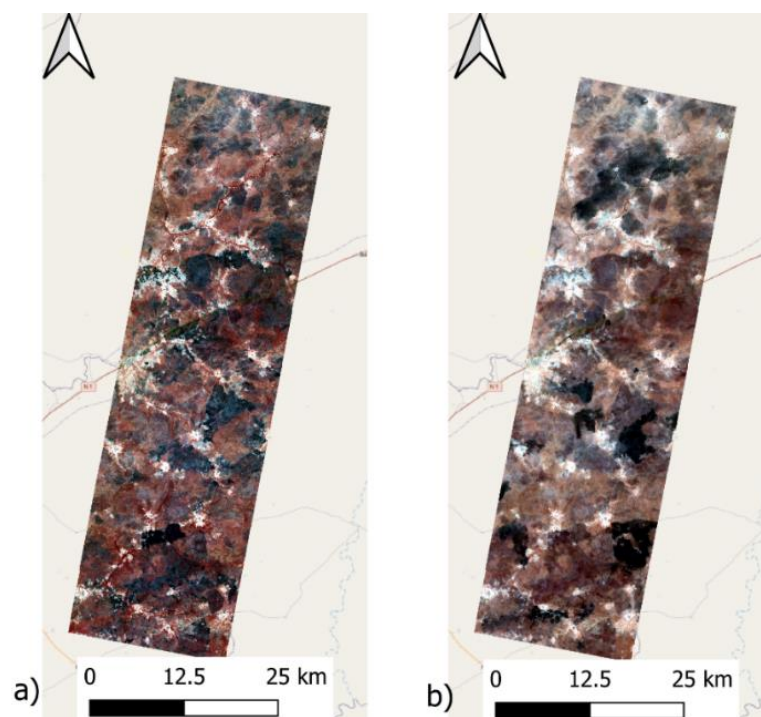


Figure 5. 1 Example Planet mosaic images over Site 1A, Senegal, displayed as RGB true colour composites for 14/01/2019 and 19/03/2019 first and last date of the site's time series, respectively.

The PlanetScope images were classified by CNR-IREA in order to identify burned areas. The classification is performed on consecutive image pairs belonging to the time series covering the predefined time interval (the information is contained in the table above). The classification produces one output, that is the map of the burned areas, per pair of images; the time interval covered by each result is defined as "short time unit". In the figure [Figure 5. 2] we see an image that contains all the layers in vector format of the classification results.

Layers referring to short time units should be merged into a single product (long unit) with the Merge tool (From the menu bar: *Vector -> Vector Management Tool -> Merge Vector Layer*). In QGIS, the data should be of the same datatype (i.e. all polygons) and each layer should have the same underlying projection

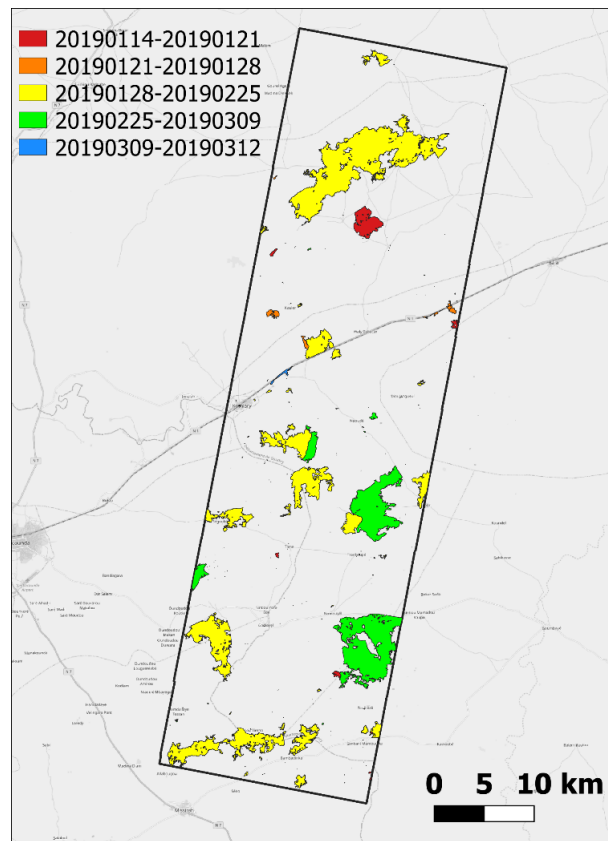


Figure 5. 2 PlanetScope output layers of the Planet Classification done by CNR-IRE for the short time unit

5.2 Sentinel-2 Multispectral Instrument

Sentinel-2 is an Earth Observation mission developed by the European Space Agency (ESA) as part of the Copernicus programme to monitor the planet's green areas and provide support in dealing with natural disasters.

Copernicus, formerly known as GMES (Global Monitoring for Environment and Security), is a complex satellite Earth observation programme launched in 1998 by the European Commission and a pool of space agencies. It is part, as a contribution of the "Old Continent", of the larger GEOSS project, which aims to develop a System of

Systems for Global Earth Observation, but its main objective is to guarantee Europe substantial independence in the collection and management of data on the state of the planet, supporting the needs of European public policies through the provision of precise and reliable services on environmental and security aspects.

The programme consists of different systems (satellites, ground stations, airborne and marine sensors) that acquire data on the planet.

The Copernicus programme is coordinated and managed by the European Commission, while the infrastructure development is under the control of ESA for the space components.

The use of Sentinel data is aimed at the study of the territory, therefore particular attention is given to the topic of land services, which are in turn divided into four areas (<http://land.copernicus.eu/>):

- Global: these services provide a set of products for studying the Earth's surface at a global level (with medium and low spatial resolutions). Products are used for vegetation monitoring, water cycle and other applications.
- Pan-European: services aimed at the study of European land cover and land use and their change over the years (water bodies, bare soil, forests, sealed areas...).
- Local: services related to specific areas in which specific peculiarities (coasts, big cities, water networks).
- In-situ: all these services need data from the study areas to support satellite products.

The Sentinel project was developed for the specific needs of the programme: the launch into orbit of 12 satellites that will form the space component of the Copernicus programme and are also the European contribution to the Global Earth Observation System (GEOSS).

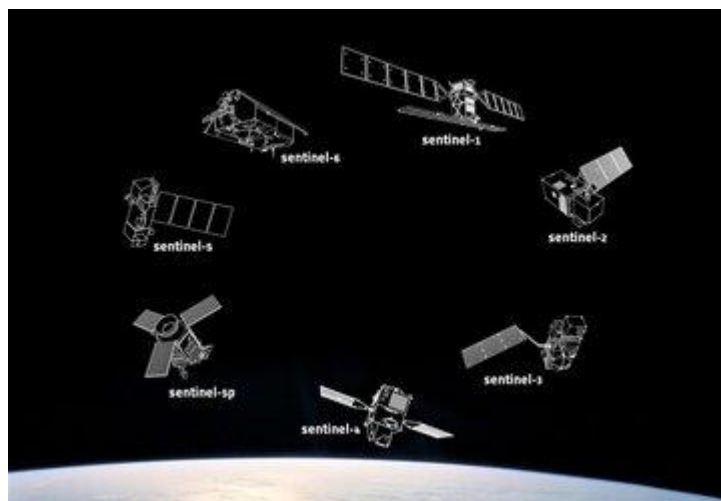
There will be six Sentinel missions [Figure 5. 3]

- **Sentinel-1:** it provides radar images, both during day and night, in all weather conditions, for land and ocean services.

Multi-scale analysis and comparison of Burned Area products derived from Earth Observation (EO) data over the African continent

- **Sentinel-2:** it produces high-resolution optical images for land monitoring services.
- **Sentinel-3:** it generates optical and radar altimetry products to study the marine environment and sea ice.
- **Sentinel-4:** it provides data for the study of atmospheric composition. Its objective is to monitor trace gases and aerosols essential for air quality over Europe, at high spatial resolution and with a short revisit time.
- **Sentinel-5:** it is entirely similar to Sentinel-4. It will provide accurate measurements of the essential constituents of the atmosphere, such as ozone, nitrogen dioxide, sulphur dioxide, carbon monoxide, methane, formaldehyde, and aerosol properties.
- **Sentinel-5P:** This is an interim mission to ensure the continuity of air quality data collection until the launch of Sentinel-5.
- **Sentinel-6:** it carries a radar altimeter to measure the sea surface, for oceanographic and climate studies.

In this work, data from the **Sentinel-2** mission are used as input for classification on Google Earth Engine (GEE) Code Editor. This dataset is stored into GEE Data Catalog (more information on Google Earth Engine will be provided in the following chapters).



*Figure 5. 3 The Copernicus Sentinel mission
(https://www.esa.int/Applications/Observing_the_Earth/Copernicus/The_Sentinel_missions)*

Sentinel-2 characteristics

The satellites weigh approximately 1200 kg and are designed for a lifespan of 7.25 years (which includes a 3 month in-orbit commissioning phase) although batteries and propellants have been charged for 12 years of operations, including end-of-life orbital manoeuvres.

They are suitable for launch with small launchers such as VEGA rocket⁴.

The MSI (Multispectral Instrument) sensors mounted on the platform work passively and collect sunlight reflected from Earth's surface. The incoming light beam is split at a filter and focused onto two separate focal plane assemblies within the instrument: one for Visible and Near-Infra-Red (VNIR) bands and one for Short Wave Infra-Red (SWIR) bands. A special mechanism, a kind of shutter, prevents the sensors from being hit by rays coming directly from the sun during the launch. The same mechanism also acts as a calibration device by collecting the sun's rays after reflection through a diffuser [Figure 5. 4].



Figure 5. 4 Schematic View of the Deployed SENTINEL-2 Spacecraft (image credit: EADS Astrium)

⁴ Vega, an acronym for *Vettore Europeo di Generazione Avanzata* (European Advanced Generation Carrier), is an operational launcher in use by Arianespace, developed in collaboration with the Italian Space Agency (ASI) and the European Space Agency (ESA) to launch small satellites (300 -1500 kg) into orbit.

Orbit and geographical coverage

The Copernicus Sentinel-2 mission comprises a constellation of twin polar-orbiting satellites (Sentinel-2A and Sentinel-2B). They operate on the same 98.62° inclined orbit (relative to the equator), heliosynchronous, at an average elevation of 786 km, offset by 180°. The heliosynchronous orbit was chosen because it minimises the impact of shading on the ground [Figure 5. 5].

The orbit is kept stable by a dedicated propulsion system and by measurements from a dual frequency GNSS receiver⁵.

Sentinel-2 aims at monitoring variability in land surface conditions with a wide swath width of 290 km and high revisit time. The revisit time is the time elapsed between observations of the same point on Earth by a satellite with the same geometry of acquisition. It depends on the satellite's orbit, target location, and swath of the sensor. For Sentinel-2, it is about 10 days at the equator with one satellite, and 5 days with two satellites (after the launch of Sentinel-2B in 2017) under cloud-free conditions, which results in 2-3 days at mid-latitudes under the same gripping conditions; this value decreases if we refer to different gripping conditions, due to the lateral overlapping of the acquisitions.

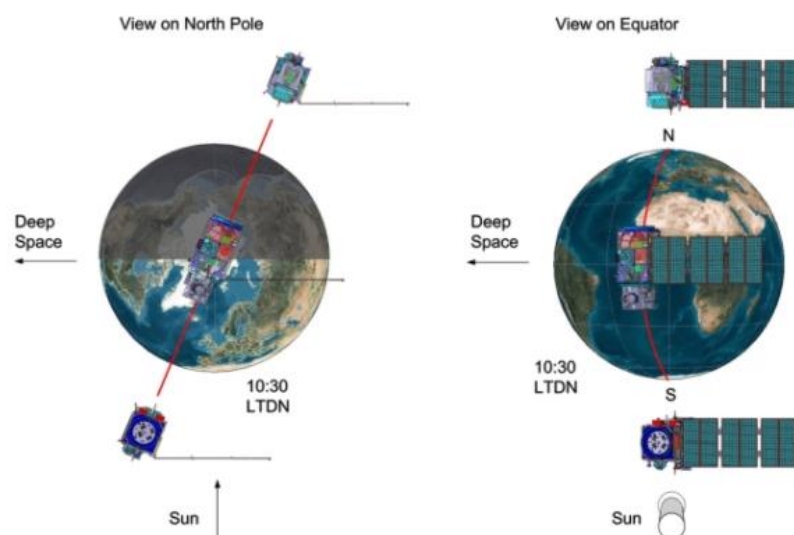


Figure 5. 5The Twin-Satellite SENTINEL-2 Orbital Configuration

⁵ Global Navigation Satellite System (GNSS) refers to a constellation of satellites providing signals from space that transmit positioning and timing data to GNSS receivers. The receivers then use this data to determine location. (<https://www.euspa.europa.eu/european-space/eu-space-programme/what-gnss>)

The Sentinel-2s systematically acquire data on land and coastal areas in a band from 56° South latitude (Isla Hornos, Cape Horn, South America) to 84° North latitude (over Greenland), including also: islands larger than 100 km², islands in the European Union, islands with a coastline of at least 20 km, Mediterranean Sea, inland water bodies (rivers, lakes, marshes...) and all enclosed seas.

Figure below shows the zones in which the satellites collect data [Figure 5. 6].

In addition to the normal acquisition zones, at regular intervals, acquisitions are made on specific zones, such as Dome-C in Antarctica, in order to calibrate the instruments.

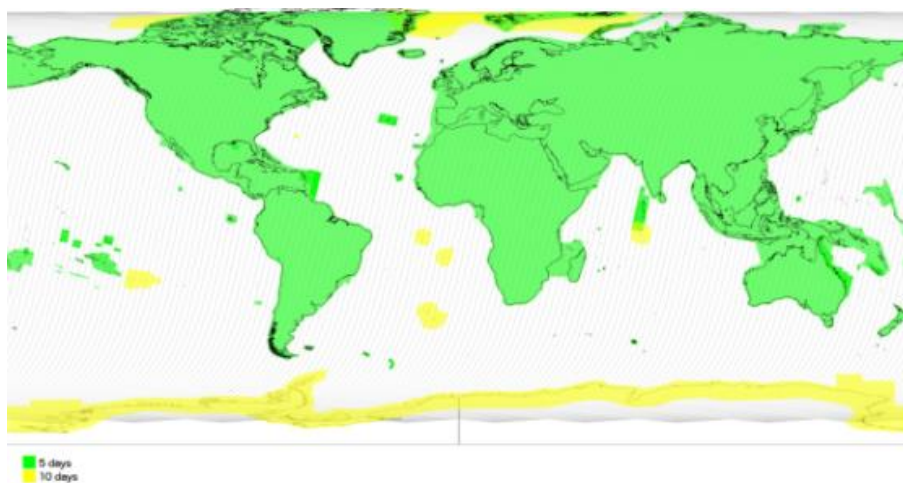


Figure 5. 6 Coverage and revisit time foreseen for Sentinel-2 MSI acquisitions

Resolution

Sentinel-2 resolution characteristics are threefold:

4. The **temporal resolution**: the revisit frequency of each single Sentinel-2 satellite is 10 days and the combined constellation revisit is 5 days.
5. The **spatial resolution**: the spatial resolution of SENTINEL-2 is dependent on the spectral band; the maximum is 10 m.
6. The **radiometric resolution**: the radiometric resolution of Sentinel-2 is 12-bit. This gives a potential range of brightness levels from 0 - 4 095.

Product Sentinel-2

Sentinel-2 data are available for users in various processing levels as follows below

[Table 5. 3]:

Name	High-Level Description	Production and distribution	Data Volume
Level 1C	Top-Of-Atmosphere reflectances in cartographic geometry	Systematic generation and online distribution	~600MB (each 100km x 100km ₂)
Level 2A	Bottom-Of-Atmosphere reflectances in cartographic geometry	Systematic and on-user side (using Sentinel-2 Toolbox)	~800MB (each 100km x 100km ₂)

Table 5. 3 Sentinel-2 product types

Data are provided in a special format (SENTINEL-SAFE) that includes, in addition to the JPEG image, quality indicators, auxiliary data and metadata.

Sentinel-2 products are a compilation of elementary granules of fixed size, along with a single orbit. A granule is the minimum indivisible partition of a product (containing all possible spectral bands); they are also called “tiles” and their dimension is 100x100 km. Each tile is in WGS84 datum in UTM projection and is aligned with the UTM zone in which it falls; adjacent tiles aligned with the same UTM zone overlap by 10 km, while the overlap area of unaligned tiles may be larger.

Sentinel-2 **Level-1C** is an orthoimage product, i.e. a map projection of the acquired image using a system DEM to correct ground geometric distortions. Pixel radiometric measurements are provided in Top-Of-Atmosphere (TOA) reflectance. These data are present in the GEE Data Catalog but are not used during this work.

The SENTINEL-2 **Level-2A** processing provides Level-2 products in Bottom-Of-Atmosphere (BOA) reflectance from Level-1C products. Level-2A can be generated by

the User from the Level-1C product using the Sentinel-2 Toolbox or the standalone version of the Sen2Cor processor. Alternatively Level-2A products can be found in the GEE Data Catalog, accessible from here: https://developers.google.com/earth-engine/datasets/catalog/COPERNICUS_S2_SR . These data were used in this thesis. BOA reflectance values in the S2 spectral bands, reflectance corrected for the atmospheric effect, are necessary to produce consistent and robust classification output especially over large areas and over multi-temporal datasets.

In the figure below there is a comparison between the TOA Level-1C image data and the Level-2A BOA image data of the same area [Figure 5. 7].

(<https://sentinels.copernicus.eu/web/sentinel/user-guides/sentinel-2-msi/product-types/level-2a>)



Figure 5. 7 TOA Level-1C image data (left) and associated Level-2A BOA image data (right) generated using Sen2Cor processor

Level-2A BOA Surface Reflectance values are derived with sub-pixel multispectral and multitemporal registration accuracy. The main characteristics of the L2A products are

listed below [Table 5. 4]. Spectral response functions of Sentinel-2A MSI bands in shown in the figure [Figure 5. 8].

<https://sentinel.esa.int/web/sentinel/sentinel-data-access/sentinel-products/sentinel-2-data-products/collection-0-level-2a>

Name	Level-2A
High-level Description	Surface reflectances in cartographic geometry for 12 spectral bands (10 m, 20 m, and 60 m depending on the wavelength; Cirrus band B10 is not included).
Algorithm used	Sen2Cor – Versions from 2.5.5 to 2.09
Data Characteristics	<ul style="list-style-type: none"> • UTM/WGS84 projection • JPEG2000 image format • 12-bit pixel values • < 12 m at 95.5% confidence of Absolute Geolocation • < 5 m at 95.5% confidence of Multitemporal Registration • < 0.3 px at 99.7% confidence of Multispectral Registration
Additional Layers	<ul style="list-style-type: none"> • Scene Classification Map at 60 m • AOT Map • Water Vapor Map
DEM used	<ul style="list-style-type: none"> • PlanetDEM at 90 m (until PB 02.14 included – March 30th 2021) • Copernicus DEM at 90 m (starting from PB 03.00 included)
Auxiliary Data used	ECMWF
Production & Distribution	Systematic generation and online distribution
Data Volume	800 MB (each 100x100 km ²)
Data Availability	Global since December 2018

Data Delivery	Available within 8 hours from sensing
----------------------	---------------------------------------

Table 5. 4 Sentinel-2 L2A Characteristics

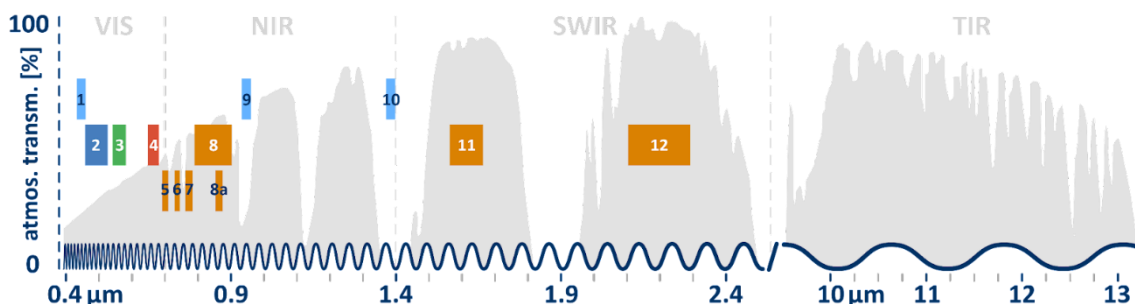


Figure 5. 8 Spectral response functions of Sentinel-2A MSI bands

Level-2A, available on the GEE archive since 2017 is composed of the layers shown in these bands [Table 5. 5] can be recalled in Google Code Editor thanks to the command `ee.ImageCollection("COPERNICUS/S2_SR")` (The Classification on of Google Earth Engine will then be explained and deepened in the section "Classification" of this thesis).

Name	Wavelength	Resolution (pixel size)	Description
B1	443.9nm (S2A) / 442.3nm (S2B)	60 m	Aerosols
B2	496.6nm (S2A) / 492.1nm (S2B)	10 m	Blue
B3	560nm (S2A) / 559nm (S2B)	10 m	Green
B4	664.5nm (S2A) / 665nm (S2B)	10 m	Red
B5	703.9nm (S2A) / 703.8nm (S2B)	20 m	Red Edge 1
B6	740.2nm (S2A) / 739.1nm (S2B)	20 m	Red Edge 2
B7	782.5nm (S2A) / 779.7nm (S2B)	20 m	Red Edge 3

Multi-scale analysis and comparison of Burned Area products derived from Earth Observation (EO) data over the African continent

B8	835.1nm (S2A) / 833nm (S2B)	10 m	NIR
B8A	864.8nm (S2A) / 864nm (S2B)	20 m	Red Edge 4
B9	945nm (S2A) / 943.2nm (S2B)	60 m	Water vapor
B11	1613.7nm (S2A) / 1610.4nm (S2B)	20 m	SWIR 1
B12	2202.4nm (S2A) / 2185.7nm (S2B)	20 m	SWIR 2
AOT		10 m	Aerosol Optical Thickness
WVP		10 m	Water Vapor Pressure. The height the water would occupy if the vapor were condensed into liquid and spread evenly across the column.
SCL		20 m	Scene Classification Map (The "No Data" value of 0 is masked out)
TCI_R		10 m	True Colour Image, Red channel
TCI_G		10 m	True Colour Image, Green channel
TCI_B		10 m	True Colour Image, Blue channel
MSK_CLDPRB		20 m	Cloud Probability Map (missing in some products)
MSK_SNOWPRB		10 m	Snow Probability Map (missing in some products)
QA10		10 m	Always empty

Multi-scale analysis and comparison of Burned Area products derived from Earth Observation (EO) data over the African continent

QA20		20 m	Always empty
QA60		60 m	Cloud mask

Table 5. 5 Bands in Sentinel-2 MSI L2A

The Sentinel-2 MSI satellites can acquire 4 bands in the visible and near infrared with 10 m spatial resolution, 6 bands in the infrared with 20 m spatial resolution and 3 bands with 60 m resolution of which one in the blue and two in the infrared. The bands are listed in the table from B1 to B12. The cirrus band B10 is excluded as it does not contain any "bottom of the atmosphere" information.

There are also additional bands which are used to improve the accuracy of surface reflectance-derived products (SCL, AOT, WVP).

In particular, SCL stands for "Scene Classification Layer"; the objective of this product is not to classify the land cover in a strict sense, but to distinguish between cloudy, clear and water pixels. It is evaluated by Sen2Cor Algorithm; the algorithm separates areas into eleven classes, shown in the table below [Table 5. 6]. The resulting SCL is applied in the cloud mask analysis, which removes unwanted cloudy pixels from the multispectral observations.

0	No data
1	Saturated_or_defective
1	Dark_Area_PIXELS
2	Cloud_Shadows
3	Vegetation
4	Bare_soils
5	WATER
6	Cloud_LOW_PROBABILITY
7	Cloud_MEDIUM_PROBABILITY
8	Cloud_HIGH_PROBABILITY
9	Thin_CIRRUS
10	Snow

Table 5. 6 Scene Classification Classes

This information is crucial for the mapping of burned areas, which will be addressed in the next chapter. SCL will be used to create the cloud cover layer, as it classifies them. A visual representation of the SCL classes is shown in figure below [Figure 5. 9]; in the left column of the composition of layer there are RGB images, in the right column there are SCL products.

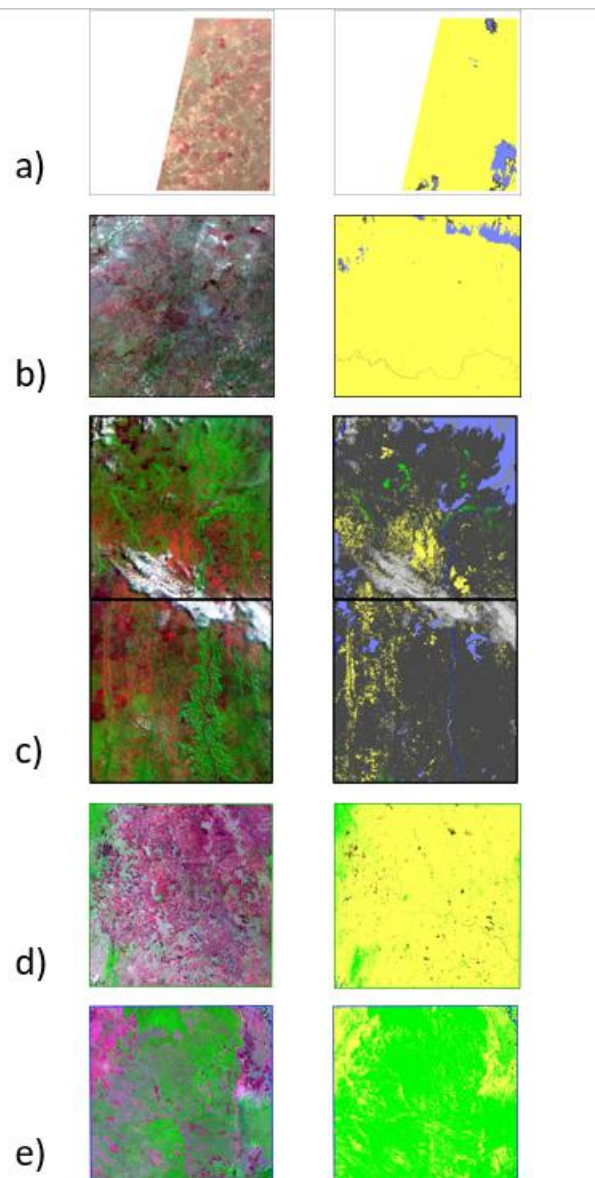


Figure 5. 9 The S² RGB and SCL layers for one example date for each AOI for the six tiles covering the regions: a)28PFA, b) 28PGV, c) 33MXQ-33MXR, d)36KWE, e)36KYE

The Sentinel-2 tiles chosen for this work are consistent with the chosen PlanetScope's sites, already stated in the previous chapter.

