



POLITÉCNICA

U N I V E R S I D A D P O L I T É C N I C A
D E M A D R I D

E.T.S. de Ingeniería de Sistemas Informáticos

**Estudio del algoritmo de
clustering DBSCAN (Density-
Based Spatial Clustering of
Applications with Noise),
aplicado en la detección de
cambios de las últimas
erupciones del volcán Etna
(Sicilia)**

P R O Y E C T O F I N D E
G R A D O

G R A D O E N I N G E N I E R Í A D E L
S O F T W A R E

C U R S O 2 0 2 2 / 2 0 2 3

A U T O R

Eduardo Mena Caballero

T U T O R A S

**Montserrat Hernández Viñas
Elvira Martínez de Icaya Gómez**

Agradecimientos

A todas aquellas personas que han estado apoyándome cada día a lo largo de este periodo de cuatro años tan especial de mi vida. Especialmente a toda mi familia y amigos.

También una mención especial a Montserrat Hernández Viñas y a Elvira Martínez de Icaya Gómez que, aparte de ser unas magníficas docentes, han sido mis dos tutoras, las cuales me han acompañado con mucho cariño en todo el proceso del desarrollo de este trabajo.

Resumen

La teledetección es una técnica que permite obtener información de alta calidad sobre la superficie terrestre sin necesidad de estar en contacto físico directo. Se basa en la captura y el análisis de datos recibidos a partir de sensores ubicados en plataformas satelitales que recogen información electromagnética reflejada o emitida por la Tierra.

El objeto de estudio de este trabajo es el volcán Etna, situado en la isla de Sicilia al suroeste de la península de Italia. El volcán Etna es uno de los más activos y con erupciones más explosivas de toda Europa.

En este trabajo, se hace uso de la teledetección para realizar un estudio de cambios del volcán Etna. Para ello se utilizan imágenes proporcionadas por la misión Sentinel-2 de la ESA. Para realizar el análisis, se ha desarrollado en *Python* el algoritmo de clasificación no supervisada DBSCAN para llevar a cabo esta clasificación en imágenes satelitales. Este algoritmo ha sido implementado en la herramienta llamada QGIS, Sistema de Información Geográfica de software libre y código abierto utilizado para el procesamiento y análisis de imágenes satelitales. A partir de esto, se realiza una comparación con otro algoritmo de clasificación no supervisada, KMEANS, y finalmente se procede al análisis cuantitativo de cambios.

Abstract

Remote sensing is a technique that allows obtaining a high-quality information about the earth's surface without the need to be in direct physical contact. It is based on the capture and analysis of data received from sensors located on satellite platforms that collect electromagnetic information reflected or emitted by the Earth.

The object of study of this work is the Etna volcano, located on the island of Sicily in the southwest of the Italian peninsula. The Etna volcano is one of the most active and explosive eruptions in Europe.

In this work, remote sensing is used to carry out a study of changes in the Etna volcano. For this purpose, images provided by ESA's Sentinel-2 mission are used. To perform the analysis, an unsupervised classification algorithm has been developed using DBSCAN to perform this classification on satellite images, which it has been implemented in QGIS, Open-Source Geographic Information System for satellite image processing and analysis. From this, a comparison is made with another unsupervised classification algorithm, KMEANS, and lastly a quantitative analysis of changes is performed.

ÍNDICE

1.	INTRODUCCIÓN	12
1.1.	MOTIVACIÓN	12
1.2.	OBJETIVOS	12
1.3.	ESTRUCTURA.....	13
2.	ESTADO DEL ARTE	14
2.1.	TELEDETECCIÓN	14
2.1.1.	<i>¿Qué es la teledetección?</i>	14
2.1.1.1.	Teledetección activa.....	14
2.1.1.2.	Teledetección pasiva	15
2.1.2.	<i>¿Cómo funciona la teledetección?</i>	15
2.1.3.	<i>Radiación y Espectro Electromagnético</i>	17
2.1.3.1.	Interacción radiación-materia	19
2.1.4.	<i>Satélites</i>	19
2.1.4.1.	Satélites Sentinel	21
2.1.4.2.	Resolución de los sensores.....	23
2.1.5.	<i>Tratamiento digital de las imágenes multiespectrales</i>	24
2.1.5.1.	Formato imágenes multiespectrales.....	25
2.1.5.2.	Índices espectrales.....	25
2.2.	VOLCANES	26
2.2.1.	<i>¿Qué son los volcanes?</i>	26
2.2.1.1.	Tipos de volcanes y erupciones	28
2.2.2.	<i>Volcán Etna</i>	29
2.2.3.	<i>Volcanes y teledetección</i>	30
2.3.	ALGORITMOS DE CLASIFICACIÓN	31
2.3.1.	<i>Clasificación supervisada</i>	31
2.3.2.	<i>Clasificación no supervisada</i>	33
2.3.3.	<i>DBSCAN</i>	35
2.3.3.1.	Ventajas y desventajas de DBSCAN	37
3.	HERRAMIENTAS UTILIZADAS.....	39
3.1.	COPERNICUS OPEN ACCESS HUB	39
3.2.	SENTINEL HUB EO BROWSER.....	41
3.3.	ERDAS IMAGINE.....	46
3.4.	QGIS.....	48
4.	METODOLOGÍA.....	50
4.1.	FUENTE DE DATOS	50
4.2.	ORGANIGRAMA	51
4.3.	METODOLOGÍA.....	51
4.3.1.	<i>Búsqueda y descarga de imágenes</i>	51
4.3.2.	<i>Desarrollo del algoritmo DBSCAN</i>	57
4.3.3.	<i>Implementación del algoritmo en QGIS</i>	59
4.3.4.	<i>Evaluación de la calidad del DBSCAN</i>	62
5.	RESULTADOS Y COMPARACIÓN DEL IMPACTO.....	63
5.1.	RESULTADOS DE LAS CLASIFICACIONES	63
5.2.	COMPARACIÓN DE LAS CLASIFICACIONES	68
5.3.	ANÁLISIS CUANTITATIVO DE LAS IMÁGENES CLASIFICADAS	69
5.4.	APLICACIÓN DEL COEFICIENTE SILHOUETTE	74

6.	IMPACTOS SOCIALES Y AMBIENTALES	78
7.	CONCLUSIONES.....	80
7.1.	TRABAJO FUTURO	82
8.	BIBLIOGRAFÍA.....	83
9.	ANEXOS	88
9.1.	ANEXO A: CÓDIGO IMPLEMENTADO EN QGIS.....	88

Índice Ilustraciones

Ilustración 1: Teledetección activa	15
Ilustración 2: Teledetección pasiva	15
Ilustración 3: Longitud de onda	17
Ilustración 4: Representación de onda de luz.....	17
Ilustración 5: Niveles de onda y sus frecuencias.....	18
Ilustración 6: Reflexión especular y reflexión difusa	19
Ilustración 7: Órbita satélites heliosíncronos	20
Ilustración 8: Órbitas satélites Geosíncronos y Geoestacionarios	20
Ilustración 9: Altura satélites GEO, MEO y LEO.....	21
Ilustración 10: Representación satélites Sentinel en órbita	23
Ilustración 11: Longitud y resolución de las bandas de Sentinel-2	24
Ilustración 12: Formato imágenes multiespectrales	25
Ilustración 13: Movimiento convergente	27
Ilustración 14: Movimiento divergente	27
Ilustración 15: Punto caliente	28
Ilustración 16: Estructura volcán	28
Ilustración 17: Volcán Etna	30
Ilustración 18: Clasificación por paralelepípedos	32
Ilustración 19: Clasificación por mínima distancia	32
Ilustración 20: Clasificación por máxima verosimilitud	33
Ilustración 21: Clasificación por KMEANS.....	34
Ilustración 22: Puntos clasificación DBSCAN	36
Ilustración 23: Resultado clasificación DBSCAN.....	37
Ilustración 24: Pantalla inicio Copernicus Open Access Hub	39
Ilustración 25: Barra búsqueda Copernicus Open Access Hub	40
Ilustración 26: Selector filtros Copernicus Open Access Hub	40
Ilustración 27: Descarga imagen Copernicus Open Access Hub	41
Ilustración 28: Pantalla inicio Sentinel Hub	42
Ilustración 29: Carga de archivo de área	43
Ilustración 30: Selección de índices en Sentinel Hub.....	44
Ilustración 31: Personalización de índices en Sentinel Hub	45
Ilustración 32: Selección formato de descarga en Sentinel Hub.....	45
Ilustración 33: Pantalla inicio Erdas Imagine	46
Ilustración 34: Spatial Modeler en Erdas Imagine	47
Ilustración 35: Módulo Python Script Erdas Imagine	47
Ilustración 36: Pantalla inicio QGIS	49
Ilustración 37: Pantalla inicio Sentinel Hub	52
Ilustración 38: Selección área Sentinel Hub	53
Ilustración 39: Selección origen de imágenes Sentinel Hub	54
Ilustración 40: Selección rango de fechas Sentinel Hub.....	54
Ilustración 41: Selección fecha Sentinel Hub.....	55
Ilustración 42: Visualización imagen Sentinel Hub	55
Ilustración 43: Selección parámetros de descarga Sentinel Hub	56
Ilustración 44: Imágenes escogidas volcán Etna a) 23 junio 2021, b) 23 febrero 2021, c) 19 mayo 2021, d) 6 junio 2022	57
Ilustración 45: Imagen una banda y con composición de bandas 12, 8A y 4	58

Ilustración 46: Pantalla inicio QGIS	59
Ilustración 47: Caja herramientas de procesos QGIS	60
Ilustración 48: Interfaz DBSCAN QGIS	62
Ilustración 49: Imagen DBSCAN eps 1,2 min 30.....	68
Ilustración 50: Resultado KMEANS 7 clústeres	69
Ilustración 51: Comparación áreas clasificaciones DBSCAN vs KMEANS	73
Ilustración 52: Coeficiente Silhouette	75
Ilustración 53: Objetivos de desarrollo sostenible.....	78

Índice de tablas

Tabla 1 : Resultados 23 febrero 2021	64
Tabla 2: Resultados 19 mayo 2021.....	65
Tabla 3: Resultados 23 junio 2021	66
Tabla 4: Resultados 6 junio 2022	67
Tabla 5: Análisis cuantitativo 23 febrero 2021	70
Tabla 6: Análisis cuantitativo 19 mayo 2021.....	70
<i>Tabla 7: Análisis cuantitativo 23 junio 2021</i>	<i>71</i>
Tabla 8: Análisis cuantitativo 6 junio 2022	71
Tabla 9: Resultados coeficiente Silhouette	76

1. Introducción

La teledetección es una herramienta fundamental en la investigación y seguimiento de fenómenos medioambientales, como puede ser el seguimiento de fenómenos geológicos.

El volcán Etna, situado en Italia, más concretamente en la isla de Sicilia, es uno de los volcanes más activos y analizados del mundo. Su actividad volcánica constante, que incluye erupciones muy explosivas, grandes vertidos de lava y emisiones de gases, representa un gran reto para todos aquellos investigadores que buscan entender su comportamiento y evaluar todos los riesgos que conlleva.

En este contexto, este proyecto de fin de grado se va a centrar en la aplicación de la teledetección para la obtención de imágenes del volcán, utilizando datos de los satélites que componen la misión Sentinel-2 de la Agencia Espacial Europea (ESA), que más tarde serán clasificadas mediante el algoritmo de clasificación no supervisada *Density Based Spatial Clustering of Applications with Noise* (DBSCAN) desarrollado e implementado en la plataforma QGIS para la identificación y clasificación de las distintas características que puedan presentar las imágenes satelitales escogidas. También se realizará la clasificación de las mismas imágenes con el algoritmo KMEANS y su posterior análisis.

1.1. Motivación

La motivación principal de este trabajo de fin de grado es explotar todo el potencial del algoritmo DBSCAN en el contexto de la teledetección aplicada al seguimiento medioambiental del volcán Etna. Se busca utilizar dicho algoritmo para el análisis de imágenes de grandes dimensiones y optimizarlo, para hacerlo en tiempos razonables, debido a la gran complejidad en memoria que presenta DBSCAN.

El desarrollo de este proyecto, no solo se basará en la aplicación del algoritmo DBSCAN para la clasificación de imágenes del volcán Etna, sino que este algoritmo desarrollado será implementado en la aplicación de código abierto QGIS. Así se podrán adquirir distintos conocimientos acerca de la propia herramienta, del algoritmo y de uno de los volcanes más activos de toda Europa, el volcán Etna.

1.2. Objetivos

Este proyecto de fin de grado pretende desarrollar, aplicar e implementar el algoritmo de clasificación no supervisada DBSCAN para el estudio del volcán Etna a partir de imágenes satelitales de la misión Sentinel-2 basándose en los siguientes objetivos:

1. Obtención de imágenes en las que se aprecie la lava del volcán Etna. Estas imágenes deben ser lo más claras posibles, es decir, sin nubes, con la menor cantidad de nieve posible y con una buena resolución a partir la cual realizar el análisis.
2. Desarrollo del algoritmo DBSCAN en *Python* y su posterior implementación en la herramienta *Erdas Imagine/QGIS*.
3. Clasificación de las imágenes escogidas para realizar el estudio con DBSCAN, previamente desarrollado e implementado y con KMEANS.

4. Estudio de la evolución de la lava emitida e impacto sobre las zonas más cercanas al volcán (superficie vegetal...) y comparación de las clasificaciones realizadas tanto por DBSCAN como con KMEANS
5. Análisis cuantitativo, a partir del SCP (*Semi-Automatic Classification Plugin*) de QGIS, de los resultados de la clasificación realizada tanto con DBSCAN como con KMEANS.
6. Evaluación de resultados de la clasificación realizada con el algoritmo de DBSCAN desarrollado y KMEANS en imágenes con una sola banda haciendo uso del coeficiente Silhouette, desarrollado en *Python*.