The PlanetScope data, having a higher resolution than the Sentinel-2 (3 m for the Planet, 10 m for the Sentinel), will be used as reference data in the validation phase of

Multi-scale analysis and comparison of Burned Area products derived from Earth Observation (EO) data over the African continent

the mapping of burned areas; for this reason, it is necessary spatial coherence (the Sentinel-2 tile must cover the whole area of the pre-selected Planet site) and temporal coherence (the dates chosen for the Sentinel images must cover the same time frame in which the Planet images were acquired).

Sentinel tiles and Planet sites are shown consistently in the following table [Table 5. 7]:

Site	S2 tile	Dates	N days
1A	28PFA	17/01/2019 27/01/2019 06/02/2019 03/03/2019 18/03/2019	61 days
1B	28PGV	17/01/2019 27/01/2019 06/02/2019 03/03/2019 18/03/2019	61 days
2C	33MXQ 33MXR	12/06/2019 17/06/2019 27/06/2019 17/07/2019 27/07/2019 06/08/2019 16/08/2019 21/08/2019	71 days
3A	36KYE	29/08/2019 13/09/2019 28/09/2019	31 days
3B	36KWE 36KWD	23/07/2019 07/08/2019	51 days

		17/08/2019	
		27/08/2019	
		11/09/2019	

Table 5. 7 Sentinel-2 tiles used for this thesis

The figure below shows the consistency between the dates on which we selected PlanetScope images and the dates on which we selected Sentinel-2 images, for the same spatial areas [Figure 5. 10].

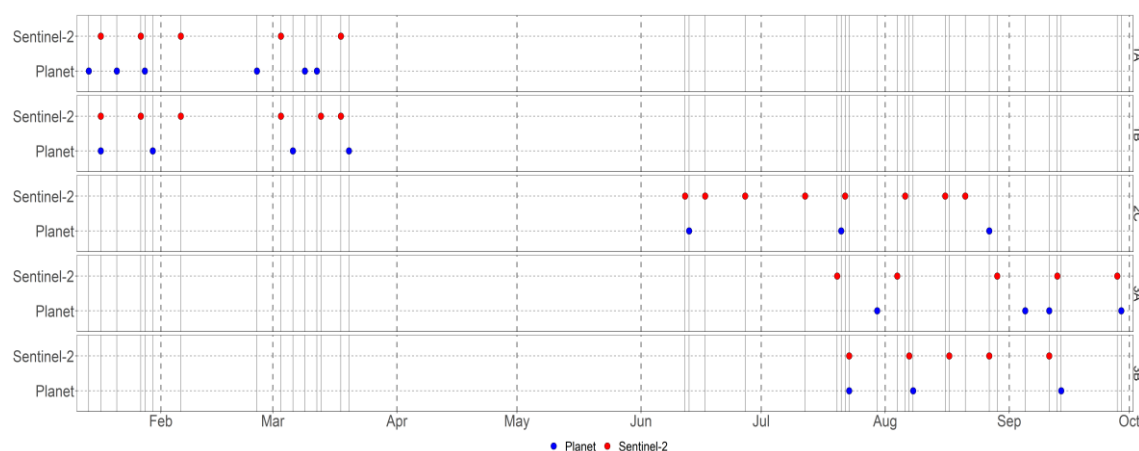


Figure 5. 10 Acquisition dates for S2 (red) and Planet time series (blue).

5.3 FireCCI51 BA product

The FireCCI51 BA global product (Lizundia-Loiola et al., 2020) is produced in the FireCCI project of the CCI program. It is an improved version of FireCCI50, the first global BA product at 250 m spatial resolution (Chuvieco et al., 2018), which, in turn, has in turn, adapted the FireCCI41 algorithm (Alonso-Canas and Chuvieco, 2015) based on the MERIS 300 m NIR and Red bands to MODIS 250 m.

The BA product consists of maps of global burned areas covering the period 2001-2020, developed and adapted for use by climate, vegetation, and atmospheric modelers, as well as researchers or fire managers. The FireCCI51 product is provided at several spatial resolutions: the full-resolution pixel product (250 m, monthly BA estimates with daily day-of-year (DOY) detection records) and the grid product (0.25 degree, total annual burned area amount). The FireCCI51 BA product is derived with a

hybrid algorithm that combines the highest resolution MODIS reflectance bands (250 m) and active fire information from MODIS thermal channels; details on the algorithm can be found in Lizundia-Loiola et al. (2020).

The next image is a Plot composition depicting the dataset for each site analysed [Figure 5. 11].

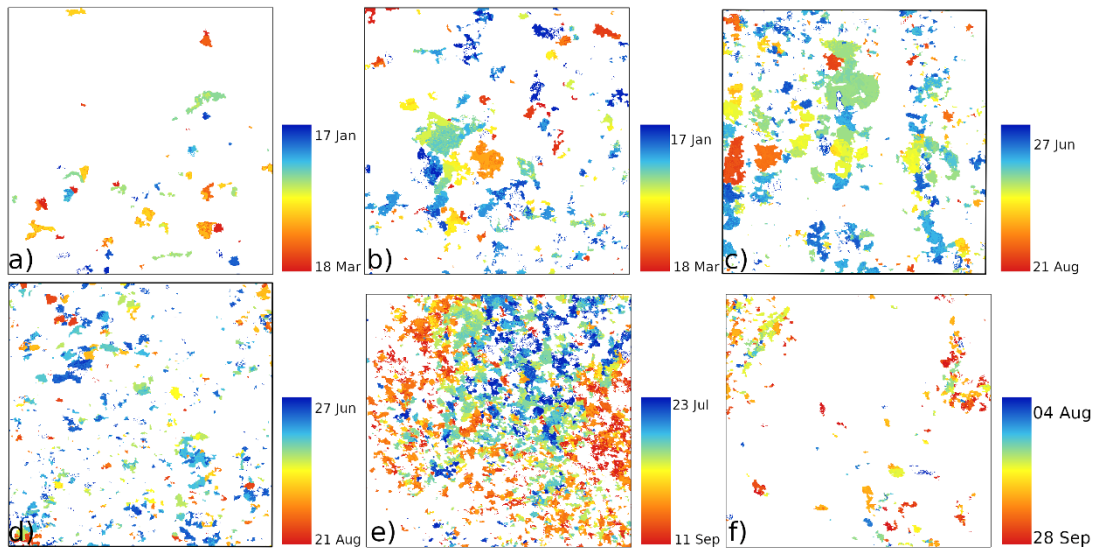


Figure 5. 11 Plots of FIRECCI51 for Sentinel study areas: a) 28PFA, b) 28PGV c) 33MXQ, d) 33MXR, e) 36KYE, f) 36KWD

The FireCCI BA product is also available in GEE catalog (https://developers.google.com/earth-engine/datasets/catalog/ESA_CCI_FireCCI_5_1).

5.4 Active Fires

Another product used in the image validation and comparison phase (Part 2 of this thesis) is Active Fire dataset. These data contain information on the place, date and acquisition time of thermal anomalies (i.e. fires) detected by satellite data. Active fires (AFs) can provide information on seasonality of wildfires and, in this thesis work they were used to confirm or deny the presence of burned areas.

Active fire data can be viewed and downloaded in NASA⁶'s FIRMS FIRE MAP (user can access the map here: <https://firms.modaps.eosdis.nasa.gov/map/>).

FIRMS stands for Fire Information for Resource Management System; it is a tool that distribute near real-time active fire data within three hours of a satellite overpass from NASA's Moderate Resolution Imaging Spectroradiometer (MODIS) and NASA's Visible Infrared Imaging Radiometer Suite (VIIRS) satellites.

FIRMS was originally developed by the University of Maryland, with funds from NASA's Applied Sciences Program and the United Nations Food and Agriculture Organization (UN FAO). It was transitioned to NASA's LANCE⁷ in 2012.

FIRMS focus and objectives include:

- PROVIDING QUALITY RESOURCES FOR FIRE DATA ON DEMAND
- WORKING WITH END USERS TO ENHANCE CRITICAL APPLICATIONS
- ASSISTING GLOBAL ORGANIZATIONS IN FIRE ANALYSIS EFFORTS
- DELIVERING EFFECTIVE DATA PRESENTATION AND MANAGEMENT

MODIS

The Moderate-resolution Imaging Spectroradiometer (MODIS) is a key instrument aboard the NASA's Terra and Aqua satellites, both of which are designed to observe the land, oceans, and atmosphere of Earth and due to its enormous coverage, it enables us to monitor activities such as active fires, land use change and volcanic activity. It has a viewing swath of 2,330 kilometers and sees the entire surface of the Earth every one to two days. The instrument's detectors measure reflected light in 36 spectral bands between 0.405 and 14.385 microns. The spatial resolution of these observations varies between 250-1000 meters. (2 bands at 250 m, 5 bands at 500 m and 29 bands at 1000m). For more details on the characteristics of MODIS data and products see: <https://modis.gsfc.nasa.gov/data/>

⁶ NASA stands for National Aeronautics and Space Administration. It is the civilian government agency responsible for the space programme and aerospace research of the United States of America.

⁷ Atmosphere Near real-time Capability for EOS (LANCE) supports users interested in monitoring a wide variety of natural and man-made phenomena using near real-time (NRT) data and imagery that are made available much quicker than routine processing allows.

VIIRS

Visible Infrared Imaging Radiometer Suite (VIIRS) data products are developed by NASA and distributed by NASA's Land Processes Distributed Active Archive Center (LP DAAC).

VIIRS is an instrument aboard two satellites: Suomi National Polar-orbiting Partnership (Suomi NPP) and National Oceanic and Atmospheric Administration-20 (NOAA-20). This instrument observes the entire Earth's surface twice each day (every 12 hours approximately): Suomi NPP crosses the equator about 1:30 a.m. and 1:30 p.m. and NOAA-20 crosses at 2:20 am and 2:20 pm, local time.

The 3,000 km swath width of the VIIRS instrument, which is 710 km greater than that of MODIS, allows for no gaps in coverage as observed in MODIS near the equator.

The VIIRS instrument provides 22 spectral bands from 412 nanometers (nm) to 12 micrometers (μm) at two spatial resolutions, 375 meters (m) and 750 m, which are resampled to 500 m, 1 km, and 0.05 degrees in the NASA produced data products to promote consistency with the MODIS heritage.

The 375 m data also has improved night-time performance thanks to the band DBN.

For more details on the characteristics of VIIRS data and products see: <https://earthdata.nasa.gov/earth-observation-data/near-real-time/firms/viirs-i-band-active-fire-data>

In detail, the MODIS- and VIIRS-derived datasets used are:

- MODIS NRT C6.1 (**MCD14DL**) are the NRT⁸ MODIS (MOD14/MYD14 respectively from Terra and Aqua Satellites) Collection 6.1 data processed by NASA LANCE FIRMS. It is available from November 2000 (for Terra) and from July 2002 (for Aqua) to the present.

Active fires are identified from MODIS data by the application of the fire detection algorithm (Giglio, L; Schroeder, W; Justice, CO (2016). The collection 6 MODIS active fire detection algorithm and fire products. *REMOTE SENSING OF ENVIRONMENT*, 178, 31-41).

⁸ NRT stands for Near real-time. NRT fire products are generated within approximately 3 hours of satellite observation to meet the needs of the applications community.

The collection 6 detection algorithm improved the performance of active fire detection, with reduced omission errors over large fires, and reduced false alarm rates in tropical ecosystems.

MODIS Collection 6.1 has a temporal coverage between November 11th 2000 and now; VIIRS S-NPP 375m has a temporal coverage between January 20th 2012 and now; VIIRS NOAA-20 375 m has a temporal coverage between January 1st 2020 and now.

- **MCD14ML** provided by FIRMS are a subset of the standard quality data processed by the MODIS Fire Team Science Computing Facility (SCF) at the University of Maryland. These are available with a 2-3 month lag through the Archive Download Tool and can be viewed and queried in Fire Map. Standard data products are an internally consistent, well-calibrated record of the Earth's geophysical properties to support science. One key difference between the MODIS/Aqua NRT and Standard (Science Quality) Fire Products is the accuracy of the Fire locations (positions or geolocation).
- VIIRS 375 m NRT active fire data are available from Suomi NPP (**VNP14IMGTDL_NRT**) and NOAA-20 (**VJ114IMGTDL_NRT**). These are processed by NASA LANCE FIRMS. The improved 375 m data provide greater response over fires of relatively small areas, as well as improved mapping of large fire perimeters.

In these datasets, each hotspot/active fire detection represents the centre of a pixel reported to contain one or more fires, or other thermal anomalies (such as volcanoes). For MODIS the pixel is about 1km and for VIIRS the pixel is about 375m. The "location" is not necessarily the coordinates of the actual fire since a fire is often smaller than the size of the pixel. Although we do not know the exact location of the fire, we do know that it is within that pixel. In *[Figure 5. 12]* and *[Figure 5. 13]* this concept is shown graphically.

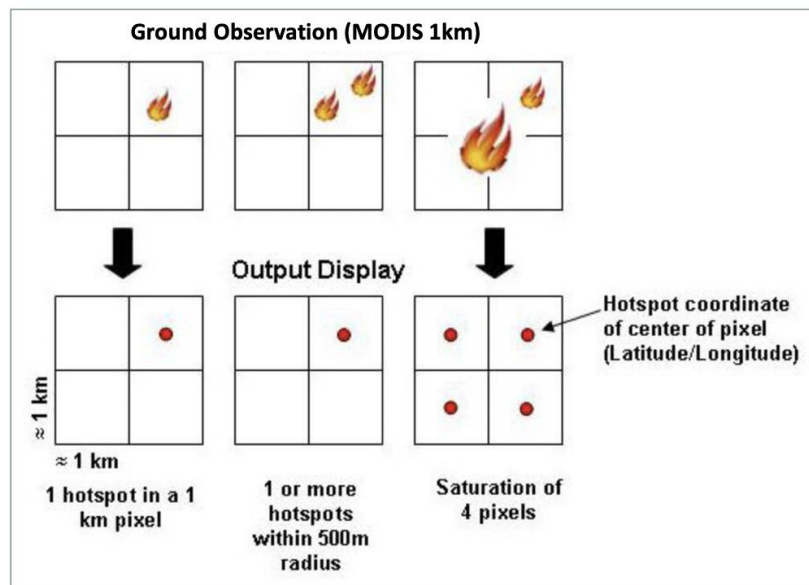


Figure 5. 12 Ground observation MODIS (<https://earthdata.nasa.gov/faq/firms-faq>)

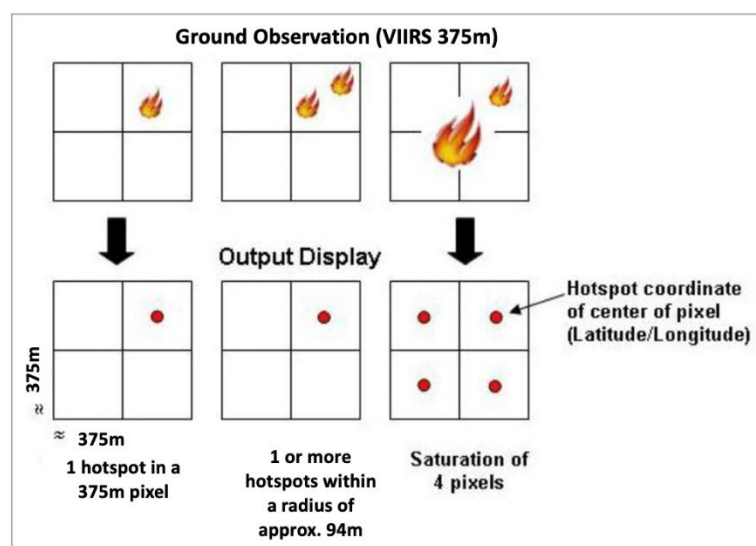


Figure 5. 13 Ground observation VIIRS (<https://earthdata.nasa.gov/faq/firms-faq>)

User can download active fire product for the last 24, 48 hours and 7 days in shapefile, KML⁹, WMS¹⁰ or text file formats. Data older than 7 days can be obtained from the Archive Download Tool on the FIRMS website (it is accessible here: <https://firms.modaps.eosdis.nasa.gov/download/>).

⁹ KML stand for **Keyhole Markup Language File**. It stores geographical modelling information in XML format. It includes points, lines, polygons and images.

¹⁰ WMS (Web Map Service) delivers a map that is a representation of spatial information in the form of a digital image file in PNG, GIF or JPEG format.

An example of Active Fires (from both MODIS and VIIRS) over sites 1A and 1B, Northern Africa is shown in [Figure 5. 14].

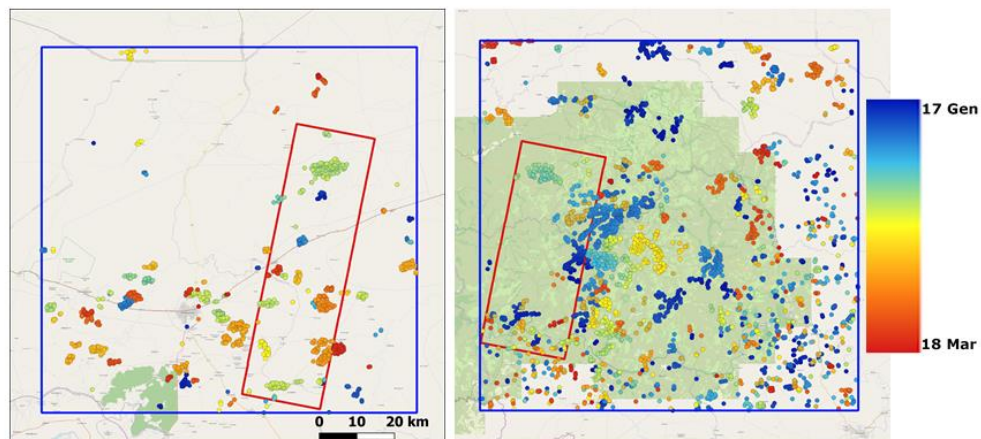


Figure 5. 14 Example of active fire points for the AOIs 1A and 1B in Northern Africa. Colours show the date of detection of Afs between January 17th and March 18th, 2019. The blue square is the size of the S2 tiles covering the regions and the red rectangles is the area

The image below is a composition of plots that shows how Active Fires are distributed in time. This figure underlines the seasonality of fires [Figure 5. 15].

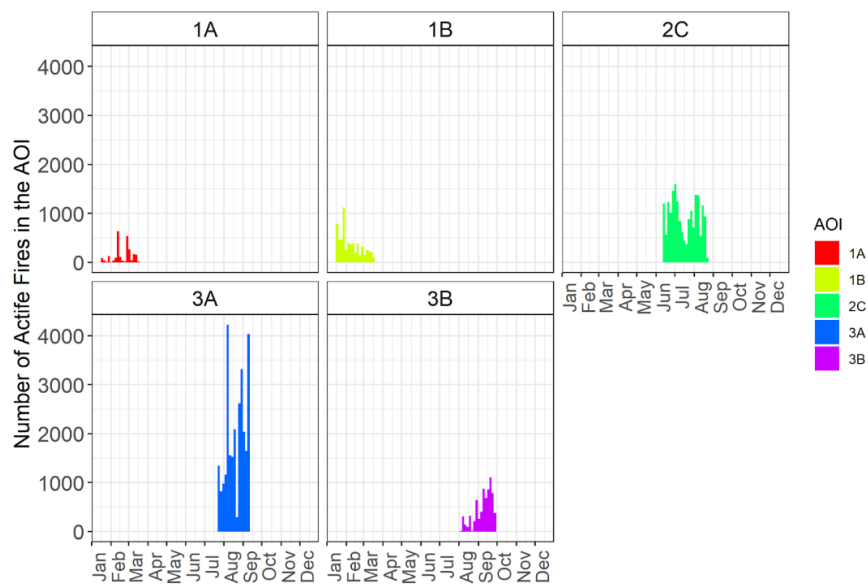


Figure 5. 15 Seasonality of fires as observed in the Afs datasets for the five AOIs in Africa and the year 2019. Active fire points for each region are subset between the first and last date of the temporal series of satellite images.

6 Burned area mapping

In this work the S2 burned areas are mapped using supervised Random Forest Algorithm with a trial and error approach for the Sentinel-2 AOIs. The algorithm works with pairs of consecutive S2 images and the output identifies the areas burned between the two dates of each pair. Burned areas (BA) from all pairs of images within each site were then combined into a synthetic map by retaining the first date of detection. These S2 BA maps were compared to fire perimeters from Plane, and the FireCCI BA maps; while the first is assumed as accuracy assessment, the second is the comparison of BA maps derived from RS data with different spatial resolution (i.e., multi-scale comparison).

The time series is a list of Sentinel-2 images, previously selected by the user, representing the same portion of territory displayed at different times. Sentinel data are stored in the GEE Data Catalog and can be viewed on the GEE Code Editor; it is in this environment that the classification algorithm will be implemented.

First of all, the training data are collected, then a classification of the burned areas between two consecutive dates is performed. The classification results are then compared with the RGB composites to assess whether the classification is consistent with the actual conditions of the area.

This interactive step is necessary to judge the quality of the results, i.e. to evaluate if the areas classified as burned by the algorithm correspond to burned areas under real conditions. If necessary, a new data set can be collected, and the algorithm run again.

6.1 Introduction to Google Earth Engine

The environment used to implement the burned area mapping algorithm from Sentinel-2 data is Google Earth Engine

Multi-scale analysis and comparison of Burned Area products derived from Earth Observation (EO) data over the African continent

Google Earth Engine (GEE) is a cloud computing open-access platform designed to store, visualize, and process geospatial datasets, for analysis and ultimate decision making.

GEE contains a catalogue of satellite images going back more than forty years, updated and expanded daily.

Therefore, Earth Engine is different from Google Earth because it is a computing platform that also provides APIs¹¹ and other tools to enable the geospatial analysis of large datasets.

An easily accessible and user-friendly front-end provides a favourable environment for the development of new algorithms by governments, independent scientists and researchers and all interested users who do not have direct access to data and professional tools.

There are two different ways to interact with the platform:

Explorer: is a simple web interface to the Earth Engine API. It allows anyone to visualize the data in the public GEE Data Catalog. Registered Earth Engine users, besides querying the catalog, can also import data from the cloud platform to the workspace, run simple analysis and export results [Figure 6. 1].

It can be access here: <https://explorer.earthengine.google.com/>

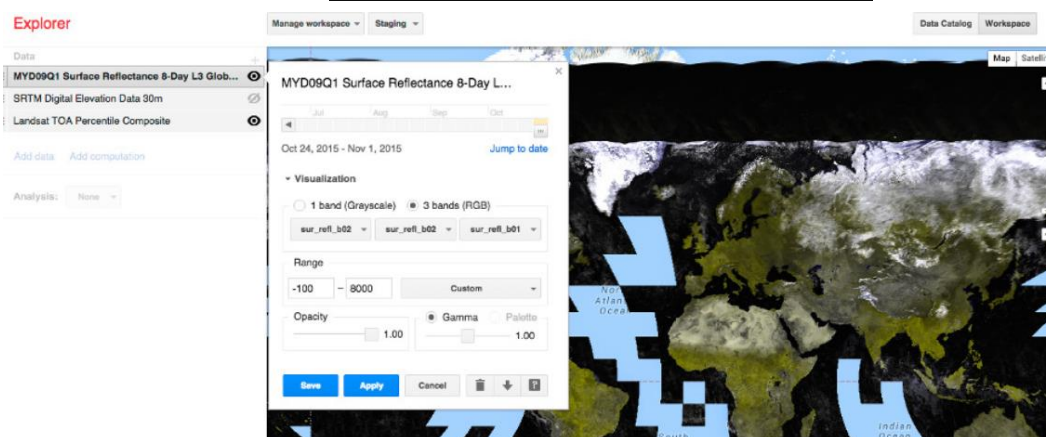


Figure 6. 1 Screenshot of Explorer Interface on the GEE platform

¹¹ "Application programming interface (API) is a set of protocols, routines, functions and/or commands that programmers use to facilitate interaction between distinct software services."

Definition by <https://technopedia.com>

Code Editor: is web-based IDE¹² for writing and running scripts for the Earth Engine JavaScript API. The interface has been designed to make the development of complex geospatial workflows very simple and fast. It can be access here: <https://code.earthengine.google.com/> and the user must log in with a Google account.

The Code Editor interface is composed as follows [Figure 6. 2]:

- **JavaScript Code Editor window:** is the window where user writes the JavaScript code to access and analyse imagery. It automatically formats the code, underlines code errors and offers fixes and correct solutions; It also autocomplete pairs of quotes, parentheses, and brackets.
There are buttons for running and saving the script above the Code Editor.
- **Search tool:** is the search bar in which the user can find data to import from the archive simply typing a name of product or keyword.
- **Map display:** is the window for visualizing geospatial datasets.
- **Asset Manager:** it is in the Asset tab in the left panel; it is used by the user to upload external dataset and to manage them in Earth Engine. In this work, this functionality is used to load a list of dates in .csv format. It will be used to select only those images from a Sentinel-2 collection whose acquisition date matches one of the dates contained in the file loaded through the Asset Manager.
- **API documentation:** it is a tab that contains the complete JavaScript API documentation. It could be searched and browsed from the tab.
- **Script Manager:** it stores private, shared and example scripts in Git repositories hosted by Google. The repositories are arranged by access level. In this work private scripts are used; they are stored in a repository called “owned folder” that is accessible only by the user.
- **Inspector:** it is a window that let user to query the map. When the inspector is activated, cursor becomes a crosshair which will display the location and

¹² “A **cloud IDE** is a web-based integrated development platform (IDE). An IDE is a programming environment that has been packaged as an application, typically consisting of a code editor, a compiler, a debugger, and a graphical user interface (GUI) builder.”

Definition by <https://searchcloudcomputing.techtarget.com>

Multi-scale analysis and comparison of Burned Area products derived from Earth Observation (EO) data over the African continent

layer values when the user clicks on the map. Briefly, the inspector functionality gives the latitude and longitude coordinates of each selected point on the map display.

- **Task Manager:** is a window in which the user can export the output files of the code. In this work it is used to download the outputs of the classification of burned areas simply clicking on the “run” button; the downloaded outputs will automatically be stored in a Google Drive folder (according to the Google account that the user inserts for the login phase)
- **Console output:** it is an interactive console that print results about outputs, error messages or more information about variables (simply writing the command `print(variable_name)` on the Code Editor window).
- **Geometry drawing tools:** are tools used to import geometries to the script drawing them directly on map screen by the user. In this work they are used to draw training polygons for the classification algorithm to identify burned areas.
- **Layer Manager:** it is a window in which the user could manage the visualization of the map layer on the map display. It allows the user also to adjust transparency and customize layer’s style; the user can also toggle the visibility of a layer added in the map.

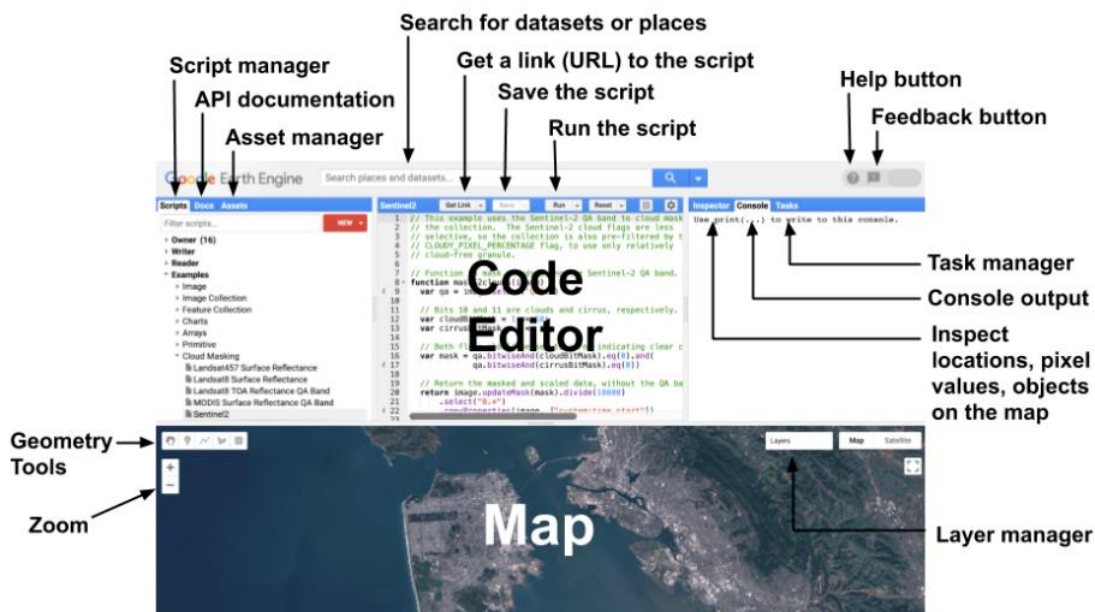


Figure 6. 2 Screenshot of Code Editor Interface on GEE platform

Earth Engine Data Catalog (accessible here <https://developers.google.com/earth-engine/datasets>) hosts a variety of standard Earth Science raster datasets; each dataset has an associated link where we can see a brief description of it including information such as temporal availability, data provider and collection ID. The collection ID allows the user to import the data into the script of the Code Editor environment for visualization and processing.

Data could be accessed, analysed and visualized thanks to the Google API's high performance. The variety of data makes it possible to address different purposes even on a global scale.

The datasets accessed through the Data Catalog are:

- I. Geophysical dataset (Digital Elevation Model, Land Cover maps etc.)
- II. Climate and Weather dataset (Surface Temperature information, weather dataset etc.)
- III. Satellites imagery (Landsat Catalogue, MODIS datasets, Sentinel data etc.)

Not only GEE stored data could be analysed on the platform; the user can import his/her own datasets into the script environment with a single click through the Asset Manager.

The dataset used to conduct the burned areas sampling - that will be explained in the next chapter - belongs to the GEE Data Catalogue. It is **Sentinel-2 MSI: MultiSpectral Instrument, Level-2A**, (here there is the official GEE page for the dataset https://developers.google.com/earth-engine/datasets/catalog/COPERNICUS_S2_SR) provided by Copernicus Programme of European Space Agency (ESA). Data are available on GEE from 28/03/2017 until now, but only data from the year 2019 are selected and used for the analysis performed in this thesis.

GEE has been used in this thesis work to implement the algorithm for classification of burned areas from time series of S2 images. GEE was chosen because it is an open-source platform and therefore easily accessible through the user's google account. It hosts all the data needed for our work in its Data Catalog and, also, has the tools needed to implement the classification algorithm. In this way the whole work is done

on the cloud platform, and not locally on the user's computer. The procedure is faster and easier, the performance is very good and the consistency of the working environment avoids errors and wasting of time.

6.2 Algorithm Modules

The algorithm is implemented on GEE and follows several steps. They are:

I. **Selection of Sentinel-2 Time Series:**

A Satellite Image Time Series (SITS) is a set of satellite images taken from the same scene at different times. In this phase the user selects the best quality images for each portion of space filtering by a Cloud Cover property (i.e. the property, contained in the metadata) that tells us how much an image is covered by clouds, and therefore, how visible it is.

This phase is divided into two parts: the first part involves the implementation of a JavaScript code to select the collection of images and filter it, the second part involves the direct intervention of the user to select the best images according to predefined parameters.

II. **Time Series and RGB Download:**

A JavaScript code has been implemented to allow the user to download the previously selected time series. The images are downloaded from the GEE cloud platform to a Google Drive folder in the user's google account; this way they can be reused for later analysis.

III. **Training data and Collection:**

Once the user has selected the time series to be analysed, he can move on to mapping the burned areas between two consecutive images. The classification algorithm, however, requires a preliminary sampling phase, that is this third step: the user draws training polygons directly on the image -via the map layer of the Code Editor-, which are also divided into classes. These polygons will be given as input to the algorithm to complete the classification.

IV. **Classification of burned area:**

Once the training polygons have been drawn, the last step is the actual classification. In this phase a script in JavaScript is implemented for the identification of burned areas taking as input the samples drawn in the previous step. The statistical algorithm used is the Random Forest Algorithm.

The outputs of the classification are different layers containing different information including burned areas, cloud cover etc. (these will be explained in detail in the next step). (these will be explained in detail in the following chapters).

Here there is a flowchart of the Algorithm Burned Areas [Figure 6. 3]:



Figure 6. 3 Flowchart of the Algorithm for burned areas mapping

6.2.1 Selection of Sentinel-2 time series

First step is the selection, for each study site, of Sentinel-2 time series that are used in the following steps for the classification/identification of burned areas.

It's important to say that each study area is divided into small units called tiles (256x256 pixels) and Sentinel-2 flies over the same point on the Earth's surface with the same viewing angle every 5 days. This means that in the Data Catalog there are images on different dates for each tile.

In order to select the satellite images with the best quality among those available in the GEE catalog, a JavaScript code has been written specifically for this thesis project [Table 6. 1]. The time series is composed of satellite images, taken at different times in a specific area of interest, and is used as input for further analysis and processing. The time period covered by the time series is identified by the start (first date in the series) and the end (last date in the series) dates.

The script is general and can be used for different time intervals and Sentinel-2 tiles by modifying the start and end date of the time interval and the name of the tile. This selecting operation must be repeated for each tile covering the region of interest.

Script 1

```
1     var START_DATE = '2019-07-20';
2     var END_DATE = '2019-09-20';
3     var MGRS_TILE = '36KWD';
4
5     var s2_unsorted = ee.ImageCollection('COPERNICUS/S2')
6                       .filterMetadata('MGRS_TILE',
7                                       'equals',MGRS_TILE)
8                       .filterDate(START_DATE, END_DATE);
9     var s2_sorted = s2_unsorted.sort('CLOUDY_PIXEL_PERCENTAGE');
10    var count=s2_sorted.size();
11    print ('SIZE OF COLLECTION:', count )
12
13    Export.table.toDrive({
14      collection: s2_sorted,
15      folder: "GEE",
16      fileFormat: 'CSV'
17    });
```

Table 6. 1 JavaScript code for selection of satellites data

The information about dates and tile is saved in three different variables: “var START_DATE”, “var END_DATE” and “var MGRS_TILE”. The variables assignment can be changed at the beginning of the processing depending on user’s needs. **(Script’s lines 1,2,3)**

Once the time period and tile to be covered are selected, the command `ee.ImageCollection()` imports a collection of images from the GEE Data Catalogue to the code environment. In GEE, an ImageCollection is a stack or sequence of images (i.e., the time series). The collection can be loaded pasting into the function an Earth Engine the ID, that identifies the type of data to be searched in the Data Catalogue. In

this case, ID is: 'COPERNICUS/S2 that identifies and loads Sentinel-2 surface reflectance collection. **(Script's line 5)**

Since the data catalog of S2 images is very large they can be filtered based on the information contained in the image metadata and with the `.filter()` function. Examples of filtering criteria for S2 images (metadata attributes) are: tile name, acquisition date, etc.

In this thesis, we chose two filtering criteria that allowed to select the time series of images to be processed. The first filter is applied to the tile name

`(.filterMetadata('MGRS_TILE', 'equals', MGRS_TILE))` by searching through the metadata the 'MGRS_TILE' attribute; if the content of the variable coincides with the value of the metadata attribute, the image is taken into account.

The second filter selects the list of images between the start and end date of the time interval (both dates are saved in the previously edited variables) `(.filterDate(START_DATE, END_DATE))`.

Only images that meet both filter requirements are loaded into the collection and the list of images satisfying the criteria are stored in the variable "s2_unsorted". **(Script's lines 6,7,8)**

Up to now the collection of images is correct, but it is unsorted. The JavaScript code sorted it in ascending order according to the property 'Cloudy_pixel_percentage', which identifies the percentage of image pixels covered by clouds. This sorting operation is useful for the user to be able to display a list of images from the most visible (i.e. lowest cloud cover) to the least visible; this saves time in searching for the best images by the user. **(Script's line 9)**

The number of S2 images contained in the collection are counted by the `size()` function to obtain the dimension of the collection (`var count`) that is printed on the console; in this way the user realises the number of images taking part in the time series. **(Script's lines 10, 11)**

The list of images is then made exportable in .csv¹³ format the file contains a table in which each row corresponds to an image (identified by its unique index) and each column to a property.

The output file is viewable in a Google Drive folder (accessible via the same Google account with which the user is logged into GEE). (*Script's lines 13-17*)

The user can run the download from the task manager window by clicking on the button 'run' [Figure 6. 4].

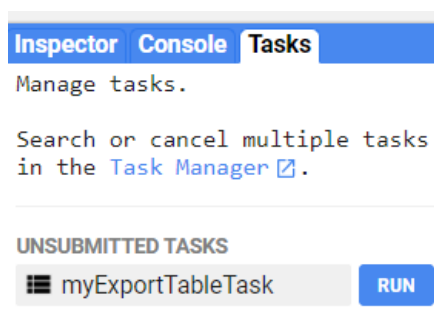


Figure 6. 4 Screenshot of Task Manager Interface on GEE platform

In the .csv file are grouped various information relating to the collection of images. These are the properties:

system:index, CLOUDY_PIXEL_PERCENTAGE, CLOUD_COVERAGE_ASSESSMENT,
DATASTRIP_ID, DATATAKE_IDENTIFIER, DATATAKE_TYPE,
DEGRADED_MSI_DATA_PERCENTAGE, FORMAT_CORRECTNESS, GENERAL_QUALITY,
GENERATION_TIME, GEOMETRIC_QUALITY, GRANULE_ID,
MEAN_INCIDENCE_AZIMUTH_ANGLE_B1, MEAN_INCIDENCE_AZIMUTH_ANGLE_B10,
MEAN_INCIDENCE_AZIMUTH_ANGLE_B11, MEAN_INCIDENCE_AZIMUTH_ANGLE_B12,
MEAN_INCIDENCE_AZIMUTH_ANGLE_B2, MEAN_INCIDENCE_AZIMUTH_ANGLE_B3,
MEAN_INCIDENCE_AZIMUTH_ANGLE_B4, MEAN_INCIDENCE_AZIMUTH_ANGLE_B5,
MEAN_INCIDENCE_AZIMUTH_ANGLE_B6, MEAN_INCIDENCE_AZIMUTH_ANGLE_B7,
MEAN_INCIDENCE_AZIMUTH_ANGLE_B8, MEAN_INCIDENCE_AZIMUTH_ANGLE_B8A,

¹³ A **comma-separated values (CSV)** is a text file that uses comma to separate values. Each line of the file is a data record and each record has one or more fields contains information separated by commas. Microsoft Excel has been used to create and read it in this project.

MEAN_INCIDENCE_AZIMUTH_ANGLE_B9, MEAN_INCIDENCE_ZENITH_ANGLE_B1,
MEAN_INCIDENCE_ZENITH_ANGLE_B10, MEAN_INCIDENCE_ZENITH_ANGLE_B11,
MEAN_INCIDENCE_ZENITH_ANGLE_B12, MEAN_INCIDENCE_ZENITH_ANGLE_B2,
MEAN_INCIDENCE_ZENITH_ANGLE_B3, MEAN_INCIDENCE_ZENITH_ANGLE_B4,
MEAN_INCIDENCE_ZENITH_ANGLE_B5, MEAN_INCIDENCE_ZENITH_ANGLE_B6,
MEAN_INCIDENCE_ZENITH_ANGLE_B7, MEAN_INCIDENCE_ZENITH_ANGLE_B8,
MEAN_INCIDENCE_ZENITH_ANGLE_B8A, MEAN_INCIDENCE_ZENITH_ANGLE_B9,
MEAN_SOLAR_AZIMUTH_ANGLE, MEAN_SOLAR_ZENITH_ANGLE, MGRS_TILE
,PROCESSING_BASELINE, PRODUCT_ID, RADIOMETRIC_QUALITY,
REFLECTANCE_CONVERSION_CORRECTION, SENSING_ORBIT_DIRECTION,
SENSING_ORBIT_NUMBER, SENSOR_QUALITY, SOLAR_IRRADIANCE_B1,
SOLAR_IRRADIANCE_B10, SOLAR_IRRADIANCE_B11, SOLAR_IRRADIANCE_B12,
SOLAR_IRRADIANCE_B2, SOLAR_IRRADIANCE_B3, SOLAR_IRRADIANCE_B4,
SOLAR_IRRADIANCE_B5, SOLAR_IRRADIANCE_B6, SOLAR_IRRADIANCE_B7,
SOLAR_IRRADIANCE_B8, SOLAR_IRRADIANCE_B8A, SOLAR_IRRADIANCE_B9,
SPACECRAFT_NAME, quality_check, system:asset_size, system:band_names,
system:bands, system:footprint, system:id, system:time_end, system:time_start,
system:version,.geo

The most interesting properties for this step of the work are the first two columns:

- 'system:index' . It is a string datatype that contains the identification code. It is important because it is unique for each image and through it the user can assume the acquisition date. This information is also necessary for selection as it is important that not too much time passes between two consecutive images in the time series so that burned areas can be better identified before the ground recovers. The lifetime of the burned area signal in multi-spectral images is very short.
- '-CLOUDY_PIXEL_PERCENTAGE'. It is a double datatype and it is a granule-specific cloudy pixel percentage taken from the original metadata that shows the

computed cloud cover percentage. This information is necessary because the less is the 'cloudy_pixel_percentage' value, the better its visibility is.

Table below shows an example of the first two columns for the .csv file downloaded for S2 tile 36KYE [Table 6. 2]:

system:index	CLOUDY_PIXEL_PERCENTAGE
20190720T072629_20190720T075356_T36KYE	0
20190913T072611_20190913T075035_T36KYE	0
20190928T072659_20190928T075119_T36KYE	0
20191008T072759_20191008T074520_T36KYE	0
20191028T073009_20191028T075426_T36KYE	0
20190730T072629_20190730T075417_T36KYE	0.0548
20191023T072951_20191023T074659_T36KYE	0.3528
20190804T072621_20190804T074810_T36KYE	0.9735
20190725T072621_20190725T074328_T36KYE	1.785
20190809T072619_20190809T074536_T36KYE	3.4583
20190829T072619_20190829T074450_T36KYE	4.3936
20190903T072621_20190903T074721_T36KYE	22.8555
20190918T072619_20190918T075013_T36KYE	24.7348
20190819T072619_20190819T075502_T36KYE	25.4313
20191018T072909_20191018T074707_T36KYE	26.0422
20191013T072841_20191013T074509_T36KYE	26.9653
20190814T072621_20190814T075029_T36KYE	27.0533
20190819T072619_20190819T074242_T36KYE	27.3828
20190908T072619_20190908T075454_T36KYE	40.32
20190824T072621_20190824T075230_T36KYE	58.4189

Table 6. 2 Example of two first attributes of .csv output related to tile 36KYE where 'system:index' is the image ID and 'CLOUDY_PIXEL_PERCENTAGE' is the percentage of image covered by clouds according to metadata information.

The best time series for mapping burned areas are selected directly by the user, following certain criteria:

Criteria1: Two consecutive S2 images should not be more than 16 days apart. This time interval has been defined based on the shortest lifetime of the burned area signal in multi-spectral images (Padilla et al., 2017). The duration of the fire spectral signal can be very short especially in grassland and savannah vegetation.

If the user takes a too long interval between successive images, he/she may not be able to detect surfaces that have been affected by fires or are in process of recover.

Criteria2: The cloud cover over the Sentinel-2 image tile must not exceed 10%. This condition assures that a sufficient portion of the image is observable with good atmospheric conditions (clouds, haze, smoke).

Image cloud cover percentage is used to select images least affected by clouds, however since it is an aggregated value over the entire image, it is recommended to inspect the individual raster image on the Map display. This task can be accomplished by using the `Map.Layer()` command, which takes the single image as input and loads it onto the map window as a layer.

Since only a portion of the Sentinel-2 image tile will be useful in the next steps of the work, the user can check if the cloud cover is accumulated in useful or discardable portions of the tile. In fact, there may be cases in which the percentage of clouds has a value close to 10% but covers an area of the tile that we do not need for the following phases of the work: this means that the portion of the territory we want to analyse is observed in good atmospheric quality conditions. On the other hand, it may happen that the percentage of clouds is lower than the threshold, but the image is still very poor due to the presence of shadows or haze.

An important role in this selection criterion is played by the radiometric resolution of the satellite image. It is measured in bit¹⁴ typically in the range of 8 to 16 bits. The

¹⁴ *In remote sensing, a bit stands for the number of grey-scale values a spectral sensor can tell apart. The greater the bit number, the greater the number of grey-scale values a spectral sensor can distinguish, and, therefore, the higher the radiometric resolution of a spectral sensor. One bit*

radiometric resolution of the Sentinel-2 MSI instrument is 12 bit, enabling the image to be acquired over a range of 0 to 4095 potential light intensity values.

A study on the impact of radiometric resolution on classification accuracy was carried out and although in some cases higher radiometric resolution resulted in up to 8% higher classification accuracy than lower resolution radiometric data, other results indicate that higher radiometric resolution does not necessarily imply better classification accuracy. Furthermore, the classification accuracy of spectral indices and texture bands is not so much related to the radiometric resolution of the original remote sensing images, but rather to their radiometric resolution itself. Overall, the results of this study suggest that data selection and classification need not always adhere to the highest possible radiometric resolution. (Natalia Verde, Giorgos Mallinis, Maria Tsakiri-Strati, Charalampos Georgiadis and Petros Patias, "Assessment of Radiometric Resolution Impact on Remote Sensing Data Classification Accuracy ", 2018, Remote Sensing). This study means that radiometric resolution is important, but that is not all: the quality of the time series images is equally important.

Here is a practical example of how two consecutive images are chosen.

A time series for tile 36KYE is shown in the figure [Figure 6. 5 Example of time series of s2 images over tile 36KYE of the three consecutive S2 imageFigure 6. 5] the dates of acquisition of the images are July 20th 2019, July 25th 2019 and August 4th 2019 respectively (such a short time span is possible due to the fact that Sentinel-2 acquires images of a point in space every 5 days). It is worth noting that all three passed the cloud cover filter of less than 10%, but it is evident that one of them (the one of July 25th) has a cloud cluster - with its shadows - at the bottom of the area).

For better accuracy, the user selected the July 20th and August 4th data as two consecutive images to be analysed during the next classification phase (the selection is

stands for a sensor that knows only black and white. 2 bit equals 4 grey-scale values and 4 bit 16 values. The equation is as follows: 2 to the power of bit = number of grey-scale values. The higher the bit, the more grey-scale values can be differentiated by a sensor.

also confirmed by the fact that they are no more than 16 days apart, so both criteria are met).



Figure 6. 5 Example of time series of s2 images over tile 36KYE of the three consecutive S2 image

6.2.2 Time Series and RGB Download

This second module of the algorithm takes as input the previously selected time series and returns RGB false colour composite images. These images will be downloaded by the user directly from the GEE Code Editor. This step is important because the RGB images in the GEE archive will be useful for comparing the burned areas visible on them with the output burned areas of the classification.

Since Earth Engine hosts an archive of geodata, they can either be viewed on the platform or downloaded to user's device.

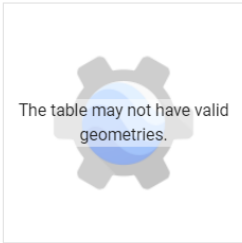
The script for the download is structured to take as input the dates of the previous images selected through a .csv file. The file is created by the user locally on his device (out of GEE environment) and then is uploaded on the Asset Manager as an external file. It has only one column containing the selected dates, one on each row.

Once the file is uploaded, the user can inspect it by clicking on it [Figure 6. 6].

Multi-scale analysis and comparison of Burned Area products derived from Earth Observation (EO) data over the African continent


Asset details DELETE SHARE IMPORT Edit

Table: dates_36KWE



The table may not have valid geometries.

Feature Index	system:index (String)	x (String)
0		23/07/2019
1		07/08/2019
2		17/08/2019
3		27/08/2019
4		11/09/2019

Table ID  users/martipastu/dates_36KWE

Date
Start date: NA
End date: NA

File Size 8.09KB

Number of Features 5

Figure 6. 6 Screenshot of Asset details for 36KYE

The .csv is imported into the script and read as a variable by the code [Figure 6. 7]. In this way, dates have no longer to be entered manually; the only variable to be changed by the user is the name of the tile.

```
Imports (1 entry)   
▶ var table: Table users/martipastu/dates_36KWE
```

Figure 6. 7 Screenshot of the imported table for 36KYE

The core of the script is a user defined function which selects one image at a time, renames it and selects the false colours ¹⁵ composite to display it (B_SWIR2, B_NIR, B_R). The function finally exports it to Google Drive with the `Export.image.toDrive()` command in 'GeoTIF' format¹⁶.

¹⁵ In contrast to a true colour image, a **false colour image**, in remote sensing techniques, is the display of data collected in a number of different wavelengths, usually longer or shorter than those perceptible to the naked eye. This facilitate the detection of features that are not easily distinguishable otherwise. The choice of triplet of spectral bands is governed by the physical properties of the object under examination. (AILSA ALLABY and MICHAEL ALLABY, "A Dictionary of Earth Sciences")

¹⁶ GeoTIFFs files are raster image file types that are commonly used to store satellite and aerial imagery data, along with geographic metadata that describes the location in space of the image. They are an extension of the raster format .TIF.

```
var show_exp = function(img) {
  var id = img.id;
  var image = ee.Image(img.id);
  var name = 'FireCCI_' + tile+ '_' + id.slice(17, 25);
  var image_rgb = image.select([B_SWIR2, B_NIR, B_R])
    .visualize({min: [0, 500, 0], max: [4000, 4500, 4000]});

  Export.image.toDrive({
    image: image_rgb,
    description: name,
    fileFormat: 'GeoTIFF',
    folder: tile,
    scale: 20,
    region: region,
    maxPixels: 1e13
  });
};
```

6.2.3 Classification (Random Forest Algorithm)

In this module, the classification of burned areas is carried out. This phase is divided into two operations: the first is the selection of training polygons, i.e. the samples that serve as input to the algorithm to classify the areas, the second is the implementation of the classification of burned areas. The statistical method used by the code to do the classification is **Random Forest Algorithm**.

Random forest (RF) is a Supervised Machine Learning¹⁷ Algorithm that is used widely in remote sensing for thematic mapping and land cover classification. This classifier has become popular within the remote sensing community due to the accuracy of its classifications. The RF classifier can successfully handle high data dimensionality and multicollinearity, being both fast and insensitive to overfitting. However, it is sensitive to the sampling design.

¹⁷“ Machine learning is a field of study that looks at using computational algorithms to turn empirical data into usable models. The machine learning field grew out of traditional statistics and artificial intelligences communities.” (Thomas W. Edgar, David O. Manz, in *Research Methods for Cyber Security*, 2017)

In this case the classification does not identify the type of soil cover (e.g. vegetation, bare soil, water etc), but separates the soil into two classes: "burned" and "unburned". This makes it possible to identify the presence of burned areas which indicate the occurrence of a fire between two consecutive satellites images of a time series.