1.3. Estructura

El documento presente está estructurado en 7 capítulos divididos en distintas secciones y un anexo. El primer capítulo, introduce el trabajo realizado, exponiendo la motivación y los objetivos. En el capítulo 2 se presenta el marco teórico en el que se ha basado el proyecto, detallando así todos los temas relacionados con la teledetección, los satélites, la misión Sentinel-2, el volcán Etna y los algoritmos de clasificación, profundizando especialmente en DBSCAN. A continuación, en el capítulo 3, se explicarán todas las herramientas utilizadas para desarrollar este proyecto. Posteriormente, en el capítulo 4, se detallará la metodología seguida para la obtención de resultados, abarcando desde la descarga de imágenes hasta la implementación del algoritmo. El capítulo 5 muestra los resultados obtenidos tras aplicar los algoritmos de clasificación en las imágenes escogidas para analizar. En este capítulo, aparte de mostrar la comparación entre los algoritmos DBSCAN y KMEANS, también muestra un análisis cuantitativo de los resultados. En el capítulo 6 se expone la contribución de este proyecto a la hora de proteger el medio ambiente y cumplir los ODS (Objetivos para el Desarrollo Sostenible) definidos en la Agenda 2030. Finalmente, en el capítulo 7, se expondrán las conclusiones extraídas de la realización de este proyecto y posibles líneas futuras de trabajo. Por último, se incluye un anexo con el código implementado en la herramienta QGIS y desarrollado en *Python*.

2. Estado del Arte

2.1. Teledetección

La teledetección es uno de los pilares fundamentales en los que se basa este proyecto, ya que se aplicará para conseguir imágenes de la superficie terrestre para realizar el estudio y seguimiento del impacto de la actividad eruptiva del volcán Etna. Es por ello por lo que se debe profundizar en este tema para poder entender e interpretar correctamente los resultados de este trabajo. Además, es una herramienta de mucho valor sobre todo en los campos del estudio de la agricultura, vegetación, temperatura terrestre o estudios de crecimiento de grandes ciudades entre otros.

2.1.1. ¿Qué es la teledetección?

La teledetección, también conocida como detección remota, puede ser definida como la ciencia o arte de adquirir información a distancia de la superficie terrestre, utilizando plataformas espaciales para la medición y observación de objetos situados en la Tierra para su posterior análisis.

En función de cómo trabaja el instrumento a partir del cual se obtiene toda la información y, según la fuente de señal que se usa para explorar el objeto, se pueden encontrar dos tipos de teledetección. La teledetección activa y la teledetección pasiva, las cuales se encuentran detalladas a continuación:

2.1.1.1. Teledetección activa

Se puede considerar como teledetección activa aquella que funciona con su propia fuente de emisión de luz. Gracias a la energía emitida, el teledetector es capaz de medir la radiación reflejada del objeto de estudio situado, en este caso, en la superficie terrestre.

El Lidar (*Light Detection And Ranging*), es un ejemplo muy claro de lo que sería un sensor que aplica la teledetección activa para escanear la superficie terrestre. Esta tecnología, a partir de impulsos luminosos, permite obtener medidas de posicionamiento mediante modelos digitales tridimensionales. Otros ejemplos de sensores que usan este principio serían los radares, las sondas para estudiar las condiciones meteorológicas o el difusómetro.

Debido a la capacidad de estos sensores de emitir su propia fuente de energía, pueden funcionar en cualquier momento del día al no depender de fuentes externas como podría ser la luz solar. Además, se pueden considerar en gran medida independientes de las dispersiones atmosféricas.

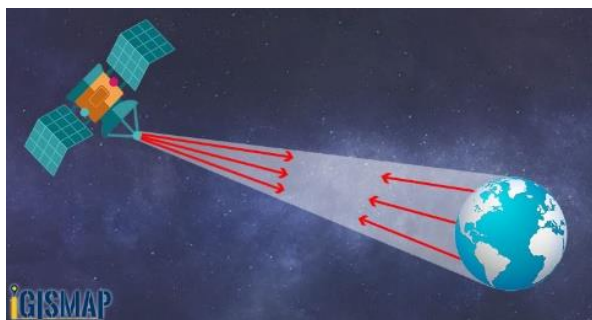


Ilustración 1: Teledetección activa [<https://www.igismap.com/what-is-remote-sensing>]

2.1.1.2. Teledetección pasiva

Se puede considerar como teledetección pasiva aquella cuya radiación primaria es el Sol, por lo que los sensores pasivos no son capaces de transmitir su propia fuente de radiación, tal y como ocurre en la teledetección activa. Estos sensores usan la energía emitida por el Sol, la cual se refleja en el objetivo situado en la superficie terrestre para luego ser captada por los sensores del satélite. Es por todo esto por lo que, a diferencia de la teledetección activa, la teledetección pasiva solo se puede aplicar en condiciones en las que haya suficiente luz solar.

La teledetección pasiva hace uso de sensores multispectrales, los cuales combinan múltiples bandas siendo así capaces de obtener datos con diferente longitud de onda. Estas bandas pueden incluir tanto espectros visibles para el ojo humano como aquellos que no lo son (IR, NIR, TIR...). Todo esto es explicado más adelante.

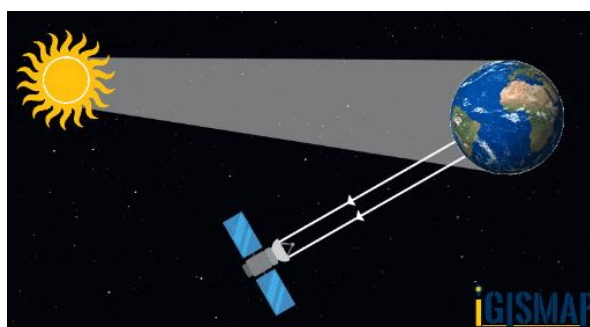


Ilustración 2: Teledetección pasiva [<https://www.igismap.com/category/remote-sensing>]

2.1.2. ¿Cómo funciona la teledetección?

Hay tres elementos fundamentales a la hora de aplicar la teledetección: Una plataforma, un objeto a observar y un sensor. Otro elemento que es importante sería la información que se obtiene con los datos de la imagen y como se analiza y almacena.

- **Una plataforma:** Como se ha comentado anteriormente, la teledetección permite la observación de objetos terrestres. Para conseguir estas imágenes necesitamos una plataforma que nos aleje de la superficie de la Tierra. Dependiendo del caso de estudio y de los medios, estas plataformas pueden ser:
 - o **Aviones:** Están equipados con una serie de cámaras muy sofisticadas para la obtención de los datos de la superficie terrestre. Los aviones, suelen volar a alturas bajas, es decir, entre unos 10,000 metros y 12,000 metros. Es por esto por lo que solo pueden obtener imágenes de áreas reducidas. A pesar de ello, las instantáneas tomadas por aviones pueden llegar a tener mucho detalle.
 - o **Satélites:** Son unos cuerpos que orbitan alrededor de la Tierra o un planeta. Un ejemplo muy claro de un satélite sería la Luna. Pero para el caso de la teledetección, los satélites que se utilizan como plataforma son los satélites artificiales. Actualmente los satélites son las plataformas principales empleadas para la obtención de información sobre la superficie terrestre a través de la teledetección. Este tipo de plataforma lleva una gran variedad de sensores para estudiar varios aspectos de la Tierra como pueden ser el tiempo atmosférico, el calor superficial o la vegetación entre otros. Al estar orbitando constantemente alrededor de la Tierra, nos permite tener imágenes de toda la superficie y en varios momentos y condiciones meteorológicas.
En nuestro caso de estudio esta va a ser la plataforma utilizada por lo que más adelante se le dedicará un capítulo para explicarla en profundidad.
- **Un objeto:** En teledetección, el principal objeto de estudio es la superficie de la Tierra. Un objeto puede ser, por ejemplo, una ciudad, un océano o, como en el caso de este estudio, un volcán. Dependiendo del tipo de objeto, la luz recogida por el sensor será diferente, por lo que se podrán obtener características propias de cada uno de ellos.
- **Un sensor:** Es el instrumento encargado de recibir la radiación reflejada por el objeto para poder obtener así información del mismo.

Además de estos elementos, se le suma como factor influyente en los resultados de la teledetección, la atmósfera. La radiación que atraviesa la atmósfera sufre diferentes variaciones como resultado de la reflexión y absorción que sufre la señal en su recorrido antes de llegar al sensor de la plataforma. Casi toda la radiación que se encuentra entre los valores de la gama de 9.5 – 10 micrómetros suele quedar absorbida por la atmósfera. Se conocen como ventanas atmosféricas los rangos de longitudes de onda en los que se puede considerar que la atmósfera es casi transparente y prácticamente no altera la luz que la atraviesa, aunque siempre sufrirá alguna modificación debido a la absorción y difusión atmosférica.

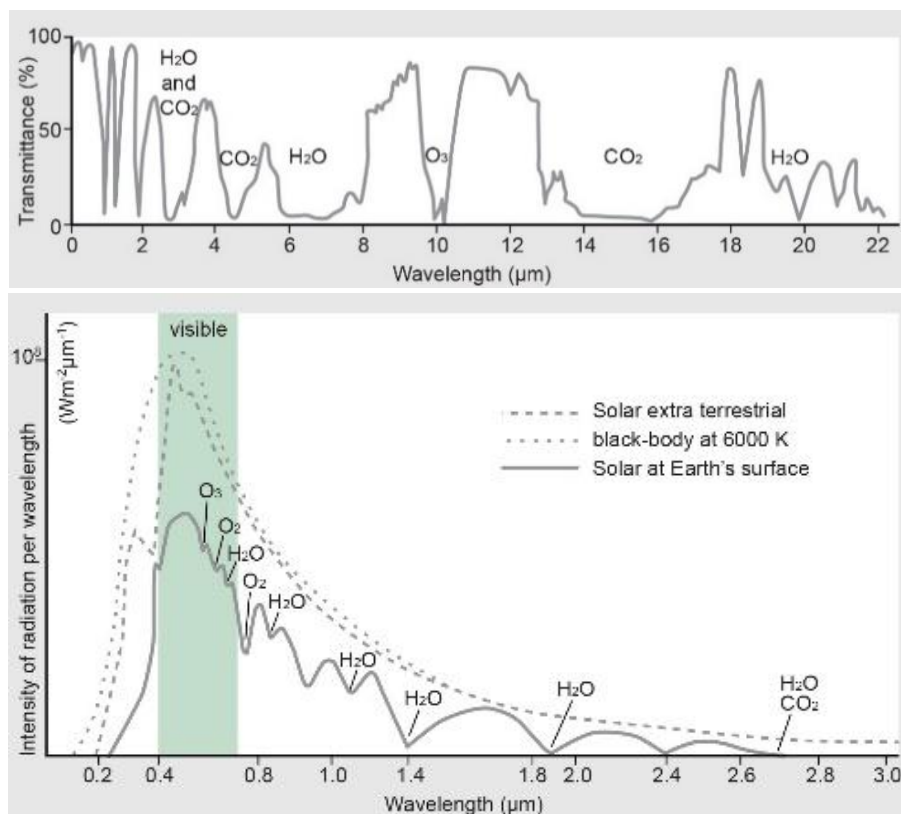


Ilustración 3: Longitud de onda [<https://ltb.itc.utwente.nl/page/498/concept/81484>]

2.1.3. Radiación y Espectro Electromagnético

La radiación electromagnética es una combinación de campos eléctricos y magnéticos oscilantes, perpendiculares entre sí, que se propagan a través del espacio a la velocidad de la luz transportando energía de un lugar a otro.



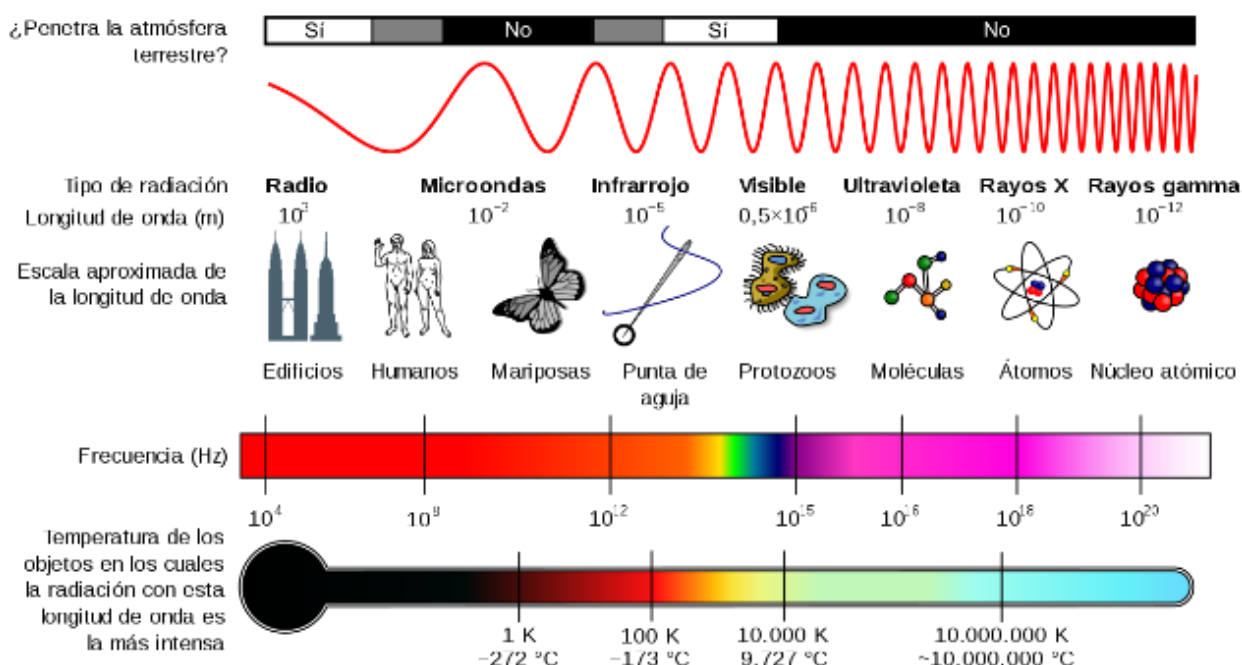
Ilustración 4: Representación de onda de luz [<https://es.wikipedia.org/wiki/Archivo:Light-wave-es.svg>]

En nuestro caso de estudio, estas ondas son captadas por los sensores que tienen incorporados los satélites. El entorno de estudio puede reflejar ondas electromagnéticas con diferente longitud de onda dependiendo de la capacidad de absorción y reflexión del objeto. Estas ondas cubren un gran rango del espectro

electromagnético, desde ondas de radio o microondas, las cuales son de baja energía, hasta ondas infrarrojas, ondas ultravioletas y ondas de rayos X o gamma, las cuales son de muy alta energía.

El espectro electromagnético se puede dividir en distintas zonas en función de la frecuencia o longitud de onda de la radiación:

- **Ondas de Radio** ($> 1\text{mm}$): Este grupo incluye todas las ondas de radio y microondas. Estas suelen ser ondas largas, pudiendo llegar a tener una longitud de onda de varios kilómetros. Este tipo de ondas suelen usarse para conexiones inalámbricas como la transmisión de radio o el Wi-Fi.
- **Ondas Infrarrojas** ($0,7\text{ }\mu\text{m} - 7\text{ }\mu\text{m}$): Se pueden clasificar en tres subgrupos. El infrarrojo lejano o térmico, el cual detecta la radiación en forma de calor. El infrarrojo medio y el infrarrojo próximo. Este tipo de ondas, aparte de ser usadas para la detección de calor, también pueden ser utilizadas en equipos de visión nocturna o la comunicación de fibra óptica.
- **Ondas visibles** ($0,4\text{ }\mu\text{m} - 0,7\text{ }\mu\text{m}$): Estas son las ondas que son detectadas por el ojo humano.
- **Ondas Ultravioleta** ($10\text{ nm} - 400\text{ nm}$): Este grupo de ondas sufre mucha absorción atmosférica, por lo que la calidad de la señal suele ser bastante baja.
- **Rayos X** ($0,01\text{ nm} - 10\text{ nm}$): Suelen ser utilizados en la medicina para la realización de pruebas radiológicas.
- **Rayos Gamma** ($< 0,01\text{ nm}$): Este grupo incluye las longitudes de onda más pequeñas, y su uso se basa principalmente en el estudio de la física de partículas y la astronomía.



[https://meteoglosario.aemet.es/es/termino/355_espectro-electromagnetico]

2.1.3.1. Interacción radiación-materia

La materia está formada por átomos y moléculas que interaccionan con la radiación electromagnética recibida. Su interacción depende de la estructura de la materia y de la longitud de onda de la radiación. Dependiendo de las condiciones dadas la onda puede ser dispersada (reflexión difusa o lambertiana), reflejada (reflexión especular), absorbida o transmitida por la atmósfera o la superficie terrestre.

La reflexión ocurre cuando las ondas electromagnéticas inciden sobre la superficie de cualquier medio natural, como puede ser la superficie terrestre o la atmósfera y parte de esa energía de la onda no llega a entrar en el material y, por lo tanto, se refleja. Existen dos tipos de reflexiones:

- **Reflexión Especular:** Este tipo de reflexión se suele producir en superficies planas siendo el ángulo de reflexión igual al de incidencia.
- **Reflexión difusa o lambertiana:** A diferencia de la reflexión especular, esta se suele producir cuando la rugosidad de la superficie es mayor que la longitud de onda, reflejándose así en diferentes direcciones.

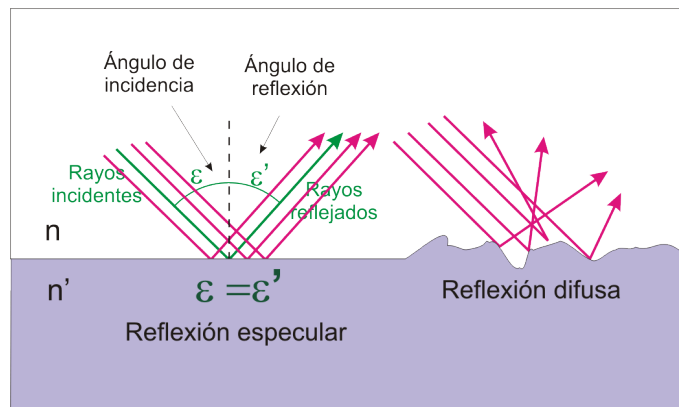


Ilustración 6: Reflexión especular y reflexión difusa
[\[https://riunet.upv.es/bitstream/handle/10251/144779/47103755P_TFG_15814112690433982411676725399313.pdf?sequence=1\]](https://riunet.upv.es/bitstream/handle/10251/144779/47103755P_TFG_15814112690433982411676725399313.pdf?sequence=1)

2.1.4. Satélites

Un satélite se puede definir como un objeto artificial puesto en órbita alrededor de la Tierra con el objetivo de recopilar datos, observar su superficie o dar servicios de telecomunicación. Para dar estos servicios, los satélites suelen estar equipados con sensores muy sofisticados y tecnología de última generación.

Se pueden realizar diversas clasificaciones según la altura de órbita o el recorrido de los satélites.

Según su recorrido:

- **Satélites Heliosíncronos:** Estos satélites, también llamados de órbita polar, viajan entre los polos a una altura de entre 200 km y 1000 km según la ESA (Agencia Espacial Europea). Aunque hacen un recorrido entre polos, su trayectoria no tiene por qué pasar exactamente por los mismos y suele tener una cierta inclinación comprendida entre 20º y 30º. Un ejemplo de satélites heliosíncronos son los LANDSAT.

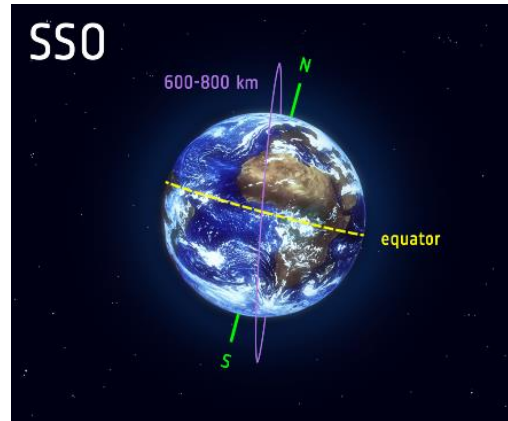


Ilustración 7: Órbita satélites heliosíncronos
[https://www.esa.int/ESA_Multimedia/Images/2020/03/Low_Earth_orbit]

- **Satélites Geoestacionarios:** Se encuentran a 36000 km de altura respecto la superficie de la tierra. Al tener el mismo periodo de rotación que la Tierra, siempre se encuentran sobre el mismo punto de su superficie. Un ejemplo de este tipo de satélites es el METEOSAT.
- **Satélites con órbitas inclinadas (Geosíncronos):** Los satélites geosíncronos se encuentran a una altura similar a los geoestacionarios, pero, a diferencia de éstos, su órbita puede tener cualquier inclinación respecto del plano ecuatorial.



Ilustración 8: Órbitas satélites Geosíncronos y Geoestacionarios [https://axessnet.com/que-son-y-para-que-sirven-los-satelites-geoestacionarios/]

Según su altura:

- **Órbita baja terrestre (LEO):** En este grupo se encuentran los satélites que orbitan a baja altura respecto de la superficie de la Tierra, en alturas en un rango entre 160 km y 1000 km- Para ponernos en situación, un avión comercial no suele volar a más de 14 km sobre la superficie terrestre. Debido a la poca altura a la que se encuentran, estos satélites orbitan a una gran velocidad para, poder así, contrarrestar la fuerza de la gravedad. Los satélites pertenecientes a este grupo tienen un periodo de órbita de entre 80 minutos y 150 minutos. Normalmente, debido a su baja altura de vuelo, son utilizados para la observación de la Tierra y la toma de imágenes en alta resolución. Algunos ejemplos de satélites que se encuentran en estos rangos de altura son el Sentinel-2 o la Estación Espacial Internacional (ISS).
- **Órbita media terrestre (MEO):** Los satélites pertenecientes a este grupo se encuentran a una altura de entre 5000 km a 20000 km. Estos, al estar a más altura que los del grupo anterior tienen un periodo de órbita menor, encontrándose entre las 10 horas y 14 horas. Estos satélites se suelen usar para realizar tareas relacionadas con el sector de la telecomunicación, como, por ejemplo, conexiones de baja latencia.
- **Órbita alta terrestre (GEO):** Esta órbita se encuentra a una altura superior a los 36000 km teniendo así un periodo orbital de unas 24 horas, para ser exactos, según la ESA, los satélites pertenecientes a este grupo tardan 23 horas, 56 minutos y 4 segundos en completar una vuelta completa a la Tierra ya que viajan exactamente a la misma velocidad que ella. Suelen ser usados por aquellos satélites que necesitan estar constantemente sobre un punto fijo de la Tierra, siendo perfecto para satélites de telecomunicación o aquellos dedicados a la monitorización del clima. Sobre esta órbita se encuentran los llamados satélites geoestacionarios.
- **Órbitas elípticas (HEO):** Estos tipos de órbitas son muy poco utilizadas. Los satélites que se encuentran en esta órbita, a diferencia de los demás, no están constantemente a la misma altura sobre la Tierra. Esto es debido a la forma elíptica de la misma.

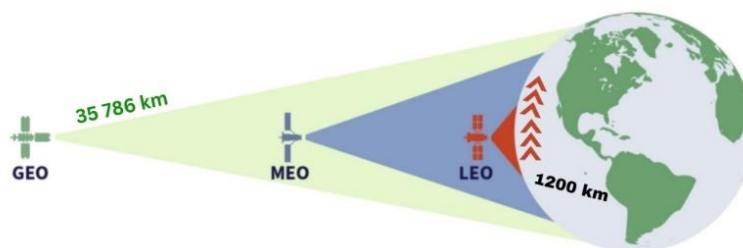


Ilustración 9: Altura satélites GEO, MEO y LEO
[<https://www.thespacereview.com/article/4504/1>]

2.1.4.1. Satélites Sentinel

Sentinel es un proyecto compuesto de varios satélites desarrollados por la Agencia Espacial Europea (ESA) dentro del marco del Programa *Copernicus*. Uno de sus objetivos es proporcionar imágenes de alta calidad que puedan ser accesibles libremente. Este proyecto lleva funcionando desde el año 2014. Desde ese mismo año, los satélites pertenecientes al programa suministran información de manera continua, incluyendo datos acerca del movimiento del terreno, control de océanos y mares, monitorización de densidad de población o seguimiento de volcanes, entre muchos otros.

Como se ha descrito anteriormente, este programa cuenta con numerosas misiones, cada una con un propósito diferente. Hasta ahora conocemos 7 misiones de la familia Sentinel, cada una de ellas con una constelación formada por varios satélites. Estas misiones se describen a continuación:

- **Sentinel-1:** Es una misión radar la cual muestra imágenes tanto terrestres como oceánicas en cualquier condición meteorológica, y en cualquier momento del día, sin importar si es de día o de noche. El primer satélite lanzado de esta misión fue en abril de 2014 y sigue una órbita polar.
- **Sentinel-2:** El objetivo de esta misión es monitorizar la superficie terrestre pudiendo proporcionar imágenes de alta resolución sobre la cobertura de suelo y agua, control de la vegetación o monitorización de la actividad volcánica. Los dos satélites que componen esta misión, al igual que en Sentinel-1, siguen una órbita polar, y serán los que utilizaremos para este proyecto. El primer satélite lanzado de Sentinel-2, denominado Sentinel-2A, fue lanzado en junio de 2015 y el segundo, Sentinel-2B, en marzo de 2017.
- **Sentinel-3:** El principal propósito de esta misión es la observación marina, aunque también realiza observaciones terrestres. Entre sus funciones se encuentra la medición de la temperatura o elevación tanto de la superficie terrestre como de océanos. Esta misión proporciona información muy valiosa para el estudio meteorológico y pronóstico marino y está compuesta por 2 satélites, donde el primero de ellos, el Sentinel-3A fue lanzado en febrero de 2016. Actualmente se encuentran en desarrollo dos satélites más que se unirán a ellos en un futuro.
- **Sentinel-4:** Esta misión, aunque en la actualidad todavía no se encuentra operativa, facilitará datos de vigilancia sobre la composición atmosférica haciendo uso de la tercera generación de METEOSAT. Se estima que el primer lanzamiento será en 2024 y el segundo en 2030.
- **Sentinel-5:** Tiene como objetivo la recopilación de datos para la monitorización de la calidad del aire. Para realizar esta medición se acoplará un espectrómetro al satélite Meteorológico Operacional MetOp aprovechando el lanzamiento de la segunda generación.
- **Sentinel-5P:** También conocido como Sentinel-5 Precursor, esta misión sirve para completar la falta de datos entre el lanzamiento de la misión Sentinel-5 y la retirada de *Envisat* y la misión Aura. Fue lanzado en 2017.

- **Sentinel-6:** Fue lanzada en 2020 y es un altímetro espacial de alta precisión cuyo objetivo es recopilar datos acerca de la altura de la superficie marina en todo el planeta para poder así hacer un control y monitoreo sobre el aumento de los niveles de los océanos y mares y ayudar a los estudios sobre el clima en general.