From a statistical and mathematical point of view, the RF builds a multitude of uncorrelated decision trees on different samples; the class with the most votes becomes the model's prediction.

In decision theory, a **decision tree** is a graph of decisions and their possible consequences, used to create a goal-oriented 'action plan'. A decision tree is constructed in order to support decision-making action. Below there is a flowchart of a decision tree for classification problem [Figure 6. 8].

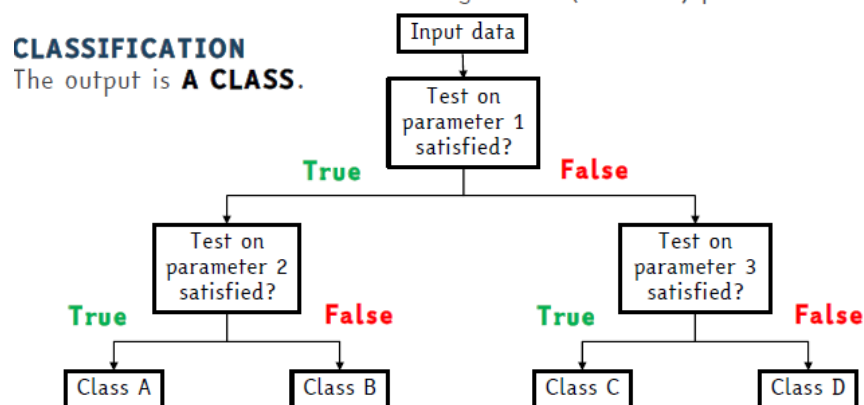


Figure 6. 8 Decision tree for classification problem

Since RF is an aggregation of decision trees, here there is a RF flowchart [Figure 6. 9].

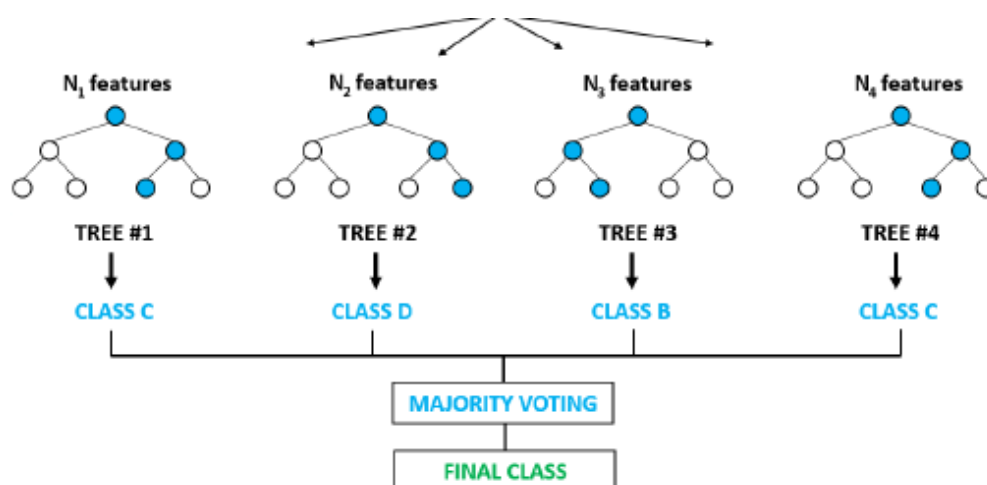


Figure 6. 9 Random Forest

Differences between Decision Trees and Random Forest are shown below [Table 6. 3 Differences between Decision Tree and Random Forest Algorithm Table 6. 3]:

Decision tree	Random Forest
One decision tree	A group of decision trees
When a data set with features is taken as input by a decision tree it will formulate some set of rules.	It selects observations, builds forest of decision trees and the average result is taken.
It is faster but low level of accuracy	It is slower but the accuracy is higher
Decision trees normally suffer from the problem of overfitting if it's allowed to grow without any control.	Random forests are created from subsets of data and the final output is based on average or majority ranking, so there is no overfitting problem.

Table 6. 3 Differences between Decision Tree and Random Forest Algorithm

Thus, the Random Forest Algorithm operates in this work as follows:

Step 1: Sampling phase.

The term 'sampling' refers to the selection of representative samples of a set of data in order to obtain results that can be generalised to the whole set. In this phase, n number of records are taken (manually) from the data set; they will be called *samples* or *training polygons*. Samples are drawn by the user on each tile for each pair of consecutive satellite images (the images considered are those previously selected for time series), visualized in the map display one pair at time. They are divided in two class of polygons: “burned” (drawing on burned areas in the time period between the two consecutive images) and “unburned” (drawing on unburned areas in the time period between the two consecutive images).

The samples used to train the supervised classifiers need to fulfil several requirements: (1) training and validation data must be statistically independent, (2) training samples must be class balanced, (3) training samples must be representative of the target classes, and (4) the training sample needs to be large enough to accommodate the increasing number of data dimensions. (Mariana Belgiu, Lucian Drăguț 2016)

Step 2: Decision tree generating

Individual decision trees are constructed for each polygon. Each node of each tree divides the model on the basis of a criterion (in this work is the difference between the two reflectance values of the first image -pre fire- and the next image -post fire- respectively. If the difference is high, it is very likely that there was a fire there and that the area can be classified as "burned area"). Each decision tree will generate an output that corresponds to one of the two classes (burned and unburned).

Step 3: Outputs

Final output is considered based on *Majority Voting or Averaging* for Classification and regression respectively. Each output refers to a pair of consecutive images for a specific tile and represents the areas burned in the time interval between the two satellite images.

The reasons why the Random Forest Algorithm is preferred to other supervised classification algorithms are:

- It is suitable to work on a mixture of feature data types (continuous, categorical, binary);
- No statistical model is assumed for data distribution;
- It mirrors human decision making (simple to understand and interpret).

The classification, therefore, is first carried out on the shorter time intervals (in this project it is called 'short time unit') which correspond to the days that elapse between the first image (considered as the pre-fire reference) and the second one (considered as the post-fire reference).

In this way it is possible to establish which areas have burned between one date and the next. Once this work has been carried out for the whole time series of a given tile, the outputs of the classification will be merged into a single shapefile representing the burned ground; this shape file refers therefore to a longer time interval which, in

comparison with the previous time unit, is called 'long time unit' (the duration of the time series is equal to the long time unit).

This procedure will be carried out later and outside the working environment of GEE.

In this thesis, classification using the Random Forest Algorithm is conducted on the Code Editor of Google Earth Engine. The mapping of burned areas is conducted with a trial and error approach. After training data collection, the classification is run and output burned areas maps are visualized to be compared with RGB false colour composites. This interactive step is necessary to judge the quality of the results, i.e., the agreement between areas identified as burned by the algorithm and the actual surface conditions in the image. If necessary, a new training set could be collected and the algorithm run again.

As seen for the previous operations performed on the GEE Code Editor, sampling is also performed by running a JavaScript code.

Steps overview

Step 1: Update variables

Step 2: First code run: layers selection and visualization

Step 3: Drawing training polygons

Step 4: Second code run: classification

Step 1: Update variables

The script is generic, but there are variables' values that must be changed.

```
var tile = '33MXR';  
var year = '2019';  
var date_pre = '2019-09-13';  
var date_post = '2019-09-28';
```

The "tile" variable is updated each time the region of interest is changed.

The variable "year" is fixed throughout the sampling, as the whole project analyses images from the year 2019.

The variables “date_pre” and “date_post” must be modified each time as the classification takes two consecutive dates to classify which areas are burned during each short time unit. The first date of the time series of a specific tile is only analysed in the “data_pre” variable at the first sampling. Dates from the second onwards in the time series are used twice: both as a value in the variable “data_post”, in pairs with the previous image, and as a value in the variable “data_pre” in pairs with the next image.

Step 2: First code run

After making the necessary changes, the run button is pressed to compile the code.

This step must be performed twice: the first run to display on the GEE Map Display the pair of consecutive satellite images (only by displaying the satellite images is it possible to select the training polygons, since the training is carried out by photointerpretation of RGB composite images with false colours necessary for the classification). The second run is used to carry out the actual classification by applying RF algorithm.

It is now analysed in detail what the JavaScript code does in this first phase, before stopping at the error message.

First of all, the code defines the region of the tile we are going to work on (area covered by the selected tile) and saves it in a variable:

```
region = ee.ImageCollection('COPERNICUS/S2_SR')
        .filterMetadata('MGRS_TILE', 'equals', tile)
        .filterDate(year+'-01-01', year+'-12-31')
        .geometry().dissolve().centroid().buffer(50100).bounds();
```

Then the code filters the 2019 Sentinel-2 image collection to select pre- and post- fire images consistent with the dates entered. These dates represent a pair of consecutive S2 images in the time series selected as described in section 1.2.2.1:

```
var post_image = ee.ImageCollection('COPERNICUS/S2_SR')
    .filterDate(date_post_1,date_post_2)
    .filterMetadata('MGRS_TILE', 'equals', tile);
var pre_image = ee.ImageCollection('COPERNICUS/S2_SR')
```

Multi-scale analysis and comparison of Burned Area products derived from Earth Observation (EO) data over the African continent

```
.filterDate(date_pre_1,date_pre_2)  
.filterMetadata('MGRS_TILE', 'equals', tile);
```

“Pre-image” and “post-image” are visualized in false colours (B_SWIR2, B_NIR, B_R) and are added on the map display via `Map.addLayer()` command.

```
var post_image_rgb = post_image.select([B_SWIR2, B_NIR, B_R])  
  .visualize({min: [0, 500, 0], max: [4000, 4500, 4000]});  
var pre_image_rgb = pre_image.select([B_SWIR2, B_NIR, B_R])  
  .visualize({min: [0, 500, 0], max: [4000, 4500, 4000]});  
  
Map.addLayer(pre_image_rgb, {}, 'Pre-fire', false);  
Map.addLayer(post_image_rgb, {}, 'Post-fire', true);
```

In addition, a third layer is also calculated and printed on map: the difference between the two consecutive satellites images, obtained with the `subtract()` function, which subtracts the values of reflectance for each pixel. This product is very useful for drawing training polygons because it immediately highlights the areas where the reflectance difference between the first and second image is so high that it is considered to be burned during the time interval.

The difference between the reflectances of the two images is also the decision criterion on which the algorithm is based, as described in the previous section. Therefore, if the difference between the two reflectances is high, it is very likely that a fire occurred in that time interval.

```
Map.addLayer(post_image_rgb.subtract(pre_image_rgb), {min:-100,  
max:100}, 'Difference', false);
```

The result is shown in the following figure [Figure 6. 10]:

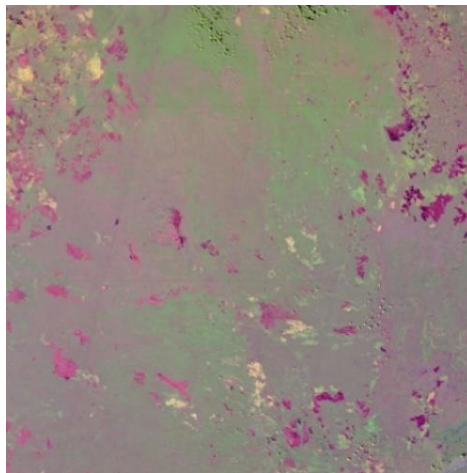


Figure 6.10 Difference between pre-fire image and post-fire image for the tile 36KYE

Step 3: Drawing training polygons

The samples, which will be given as input to the Random Forest Algorithm, are drawn directly by the user on the RGB false colour composite images visualized in the map layer.

When the user selects training polygons, he/she has to follow some rules:

- The analysts select training polygons according to their knowledge of the territory (not as a random selection). Training data must describe the characteristics of the data to classify.
- All land cover classes (information classes) must have training samples (spectral classes); in this case the land cover classes are “burned” and “unburned”.
- Select training samples in homogenous areas (unimodal class)
- Select training samples in multiple locations to account for its natural (spectral) variability
- DO NOT SELECT outliers = samples with “strange” spectral values
- DO NOT SELECT samples across different land cover classes (mixed class)
- Samples have to be enough. The optimal number of samples depends on: data used, territory studied and classification algorithm used. It’s important to select enough samples to create a link between spectral classes and information classes. The more they are, the better is, but the user has to consider the cost benefit analysis.

The user creates two different geometries (each geometry corresponds to a class): burned and unburned. There is an example of map display with two training polygons below [Figure 6. 11].

The selection of a geometry enables the drawing option; each new drawn polygon is a feature that is added to the selected class.

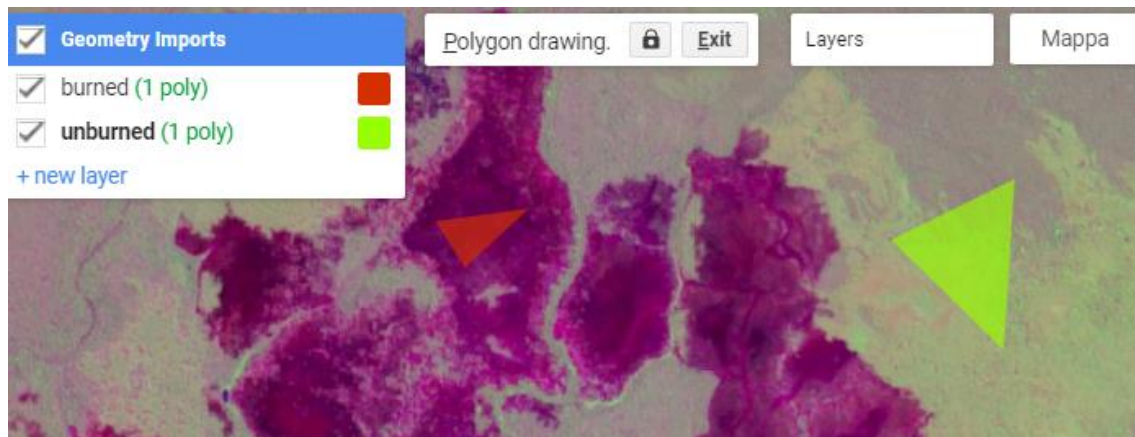


Figure 6. 11 Training polygons on difference layer for the tile 36KYE

Training polygons must be selected manually according to predefined criteria:

- It is important to choose the appropriate number of samples. Although the overall accuracy of the RF classifier increased as the training sample size increased, the overall RF accuracy started to stabilise when the sample size reached 1250 (Ramezan, Warner, Maxwell, Price, 2016). This shows that taking too many samples will not continue to improve an already stabilised accuracy but will only waste time.
- It is important to take samples of unburned areas on all land types (bare ground, vegetation, water, area already burned etc.). This increases the accuracy of the classification.
- It is important to take burned class samples of different severity (the more intense is the colour, the more recent is the fire).
- If the preliminary work is done on the difference layer, it is important to check the pre-fire and post-fire satellites images to ensure that the difference between the pixels is indeed a burned area (and not a difference caused by cloud shadows or haze).

NB: Areas already burned in the pre-fire image are considered in the 'unburned' class. The 'burned' class only counts the burned areas in the time interval between the two images, and not before.

Step 4: Second code run and results of classification

Once the code is executed again, the training geometries are imported automatically and saved in the respective variables [Figure 6. 12]. This operation is possible only if the names of the geometries and variables match.

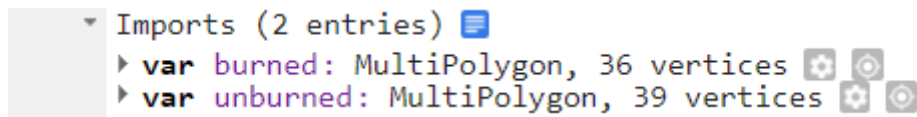


Figure 6. 12 Screenshot of the two geometries imported as variables

All samples are then merged into a single variable defined as 'training', within which the two classes still remain distinct.

The Random Forest algorithm is contained in GEE's "Classifier" package; thanks to this package, the core of classification is reduced to very few commands:

-Get a classifier with the command `ee.Classifier` and save it into a variable called "RF".

-Train the classifier using `RF.train()`, with default parameters.

-Classify an Image with the same bands used for training with the command `classify()`.

Below there is the JavaScript code:

```
var RF = ee.Classifier.smileRandomForest({numberOfTrees:100,  
    minLeafPopulation:10});  
RF = RF.setOutputMode('PROBABILITY');  
var trained = RF.train(training, 'class', bands);  
var classified = diff_image.select(bands).classify(trained);
```


The output of the classification is a layer containing the polygons of the areas classified as burned. Only classified burned areas will appear on the map display; anything that is not part of this layer will be interpreted as not burned during that time interval.

The figure shows an example of the burned area layer overlaid on the difference layer [Figure 6. 13]:

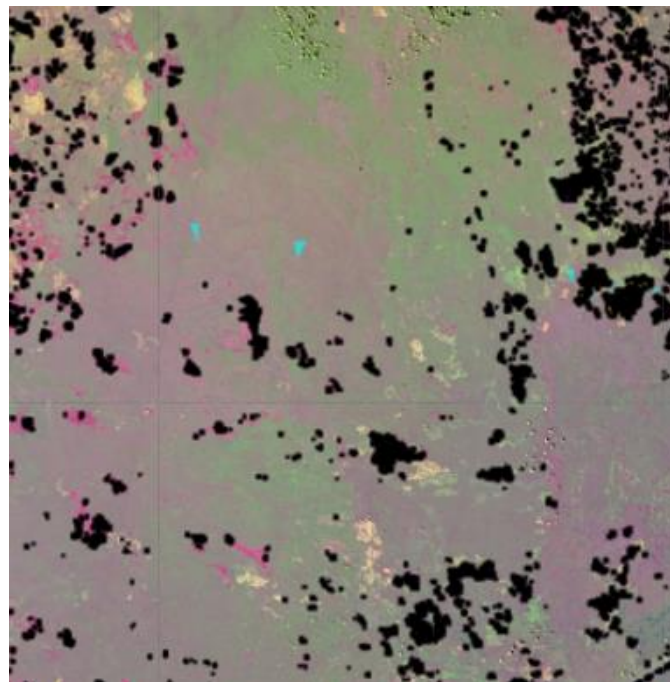


Figure 6. 13 Example of an inaccurate first output

It's easy to notice that many slightly burned areas, which are visualised in the difference layer in a light pink, are not part of the classification's result, and this will produce an inaccurate result. For them to be considered as burned areas, it is necessary to take new samples on the wrongly classified areas and to perform a new run; the new output obtained can be better, although 100% accuracy is impossible.

Here there is the result obtained after manual correction by the user [Figure 6. 14], a zoom on the top left of the tile is also shown, to better observe details [Figure 6. 15].

Multi-scale analysis and comparison of Burned Area products derived from Earth Observation (EO) data over the African continent

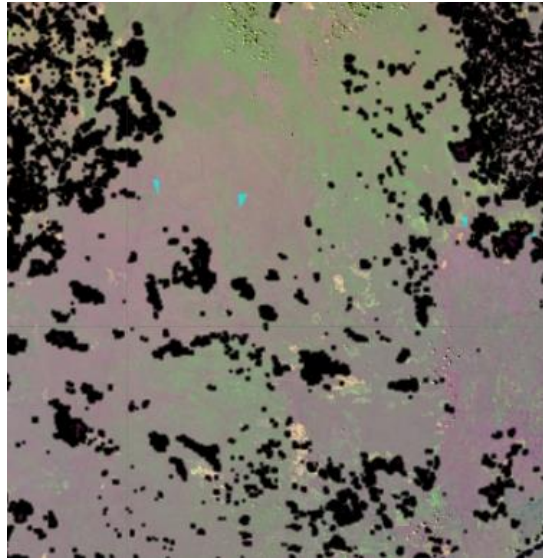


Figure 6. 14 Example of output of classification after manual correction

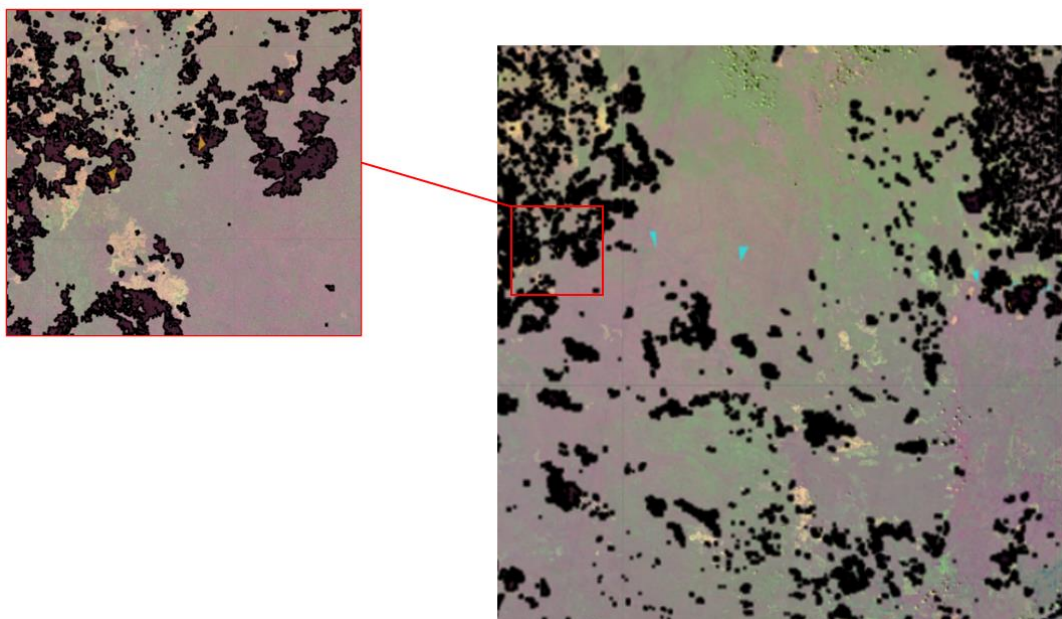


Figure 6. 15 Zoom on the top left part of the previous figure

It follows that it is advisable to carry out a visual check on the results in order to make sure that the classification is as correct as possible.

Another common error, as already mentioned, is that low albedo surfaces (cloud shadows, water courses) can be mistaken by the algorithm as burned areas. Again, new training polygons need to be added to raise the level of classification accuracy.

It can therefore be deduced that one of the main advantages of manual sampling is that it is more accurate than automatic sampling because it allows the user to instruct the algorithm. In fact, the first display check greatly reduces errors, making the classification more accurate.

6.3 Classification outputs

The burned area classification algorithm implemented in GEE allows the user to view and download a file containing the classified burned areas (as polygons) and other ancillary products, useful for the analysis. Output shape files are Cloud Mask, Region (area of the selected tile) and Training Polygons.

Here all outputs will be analysed in detail:

i. Burned Areas (BA)

The product of the algorithms is the result of the classification of a pair of consecutive S2 images: this is a set vector files (one for each pair of images) in which polygons represent burned areas and clouds. It can be downloaded in .KML format. Each shapefile associated to a short time unit is merged to form a single product covering the whole long time unit, which will then be validated with the Planet data and analysed. In the following chapters all the steps of this work will be explained.

The images represent an example of the processing of the time series from a series of consecutive images to a single layer containing the burned areas in that time interval [Figure 6. 16].

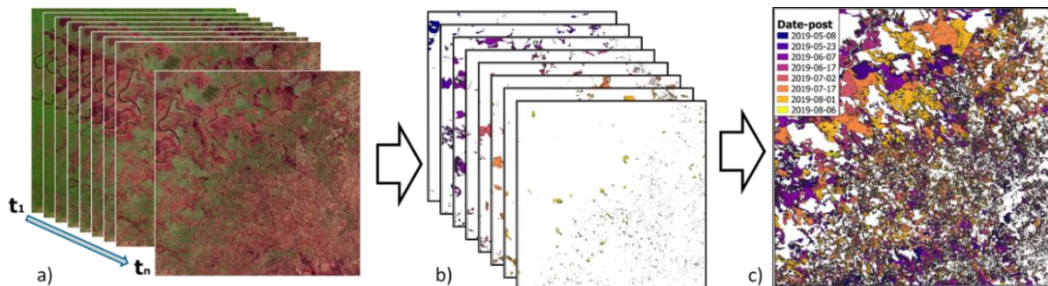


Figure 6. 16 a) Time series of S2 images over a sample tile (35LMD, Zambia, Tropical savanna High); b) fire polygon derived with RF classification for each S2 image pair; c) synthetic fire reference perimeter products over the series length.

Sentinel-2 products, referred to short time units, must be merged into a single vector layer to create the long unit with the Merge tool (From the menu bar: *Vector -> Vector Management Tool -> Merge Vector Layer*).

The long unit is a shapefile (one for each AOI), which contains all information about burned areas and clouds.

ii. Cloud mask:

The cloud mask file contains geometries of clouds present for each consecutive image pair. The JavaScript code associates the pixels to the clouds according to the value of the SCL parameter. An example below, in which red polygons are clouds [Figure 6. 17].

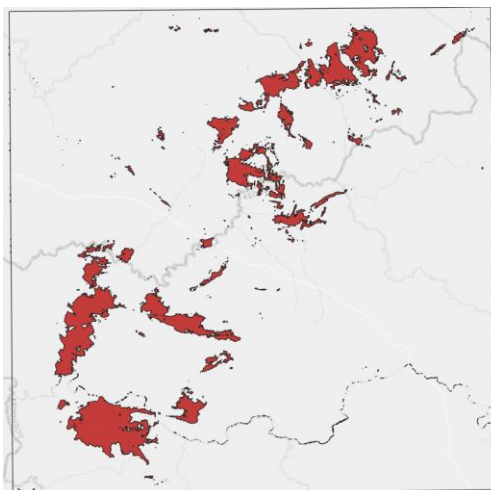


Figure 6. 17 Cloud Mask layer for the tile 28PVG 17 Jan 2019 - 27 Jan 2019

As already seen in the description of the datasets used for this work, the MSI Sentinel-2 Level-2A dataset has fifteen bands of data: twelve data bands of multispectral observations and three data bands which are used to improve the accuracy of surface reflectance-derived products (SCL, AOT, WVP).