Ilustración 10: Representación satélites Sentinel en órbita [<https://www.ceit.es/noticias/-/contents/20/03/2023/tecnologia-espacial-y-blockchain-para-reforzar-la-trazabilidad-de-la-cadena-agroalimentaria/content/1tZ9oin6Nj8k/44496964>]

2.1.4.2. Resolución de los sensores

La resolución de los sensores se refiere a la unidad mínima discernible en el proceso de captación remota de la información. Existen cuatro tipos de resolución:

- **Resolución espacial:** Esta resolución es el tamaño de un píxel. Para entenderlo mejor, un píxel, en términos de teledetección, es el punto más pequeño que somos capaces de distinguir sobre la superficie de la Tierra. Dependiendo de la distancia del sensor al objeto y de la tecnología del equipo se pueden diferenciar: resolución espacial alta (1 a 5 metros por píxel), resolución espacial media (5 a 30 metros por píxel) y resolución espacial baja (30 a 250 metros por píxel). Una resolución espacial baja muestra una imagen con un tamaño de píxel grande por lo que se ve borrosa y con poco detalle, mientras que una resolución espacial alta tiene mucho más detalle al tener un tamaño de píxel más pequeño. Para resumir, la resolución espacial define la claridad y precisión con la que una imagen muestra los detalles de la superficie terrestre.
- **Resolución espectral:** Indica el número y ancho de las bandas en las que el sensor capta radiación electromagnética, es decir, el número de canales espectrales. Cuanto mayor sea la resolución espectral, mayor será el número de bandas de la imagen. Por ejemplo, las imágenes tomadas por la misión Sentinel-2, cuentan con 12 bandas, por lo que tiene una resolución espectral de 12.
- **Resolución radiométrica:** También conocida como resolución dinámica, indica la sensibilidad del sensor, es decir, la capacidad del sensor para medir con precisión la intensidad de la radiación. Se

suele expresar mediante el número de bits necesarios para almacenar cada píxel. Por ejemplo, Sentinel 2 utiliza 12 bits.

- **Resolución temporal:** Indica la frecuencia de paso del satélite por un mismo punto de la superficie terrestre. Este tipo de resolución depende exclusivamente de las características de la órbita que sigue el satélite. Por ejemplo, los satélites pertenecientes a la misión Sentinel-2 tienen una resolución temporal de 5 días, es decir, tarda 5 días en volver a observar un mismo punto en la superficie terrestre.

Para la realización del proyecto se hará uso de las imágenes proporcionadas por la misión Sentinel-2. Esta misión, aparte de lo que se ha comentado con anterioridad, ofrece imágenes con una resolución espacial de 10m x 10m por píxel, 20m x 20m por píxel o 60m x 60m por píxel, dependiendo de la banda espectral. En la siguiente imagen se pueden observar, para cada banda, la longitud de onda central y su resolución.

Sentinel-2 Bands	Central Wavelength (µm)	Resolution (m)
Band 1 - Coastal aerosol	0.443	60
Band 2 - Blue	0.490	10
Band 3 - Green	0.560	10
Band 4 - Red	0.665	10
Band 5 - Vegetation Red Edge	0.705	20
Band 6 - Vegetation Red Edge	0.740	20
Band 7 - Vegetation Red Edge	0.783	20
Band 8 - NIR	0.842	10
Band 8A - Vegetation Red Edge	0.865	20
Band 9 - Water vapour	0.945	60
Band 10 - SWIR - Cirrus	1.375	60
Band 11 - SWIR	1.610	20
Band 12 - SWIR	2.190	20

Ilustración 11: Longitud y resolución de las bandas de Sentinel-2
[https://www.researchgate.net/figure/a-Landsat-8-and-b-Sentinel-2-band-characteristics-6-9-10_fig2_334675844]

2.1.5. Tratamiento digital de las imágenes multiespectrales

Como resultado de la toma de imágenes por parte de los distintos satélites que se encuentran en órbita alrededor de la Tierra, se obtienen una serie de imágenes del área de estudio. Estas imágenes son preprocesadas por el propio satélite y enviadas a distintas estaciones que se encuentran en la Tierra para su posterior procesamiento.

2.1.5.1. Formato imágenes multiespectrales

Estas imágenes multiespectrales cuentan con una capa distinta por cada banda que capte el satélite, tal y como se explica en el capítulo 1.4.1, teniendo así una imagen de resolución $i \cdot j \cdot k$, siendo i , la altura de la imagen, j el ancho y k el número de capas o bandas de la imagen.

Hay que tener en cuenta que para que la imagen pueda ser visualizada en color debe tener como mínimo 3 canales. A estos canales se les llama RGB. Cuando queremos mostrar "color verdadero", en estos canales se introducen las bandas o espectros rojo, verde y azul. No obstante, los canales RGB se pueden combinar de distintas formas con las bandas espectrales dependiendo del resultado deseado. Por ejemplo, en el Sentinel 2 la combinación RGB de bandas 8, 4 y 3 permite resaltar la vegetación sana que aparece coloreada en rojo intenso debido a su alta emisión en el infrarrojo.

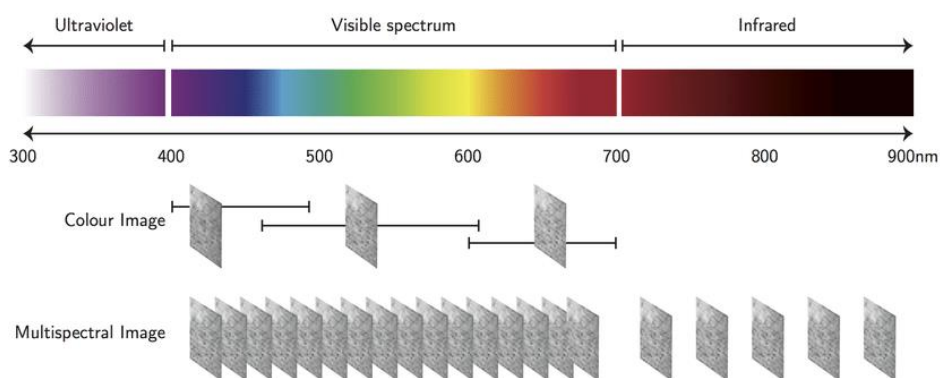


Ilustración 12: Formato imágenes multiespectrales
[\[https://www.researchgate.net/figure/Multispectral-images-are-captured-in-a-similar-process-to-colour-images-A-colour-image_fig2_295856116\]](https://www.researchgate.net/figure/Multispectral-images-are-captured-in-a-similar-process-to-colour-images-A-colour-image_fig2_295856116)

2.1.5.2. Índices espectrales

Los índices espectrales son combinaciones aritméticas de distintas bandas correspondientes a una misma imagen multiespectral, que nos permiten identificar los elementos de la superficie terrestre (agua, vegetación, suelos, nieve, lava, zonas quemadas, etc.) contenidos en dicha imagen.

Existen bastantes índices espectrales. En el caso de Landsat y Sentinel, comparten el mismo tipo de índices, aunque la composición de bandas es distinta dependiendo de la misión. A continuación, se muestran los índices más conocidos y utilizados por estas dos misiones:

- **Normalized Difference Vegetation Index (NDVI):** Sirve para el estudio de la vegetación. La fórmula de composición es la siguiente:

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

Dónde NIR es la banda del infrarrojo cercano y Red es la banda roja del espectro visible.

- **Normalized Difference Water Index (NDWI):** Se utiliza para el estudio y análisis de masas de agua, como mares, lagos u océanos. Su fórmula de composición es distinta según el método que sigamos:

- Método de McFeeters (1996):

$$NDWI = \frac{Green - NIR}{Green + NIR}$$

- Método de Gao (1996):

$$NDWI = \frac{NIR - SWIR}{NIR + SWIR}$$

Dónde *Green* es la banda verde del espectro visible y *SWIR* es la región del espectro infrarrojo cercano definido entre 1 y 2,5 micras.

- **Normalized Burned Ratio Index (NBRI):** Destaca el área de zonas que han sufrido el efecto de un incendio. La fórmula de composición del índice es la siguiente:

$$NBRI = \frac{NIR - SWIR}{NIR + SWIR}$$

Se puede destacar que la composición de este índice es exactamente igual al NDWI definido por el método Gao.

- **Normalized Difference Snow Index (NDSI):** Este índice destaca sobre todo áreas en las que hay nieve. Su fórmula es:

$$NDSI = \frac{Green - SWIR}{Green + SWIR}$$

En este trabajo, se utilizarán las bandas 12 (SWIR 2), 8A (NIR) y 4 (RED) de las imágenes ofrecidas por la misión Sentinel-2 para representar las imágenes RGB y poder analizar y estudiar los efectos medioambientales del volcán Etna en su entorno. Esta composición destaca en rojo la lava ardiente debido a su fuerte emisión en la banda 12 (SWIR 2).

En cuanto a índices, en el contexto de este trabajo no se le dará uso a ninguno, pero en el caso en el que, por ejemplo, se quisiera resaltar la vegetación, se usaría el índice NDVI mencionado con anterioridad.

2.2. Volcanes

2.2.1. ¿Qué son los volcanes?

Según el Instituto Geográfico Nacional (IGN), “un volcán es el resultado visible en la superficie terrestre de un proceso geológico el cual expulsa material rocoso fundido y gases del interior de la Tierra”. El resultado de esta sucesiva emisión de magma y de gases produce un relieve que, en la mayoría de los casos, es de forma cónica y, dependiendo de distintos factores, puede tener una altura u otra. Hoy en día se puede estimar que hay 1350 volcanes potencialmente activos alrededor de todo el planeta.

Los volcanes se pueden formar por distintas causas.

- Movimiento de placas tectónicas:** Este tipo de volcanes se producen en los bordes de las placas tectónicas. Dependiendo del movimiento producido por estas placas puede ser convergente o divergente.
Se dice que un movimiento de placas es **convergente** cuando dos placas, una oceánica y otra continental convergen, es decir, chocan entre ellas. Al haber una placa más densa que la otra, la menos densa, que en este caso es la placa continental, se hunde creando una gran fricción y originando así la salida del magma.

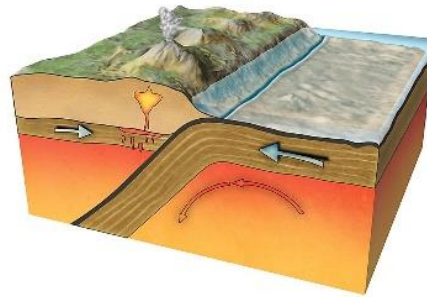


Ilustración 13: Movimiento convergente [<https://cordis.europa.eu/article/id/36050-understanding-the-forces-that-shape-the-earth/es>]

En cambio, un movimiento de placas **divergente** se produce cuando dos placas tectónicas se separan, en este caso, la litosfera sufre una fisura a través de la cual el magma emerge a capas superiores produciendo una erupción volcánica.

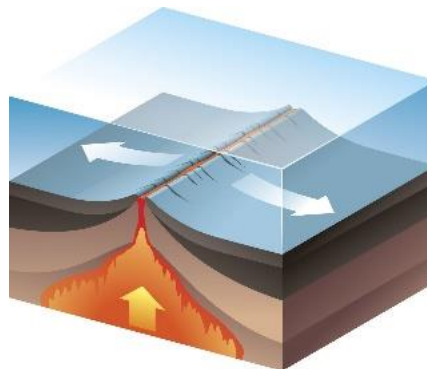


Ilustración 14: Movimiento divergente [<https://www.istockphoto.com/es/vector/situaciones-l%C3%ADmite-de-placa-gm483766933-37778124>]

- Puntos calientes:** Se conocen como puntos calientes las zonas de ascenso de magma que llegan hasta la corteza de la tierra, generando así volcanes que no son producidos por movimiento de placas.

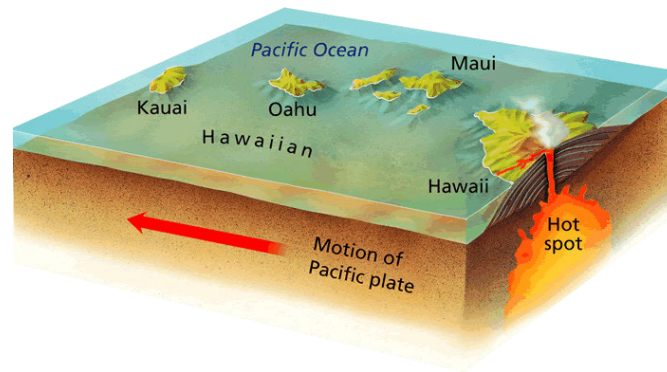


Ilustración 15: Punto caliente [<https://www.pngwing.com/es/free-png-ztssz>]

En un volcán se pueden describir distintas partes. Las más importantes son:

- **Cráter:** Es la apertura que se encuentra en la cima del volcán a través del cual se expulsa el magma, las cenizas, los gases y todo el material piroclástico que es proyectado por el mismo.
- **Cono volcánico:** Se puede definir como el producto de toda la acumulación de lava que se solidifica en el cráter del volcán.
- **Chimenea:** Es el conducto por el cual asciende la lava y que conecta la cámara magmática con el cráter.
- **Cámara magmática:** Es una bolsa encargada de almacenar todo el contenido magmático que procede del interior de la Tierra.