In particular, SCL stands for “Scene Classification Layer; the objective of this product is not to classify the land cover in a strict sense, but to distinguish between cloudy, clear and water pixels.

The resulting SCL is applied in the cloud mask analysis, which removes unwanted cloudy pixels from the multispectral observations.

In the Javascript code, the operation to identify the clouds is carried out by the function `mask_sentinel(image)` that selects SCL values other than 8 (Cloud medium probability), 9 (Cloud high probability), 10 (thin cirrus), which are the classes associated with the clouds).

The part of the script that contains this function is as follows:

```
function mask_sentinel(image) {  
  var mask = image.select('SCL').neq(8)  
    .and(image.select('SCL').neq(9))  
    .and(image.select('SCL').neq(10));  
  var masked = image.updateMask(mask.eq(1));  
  return masked;  
}
```

The code processes the pixels classified as clouds to produce a user-downloadable output file; It is a vector file in KML format containing all the clouds classified by the automatic algorithm. There is one product for each analysed image pair, and it is downloaded by pressing the appropriate button in the Console [*Figure 6. 18*] created by the following lines of JavaScript code:

```
var button_export_cloud_mask = ui.Button({  
  label: 'Export Cloud Mask',  
  onClick: download_CloudMask  
});  
print(button_export_cloud_mask);
```

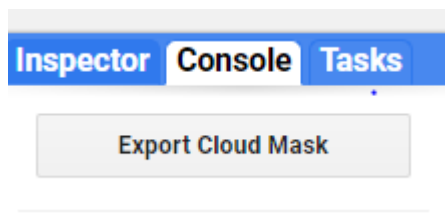


Figure 6. 18 Cloud Mask Button on Console Interface

iii. Region

The file product “Region” represents a shapefile covering the extent of the area of the tile we are considering. It is the same for the entire time series of a selected tile. It is therefore only necessary to download it once when sampling any of the consecutive image pairs.

There is a function in the script that takes as input the region of interest already selected in a previous stage of the code, renames it and transforms it into a format that can be downloaded to Google Drive.

The function used to export the region is as follows:

```
function download_region() {  
  var BA_region =  
  ee.FeatureCollection(ee.Feature(region2));  
  var name = 'Region_' + tile;  
  var folder = tile;  
  Export.table.toDrive({collection: BA_region,  
description: name, folder: folder, fileFormat: 'SHP'});  
}
```

The file is downloaded, as in the previous case, from a button in the Console. The portion of code for creating the button is consistent with that reported for the Cloud Mask product.

iv. Training

The algorithm has been coded so that also the training datasets used for classification can be exported. The product “Training” is composed of two files containing the polygons drawn by the user during the training phase of classification algorithm. These files are vector files, in shapefile format, and contains the geometries for the burned and unburned training polygons defined by the user.

This product is useful for conducting post-classification checks. By reloading the file and importing it into the script, it is possible to recreate the same samples and reproduce the same classification performed previously.

This can also be downloaded from the task manager after pressing the 'Export Training' button on the console. The GEE code used to create the button and to download the product is consistent with that shown in the previous paragraphs.

An example of training polygons and region in figure below [Figure 6. 19]:

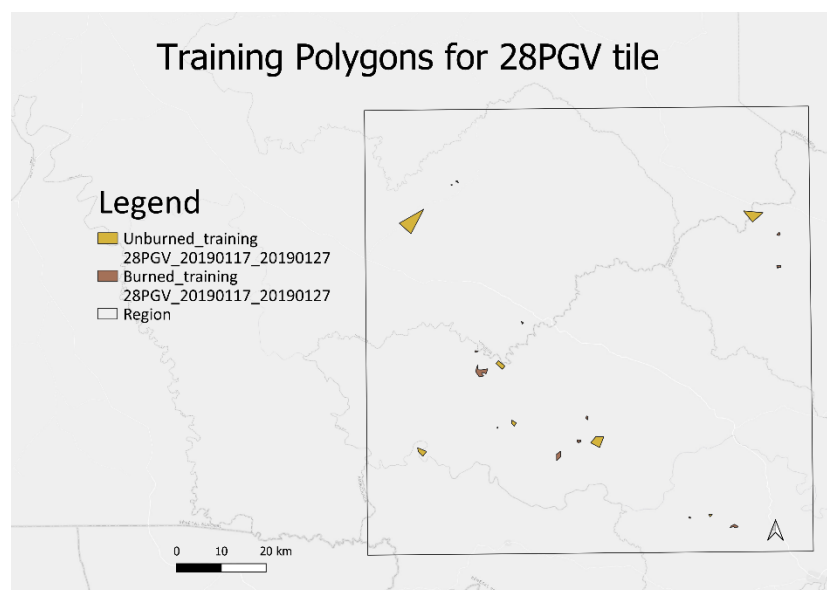


Figure 6. 19 Training Polygons vector layer and region for 28PGV 17 Jan 2019 - 27 Jan 2019

7 Validation

Validation can be defined as “The process of assessing, by independent means, the quality of the data products derived from the system outputs” (European Space Agency, 2007; Morisette et al. 2006). A validation phase is required after the classification phase to evaluate product’s accuracy and uncertainty and it is a key information for scientists and end-users (Congalton and Green 1999). For large-area (global and/or continental) remote sensing products, validation is even more challenging due to the wide variety of ecosystems and climatic conditions that may affect the mapping accuracy of the mapping (Chuvieco et al., 2008). The term “accuracy” refers to the closeness to ground truth or the surface condition to be assumed as ground truth (Justice et al. 2000), whereas "uncertainty" generally refers to the probability that a pixel is burned based on the sources of error within the input data and the algorithm. In remote sensing, uncertainty assessment is the quantification of the error associated with all phases of data acquisition and processing even though the sources of uncertainty are often unknown and difficult to quantify.

Validation refers to the comparison of remotely sensed products with ground truth or information that is assumed as the true state of a target. Ideally, ground truth should be collected by field surveys, but ground based burned area measurement is time consuming and difficult to undertake over large regions (Cardoso et al., 2005). Since it is not possible to collect samples of burned area independent reference data from ground measurements over large areas and/or in remote regions (i.e., Africa), it is widely accepted to use as reference data derived from remotely sensed images (Boschetti et al., 2016). In this thesis we compare the results of the classification of burned areas mapped from Sentinel-2 data (as explained in the previous chapter) with burned areas classified from Planet data. Given the higher spatial resolution of Planet images, fire perimeters can be assumed as ground truth. The results of the validation are the so-called accuracy metrics derived from the confusion matrix, that shows the agreement/disagreement between reference and classification, and a subsequent evaluation of the sources of error. The validation protocol applies in this thesis for the validation of Burned Areas (BA) products consisted of the following steps:

1. *Generation of reference fire perimeters*: reference fire perimeters are the fire perimeters derived from Planet very high resolution images (as described in chapter 5);
2. *Comparison to S2 classification of burned areas and computation of confusion matrices*: S2- and Planet perimeters of the areas affected by fires are compared to derive for each of the five test sites a confusion matrix (Congalton et al., 1983) that is a two-way matrix composed of proportions of burned and unburned correspondences between the product and the independent reference over the common area;
3. *Calculation of accuracy metrics*: accuracy metrics are used to quantify the following accuracy metrics were calculated from the error matrix computed on reference regions, the accuracy metrics: commission error ratio, omission error ratio, Dice coefficient (DC) were calculated. This will be explored in the Validation chapter.

As this thesis is part of the larger FIRE_CCI project, the guidelines of validation activities are fully described in the Production Validation Plan (PVP) that are followed. The PVP can be consulted here:

https://climate.esa.int/media/documents/Fire_cci_D1.3_PVP_v2.0.pdf

7.1 Validation Methods

The comparison between the burned areas classified on Sentinel-2 data and the reference Planet products is carried out in the QGIS¹⁸ workspace. Technically, the comparison is done between burned area maps in vector format (as shapefile ©).

Two input layers are required to perform the comparison: shapefiles contain polygons of areas classified as burned for the time interval covered by the temporal series of S2 and/or Planet images (i.e., long unit). In order to have the correct input layers, a data pre-processing step is required.

¹⁸ “Quantum GIS (QGIS) is an open-source Geographic Information System that supports most geospatial vector and raster file types and database formats. The program offers standard GIS functionality, with a variety of mapping features and data editing.” Definitio on QGis by Center of Research and Capacity Development.

The output result of this validation steps as implemented in QGIS are three different vector layers:

- **Agreement areas** (shapefile that contains the burned area polygons common to both input files, i.e. the areas where the classification product agrees with the reference);
- **Omission areas** (shapefile that contains the burned area polygons not classified as burned but burned according to the reference);
- **Commission areas** (shapefile containing the polygons classified as burned areas, but records as unburned in the reference).

The total burned areas in square meters is quantitatively calculated for each product. This value is needed for the next step of comparison, which estimates errors and thus calculates the accuracy of the classification.

Once the above files are derived in QGIS, the computation of the accuracy metrics is carried out on Excel; it previews the calculation of the error of **commission** and **omission** (percentage of area mistakenly classified) and the **coefficient of Dice** (indicator of the level of similarity between the two inputs of the comparison). Once the errors have been calculated, the user will make considerations about the sources of error and estimate their influence on the result. Some errors of considerable importance and uncertainty are investigated by studying the sources of error by analysing the Active Fire dataset and the spectral signature of the surface erroneously classified.

Classification errors, as the name suggests, refer to pixels that are not identified as the same category in the reference data and in the classified image, i.e. pixels that are not burned according to both products. There are two types of errors: underestimation, i.e. **omission error**, and overestimation, i.e. **commission error**.

The error of omission is the error of classification, in which pixels that are burned according to the reference are classified as burned (Eq. 2). In other words, errors of omission quantify the number of pixels left out of the BA class: areas that are burned by the PlanetScope reference data, but not by the Sentinel-2 data classification.

The commission error is the misclassification error, where pixels are classified as burned but do not belong to this class according to the reference (Eq. 1). Both metrics are expressed in terms of percentage: the smaller the error, the better the result of classification.

$$COMMISSION\ ERROR = \frac{Sample's\ Area - Agreement\ Area}{Sample's\ Area} * 100 \quad (1)$$

$$OMISSION\ ERROR = \frac{Reference's\ Area - Agreement\ Area}{Reference's\ Area} * 100 \quad (2)$$

$$DICE\ COEFFICIENT = \frac{2 * Agreement\ Area}{2 * Agreement\ Area + Omission\ Area + Commission\ Area} * 100 \quad (3)$$

Another accuracy metric included in this study is the Dice coefficient (DC) that summarises omission and commission errors in a single metric (Dice, 1945): given that one classifier (product or reference data in our case) identifies a burned pixel, DC is equal to the conditional probability that the other classifier will also identify it as burned (Padilla et al., 2015). This index is expressed in terms of percentages, and unlike errors, the higher the DC, the better the validation result.

Finally, the Pareto Boundary (Boschetti et al., 2004) is also calculated to graphically display and analyse the accuracy of the classification as a function of the spatial resolution of the input data (S2, in this case). The Pareto Boundary can help to understand if the accuracy of the S2 classification is due to the spatial resolution or to the poor performance of the algorithm.

Below there is the Flowchart of the Validation phase [Figure 7. 1]:

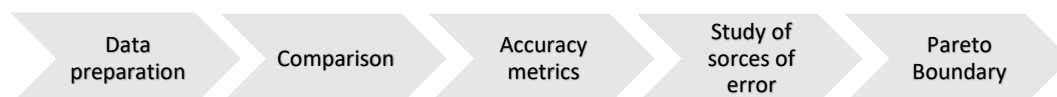


Figure 7. 1 Flowchart of Validation phase

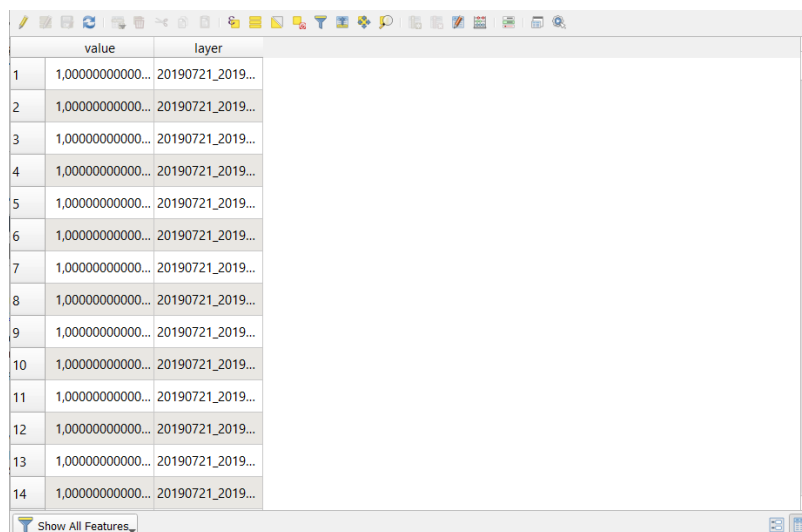
7.1.1 Data Preparation

The objective of these operations are two vector layers (one derived from the Planet reference classification and one from the Sentinel-2 sample classification) which contain information on the areas burned in approximately the same time interval and over the same area.

Planet BA maps

The Planet BA products that were used as reference are vector files (shapefiles @ format) that contain polygons of the areas burned. The attribute table associated with the shapefile is composed of the following fields: **layer** (which contains integer values between 1 and 2, with 1 indicating the polygons for burned areas and 2 indicating those for unburned areas) and **Area** (contains the area of the polygon in square meters). The BA shapefile is prepared for validation by deleting the polygons referring to unburned areas and keeping, instead, only the information related to the polygons of burned areas.

This operation is carried out with the QGIS "select by expression" command which can be found directly on the attributes table; all the selected polygons are saved in a new shapefile. Keeping only polygons referring to burned areas makes further steps (comparison with reference) quicker and easier. In the figure below there is the attribute table of the product [Figure 7. 2]:



	value	layer
1	1,000000000000...	20190721_2019...
2	1,000000000000...	20190721_2019...
3	1,000000000000...	20190721_2019...
4	1,000000000000...	20190721_2019...
5	1,000000000000...	20190721_2019...
6	1,000000000000...	20190721_2019...
7	1,000000000000...	20190721_2019...
8	1,000000000000...	20190721_2019...
9	1,000000000000...	20190721_2019...
10	1,000000000000...	20190721_2019...
11	1,000000000000...	20190721_2019...
12	1,000000000000...	20190721_2019...
13	1,000000000000...	20190721_2019...
14	1,000000000000...	20190721_2019...

Figure 7. 2 Attribute table for Planet BA ng unit product, site 2C as example

The result of this data preparation is shown in the figure below [Figure 7. 3]:

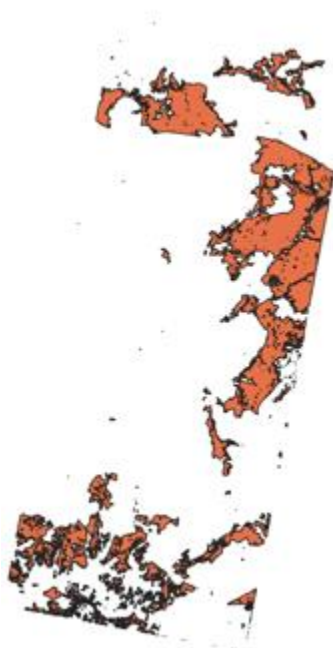


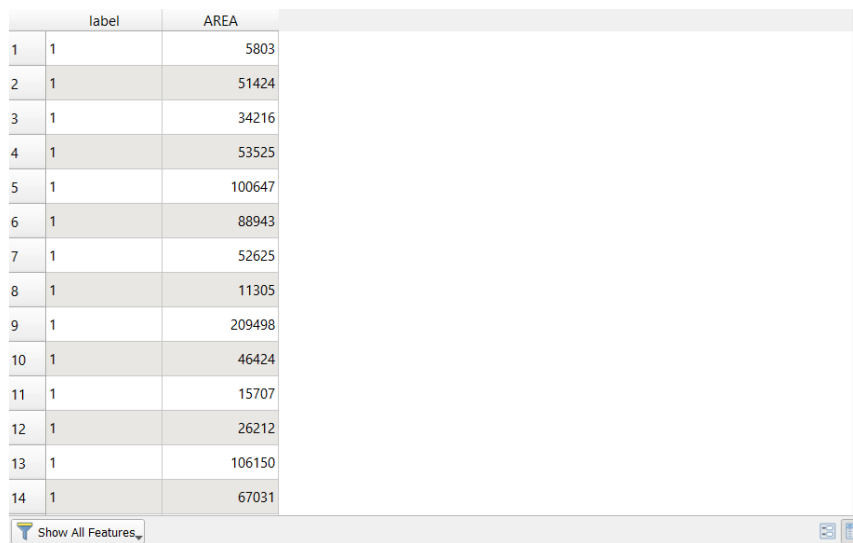
Figure 7. 3 Planet BA product after Data Preparation phase

Sentinel-2 BA maps

As seen in the chapter related to the presentation of the data, the area covered by the Planet data is smaller than that of the Sentinel-2 tile. Sometimes the Planet is contained in a single tile, other times -as in the case of sites 2C and 3B- it is contained in two different tiles. In any case the Sentinel-2 BA product is clipped on the Planet mosaic area since validation is carried out over the common overlapping area (Figure 5.6): the result is a smaller shapefile that is spatially consistent with the reference against which it is to be compared. This is done in QGIS with the following commands: vector-> Geoprocessing Tool -> Clip. The tool clips an input vector layer (Sentinel-2) using the feature of an additional overlay layer (Planet). Only the parts of the features in the Input layer that fall within the polygons of the overlay layer will be added to the resulting layer, but attributes of the features are not modified.

As seen for Planet, Sentinel shapefiles contain two columns in the attribute table: one is **Area** (Area of polygons in square meters) and the other is **label** ("label=1" stands for burned areas and "label=2" stands for cloud mask) [Figure 7. 4]. Unlike the Planet BA product, the output S2 BA maps depict burned and unburned areas as well as cloudy

regions; since clouds prevent from observing the surface conditions cloud cover in any of the scenes composing the temporal series for each AOI is preserved in the final map. The BA shapefile on the long time unit, is cleaned up by removing the polygons that refer to clouds. This operation is done again by the "select by expression" command, where the user selects the areas that have the attribute "label=1", then right-click on the *layer* -> *Export* -> *Save selected features as* -> *save all selected polygons* in a new shapefile.



	label	AREA
1	1	5803
2	1	51424
3	1	34216
4	1	53525
5	1	100647
6	1	88943
7	1	52625
8	1	11305
9	1	209498
10	1	46424
11	1	15707
12	1	26212
13	1	106150
14	1	67031

Figure 7. 4 Attribute table for Sentinel BA product long unit, 2C

Notice that the time interval may vary across AOIs according to the availability of satellite images of the two different EO missions; the dates of acquisition of the images are always different, although they are chosen close as possible, and this can become a source of error between classified and reference BA products as we will see later.

In the figure below an example of the data preparation results, from Sentinel-2 short unit to clipped Sentinel-2 long unit for the site 1A [Figure 7. 5].

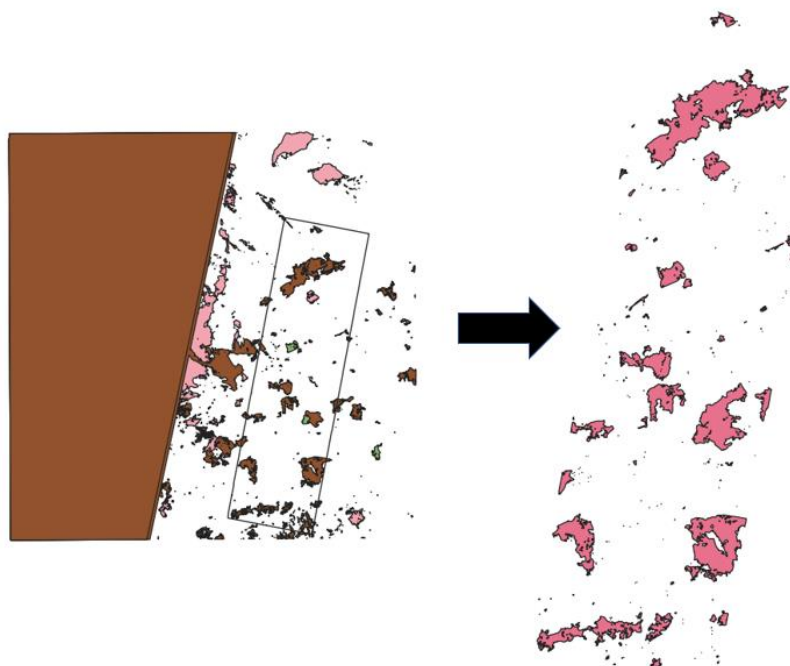


Figure 7. 5Data preparation results from short unit to clipped long unit for the site 1A

7.1.2 Comparison and accuracy metrics computation

After preparing the BA maps from Planet and Sentinel-2 for comparison, the following paragraphs explain the heart of validation: comparison between two products and error evaluation.

Comparison

Once the two shapefiles with the burned area polygons are ready, the first operation the user performs is to calculate the burned area common to both products. This operation is carried out with the Intersect Tool, an algorithm that extracts the overlapping portion of polygon in the two Input shapefiles. The path to perform this operation is *Vector -> Geoprocessing Tools -> Intersection*. The result is a shapefile with the common burned areas, previously called **Agreement Areas**.

Once the common burned area polygons have been extracted, the next step is to calculate the commission and the omission areas; both can be extrapolated in QIS using the Difference Tool (*Vector -> Geoprocessing Tools -> Difference*). This algorithm extracts features from the Input layer that fall outside, or partially overlap, features in

the Overlay layer. Input layer features that partially overlap feature(s) in the Overlay layer are split along those features' boundary and only the portions outside the Overlay layer features are retained.

Therefore, since the **Commission Areas** was calculated by making the difference between the S2 BA sample layer (entered in the algorithm as input) and the Agreement Area (entered in the algorithm as overlay layer).

For the **Omission Areas**, the difference between the reference product (entered as input) and the Agreement Area was calculated.

Areas Computation

Once we have obtained all the layers, it is important not only to display the burned area, but also to calculate a quantitative value in square metres, so this information is extrapolated for all the layers: the S2 BA product, the reference BA product, the Agreement burned area, the omission and the commission burned areas (in total there are 5 numerical values). This operation is very simple and is carried out in the QGIS environment.

Before calculating the area, the "dissolve" operation is performed on all products; the algorithm converts all polygon geometries into a multiple geometry where common boundaries of adjacent polygons being dissolved are erased. The path is *Vector -> Geoprocessing Tool -> Dissolve*. The Dissolve tool's results are five shapefiles with a single polygon each (only one row in the attributes table). After this operation, Burned Area calculation is faster and it is performed by simply selecting the area function from the QGIS *Field Calculator*, which is opened from the attribute table; this last operation creates a new column containing the area in square metres.

3 Metrics computation

Once all the polygon area values have been calculated, accuracy metrics (Eqs. 1,2,3) are computed in MS Excel spreadsheet following equation 1 to 3.

Although these assumptions are correct, it is difficult to establish what is the threshold percentage for which a result can be considered valid, because there is no

unambiguous rule, but it goes to the discretion of the user according to the conditions of the case study. In this case, no numeric threshold was fixed; results are compared with literature.

Analysis of the sources of classification errors

Once the metrics have been calculated, the user selects some larger error areas to be analysed in detail to understand the likely source of the errors and consequently its impact on validation. In some cases, it is considered appropriate to take samples of ambiguous or very large errors to study them in more detail, as follows:

- i. Investigating whether there are active fires in the misclassified area and in the time interval considered. To perform this operation the user intersects the work area with the shapefile of active fires points (introduced in chapter 3.3 Active Fires), filtered for the dates of interest. This operation is carried out in the QGIS.
- ii. Analysing the reflectance mean values in the S2 wavebands over the error region. This analysis is carried out by drawing a polygon in QGIS over the error area to be analysed and then giving it as input to a GEE algorithm that extracts the reflectance values from S2 images. The algorithm is written ad hoc for this purpose as a JavaScript code. The core of the algorithm, however, is given below:

```
1 //convert image collection in a list of features
2 var maxFeatures = 1000;
3 var data = S2_collection.toList(maxFeatures)
4   .map( function (img) {
5     img = ee.Image(img);
6     var meanValues = img.reduceRegion({
7       reducer: ee.Reducer.mean(),
8       geometry: area,
9       scale: 30
10    });
11 return ee.Feature(img).select(['system:time_start']).
12     set(meanValues);
13 });
14 //plot the chart
```

```
15 var chart= ui.Chart.feature.byFeature({
16   features: data_,
17   xProperty: 'system:time_start'
18 })
19 .setOptions(
20   {
21     title: 'title',
22     hAxis: {
23       title: 'Dates',
24       titleTextStyle: {italic: false, bold: true}
25     },
26     vAxis: {
27       title: 'Reflectance',
28       titleTextStyle: {italic: false, bold: true}
29     },
30     lineWidth: 2,
31     series: {
32       9: {lineWidth: 0, visibleInLegend: false},
33     }
34   });
35 print(chart);
```

The first part of the code concerns the import of the error sample area and the import of the time series in which it falls, filtered by tile and date. This script is not shown as it is very similar to the scripts in the previous chapter.

The code **lines [1-13]** transform the image collection into a list of features thanks to the function `toList()`. Each element of the list contains several pieces of information: the date of acquisition of the image ('system:time:start') and the mean reflectance for each bands that is calculated with the command `ee.Reducer.mean()`.

In code **lines [14-34]** a graph is created using the function `ui.Chart.feature.byFeature()`, which directly takes as input the list of features previously created and plots it on a Cartesian plane (x-axis is the attribute 'system:time_start' and y-axis is the reflectance). Hence, each line in the graph represents the reflectance trend over time for each band. Below there is an example of a result for a sample on the Agreement Area for tile

28PGV [Figure 7. 6] (Agreement Area is chosen as example of how the reflectance of a burned area behaves.

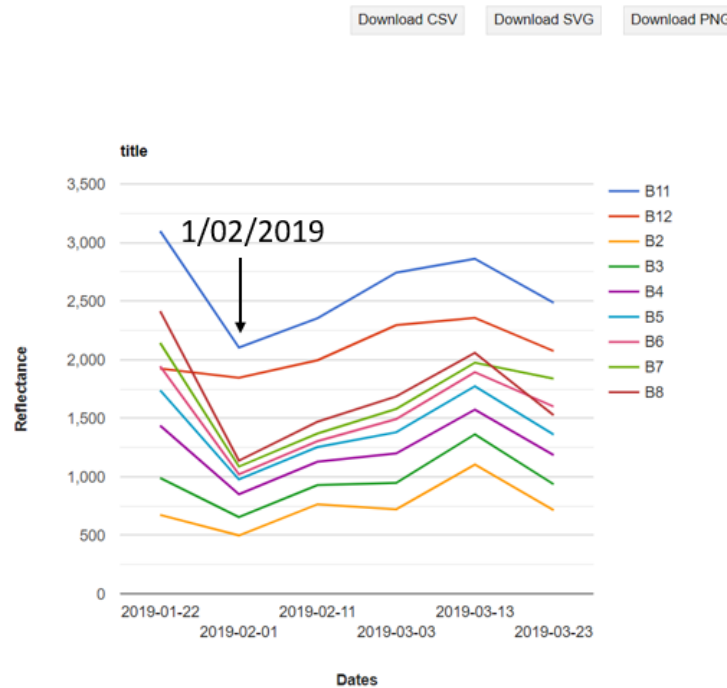


Figure 7. 6 Plotted chart on GEE Code Editor Console. Reflectance trends as function of dates for each band for 28PGV tile, agreement sample area

Line 35 is used to print the chart in the GEE console.

Outputs can be saved either as an image or in .csv format. In this thesis the plot is displayed to observe the behaviour of the reflectance and, if the user feels that the displayed information is useful for further analysis, it is downloaded both as an image and in .csv format to then draw the spectral signature. The spectral signature can also be visualized to compare reflectance values as function of the wavelength for different sources and/or for different acquisition dates [Figure 7. 7].

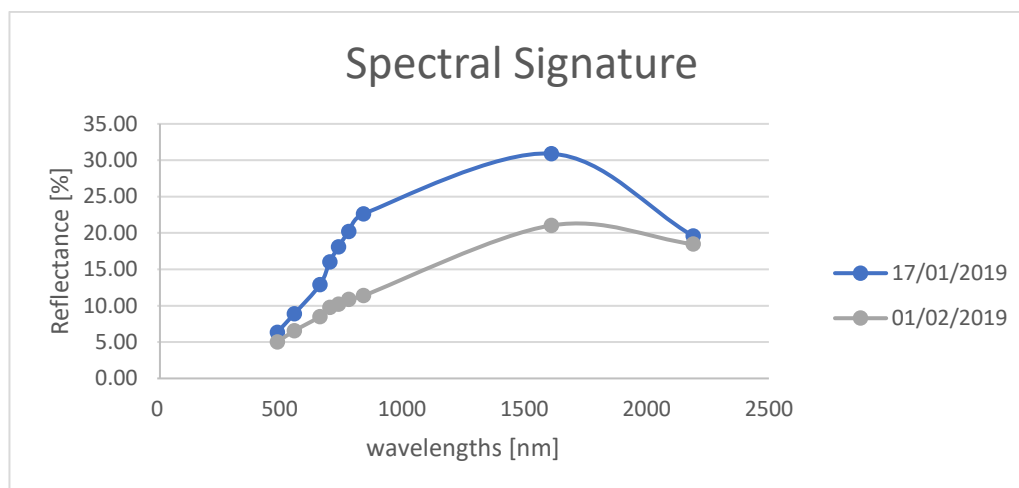


Figure 7. 7 Spectral Signature for 28PGV derived for a region of Agreement (both S2 and PPlanet identified the area as burned). Spectral signature as a function of the wavelength of S2 bands (x-axis) is displayed for different dates (colors).

By observing the trend of the spectral signature, it is possible to confirm whether the area is burned or not, or whether the surface is covered by clouds or cloud shadows. From this example it is possible to see that the area is actually burned by the negative reflectance peak on 1/02/2019 in **Figure 7.6**; the same negative reflectance peak is visible in the spectral signature **[Figure 7.7]** of that date, with the lowest reflectance values of all. On previous dates the reflectance is higher (especially on January 17th, when the area is an unburned dry vegetation into a Savana zone). On subsequent dates the reflectance rises gradually and signifies the recovery of the site. Other practical examples will be reported together with the results in the next section.

7.1.3 Computation of the Pareto boundary

The validation and accuracy assessment described in the previous paragraphs referred to a comparison between the classified and reference vector layers to evaluate the performance of the algorithm in identifying the burned areas. However, one major source of error in the classification process is due to the lower spatial resolution of the input data (S2 in this case) compared to the reference (Planet, in this case). Even if the algorithm performs best, the error due to the difference in spatial resolution cannot be

reduced due to the presence of partially burned pixels. The Pareto Boundary makes it possible to determine the maximum accuracy values that could be achieved under optimal conditions and represent them graphically in a two-dimensional space. In this thesis it is applied to an image dataset and shows how the spatial resolution of medium (Sentinel-2) and low resolution (MODIS) sensors limits the accuracy of the results.

The concept of omission and commission errors (errors due to algorithm performance) relies on the assumption that classified and reference data have the same spatial resolution, but this is not at all common in the validation phase, which involves comparing a low/medium resolution data with a high resolution data that serves as a reference (in this thesis the data has a maximum spatial resolution of 10/20 m (Sentinel) and 3m (Planet)). The result of the difference in resolution is that while some pixels are entirely correctly classified and others are entirely misclassified, all remaining pixels were only correctly classified for a fraction of their extent. Whenever a low resolution pixel is partially covered by more than one class in the high resolution map, the commission error or omission error will increase (L. Boschetti, 2004).

To create the boundary, we need to use:

- i) the Planet high resolution reference map, which contains the perimeters of the BA products.
- ii) the low resolution cell size (not the product, just the pixel grid), i.e., Sentinel-2 and MODIS.

Once the two inputs have been retrieved, the procedure for generating the Pareto Boundary is:

- a) The low resolution grid is superimposed on the high resolution reference vector.
- b) The percentage of burned areas is calculated for each Sentinel-2/MODIS grid cell.
- c) A set of threshold-dependent products is generated; the threshold t varies in the range (0, 100) and represents the percentage of burned areas (according to the reference) in the grid pixel for that cell to be considered fully burned. Varying the threshold varied the omission and commission error rates. The map generated with $t =$

100% will have no commissions (no mixed pixels included) but large omissions; this is because the grid cell is only considered burned if the area of the Planet product superimposed on it is also entirely burned. The map generated with $t = 1\%$ will have no omission (all mixed pixels are included) but a large commission. In general, the higher it is, the lower the commission and the higher the omission.

d) The results are plotted on a two-dimensional graph describing the omission and commission for each threshold. Since all maps are efficient solutions, the line connecting the corresponding points is a discretization of the Pareto Boundary.

All these schemas are shown in the figure below taken from Boschetti et al. (2004) [Figure 7. 8].

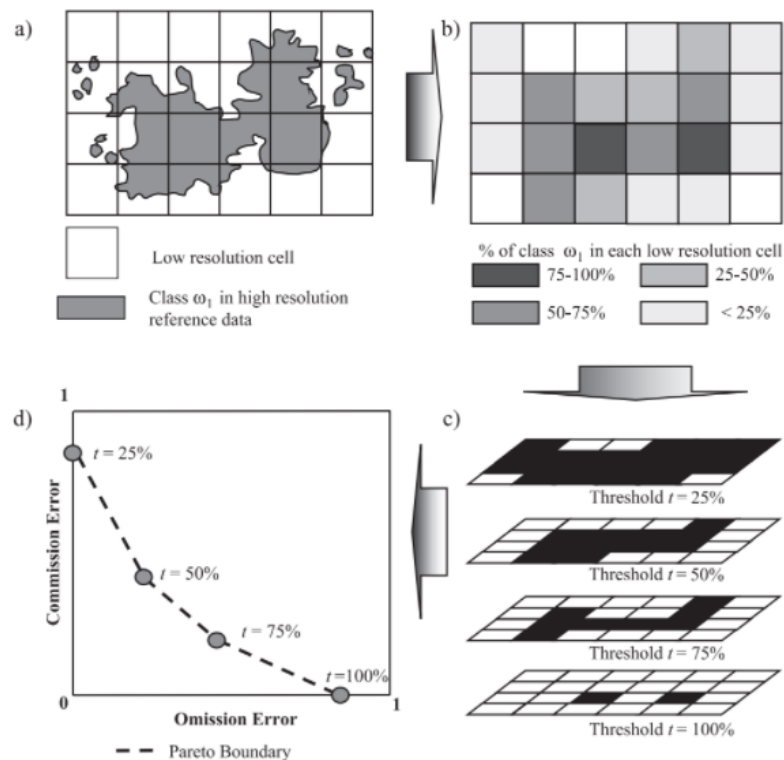


Figure 7. 8 Step for the Pareto Boundary Computing. a) Sentinel Grid and overlapping Planet BA product as input; b) intersection between cells and BA in terms of percentage; c) Evaluating omission and commission for different threshold; d) Pareto Boundary Plot (Boschetti et al., 2004)

To achieve this analysis, an algorithm was implemented in the R Studio working environment (code in appendix) with the following packages:

- *sf*: it represents simple features as records in a data.frame with a geometry list-column;
- *raster*: it implements basic and high-level functions for raster data and for vector data operations;
- *fasterize*: There is support for the main options provided by the `rasterize()` function, including setting the field used and background value, and options for aggregating multi-layer raster;
- *rgeos*: that implements functionality for the manipulation and querying of spatial geometries using the Geometry Engine — Open Source (GEOS);
- *dplyr*: it provides a set of tools for efficiently manipulating datasets;
- *ggplot2*: it's a package for data visualisation.

Once the working environment is ready, the Planet BA product is loaded with the `st_read()` command, which reads the shapefile. The raster of the Sentinel BA product is loaded with the `raster()` command and then cropped to the size of Planet mosaic. Since the BA products are very large, it was decided to run the code by tiles (i.e, portions of the input files) that were processed separately, in order to make faster run of the code and avoiding the code getting stuck during the running phase. The raster is, therefore, divided into blocks where the chunk size is suggested by the `blockSize()` command. Within the for loop, each portion of the raster is first converted to the grid by the `as()` command. Both the grid and the shapefile of the burned areas are in Spatial Polygons DataFrame format, so as to be comparable; the `intersect()` function is then applied, which calculates the intersection of the Planet polygons with those of the grid. This is followed by an if condition that checks that the intersection result is not null; if so, it calculates the area of the intersection result (by adding a column to the DataFrame), if not, it rebuilds the DataFrame with all the blanks. Also in the same for loop, the number of cells of burned and unburned areas are counted and saved in the variables “nfratures1” and “nfeatures2”. Once the for loop is finished the result is saved in a DataFrame called “df_intersect” [Figure 7. 9].

Multi-scale analysis and comparison of Burned Area products derived from Earth Observation (EO) data over the African continent

	date_planet	JD_planet	value	Area	nfeatures1	nfeatures2
1	20190225	56	160	135	407296	4692
2	20190225	56	153	117	407296	4692
3	20190225	56	153	197	407296	4692
4	20190225	56	160	136	407296	4692
5	20190225	56	155	260	407296	4692
6	20190225	56	158	76	407296	4692
7	20190225	56	159	170	407296	4692
8	20190225	56	155	259	407296	4692
9	20190225	56	152	15	407296	4692
10	20190225	56	156	39	407296	4692
11	20190225	56	153	38	407296	4692
12	20190225	56	153	400	407296	4692
13	20190225	56	158	87	407296	4692
14	20190225	56	156	93	407296	4692
15	20190225	56	159	21	407296	4692
16	20190225	56	154	80	407296	4692
17	20190225	56	154	373	407296	4692
18	20190225	56	156	346	407296	4692

Figure 7. 9 Example of the first 18 rows of DataFrame output of the intersection

Since the output dataframe lists all intersections with information on the area covered by the intersection: these values have a maximum of 20x20 m² (S2 pixel size) when the pixel is fully covered by reference BA. These partially burned pixels are assigned to either of the two categories (burned, unburned) depending on the threshold t . Hence, another for loop was implemented where the iteration is based on the value of t threshold. The threshold represents the percentage of the grid cell size above which the grid cell is assigned to the burned class. Within the loop, the code calculates the number of cells where area from intersection between burned area and grid cell area is greater than the threshold. With the quantitative information of the area contained in the DataFrame, the omission error and commission are calculated for each threshold. The results are plotted in a graph with the command `ggplot()` to draw the Pareto Boundary, a curve of maximum accuracy. The algorithm is a script in R and is compiled for each of the five sites in the R Studio workspace.

7.2 Results and sources of error

Following there are results of validation in the five sites of the African continent.

Site 1A

Validation for site 1A showed very satisfactory results (they are 5.9% for commission and 1.6% for omission) and the Dice coefficient is 96.1%. Following there is a table that resumes all the areas' values and the percentages of errors and similarity index [Table 7. 1].

	Area [m ²]	Percentage [%]
Reference Data Burned Area	176183984	
Sample Data Burned Area	184286127	
Agreement Burned Area	173363181	
Omission Error	2820803	1,60
Commission Error	10922946	5,93
Dice Coefficient		96,19

Table 7. 1 Areas' values and Percentage of errors and similarity, site 1A

The image is a thematic categorical map showing the Agreement Area as green, the Omission Area as red, and the Commission Area as blue. Square rectangles show two areas of the larger error where a more detailed analysis was carried out to investigate the source of error [Figure 7. 10]. This scheme of work is also applied to all subsequent sites.

Multi-scale analysis and comparison of Burned Area products derived from Earth Observation (EO) data over the African continent

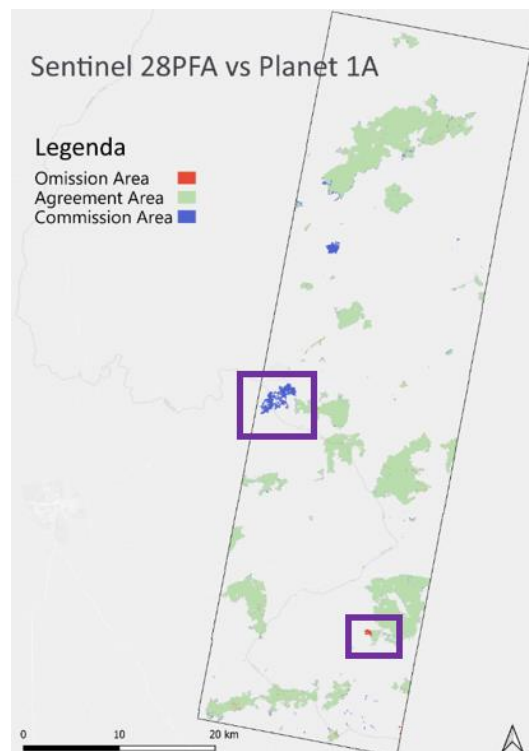


Figure 7. 10 Map of agreement over site 1A showing the spatial distribution of correctly classified burned areas (green) and omission (red) and commission (blue) errors.

Analysing the omission error, it was noted that it could be due to a temporal different in image date acquisition: the datasets on which the burned areas were classified are acquired by different satellites (S2 and Planet) with different orbit configuration therefore simultaneous (same day) acquisitions are not easy to obtain. Moreover, in some cases when fires are on-going even different times of acquisitions during the same day could cause discrepancies between classified BA maps and the reference. In particular, for site 1A, it was observed that the difference in the first date of the temporal series (January 14th for Planet and January 17th for Sentinel-2) could have determined omission errors over areas that burned between those two dates. Since the algorithm is trained to classify as BA portions of land that burned between the two consecutive images in the time series (i.e., the algorithm is based on change detection from unburned to burned surface conditions), areas that burned between January 14th and January 17th are classified as unburned (already burned) by S2 classification. The area analysed covers approximately 336384 sqm (0.18% of total Sentinel area and 11% of total omitted area) and it is shown in the zoom figures [Figure 7. 11].

Multi-scale analysis and comparison of Burned Area products derived from Earth Observation (EO) data over the African continent

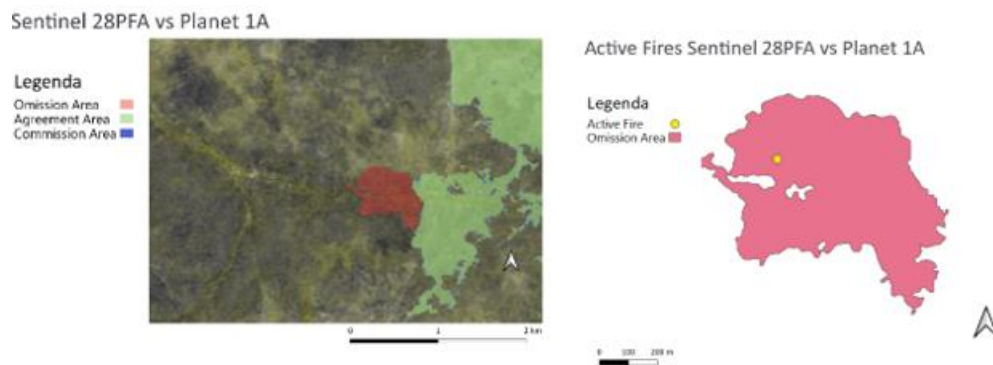


Figure 7. 11 Zoom on Omission Error due to difference in date of acquisition of the first image in the time series, site 1A (left) and Active Fires (right).

To confirm the source of the error, an intersection is performed between the Active Fires dataset and the selected error area, filtered in the time interval of interest. The result is shown in the same figure [Figure 7.11].

As mentioned in the introduction to the input dataset, Active Fires are point geometries and they refer to a midpoint where a flame front has been recorded. Furthermore, the flame front is the region of separation from the burned area and the unburned area, in which combustion takes place; fire front can be observed only at the time of satellite overpass and it is not a good indicator of the amount of the area burned while given the intense signal it can be a good indicator of the presence of a fire. That just said, it is visible from the fact that in the figure there is one fire ground for a very large burned area.

As it turned out, an active fire was registered on January 14th, which is why the Planet - which was acquired before the active fire - registered it as a burned area and the Sentinel did not (it sees it as an “already burned” area).

The same kind of detailed analysis is also carried out on the largest commission area. Exactly as before, the error could be due to a time difference between dates of acquisition. The last date of acquisition of Planet images for this site is March 12th, while the last date of the Sentinel-2 time series is March 18th. This difference might lead to commission errors if there are fires that burn in this time frame. The extent of the error zone shown in the figure is approximately 4376793 sqm (2% of the total Sentinel area, 40% of the commission area) [Figure 7. 12].

Multi-scale analysis and comparison of Burned Area products derived from Earth Observation (EO) data over the African continent

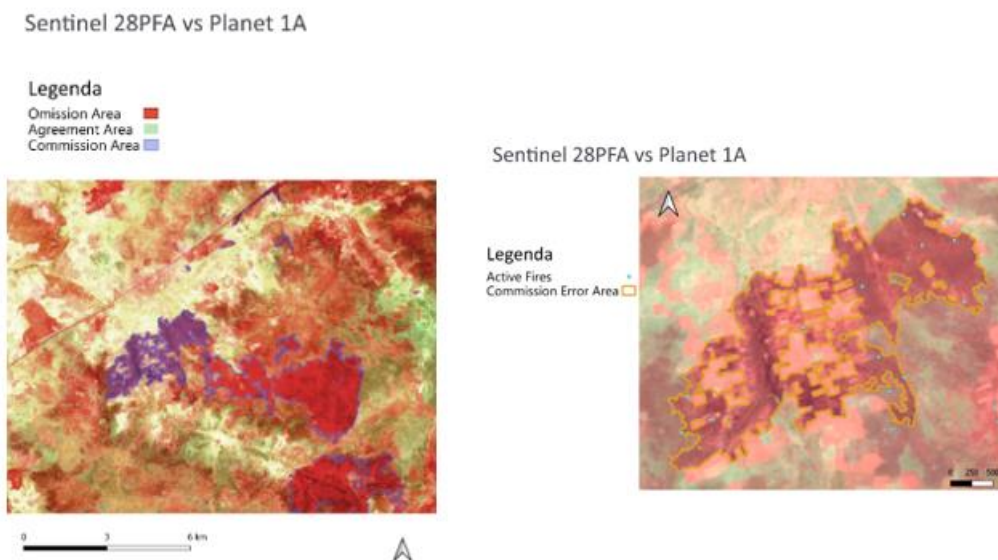


Figure 7.12 Zoom on Omission Error, site 1A (left) and the active fires detected over the same area (right)

Once again, the intersection of the Active Fires dataset filtered by the dates in the time interval of interest and by the area under analysis, shows that our hypothesis is true. The fires were recorded between March 12th and March 13th, so they were only acquired from Sentinel images; this is the reason why there is a commission error [Figure 7.12].

These detailed analysis on classification errors show that the classification could be improved with simultaneous acquisition of the images used for classification and the reference dataset.

Another source of error is the different spatial resolution between S2 (10/20 m) and Planet (3 m) images used for deriving BA map products. Zooming in on the perimeter boundaries of the burned areas clearly shows that errors mainly occur along the border of the fire polygon where the effect of the difference in the spatial resolution of input images (i.e., size of the pixels) is greatest. [Figure 7.13]. The disagreement is only evident on the boundary, but the resolution affects the accuracy of the whole classification. This will be discussed in more detail in the next step: the Pareto Boundary.

Sentinel 28PFA vs Planet 1A

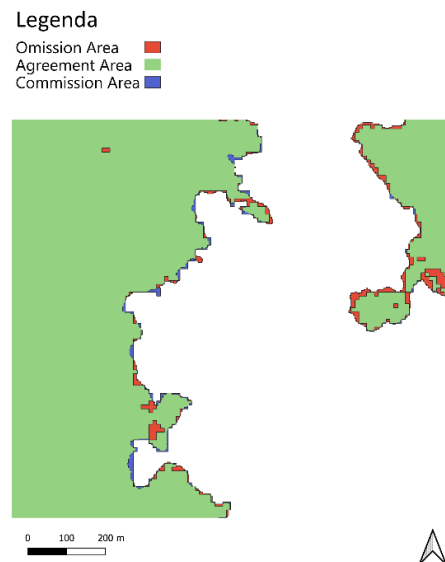


Figure 7. 13 Zoom on the boundary on classification polygons, site 1A

Site 1B

As with the previous site, the table shows the results of the comparison between the two products [Table 7. 2]. The next figure shows the thematic map of the areas (Agreement, Omission and Commission), highlighting the most extensive error that will be analysed in detail [Figure 7. 14].

	Area [m ²]	Percentage [%]
Reference Data Burned Area	228057984	
Sample Data Burned Area	218567087	
Agreement Burned Area	206466222	
Omission Error	21591762	9,47
Commission Error	12100865	5,54
Dice Coefficient		92,46

Table 7. 2 Areas' values and Percentage of errors and similarity, site 1B

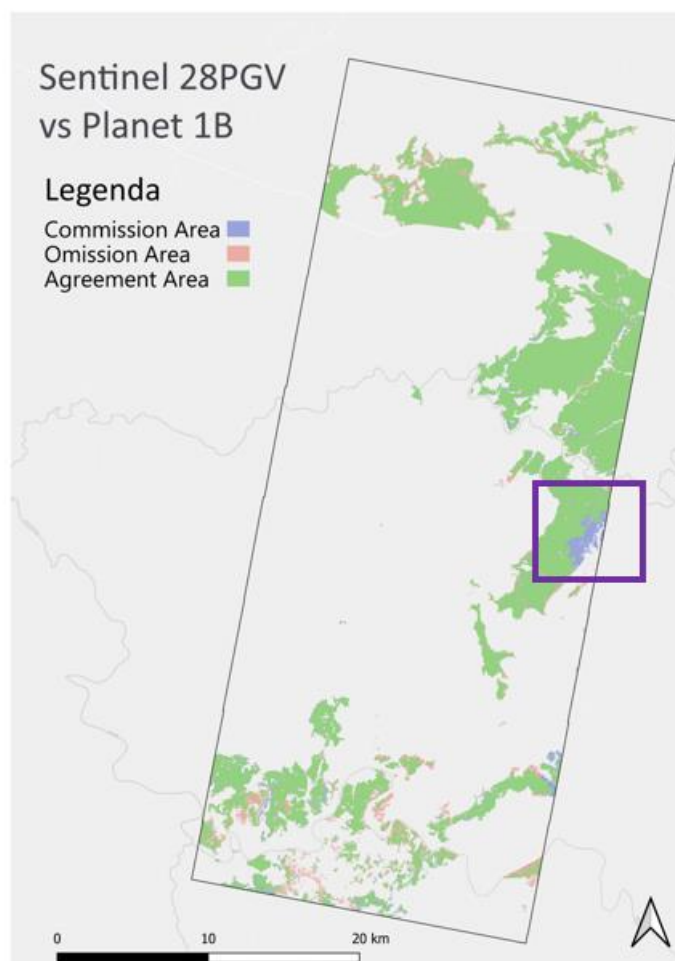


Figure 7. 14 Map of agreement over site 1B showing the spatial distribution of correctly classified burned areas (green) and omission (red) and commission (blue) errors.

In this case, there should be no commission errors related to the dates of image acquisition since the first date of the time series is the same for both S2 and Planet January 17th (10:46 is the Sentinel Timetable and 10:56 the Planet) while the last Sentinel date is recorded earlier than Planet; this could cause errors of commission, i.e. areas recognized as burned only by Sentinel, not the other way around. However, visual analysis of the fire perimeters and S2 and Planet images showed that some errors are nevertheless due to the fact that fires are ongoing during the period covered by the images. Indeed, the fire season in Northern Africa savannas is November to March, during the dry season.