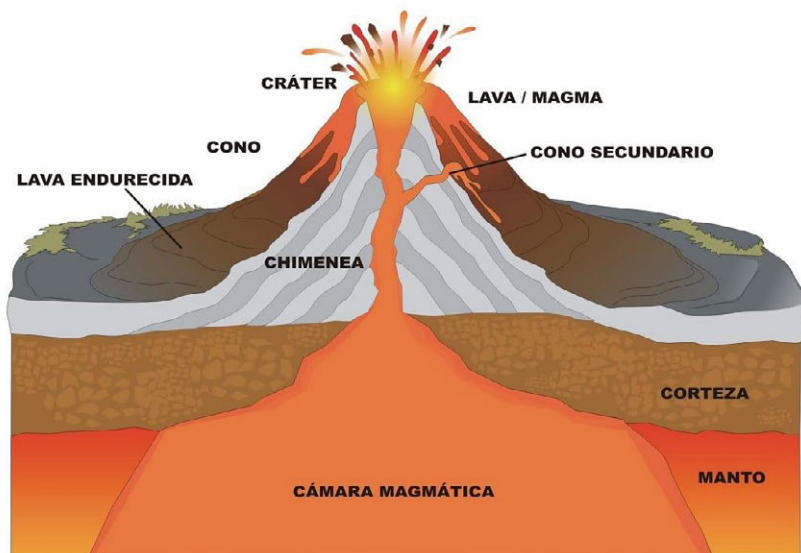


Ilustración 16: Estructura volcán
[http://agrega.juntadeandalucia.es/repositorio/31102018/91/es-an_2018103112_9131849/12_tipos_de_volcanes_y_estructuras_volcnicas.html]

2.2.1.1. Tipos de volcanes y erupciones

Los volcanes se pueden clasificar en distintos grupos dependiendo de su actividad, su forma o el tipo de erupción que produce. A continuación, se enumeran los distintos tipos de volcanes según diferentes características.

Según su actividad, los volcanes se pueden clasificar en tres grupos distintos. Los volcanes activos son aquellos que pueden entrar en estado de erupción en cualquier momento, como por ejemplo el volcán Etna. Los volcanes inactivos son aquellos que llevan sin sufrir ninguna erupción durante siglos. Finalmente, los volcanes extintos son aquellos que llevan sin erupcionar miles de años, por lo que se consideran apagados por completo.

Dependiendo de la morfología que presente un volcán, se podrá clasificar en cinco grupos. Los volcanes en escudo, domos de lava, estratovolcanes, conos de escoria o calderas volcánicas.

Finalmente, dependiendo de la erupción producida, podemos encontrar volcanes hawaianos, volcanes vulcanianos, volcanes islándicos, volcanes estrombolianos, volcanes hidromagmáticos, volcanes submarinos o volcanes peleanos.

2.2.2. Volcán Etna

El volcán Etna se encuentra en la isla de Sicilia en Italia. Más concretamente se sitúa en la costa este de la isla, cerca de la población de Catania, la segunda ciudad más poblada de la región de Sicilia. Su historial se puede remontar hasta el año 1500 AC, año a partir del cual se tiene constancia de su existencia.

El también conocido como el monte Etna o “Mongibello” en el idioma regional de la isla, tiene una altura de algo más de 3300 metros sobre el nivel del mar, pudiendo variar varios metros dependiendo de su actividad, considerándose el volcán activo más alto de toda Europa.

Este volcán es de tipo estratovolcán por lo que tiene una forma cónica con un cráter principal situado en el centro, además tiene otros cráteres repartidos a lo largo de su superficie sumando un total de cuatro y un alto número de respiraderos. Al ser un estratovolcán, tiene varias capas y cámaras que contienen lava.

Las erupciones producidas por el Etna son muy explosivas y se denominan erupciones estrombolianas. La principal característica de este tipo de erupciones es que el material magmático es muy denso. Tal y como se ha comentado con anterioridad, las erupciones son bastante explosivas, esto quiere decir que proyectan bastante material piroclástico como resultado de la cristalización de la lava a lo largo de la chimenea durante su ascenso al cono volcánico.

A lo largo de la historia, este volcán ha producido bastantes catástrofes debido a erupciones muy violentas y explosivas, pero sin duda alguna, la que más daño causó fue la producida en el año 1669, la cual arrasó y enterró a varias localidades situadas en la periferia del volcán llegando incluso a afectar a la ciudad de Catania. Otras erupciones históricas son las producidas en el 122 AC y en el año 1787. Desde el inicio del siglo XXI, las erupciones más potentes se produjeron entre los años

2002 y 2003 las cuales liberaron una enorme cantidad de lava y cenizas. Estas últimas se llegaron a ver al otro lado del Mar Mediterráneo.



Ilustración 17: Volcán Etna [https://www.peakpx.com/es/hd-wallpaper-desktop-guyxh]

2.2.3. Volcanes y teledetección

En este proyecto, se va a aplicar el algoritmo de clasificación DBSCAN, el cual se explicará más adelante, para la detección de cambios en las últimas erupciones del volcán Etna a través del uso de imágenes de teledetección.

En el contexto de este trabajo, daremos especial importancia a la información sobre la lava y el área que ocupa. Para conseguir esto, se necesita analizar una banda que resalte la luz reflejada por los materiales que se encuentran a una alta temperatura en la superficie de la Tierra. En los satélites Sentinel-2, esa banda es la conocida como banda SWIR o banda infrarroja de onda corta.

La banda SWIR, a parte de la lava incandescente también capta muy bien la nieve, agua o nubes, por lo que al realizar la obtención de imágenes para su posterior estudio se debe evitar en ellas la presencia de elementos que puedan corromper la clasificación resultante. En nuestro caso particular, elegir una imagen del volcán Etna es todo un reto por varios factores.

- Al encontrarse a una gran altura (3300 metros) es muy propenso a que una gran parte de la cima del volcán presente una importante masa de nieve durante varios meses a lo largo del año que suelen comprender, de media, entre los meses de octubre hasta mayo. Esto limita mucho la búsqueda de imágenes.
- Otro factor que dificulta la selección de imágenes es la presencia de nubes, pues al encontrarse a una gran altura y al estar situado en una isla que sufre constantes ráfagas de viento, es muy común que haya nubes cubriendo parcialmente el cielo.
- Finalmente, al tratarse del estudio de un volcán en erupción, las imágenes que interesan son aquellas en las que la lava ardiente está presente. El estado de erupción asimismo conlleva la emisión de gases y ceniza por parte del volcán por lo que se debe evitar también su presencia al realizar la selección de las imágenes del estudio.

2.3. Algoritmos de clasificación

La clasificación es una tarea fundamental en el campo del aprendizaje automático. El principal propósito de los algoritmos de clasificación es, como bien dice su nombre, clasificar o categorizar datos de manera automática en distintos grupos. Este tipo de algoritmos suele ser usado cuando el conjunto de datos que tenemos que clasificar es de tipo discreto, como ocurre en nuestro caso.

El objetivo de la aplicación de algoritmos de clasificación en este estudio es delimitar grupos de píxeles para así poder detectar cambios en la superficie terrestre y realizar un estudio de la zona.

Se pueden identificar dos tipos de algoritmos de clasificación, los supervisados y los no supervisados. En nuestro caso se va a hacer uso del algoritmo de clasificación no supervisada DBSCAN, el cual se explicará con detalle en este capítulo.

2.3.1. Clasificación supervisada

La clasificación supervisada es una forma de agrupar datos en la que el algoritmo aprende a partir de un conjunto de datos de entrenamiento ya previamente etiquetados. Después de estudiar dicho conjunto, el algoritmo trata de agrupar datos según distintos patrones que ha sido capaz de establecer como consecuencia del entrenamiento inicial.

Dicho esto, se puede definir el proceso de clasificación supervisada en tres pasos. El primer paso es la selección de las áreas a entrenar, es decir, los ejemplos a partir de los cuales el algoritmo va a tratar de aprender. El segundo paso sería que el algoritmo se encargase de buscar un patrón en el conjunto de datos que pertenecen al área de entreno. Y finalmente, el último paso es clasificar el resto de los datos a partir del aprendizaje obtenido como resultado de la búsqueda de patrones en las áreas o conjuntos de datos seleccionados en el primer paso. Una vez realizada la clasificación, hay que estudiar la precisión con la que se ha producido la agrupación y la calidad del aprendizaje obtenido por parte del algoritmo y, por tanto, su grado de fiabilidad.

Tal y como se ha descrito con anterioridad, el algoritmo de clasificación supervisada debe hacer una detección de patrones en el conjunto de datos seleccionado. Existen varios algoritmos y procedimientos para analizar patrones en datos. Hay que tener en cuenta que, dependiendo de las características de los datos, el resultado de la aplicación de un algoritmo u otro puede variar de manera significativa. A continuación, se pueden encontrar los más utilizados.

- La **clasificación por paralelepípedos** se basa principalmente en la definición de distintas áreas delimitadas por paralelepípedos. Cada paralelepípedo define o representa a una clase distinta del conjunto de datos. Una vez definidas todas estas regiones los píxeles que se encuentren dentro de cada área de datos pertenecerá a una clase en concreto, en caso contrario los datos se quedarían fuera de cualquier clase. Este tipo de clasificación obtiene muy buenos resultados en

situaciones en las que el número de características de la imagen no sea muy extenso. Sin embargo, no es tan eficiente en situaciones en las que los datos tengan una elevada variabilidad.

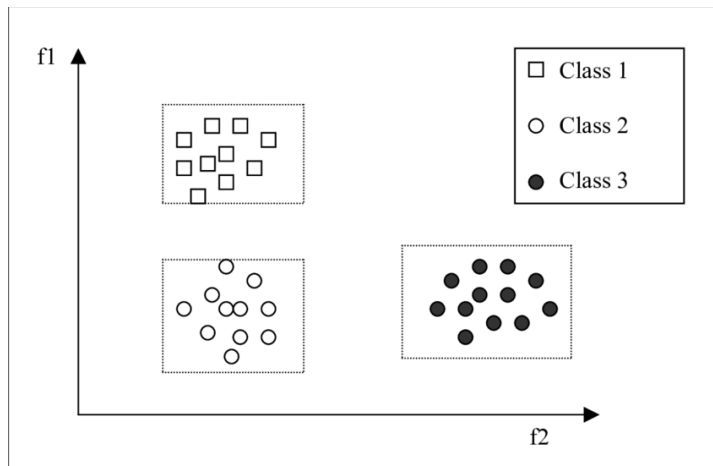


Ilustración 18: Clasificación por paralelepípedos [https://www.researchgate.net/figure/3-Parallelepiped-classification-strategy_fig3_37245668]

- La **clasificación por mínima distancia** se basa principalmente en la distancia euclidiana entre las muestras etiquetadas en el entrenamiento inicial y las muestras que todavía no se han etiquetado. Después del cálculo de esta distancia, el píxel se asignará a una clase dependiendo de la distancia mínima a la que se encuentre. La clasificación por mínima distancia es bastante adecuada para datos que presentan una baja dimensionalidad. Pero, al igual que la clasificación por paralelepípedos, no resulta muy efectivo para datos que presenten una gran variabilidad.

A partir de este tipo de clasificación han surgido algunas similares como la clasificación por mínima distancia ponderada, que aparte de utilizar simplemente la distancia mínima también hace uso de una ponderación sobre esa distancia.

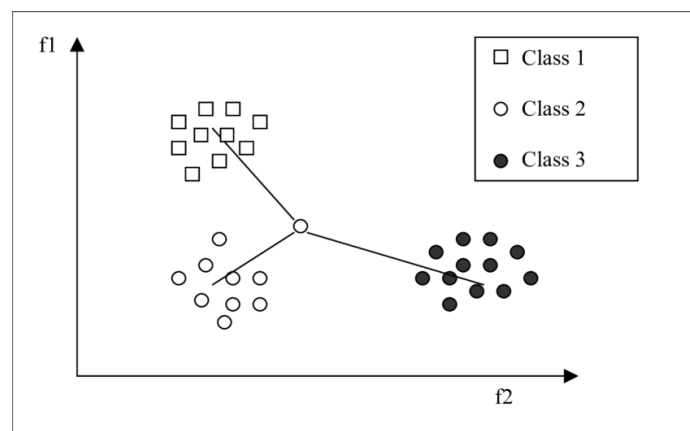


Ilustración 19: Clasificación por mínima distancia [https://www.researchgate.net/figure/4-Minimum-distance-to-mean-classification-strategy_fig4_37245668]

- La **clasificación por máxima verosimilitud** hace una asignación de cada píxel dependiendo de la probabilidad que tiene de pertenecer a una clase u otra. Este método de clasificación asume que la distribución de los datos es normal. Para realizar este cálculo de probabilidad hace un cálculo a partir de las varianzas y covarianzas de cada dato sin clasificar para posteriormente, dependiendo de los resultados, asignarlo a una clase determinada con la que encuentra mayor probabilidad de pertenecer. Este tipo de clasificación muestra muy buenos resultados sobre todo cuando el conjunto de datos a clasificar es multivariable. Por otro lado, tal y como se ha mencionado con anterioridad, esta clasificación asume que los datos siguen una distribución normal, lo cual no tiene por qué funcionar correctamente en todos los casos de análisis.

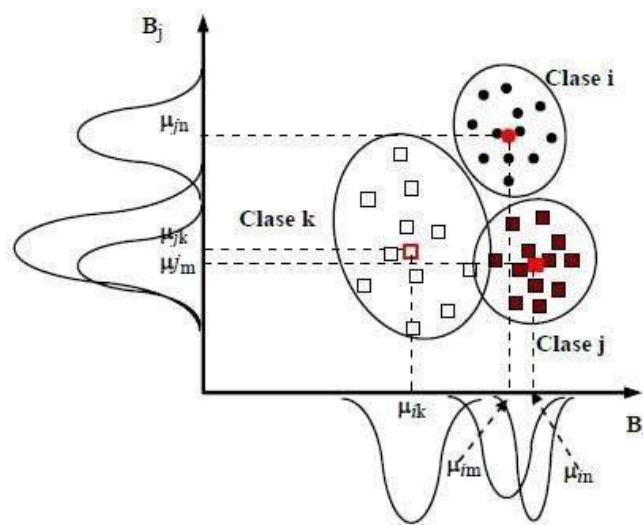


Ilustración 20: Clasificación por máxima verosimilitud
[\[https://ingeoexpert.com/articulo/clasificaciones-de-imagenes-de-satelite/\]](https://ingeoexpert.com/articulo/clasificaciones-de-imagenes-de-satelite/)

2.3.2. Clasificación no supervisada

La clasificación no supervisada es otro tipo de aprendizaje automático en el que, a diferencia de la clasificación supervisada, no etiqueta los datos, es decir, no es necesario hacer uso de un conjunto de estos para tener una referencia y poder así realizar la segmentación. Es una manera más sencilla de clasificar los distintos datos de una imagen o cualquier otro conjunto, aunque puede resultar algo más imprecisa.