As shown in the figure, there is an error area with an extension of 3792745 m² (~380 ha) (1.7% of the total S2 area of the sample and 31% of the area covered by the

commission error). It is evident from the Sentinel RGB image that the area is burned [Figure 7. 15]. The spectral signature of this region confirms the classification; the data from January 17th is the one with the highest reflectance value and the next image, referred to 27th, is the one with the lowest reflectance value. In fact, the occurrence of a fire reduces the reflectance in the visible and near infrared S2 bands due to the removal of vegetation and the darkening of the surface covered by combustion products. Over this area, fires must have happened in this time interval (visually the lines on the spectral signature graph [Figure 7. 16]).



Figure 7. 15 Zoom on Commission Error, site 1B

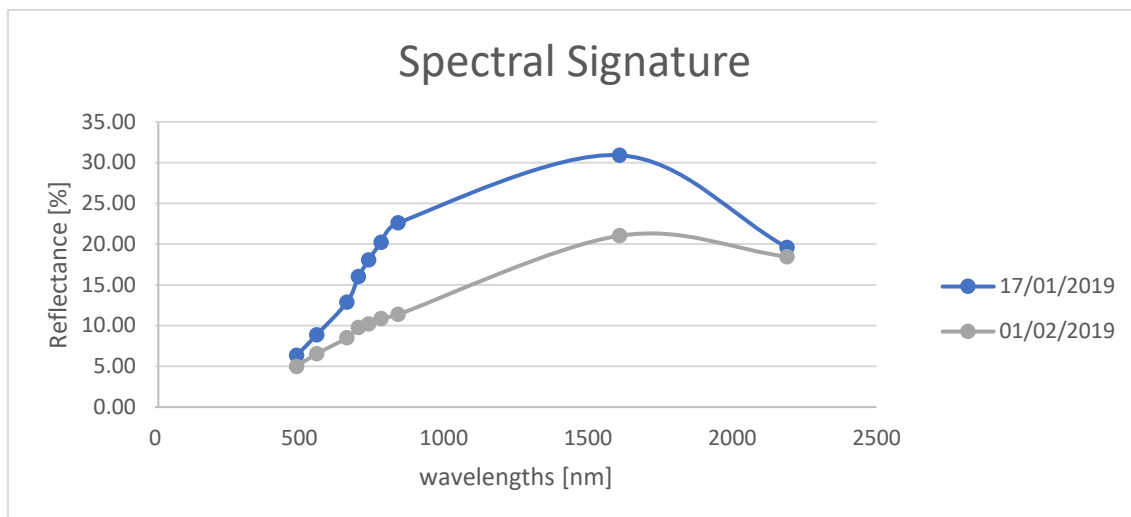


Figure 7. 16 Spectral Signature for the Commission Error zone, site 1B in the s2 wavebands

Once it was that the area was burned, the sources of error in the classification were investigated. Since the source related to time difference has been ruled out, we look at the Planet image to understand the reason for the misclassification [Figure 7. 17]. The figure below shows RGB false colour composites of both S2 and Planet images acquired on January 17th over the region of commission error (blue polygon). From the image it is evident the presence of smoke (and its shadow) from the active front of the fire that can alter reflectance values; since the fire is on-going and Planet is acquired later in the morning than S2, smoke and shadows are thicker in the Planet image thus lowering reflectance values that are spectrally similar to a burned surface. Hence, this area being identified as burned (“already burned”) in the first Planet image, it is not classified as “new burn”. The same figure clearly shows that in the S2 image the unburned area below the smoke is unburned and the smoke is probably more transparent/less thick compared to the conditions of the Planet image where shadows can also be observed.

Multi-scale analysis and comparison of Burned Area products derived from Earth Observation (EO) data over the African continent

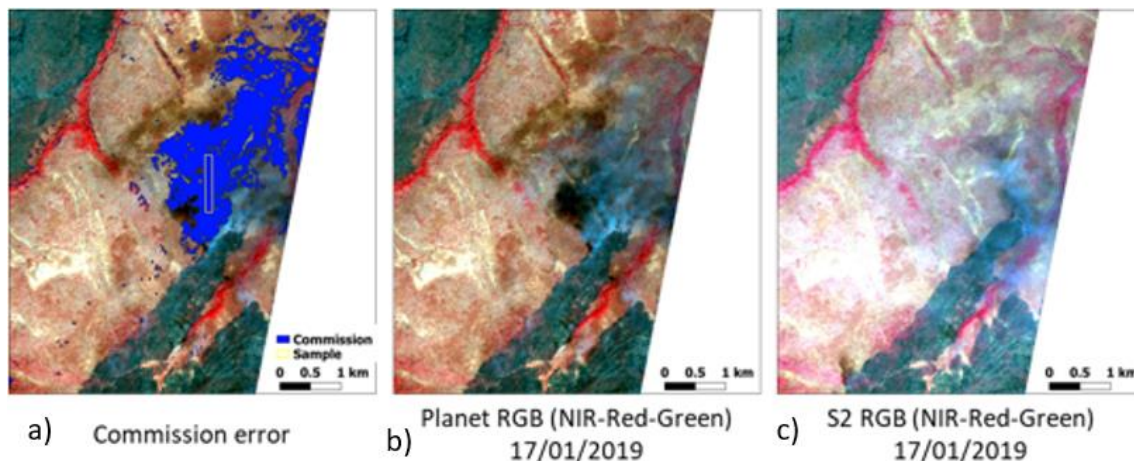


Figure 7. 17 site 1B a) region of commission error (blue) over the RGB false color composite of the Planet images acquired on January 17th; b) RGB false color composite of the same image showing on going fire; c) RGB false color composite of the S2 images on the sam

It is important to reiterate that also in this site, as in the others, a source of error - even if numerically less relevant on the imprecision estimate - derives from the different spatial resolution of the two products.

Site 2C

As seen in the previous sites, the table shows the results of the comparison between the two products [Table 7. 3].

	Area [m ²]	Percentage [%]
Reference Data Burned Area	432874189	
Sample Data Burned Area	436691014	
Agreement Burned Area	350003380	
Omission Error	82870809	19,14
Commission Error	866687634	19,85
Dice Coefficient		80,50

Table 7. 3 Areas' values and Percentage of errors and similarity, site 2C

The next figure presented the agreement thematic map of validation product's areas (Agreement, Omission and Commission) [Figure 7. 18]:

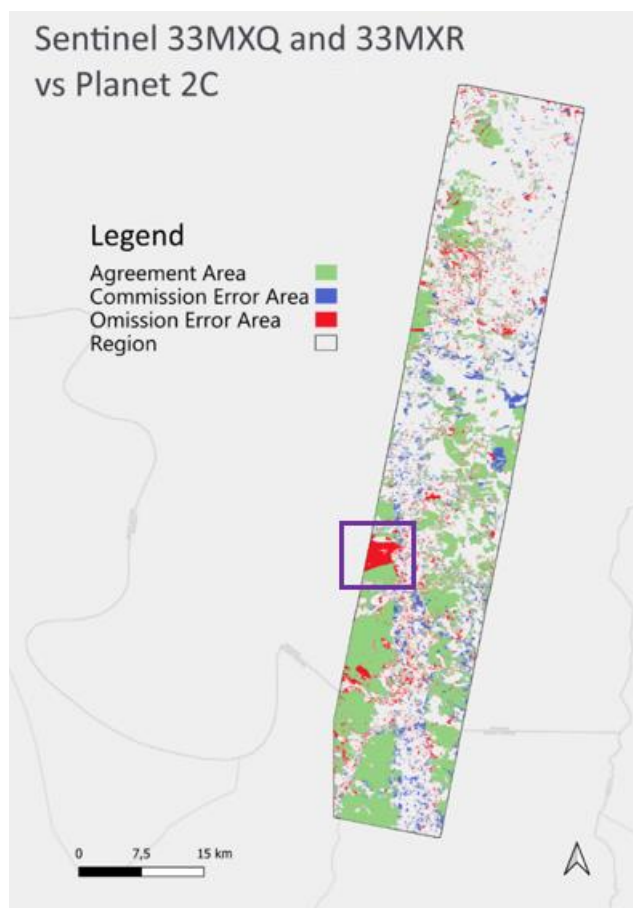


Figure 7. 18 Map of agreement over site 2C showing the spatial distribution of correctly classified burned areas (green) and omission (red) and commission (blue) errors.

The region covered by the site is very huge (it joins two different Sentinel tiles) and the S2 classified BA polygons are numerous with several smaller polygons. These conditions increase the probability of errors; small and irregular BA polygons suffer even more from the error due to the different spatial resolution between the two products.

Looking in detail at the largest area of omission highlighted in the figure by the blue square, it can be observed that. As a first analysis the error area was intersected with the Active Fires dataset, but no fire point falls into this area in the time interval of interest. Then the S2 spectral signature of the region (10607452 sqm area, 12% of the omission area and 2,4% of the total Sentinel BA) was analysed in depth. Three sample polygons were drawn around the error over: unburned vegetation, omission error (yellow) certainly burned area as confirmed by both the classifications and the Active

Fires dataset (blue). Reflectance values in the S2 bands were extracted and plotted as a function of the wavelength for the three sampled surface conditions [Figure 7. 19].

Sentinel 33MXR vs Planet 2C

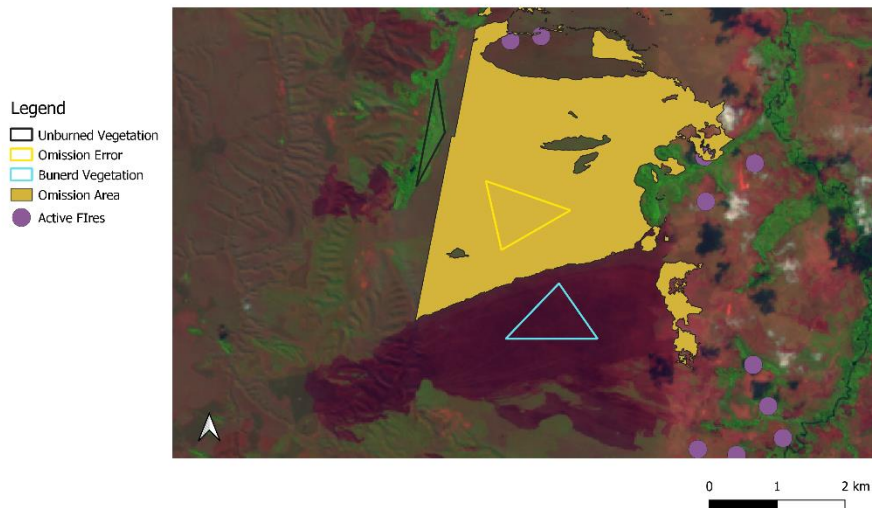


Figure 7. 19 The three sample regions in site 2C selected for the extraction of the spectral signature over unburned (black line), burned (blue) and error (yellow) regions

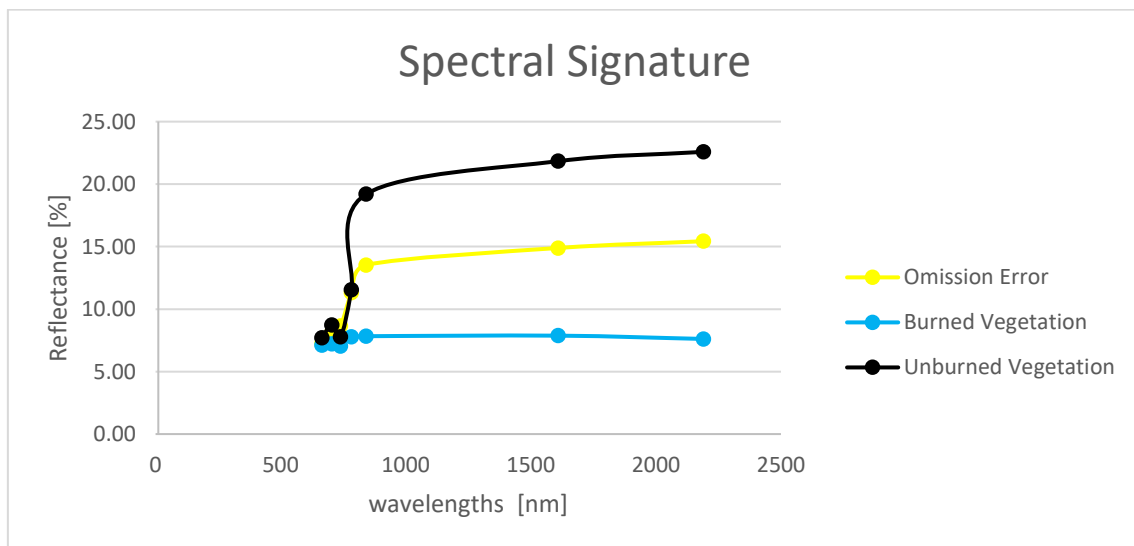


Figure 7. 20 Spectral Signature for the three sampled regions shown in the figure above for Omission Error, site 1C 27/07/2019

As is evident from the spectral signature [Figure 7. 20], the burned polygon (blue line) has a very low reflectance in all S2 spectral bands in the visible to near and shortwave

infrared wavelengths, while the unburned vegetation has a higher reflectance (black line). The area of omission error is exactly in the middle; this area has lower reflectance in the NIR to SWIR wavelength (> 900 nm) than unburned vegetation, but from the analysis of the spectral signature, it is clear that reflectance it is not as low as a burned surface. It is likely that the area may have been burned earlier thus being classified as unburned (not new burn).

Site 3A

As with the previous site, the table shows the results of the comparison between the two products [Table 7. 4].

	Area [m ²]	Percentage [%]
Reference Data Burned Area	92728251	
Sample Data Burned Area	94022976	
Agreement Burned Area	77123514	
Omission Error	15604737	16.83
Commission Error	16899462	17.97
Dice Coefficient		82.60

Table 7. 4 Areas' values and Percentage of errors and similarity, site 3A

The S2 BA product for the 3A site was derived by repeating the classification with various training set combinations and images to achieve the highest level of accuracy. The hypothesis about this uncertainty is that the more influential source of error is the different dates of acquisition of S2 and Planet images. The Sentinel time series starts ten days after the Planet one and finishes one day earlier. The time series analysed for this site is shorter compared to the other sites, because it covers only one month, but as it can be observed in [Figure 7. 21], it is dense with fires. The fact that the analysed period falls in a fire season also leads to the creation of many errors in case there is a temporal inconsistency, because reflectance of burned areas classified changes quickly

depending and the classification performance changes depending on the images selected as input.

It should also be noted that this site was often covered by clouds, so fewer images were available that fit the criteria for the selection of time series (introduced in the previous chapter).

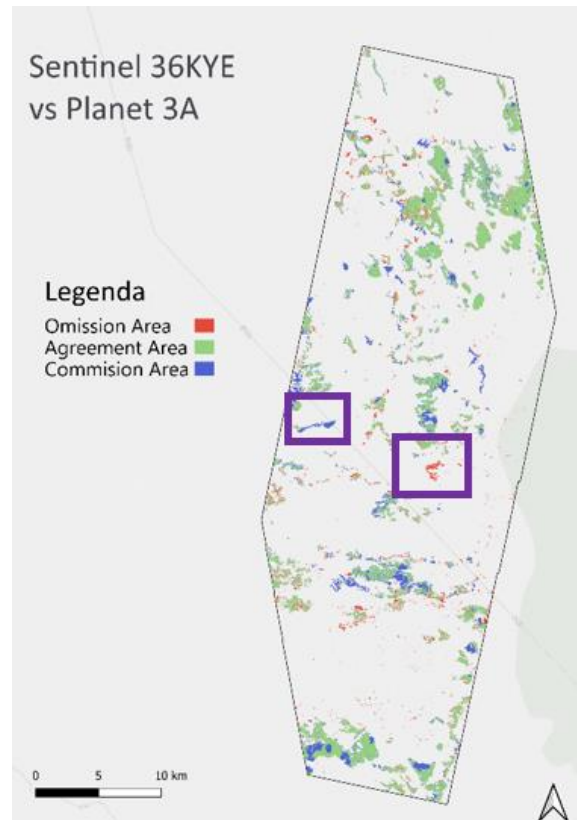
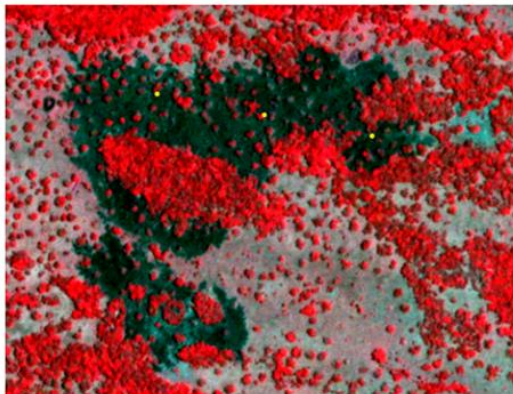


Figure 7. 21 Map of agreement over site 3A showing the spatial distribution of correctly classified burned areas (green) and omission (red) and commission (blue) errors.

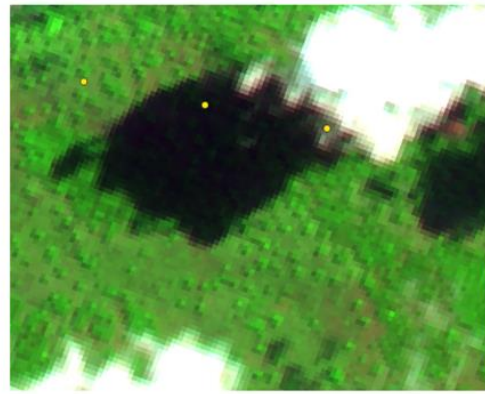
The analysis of the sample omission area [Figure 7. 22] (blue region) overlaid on the Sentinel-2 and Planet RGB images highlighted that the no error region appeared as unburned in the S2 on September 28th, even if the presence of cloud shadows prevents from observing the surface. In the Planet image, acquired the day after, the same region appears as burned (it is recorded on September 29th). On the RGB images there are yellow points that are Active Fires in this time interval; they confirm the presence

of burned area. (The area is 597747 sqm 0.55% of the total area burned according to sentinel data and 4.3% of the total area omitted).

This example underlines that even taking images on consecutive days produces different results. Two sample areas were then taken to investigate the source of error and thus justify this level of uncertainty.



29/09/2019 Planet



28/09/2019 Sentinel

Figure 7. 22 Zoom on omission error, site 3A for Planet RGB (left) and Sentinel RGB (right)

Another source of error in the S2 classification could be poor image radiometric quality that can affect the classification. One example is given in [Figure 7. 23] below where output burned perimeters according to the Sentinel dataset are not accurate. In fact, the Sentinel-2 RGB image shows that the burned area is covered by clouds and cloud shadows and it is difficult to identify by the algorithm despite repeated training tests. This example region does not have a large area in square metres and therefore has little influence on the overall accuracy taken individually, but cloud and cloud shadows can have a large impact on classification accuracy especially over geographical regions often covered by clouds.

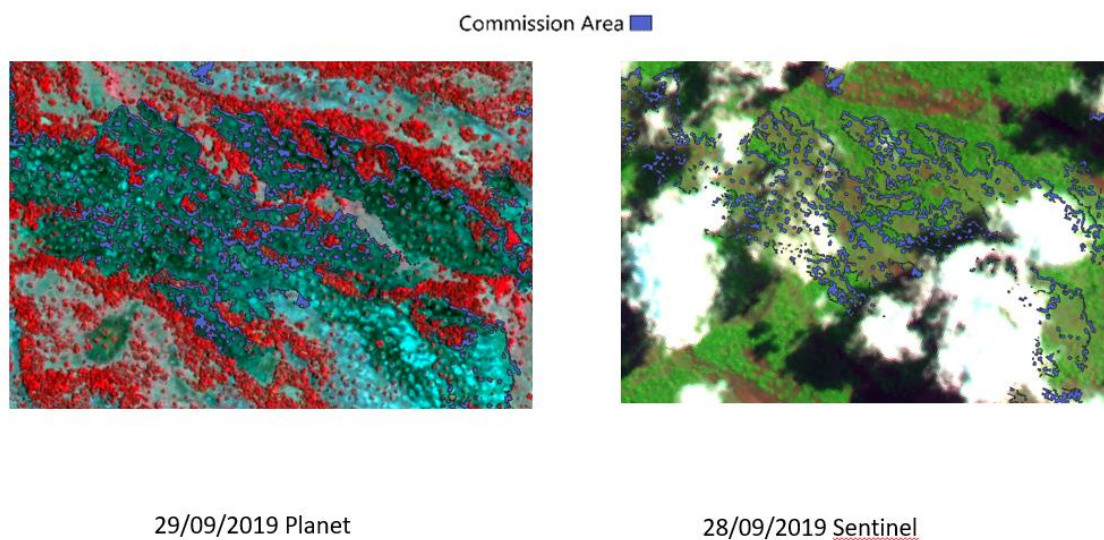


Figure 7. 23 Zoom on commission error, site 3A for Planet RGB (left) and Sentinel RGB (right)

This is followed by a focus on the spectral signatures of clouds and its shadows and how they affect classification. Sample polygons were taken on an RGB Sentinel acquired on August 29th 2019 (the blue polygon for the burned area, the grey polygon for the cloud and the purple polygon for the cloud shadow) [Figure 7. 24].

The spectral signature shows that the cloud has very high reflectance values (grey line), while its shadow has low reflectance values similar to the spectral signature of the burned area. This similarity can be a problem during classification in fact one large source of uncertainty in burned area classification is the spectral confusion with low albedo regions such as shadows (cloud and topographic) and water bodies (Roy et al., 2019).

The product contained a mask for shadows (as well as a mask for clouds), but it was decided not to use it because it also masked out many burned areas, thus increasing the omission error. Thanks to the trial and error approach, it turns out that the best method for accurate classification is to take as much unburned training as possible on shadows and clouds on different landcovers until the most accurate result is achieved.

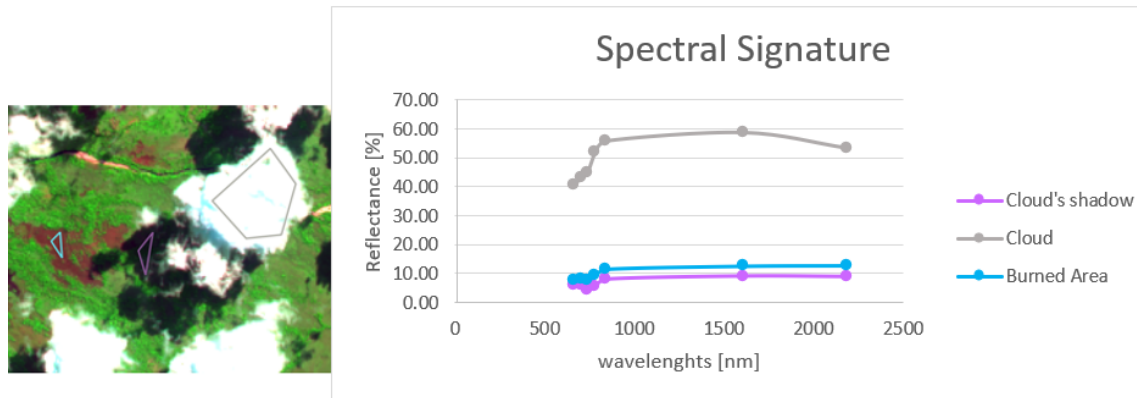


Figure 7. 24 Spectral Signature for Omission Error, site 3A 29/08/2019

Also, for this site, it was analysed the uncertainty due to the different spatial resolution of S2 and Planet. As also presented for the previous site, small and very irregular BA polygons can increase the error due to spatial resolution and this can be seen graphically by the amount of small errors recorded along the perimeter of the areas affected by the fires [Figure 7. 25].

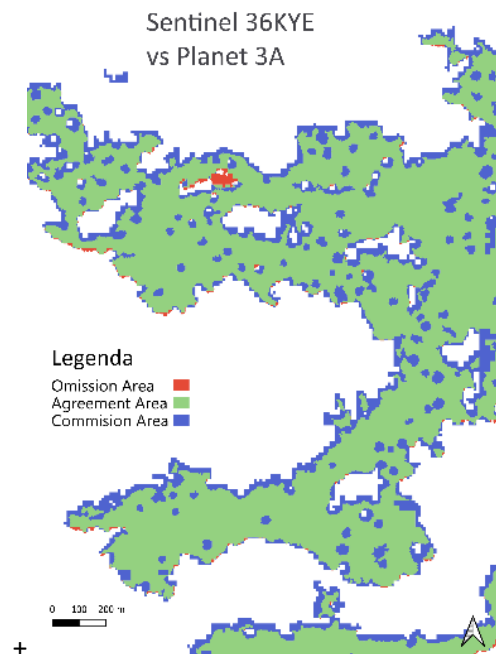


Figure 7. 25 Zoom on thematic map, site 3A

Site 3B

As with the previous sites, the table shows the results of the comparison between the two products [Table 7. 5].

	Area [m ²]	Percentage [%]
Reference Data Burned Area	525978159	
Sample Data Burned Area	460027065	
Agreement Burned Area	415678912	
Omission Error	110299247	20,97
Commission Error	44348153	9,64
Dice Coefficient		85,32

Table 7. 5 Areas' values and Percentage of errors and similarity, site 3B

It should be noted that the area estimates in square metres are the largest, this is because this site has large extension and it is dense with fires in the time interval considered (as it is already seen for site 2C). Although the commission error is small (9%) the omission error is the highest error metric in this thesis (20,97%); portions of the territory will now be analysed in detail to understand the cause. Below is shown the thematic agreement map that shows the results for this region with one selected area of omission that was further investigated [Figure 7. 26].