Al no hacer uso de etiquetas o de un entrenamiento previo de un subconjunto de los datos a clasificar, los algoritmos de clasificación no supervisada clasifican en grupos de datos con características similares, es decir, buscan patrones o estructuras ocultas en los datos sin la ayuda de etiquetas. En algunos algoritmos utilizados, para clasificar una imagen o conjunto de datos deseado mediante clasificación no supervisada, hará falta indicar el número de clases, también conocidas como clústeres, que queremos establecer, como por ejemplo KMEANS.

En otros algoritmos, como DBSCAN, no hará falta indicar el número de clústeres ya que los creará en base a otros parámetros.

Existe una variedad muy grande de algoritmos de clasificación no supervisada. Los más utilizados son:

- KMEANS es un algoritmo de clasificación no supervisada cuyo objetivo es agrupar los datos (representados por puntos) que se presentan en un número fijo de clústeres, de modo que se minimiza la suma de las distancias entre los puntos y el centroide del clúster al que pertenecen. Inicialmente, el centroide se establece sobre uno de los puntos escogido de forma aleatoria. A partir de estos centroides, los datos son asignados al que se encuentre más cercano. A continuación, se vuelve a recalcular la posición de los centroides de modo que quede en el centro de cada clúster y se realiza el mismo proceso hasta que no se produzcan más cambios en la reasignación de puntos.

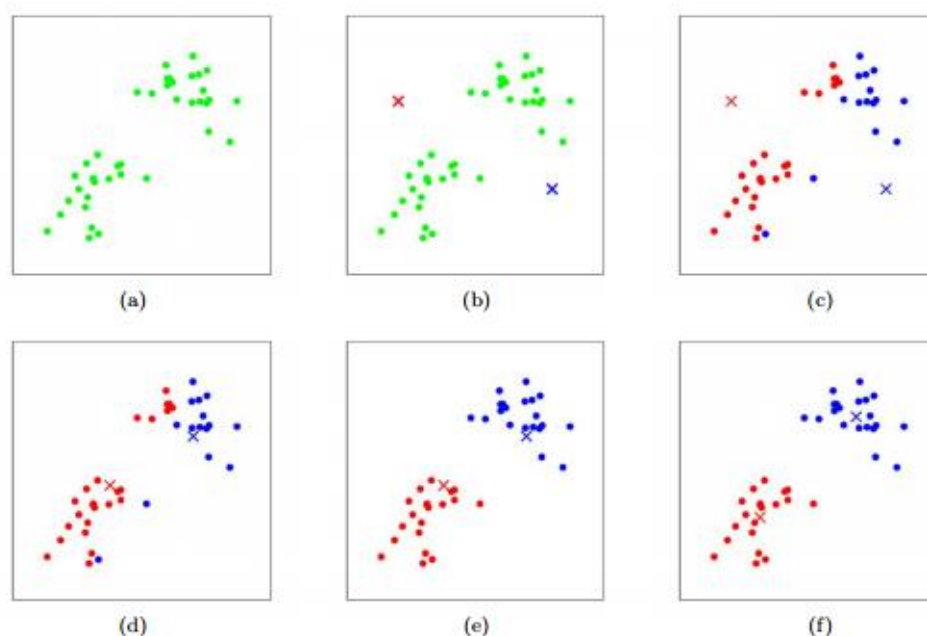


Ilustración 21: Clasificación por KMEANS [https://www.researchgate.net/figure/Example-of-K-means-algorithmircles-Adaptedfrom-website-Stanfordedu-by-Chris-Piech_fig2_313879313]

- DBSCAN (*Density Based Spatial Clustering of Applications*) es otro algoritmo de clasificación no supervisada que se va a utilizar en este trabajo. A este algoritmo, a diferencia de KMEANS, no se le pasa como parámetro el número de clústeres que se quiere tener como resultado de la clasificación ya que el algoritmo, a partir de un valor épsilon (*eps*) y un parámetro que se llama *min samples*, calculará los clústeres necesarios. Todo esto se comentará con detalle en la siguiente sección de la memoria.

La clasificación no supervisada es una herramienta de agrupamiento de datos muy valiosa, sobre todo, para conseguir explorar y entender los distintos patrones ocultos que pueden tener los conjuntos de datos.

2.3.3. DBSCAN

DBSCAN o Density Based Spatial Clustering of Applications with Noise es un algoritmo de clasificación no supervisada bastante utilizado y popular en el área del aprendizaje no supervisado. Fue creado en el año 1996 por un grupo de científicos en computación llamados Martin Ester, Hans-Peter Kriegel, Jörg Sander y Xiaowei Xu.

La idea clave que hay que tener en cuenta al hablar de este algoritmo es que un punto pertenece a un clúster determinado si está cerca de muchos puntos de ese mismo clúster. Es un algoritmo de *clustering* basado en densidad.

DBSCAN hace uso de dos parámetros para poder identificar los clústeres:

- **Épsilon (*eps*):** Este valor indica lo cerca que deben estar dos puntos para ser considerados vecinos y, por lo tanto, miembros de un mismo clúster. Dicho en otras palabras, dos puntos se pueden considerar vecinos si la distancia que hay entre ellos es menor o igual al valor de épsilon. Para obtener el valor de épsilon más adecuado para nuestros datos, se puede hacer uso del algoritmo *K-Nearest Neighbors* (KNN). A partir del resultado de este algoritmo, se puede determinar el valor de épsilon, aunque es bastante complicado encontrar el óptimo. Este método también es conocido como el método del codo, debido a la forma que aparenta la gráfica resultante.
- **Puntos mínimos (*min samples*):** Se refiere al número mínimo de puntos de datos necesarios para poder definir un clúster, es decir, necesitamos como poco un número de puntos igual al de puntos mínimos para formar un clúster. Hay que tener en cuenta que a la hora de asignar un valor al parámetro se tiene que seguir la siguiente norma:

$$\text{min samples} \geq \text{Dimension} + 1$$

Lo cual quiere decir que el número de puntos mínimos debe ser mayor que la dimensión del conjunto de datos (número de características con las que se están trabajando) más 1. Esto se hace para que el algoritmo pueda detectar al menos más de un clúster distinto dentro del conjunto dado.

Una vez está en funcionamiento el algoritmo, los puntos identificados se pueden clasificar en tres tipos:

- **Puntos centrales:** son aquellos que satisfacen las condiciones de los parámetros especificados, es decir, están dentro del radio definido por *eps* y contienen más de *min samples* a su alrededor.

- Puntos de borde o alcanzables: son aquellos que, aunque no satisfacen las condiciones de los parámetros especificados, tienen algún punto central cerca, es decir, dentro del rango ϵ .
- Puntos de ruido o atípicos: es cualquier punto que no es un punto central ni de borde, es decir, que no cumple con ninguna de las condiciones de los parámetros.

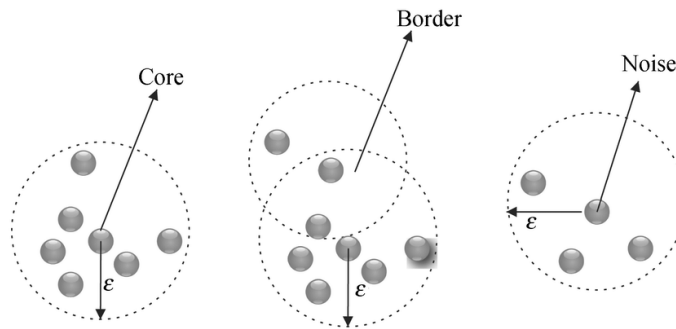


Ilustración 22: Puntos clasificación DBSCAN [https://www.researchgate.net/figure/DBSCAN-core-border-and-noise-points_fig1_258442676]

Como bien se ha mencionado con anterioridad, para realizar la clasificación mediante DBSCAN, necesitamos dos parámetros.

El primer paso es dar un valor tanto a la variable ϵ como al número mínimo de datos (*min samples*). Esta es una tarea muy importante para que el algoritmo consiga realizar una correcta clasificación a partir de nuestro conjunto de datos. Una mala elección de los parámetros da paso, a la creación de menos clústeres de los necesarios, a la creación de muchos más clústeres de los deseados. Los valores de estos parámetros variarán dependiendo del conjunto de datos presente.

El siguiente paso es la elección aleatoria, por parte del algoritmo, de un punto de partida. Este punto se considera el primer núcleo.

A continuación, a partir de ese punto que se ha elegido de forma aleatoria, se determinan cuáles son los puntos que están a una distancia menor a la determinada por el parámetro ϵ . Una vez determinados estos puntos, se comprueba que dentro de esa área haya como mínimo el valor del parámetro *min samples*. Si satisfacen ambas condiciones, ese punto se considerará como un punto central y se creará un nuevo clúster.

En el caso en el que no cumpla alguna de las condiciones, hay varias opciones que pueden ocurrir:

- Si el punto contiene menos valores que el indicado por el parámetro *min samples* pero, dentro de ese radio se encuentra algún punto central, decimos que ese punto va a ser un punto de borde.
- En el caso en el que el punto no cumpla ninguna de las condiciones establecidas por los parámetros y, además, dentro del área establecida por ϵ no se encuentre ningún punto central, ese punto se consideraría como un punto de ruido.

Estos pasos se realizan con todos los puntos del conjunto de datos hasta que todos estén clasificados. Una vez que todos los puntos se encuentren etiquetados, el siguiente paso es conectar los puntos centrales que son vecinos, y colocarlos en el mismo clúster. Es decir, un clúster se forma con al menos un punto central, además de todos los puntos centrales accesibles, además de todos sus puntos fronterizos.

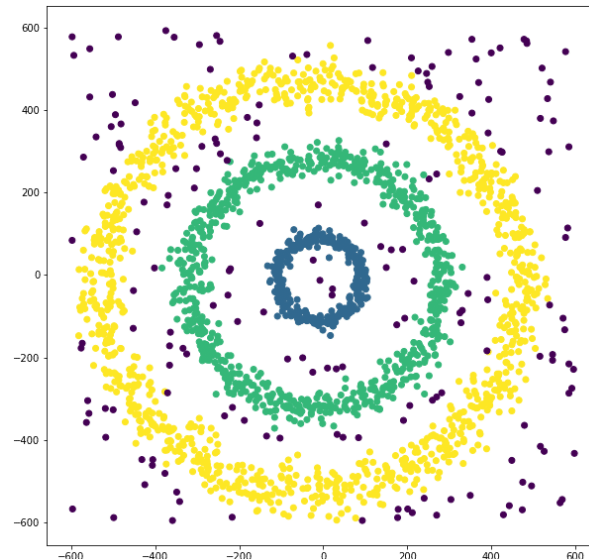


Ilustración 23: Resultado clasificación DBSCAN

2.3.3.1. Ventajas y desventajas de DBSCAN

Tal y como se ha descrito con anterioridad, el funcionamiento de este algoritmo es bastante sencillo y muy útil a la hora de clasificar conjuntos de datos. A pesar de esto, DBSCAN, como todos los algoritmos, tiene una serie de ventajas y desventajas.

Ventajas:

- Es capaz de detectar puntos aislados o atípicos, es decir, puntos que no pertenecen a ningún clúster. Es por esto por lo que tiene muy buenos resultados en conjuntos de datos con bastante ruido.
- No hace falta definir el número de clústeres que se quiere que se creen. Esto puede ser tanto una ventaja como una desventaja, dependiendo de qué tipo de conjuntos de datos queramos clasificar.
- Es un algoritmo fácil de entender, tanto el proceso en sí mismo como sus parámetros de entrada.
- En los casos en los que el conjunto de datos no está basado en una matriz, no tiene una complejidad muy elevada siendo de $O(n)$.

Desventajas:

- El resultado es muy variable dependiendo de los valores dados a ϵ y $min\ samples$. Por esta razón, es importante que el

responsable de ejecutar el algoritmo esté familiarizado con el conjunto de datos a clasificar.

- No es recomendable para un conjunto de datos con una elevada dimensionalidad. Cuanto mayor es la dimensionalidad del conjunto de datos, la complejidad del algoritmo aumenta, tal y como ocurre en todos los algoritmos.
- A parte de la dimensionalidad, si el conjunto de datos está basado en una matriz, la complejidad puede ser de $O(n^2)$.
- La clasificación puede variar ligeramente entre una ejecución y otra.

3. Herramientas Utilizadas

Para la realización de todas las tareas necesarias para el progreso del trabajo, principalmente, se va a hacer uso de tres herramientas: *Copernicus Open Access Hub*, *Sentinel Hub EO Browser* y *Erdas Imagine*. Estas herramientas se utilizarán para obtener y analizar las imágenes satelitales sobre el volcán Etna. También se utilizarán otras herramientas para implementar el algoritmo DBSCAN, como *Visual Studio Code*, *QGIS* y, de nuevo, *Erdas Imagine*.

3.1. Copernicus Open Access Hub

Copernicus Open Access Hub es una plataforma creada por la Agencia Espacial Europea (ESA) con el objetivo de ofrecer información completa, gratuita y de acceso libre sobre las misiones Sentinel-1, Sentinel-2, Sentinel-3 y Sentinel-5P.

Esta plataforma es sobre todo utilizada para la observación y estudio de distintos entornos de la superficie terrestre, por lo que es una aplicación muy potente y robusta, pero, con el paso de los años se ha convertido en una aplicación desactualizada en cuanto a funcionalidad e interacción con el usuario debido a las limitaciones que tiene a la hora de la obtención y búsqueda de imágenes. Por eso, la ESA, desde el 2023, ha decidido realizar una migración de toda la plataforma a otra nueva llamada *Sentinel Hub EO Browser*.

El proyecto se inició utilizando *Copernicus Open Access Hub* para la obtención y descarga de las imágenes idóneas para la realización del análisis, pero debido a su dificultad, se valoró y decidió utilizar la nueva plataforma desarrollada por la ESA llamada *Sentinel Hub*.

Al iniciar *Copernicus Open Access Hub* en cualquier navegador nos encontramos con una interfaz que presenta un mapa completo de la Tierra y una barra de búsqueda en la parte superior izquierda tal y como se muestra a continuación. A partir de este mapa se puede buscar o seleccionar cualquier zona que queramos estudiar.



Ilustración 24: Pantalla inicio Copernicus Open Access Hub

Si analizamos con más detenimiento la barra de búsqueda, se pueden encontrar 4 iconos que se explicarán a continuación.



Ilustración 25: Barra búsqueda Copernicus Open Access Hub

El primero de ellos despliega un menú, en el que se puede indicar varios filtros sobre las imágenes que queremos obtener. Entre estos filtros, están las fechas entre las cuales se quieren obtener imágenes o la misión y plataforma de las que se quieren obtener.

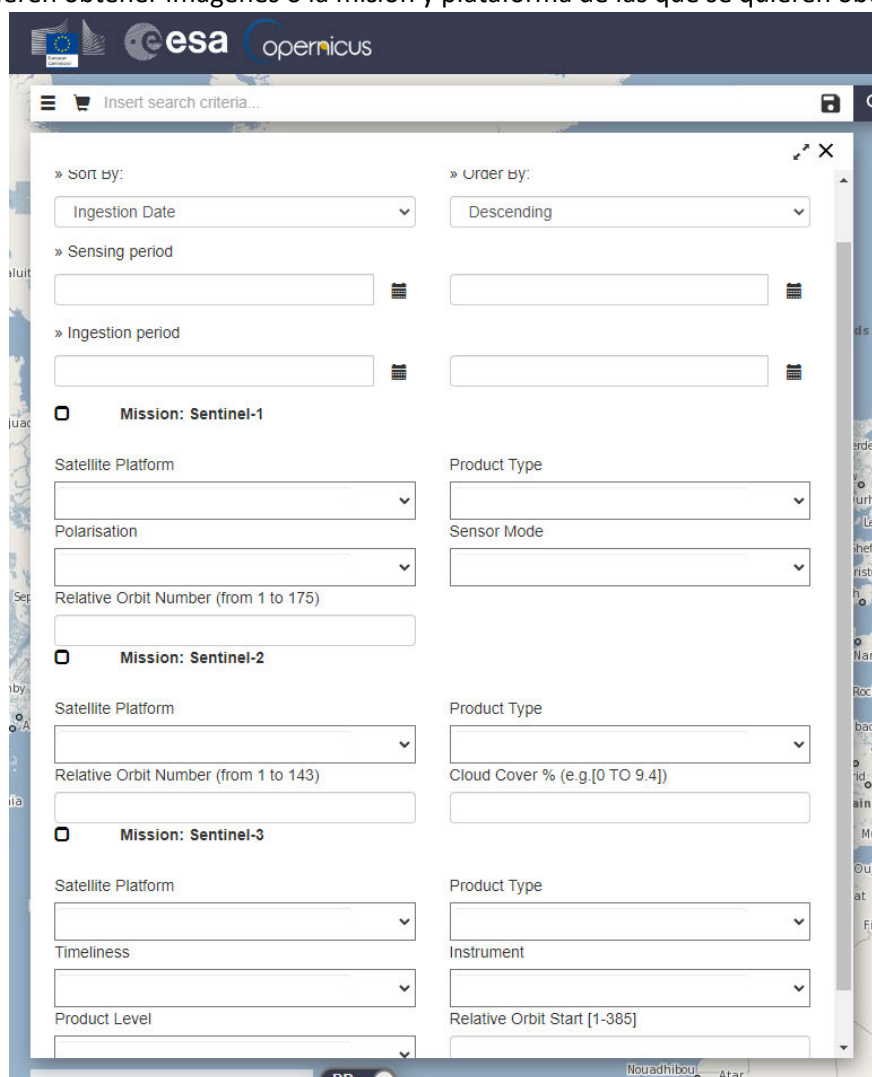


Ilustración 26: Selector filtros Copernicus Open Access Hub

El siguiente icono que aparece en la barra de búsqueda es un carrito, si se pulsa aparecerán las imágenes solicitadas que no estaban en ese momento en la base de datos de la plataforma para poder descargarlas de manera gratuita.

También hay un icono de un disco el cual guarda una búsqueda realizada.

Finalmente encontramos el símbolo de la lupa utilizado para ejecutar la búsqueda con todas las preferencias indicadas en el menú.

Una vez encontrada la imagen ideal para el estudio, se procede a descargarla pulsando en el icono de descarga que aparece en la esquina inferior derecha tal y como se muestra a continuación.

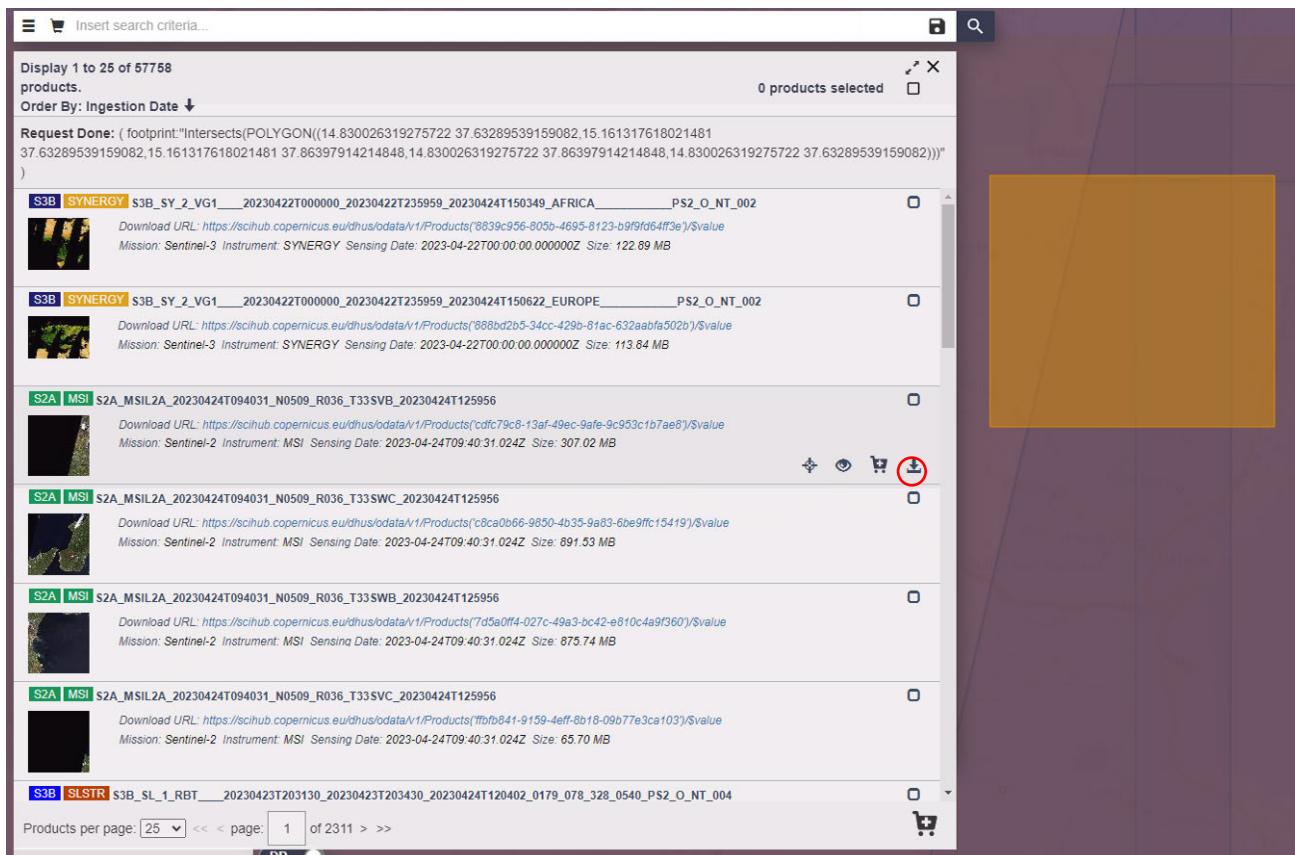


Ilustración 27: Descarga imagen Copernicus Open Access Hub

3.2. Sentinel Hub EO Browser

Sentinel Hub es una plataforma web que, al igual que Copernicus Open Access Hub, permite la búsqueda y posterior descarga de imágenes satelitales de diversas misiones como, por ejemplo, Sentinel-2, Sentinel-3, Landsat 8-9, Landsat 7 o MODIS, entre muchos otros. Esta herramienta, a diferencia de la explicada anteriormente, es mucho más útil a la hora de realizar la búsqueda y descarga de las imágenes gracias a su interfaz de usuario la cual la hace más visual y fácil de utilizar.

Nada más abrir la aplicación se presenta una interfaz muy similar a la anterior herramienta, con un mapa en la parte central, una barra de búsqueda y un menú lateral en el que seleccionar las características de la búsqueda.

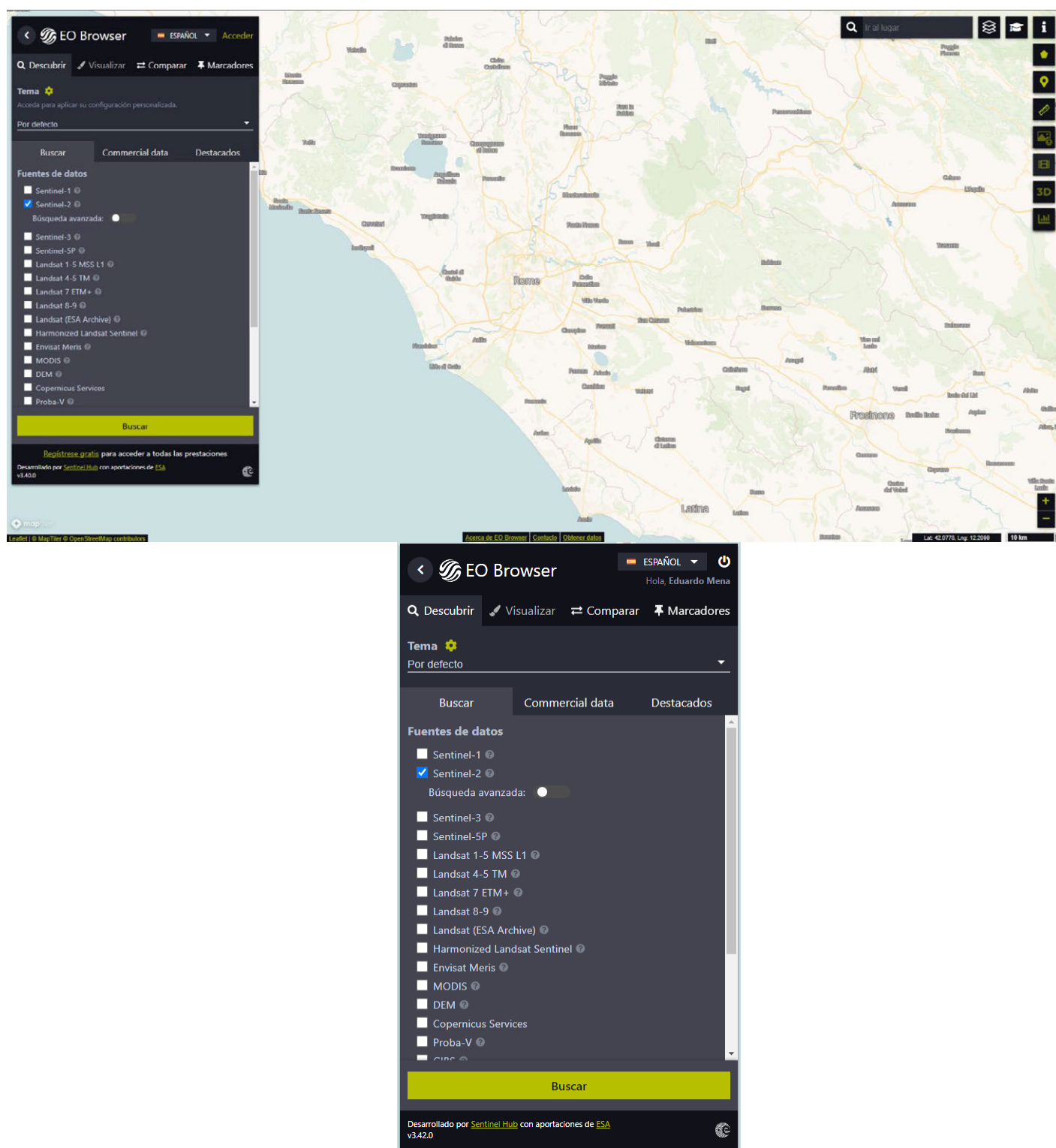


Ilustración 28: Pantalla inicio Sentinel Hub

A la hora de buscar imágenes de un área en específico tenemos la opción de delimitar un área de búsqueda a través de la importación de un fichero KML, GPX, WKT o JSON o crear una nueva directamente en la propia herramienta.