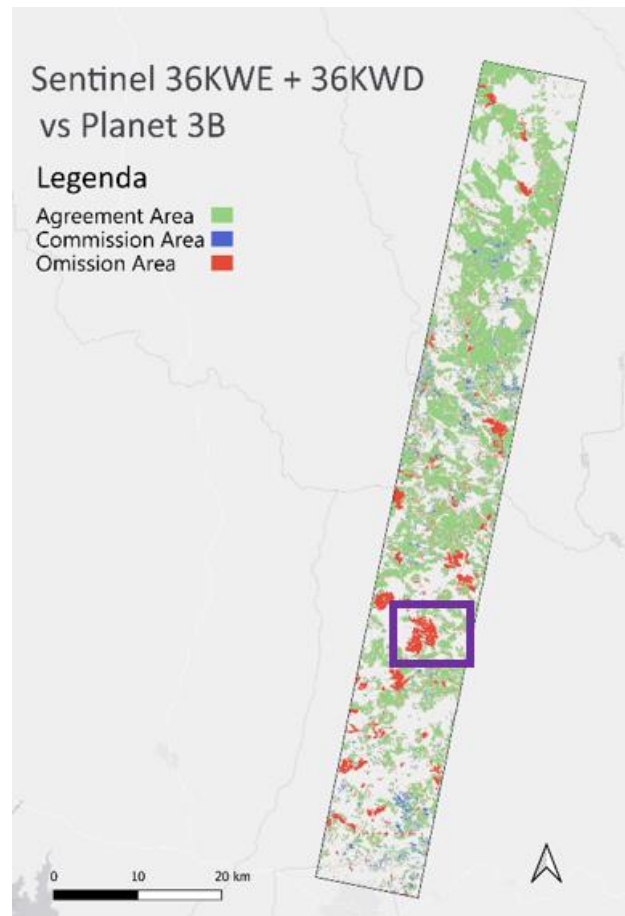


Figure 7. 26 Comparison's errors, site 3B

Looking at the acquisition dates of the Sentinel-2 satellite images, there may be errors of omission due to temporal gap: the last Planet image of time series was acquired on September 14th, while the last Sentinel-2 image on the 11th of the same month. Performing a depth analysis on this error area (9037589 sqm, 2% of the total Burned Area calculated with Sentinel data and 8.1% of the Omitted Area) it is observed that Active Fire points intersecting that area are recorded between September 11th and 12th, therefore after, the last Sentinel-2 acquisition thus leading to an omission error [Figure 7. 27].

Sentinel 36KWE + 36KWD
vs Planet 3B

Legenda

Active Fires ●
Omission Area ■

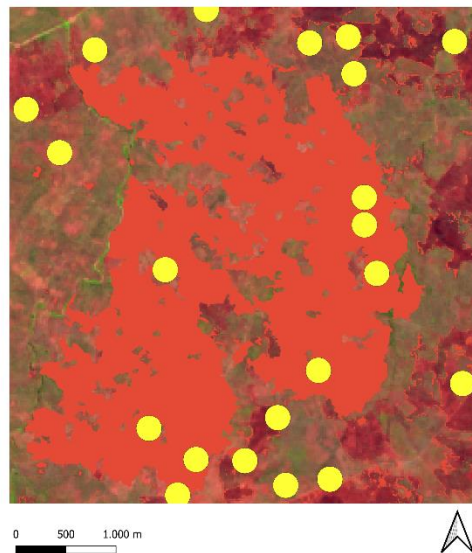


Figure 7. 27 Zoom on omission error area, site 3B

By looking at the RGB images of a portion of the area, active fires are easily visible. [Figure 7. 28]. The left image was acquired on 27th August and registers two flaming fire fronts (highlighted by the yellow circles). The left image is recorded September 11th and shows one flaming fire front (that is the one of the active fires in the error area shown in the Figure 7.27). Note that the areas around the flame fronts in a), are burnt areas in image b); this example is useful to understand the difference between burned areas and active fire: the one is the effect of the second. Moreover, burned area signal is persistent in time since it can be observed days after the fire event (in some cases such as in the boreal forests the effect of a fire can be observed even years after the fire event) while an active fire can be detected only if it is burning at satellite time overpass.

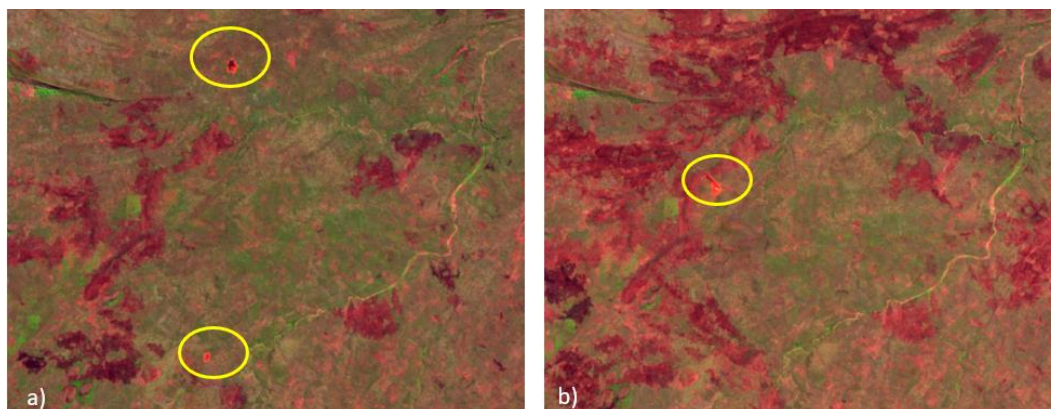


Figure 7. 28 a) RGB on 27/08/2019; b) RGB on 11/09/2019, both for a zoom area of the site 3B

7.3 Comparison between BA maps from S2 and the FireCCI product

The FIRECCI51 BA products were compared to BA Sentinel-2 products from the supervised RF classification. This comparison is an additional analysis to compare the maps derived in this thesis with BA maps obtained by an operational product such as the FireCCI51. This study is not part of the validation phase since neither of the two products can be assumed as representative of the ground truth.

The two BA products are compared to study their similarity. The table below shows the metrics, results of the comparison: the area of Agreement Burned and the area of Agreement Unburned and the Dice similarity coefficient (which studies how much of the product is overlapping). The agreement area values shown in the table are in square meters [Table 7. 6].

	Unburned agreement	Burned-Unburned	Unburned-Burned	Burned Agreement	Dice
28PGA	4502091800	126164400	40772000	213726400	71,91 %
28PGV	6917318500	829654500	428113300	858565800	57,72%
33MXQ	4820103400	1038144200	541246800	1675381600	67,96%
33MXR	5895698300	1030320000	368151900	690912400	49,70%

Multi-scale analysis and comparison of Burned Area products derived from Earth Observation (EO) data over the African continent

36KWE	4219738400	1417618400	1639007000	2596193000	62,94%
36KYE	7475218100	813172800	152932500	406400200	45,69%

Table 7. 6 Metrics about comparison between FIRECCI51 and Sentinel-2 BA products

The histogram below shows the burned areas mapped in S2 BA product (represented with blue) and the burned areas in the FIRECCI51 product [Figure 7. 29].

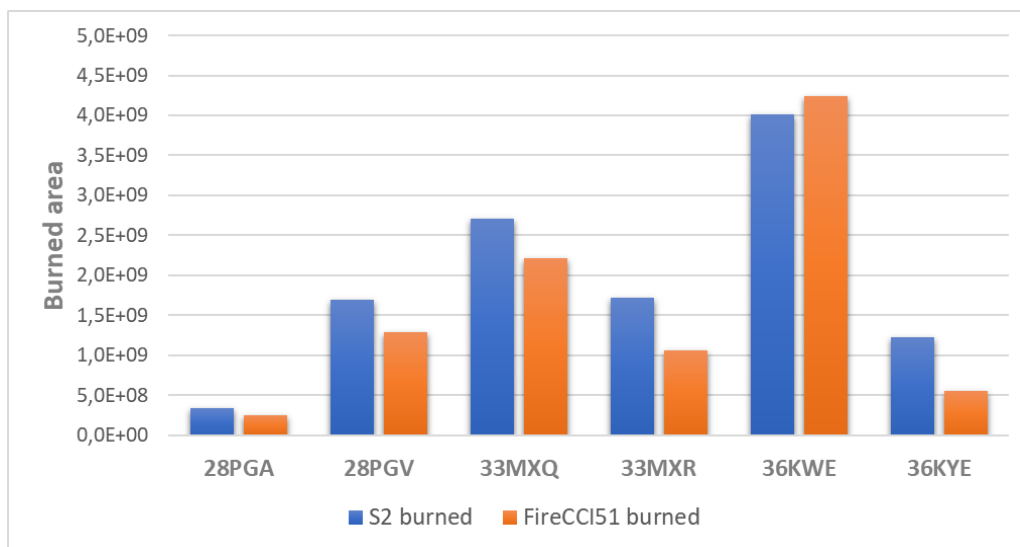


Figure 7. 29 Histogram for BA products (S2 products are blue and FireCCI51 products are orange)

Observing results, it can be seen that metrics have a much lower BA agreement values than the coefficients calculated during the Validation phase (Comparison Sentinel-2 and Planet). This happens because the spatial resolution of FIRECCI51 data is 250 m, unlike that of Sentinel, which has a maximum spatial resolution of 10m. The FIRECCI51 products, therefore, are much less accurate.

Visually, the result of this multi-scale comparison is shown in the figure below [Figure 7.29].

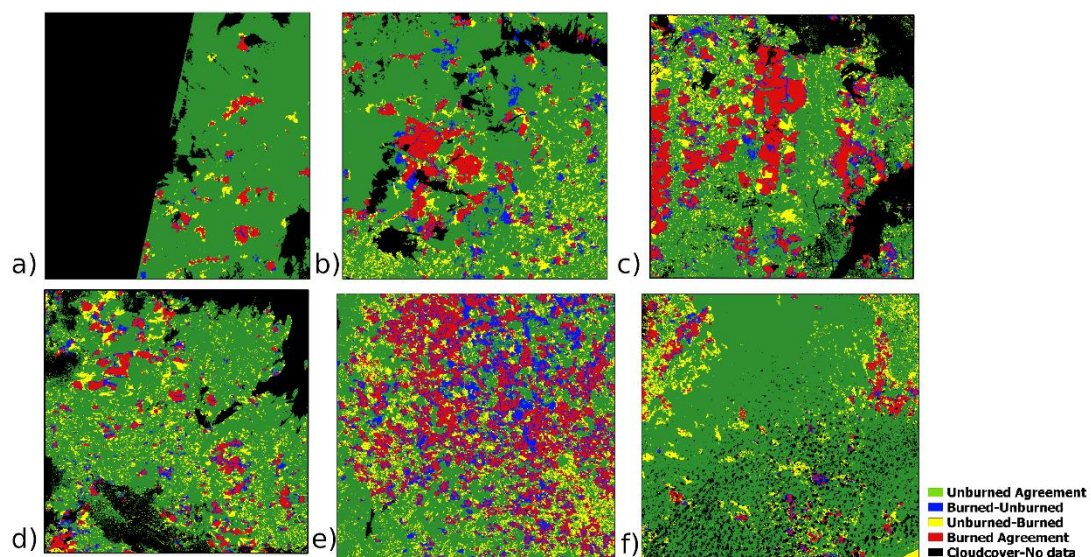


Figure 7. 30 The results of the comparison between BA maps from S2 and FireCCI51 for the six S2 tiles of the AOIs:28PFA (a), 28PGV (b), 33MXQ (c), 33MXE (d), 36KWE (e) and 36KYE (f). Colors represent agreement and disagreement areas.

7.4 Pareto Boundary

The Pareto boundaries for the five sites are displayed in [Figure 7. 31] together with actual omission and commission errors obtained from the validation. The curve of the Pareto boundary shows the lowest errors that can be obtained with the best performing algorithm and given the resolution: if no errors are committed by the algorithm, a certain error level is intrinsic due to the lower spatial resolution of the input data: if no errors are committed by the algorithm, a certain error level is intrinsic due to the lower spatial resolution with respect to reference. The number of partially burned pixels in the classified BA map is linked to the characteristics of the features on the ground such as shape, size and fragmentation (Boschetti et al., 2004).

Results shows that the site least affected by this effect is site 1A; its Pareto Boundary (red line) is the closest to zero and, in addition, it is also the site that had the lowest error metrics (OE=1,60%, CE=5,92%) (the burned areas were few in numbers and much more regular than the other AOIs). Second best results were obtained for site 1B (yellow curve), with less good results, but still similar to those of the previous site.

Next are sites 2C and 3B: both are the product of the union of two different Sentinel tiles, they are very large and present many polygons of small and irregular burned areas; as already hypothesised in the previous paragraph, shape, size and fragmentation of the burned polygon as depicted in the reference dataset affects the accuracy of the classification obtained from lower resolution data.

For site 3A errors related to the spatial resolution are larger. In fact, the position of the Pareto frontier depends mainly on the geometry of the reference (BA Planet classification), while the position of the point depends on the actual accuracy (and therefore also depends on the algorithm). As in the two previous sites, this type of geometry affects the accuracy.

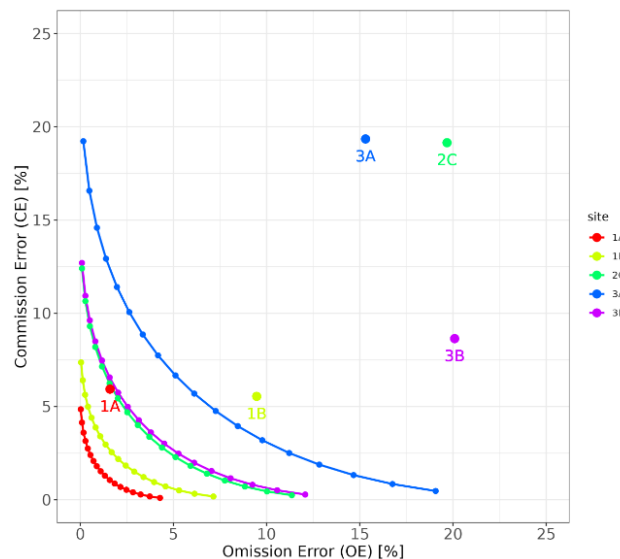


Figure 7.31 Pareto Boundary for all the five sites

The same approach of the Pareto Boundary was applied with a low resolution grid (MODIS grid with cells of 250mx250m) as input. The result is shown in the figure [Figure 7.32] The position of the Pareto is shifted towards greater omission and commission values due to the lower resolution of the MODIS grid cells. What follows from this comparison is that the spatial resolution of the data is actually a limitation accuracy factor of the resulting maps. If spatial resolution decreases, the omission/commission error will grow.

Multi-scale analysis and comparison of Burned Area products derived from Earth Observation (EO) data over the African continent

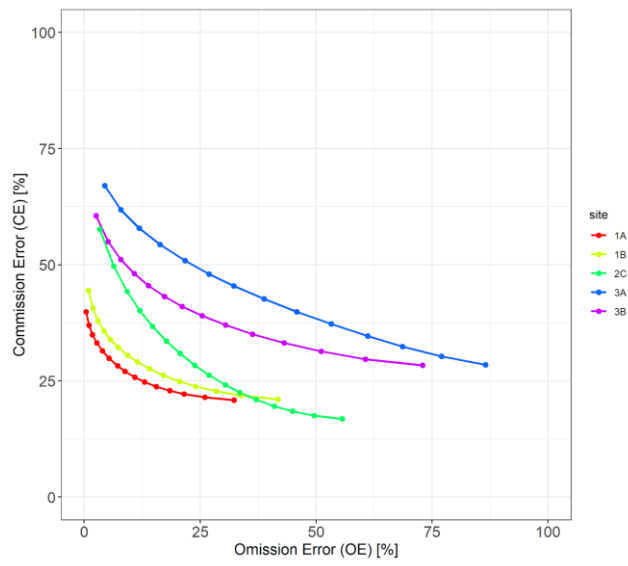


Figure 7. 32 Pareto Boundary with MODIS grid as input

Discussion and Conclusion

For this thesis, a multiscale comparison was conducted on burned area mapping products at five sites on the African continent. First, the Burned Area (BA) maps were created from a collection of Sentinel-2 images, in the tiles that overlay the regions of interest. The product was created by implementing a Random Forest supervised classification algorithm in Google Earth Engine (GEE); the script for the algorithm was written using the JavaScript programming language.

GEE was the optimal working environment because it has a data repository (Data Catalog) and a platform (Code Editor) on which to import, manage and analyse satellite images. GEE, therefore, has enabled unprecedented access to petabytes of geospatial data (i.e., the entire Sentinels archive) with multiple processing functions. GEE's powerful parallel computing infrastructure facilitates on-demand and on-the-fly production of complex geospatial products at various geographic and temporal scales without the need for supercomputers. Its capabilities have been demonstrated in various applications among which Burned Area mapping (Roteta et al., 2019).

After the mapping phase of the burned areas was carried out the Validation phase: Sentinel-2 BA products were compared with Planet BA fire perimeters (considered as ground truth for their excellent spatial resolution). The results are satisfactory: the Sentinel-2 BA maps have omission error in the range [1,60-20,97 %] and commission error in the range [5,93-19,85 %] across sites as summarised in the Table below. The site where the algorithm best performs is 1A with lowest error rates. Generally commission is greater than omission as expected in lower resolution BA maps. Results are collected in the table below *[Errore. L'origine riferimento non è stata trovata.]*:

Site	OE [%]	CE [%]	Dice [%]
1A	1,60	5,93	96,19
1B	9,47	5,54	96,46
2C	19,14	19,85	80,50

Multi-scale analysis and comparison of Burned Area products derived from Earth Observation (EO) data over the African continent

3A	16,83	17,97	82,60
3B	20,97	9,64	85,32

Table 7. 7 Results of Validation phase for all five sites

The validation results were also confirmed to be more than satisfactory compared to other studies. Comparison is not always easy as the reference data are often obtained in different ways; moreover, not all studies on the validation of burnt area mapping are conducted with an independent reference BA dataset at higher resolution.

Here are some examples from the literature to compare results: Pulvirenti et al. (2020) proposed an automated algorithm based on S2 spectral indices over forest areas and achieved average commission error of 6.3% and omission error of 12.7%. Similarly, Smiraglia et al. (2020) obtained commission error = 33% and omission error = 24% by also exploiting S2 spectral indices. Furthermore, Seydi et al. (2021) [50] mapped burned areas with a Random Forest algorithm (proved to be the algorithm providing greatest accuracy) with ce= 8.7% and oe= 9.2%.

It is evident that the result of these studies -with different algorithms- is consistent with the result obtained in this thesis.

Studying the sources of error with respect to the validation results, it was determined that the factors that most influence the result are:

- Difference in the dates of acquisition of satellite images for the source data for classification and for reference;
- The different spatial resolution of the two instruments (Sentinel-2 has a resolution of 10m, while Planet 3m);
- The presence of clouds (and their shadows) and smoke from the flame fronts that disturb the image and, therefore, bring errors and inaccuracies to the result.

Some of these errors could be reduced: this is the case of errors due to the difference in dates of acquisition of satellite images. Although obtaining simultaneous source images is very difficult. Other errors cannot be corrected, because they are an integral

part of the characteristics of the image on which you are working such as in the case of spatial resolution.

The accuracy of the results was not only studied by comparing the vector layers (format of BA products) of Planet and Sentinel-2, but also analysing the direct effect of different spatial resolution on classification. The result of this comparison is the Pareto Boundary.

The strong point of this thesis is to span different working environments (GEE, QGIS, R Studio) using multiple types of programming languages (JavaScript, R) to obtain a complete study not only on mapping and comparison of burned areas, but also a critical investigation on accuracy and sources of error.

Scientific inquiry allows us to understand what works and what doesn't, but it is only by understanding why that we can take action, making a difference.

Bibliography

Belgiu M., Drăguț L., Random forest in remote sensing: A review of applications and future directions, *ISPRS Journal of Photogrammetry and Remote Sensing*, Volume 114, 2016, Pages 24-31,)

Bonsor, K. (2001, Maggio 29). *How Wildfires Work*. Retrieved from How stuff works: HowStuffWorks.com

Boschetti L., Flasseb S.P.; Analysis of the conflict between omission and commission in low spatial resolution dichotomic thematic products: The Pareto Boundary; CNR-Istituto per il Rilevamento Elettromagnetico dell'Ambiente, via Bassini 15, I-20133 Milan, Italy; Received July 8th 2003; received in revised form February 26th 2004; accepted February 28th 2004

Boschetti L., Stehman S. V., Roy D. P.; A stratified random sampling design in space and time for regional to global scale burned area product validation *Remote Sensing of Environment* 186 (2016) 465–478

Cardoso, M.F., Hurtt, G.C., Moore, B., Nobre, C.A., Bain, H., 2005. Field work and statistical analyses for enhanced interpretation of satellite fire data. *Remote Sens. Environ.* 96, 212–227.

Chuvieco, E., Opazo, S., Sione, W., Valle, H. d., Anaya, J., Bella, C. D., Cruz, I., Manzo, L., López, G., Mari, N., González-Alonso, F., Morelli, F., Setzer, A., Csizar, I., Kanpandegi, J.

A., Bastarrika, A. and Libonati, R. (2008), Global burned-land estimation in Latin America using MODIS composite data. *Ecological Applications*, 18: 64-79

Congalton, R.G., Oderwald, R.G., Mead, R.A., 1983. Assessing Landsat classification accuracy using discrete multivariate analysis statistical techniques. *Photogramm. Eng. Remote Sens.* 49, 1671–1678.

Multi-scale analysis and comparison of Burned Area products derived from Earth Observation (EO) data over the African continent

Congalton, R.G., & Green, K. (1999). *Assessing the Accuracy of Remotely Sensed Data: Principles and Applications*. Boca Raton: Lewis Publishers.

David M. Szpakowski D. M., J.L.R. Jensen A review of assessing the accuracy of classifications of remotely sensed data

Department of Forestry and Resource Management, University of California, Berkeley USA Received October 15th 1990, Revised April 14th 1991, Available online April 14th 2003.

Dice, L.R. (1945). Measures of the Amount of Ecologic Association Between Species. *Ecology*, 26, 297-302.

Giglio, L; Schroeder, W; Justice, CO (2016). The collection 6 MODIS active fire detection algorithm and fire products. *REMOTE SENSING OF ENVIRONMENT*, 178, 31-41.

Justice, C.O., Belward, A., Morisette, J., Lewis, P., Privette, J. & Baret, F., (2000), Developments in the 'validation' of satellite sensor products for the study of land surface. *International Journal of Remote Sensing*, 21, 3383-3390.

Knorr, W., Jiang, L., Arneeth, A., 2016. Climate, CO₂ and human population impacts on global wildfire emissions. *Biogeosciences* 13, 267.

Lillesand T.M. and Kiefer R. W., "Remote Sensing and Image Interpretation," John Willey and Sons, New York, 1999.

Lizundia-Loiola, J., Otón, G., Ramo, R. and Chuvieco, E., 2020. A spatio-temporal active-fire clustering approach for global burned area mapping at 250 m from MODIS data. *Remote Sens. Environ.*

Mallinis G. , Tsakiri-Strati M., Georgiadis C. and Patias P.; 2018, Assessment of Radiometric Resolution Impact on Remote Sensing Data Classification Accuracy - Natalia Verde 1

Meinrat. (2019, July 4). Emission of trace gases and aerosols from biomass burning- an updated assessment. *Atmospheric Chemistry and Physics*, pp. 8523-8546.

Mutanga, O.; Kumar, L. Google Earth Engine Applications. *Remote Sens.* 2019, 11, 591.

Padilla M., Olofsson P., Stehman S.V., Tansey K., Chuvieco E., 2017. Stratification and sample allocation for reference burned area data, *Remote Sens. Environ.*, 203, 240-255.

Planet Team. Planet Application Program Interface: In Space for Life on Earth. San Francisco, CA. <https://api.planet.com>, 2017.

Pulvirenti, L.; Squicciarino, G.; Fiori, E.; Fiorucci, P.; Ferraris, L.; Negro, D.; Gollini, A.; Severino, M.; Puca, S. An Automatic Processing Chain for Near Real-Time Mapping of Burned Forest Areas Using Sentinel-2 Data. *Remote Sens.* 2020, 12, 674.

Roy D.P.; Huang H., Boschetti L., Giglio L., Yan L., Zhang H. H., Li Z., Landsat-8 and Sentinel-2 burned area mapping - A combined sensor multi-temporal change detection approach, *Remote Sensing of Environment*, Volume 231, 2019

Seydi, S.T.; Akhoondzadeh, M.; Amani, M.; Mahdavi, S. Wildfire Damage Assessment over Australia Using Sentinel-2 Imagery and MODIS Land Cover Product within the Google Earth Engine Cloud Platform. *Remote Sens.* 2021, 13, 220.

Smiraglia, D.; Filipponi, F.; Mandrone, S.; Tornato, A.; Taramelli, A. Agreement Index for Burned Area Mapping: Integration of Multiple Spectral Indices Using Sentinel-2 Satellite Images. *Remote Sens.* 2020

Zhao H., Emissioni di combustione di biomassa e impatti sull'inquinamento atmosferico in Cina, in *Inquinamento atmosferico asiatico* , 2022

Sitography

<https://climate.esa.int/en/projects/fire/>

[https://sentinel.esa.int/documents/247904/685211/Sentinel-2 User Handbook](https://sentinel.esa.int/documents/247904/685211/Sentinel-2_User_Handbook)

https://developers.google.com/earth-engine/datasets/catalog/COPERNICUS_S2_SR

<https://api.planet.com>

<https://earthdata.nasa.gov/earth-observation-data/near-real-time/firms/active-fire-data>

<https://earthdata.nasa.gov/earth-observation-data/near-real-time/download-nrt-data/viirs-nrt>

<https://earthdata.nasa.gov/earth-observation-data/near-real-time/download-nrt-data/modis-nrt>

https://developers.google.com/earth-engine/datasets/catalog/ESA_CCI_FireCCI_5_1

<https://code.earthengine.google.com/>

https://climate.esa.int/media/documents/Fire_cci_D4.1_PVIR_v1.1.pdf
https://eco-ke-staging.com/media/documents/Fire_cci_D3.1_PVP_v1.1.pdf

Multi-scale analysis and comparison of Burned Area products derived from Earth Observation (EO) data over the African continent