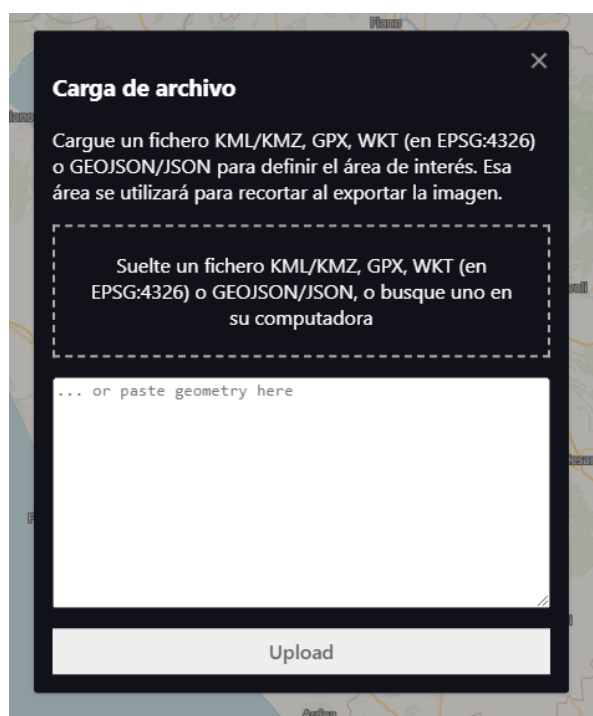


Ilustración 29: Carga de archivo de área

Una vez seleccionada el área, se pueden escoger y filtrar los resultados de las imágenes que vamos a obtener. A parte de poder elegir la plataforma de la que se quieren recibir las imágenes, se puede especificar la cantidad de nubosidad que se quiere que presente la imagen o la fecha de esta. Al seleccionar una imagen, la plataforma da la opción de visualizar directamente la imagen sin descargar con una combinación distinta de índices. Algunas composiciones de bandas que están incluidas en la herramienta son la NDVI, SWIR, NDWI o NDSI. Además, se puede crear cualquier índice a partir de las bandas de la imagen, incluso crear un *script*.

Finalmente, a la hora de la descarga, la aplicación permite escoger el formato de la imagen (TIFF, JPG, KMZ...), la resolución en metros/píxel, el sistema de coordenadas y las capas que se desean descargar. Estas capas pueden ser las bandas directamente crudas o con los índices preestablecidos por el propio *Sentinel Hub*.

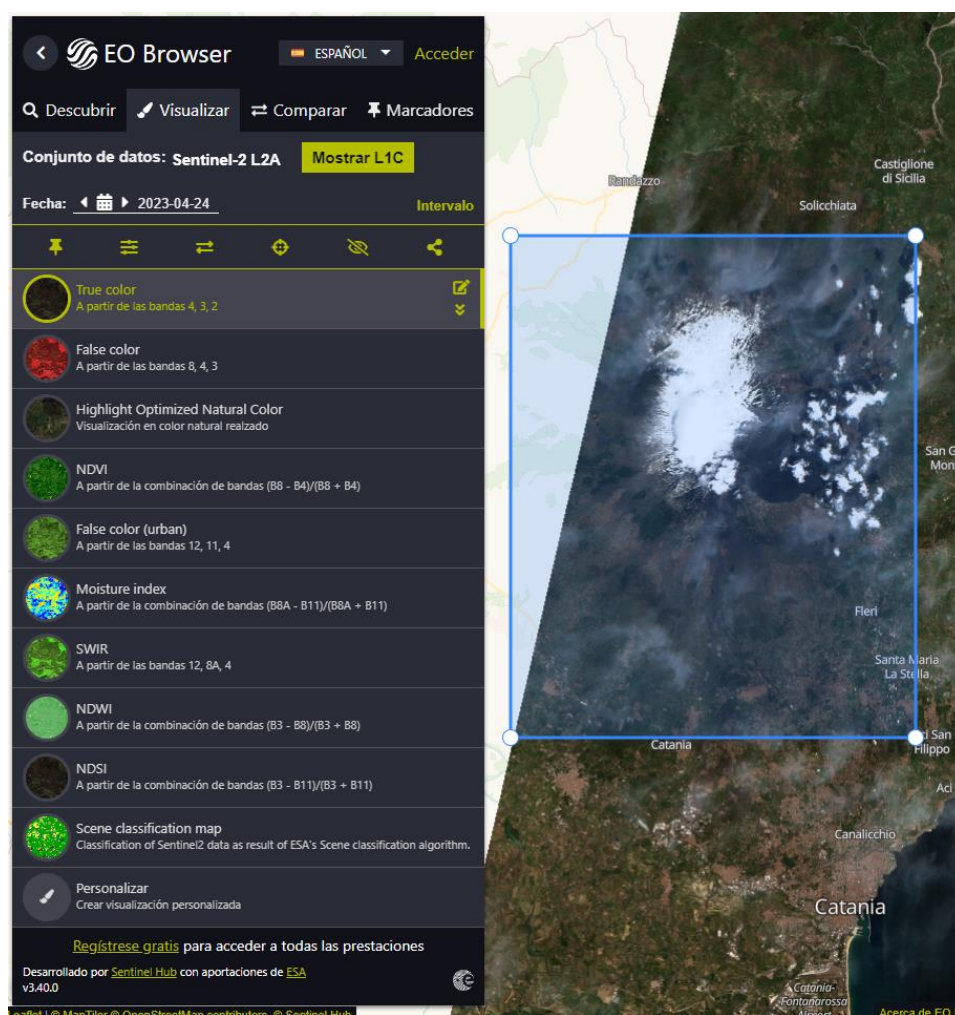


Ilustración 30: Selección de índices en Sentinel Hub



Ilustración 31: Personalización de índices en Sentinel Hub

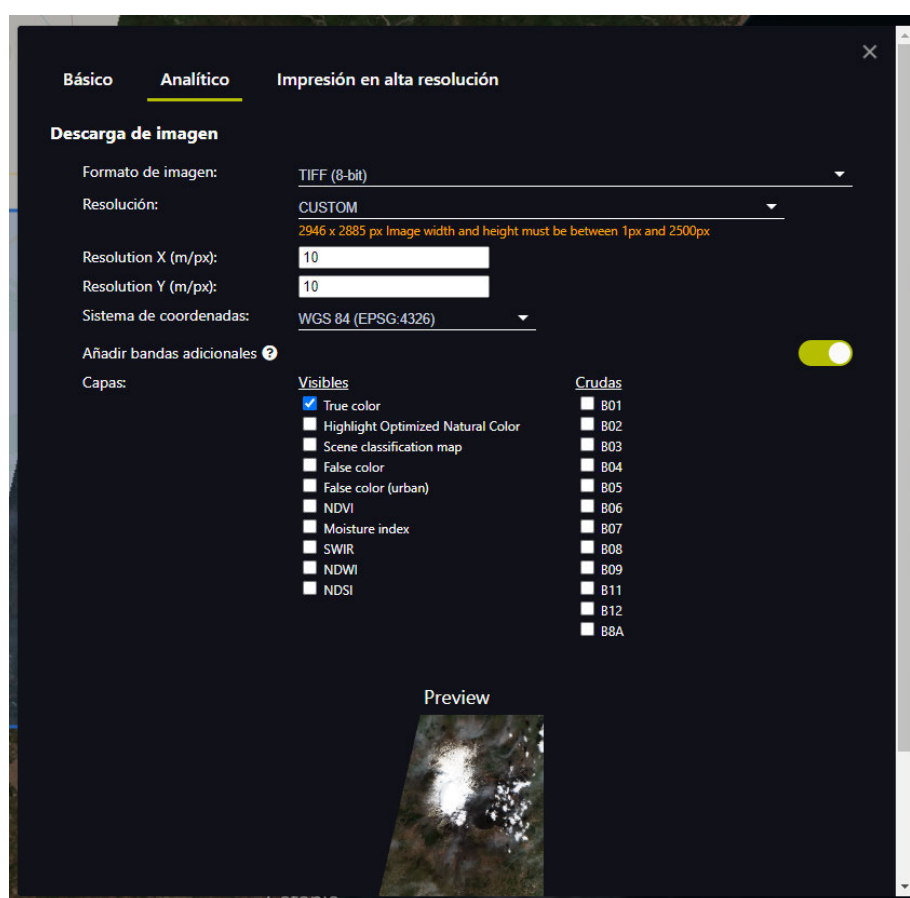


Ilustración 32: Selección formato de descarga en Sentinel Hub

Teniendo en cuenta las funcionalidades básicas de la herramienta y sus ventajas a la hora de la descarga de imágenes satelitales, se ha decidido utilizar esta herramienta para escoger y descargar las imágenes que se van a estudiar en este trabajo.

3.3. Erdas Imagine

Erdas Imagine es un software desarrollado por *Hexagon Geospatial* dedicado exclusivamente al procesamiento de imágenes y análisis geoespacial. Esta herramienta se utiliza sobre todo en el campo de la teledetección, cartografía, agricultura y geología.

Erdas permite procesar, analizar y visualizar un gran abanico de datos a partir de imágenes satelitales. Se trata de una aplicación muy extensa y con una inmensa variedad de funcionalidades, aunque está especialmente pensada y optimizada para su uso en el ámbito de la teledetección.

Nada más abrir la aplicación, se presenta una interfaz con un visualizador de imágenes en el centro, una barra superior con una gran variedad de opciones y otra barra lateral la cual muestra un esquema de las capas con las que se están trabajando.

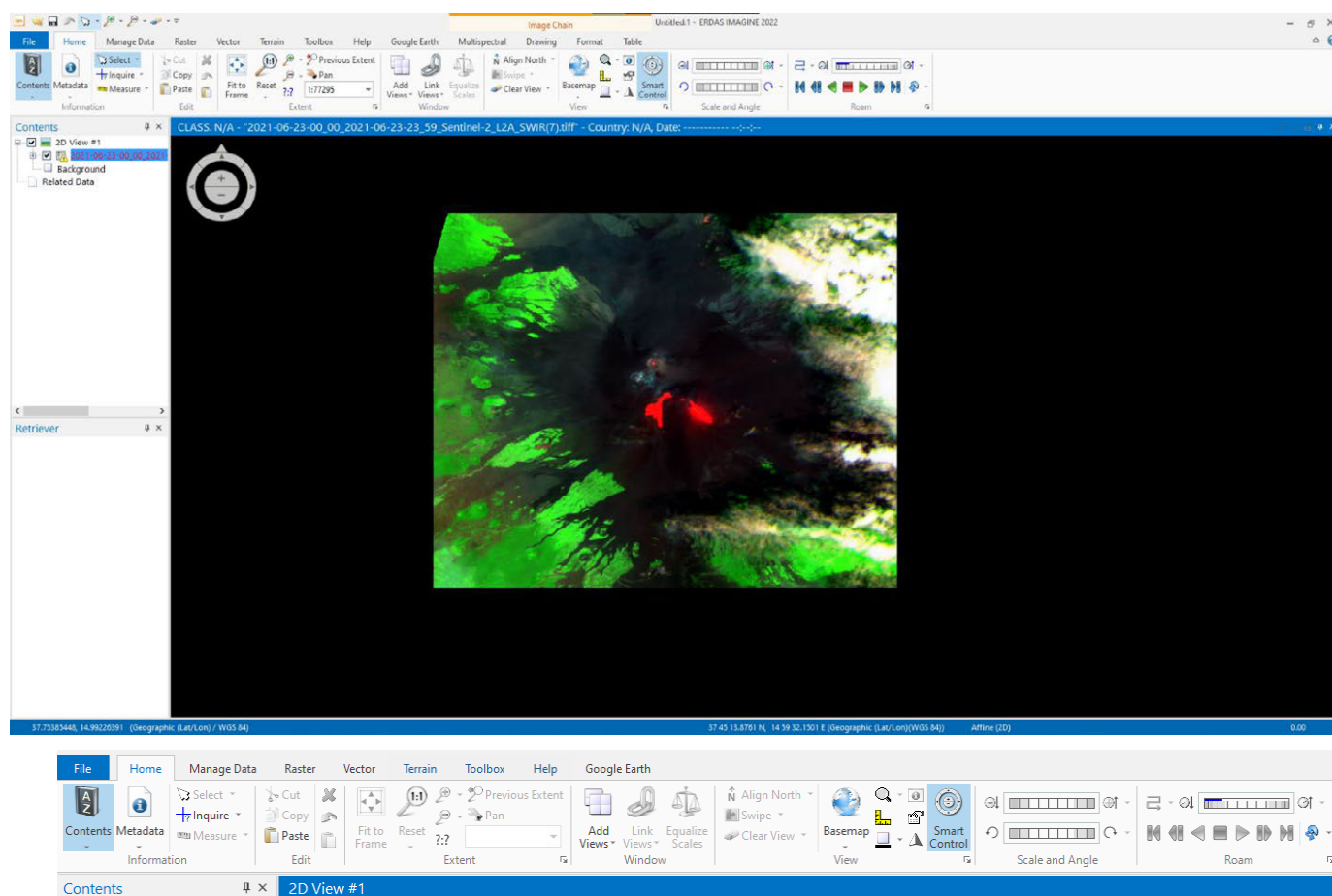


Ilustración 33: Pantalla inicio Erdas Imagine

Una vez cargada cualquier capa, algunas de las principales funcionalidades que se pueden realizar son la visualización de los metadatos y de los distintos atributos que

presente la imagen, además esta herramienta también permite la clasificación de la imagen tanto con algunos algoritmos de clasificación supervisada como no supervisada.

Además de poder realizar casi cualquier ajuste tanto de edición como de visualización sobre la imagen, también se pueden definir modelos para realizar operaciones sobre las distintas capas y así poder obtener un resultado según lo requerido. Para realizar estos modelos, en la pestaña de *Toolbox*, hay una opción llamada *Spatial Model Editor*.

Al seleccionar esta opción se nos muestra una barra lateral con distintos módulos ya predefinidos y un tablón sobre el que colocarlos. Estos módulos deben estar interconectados unos a otros para definir así el flujo del modelo.

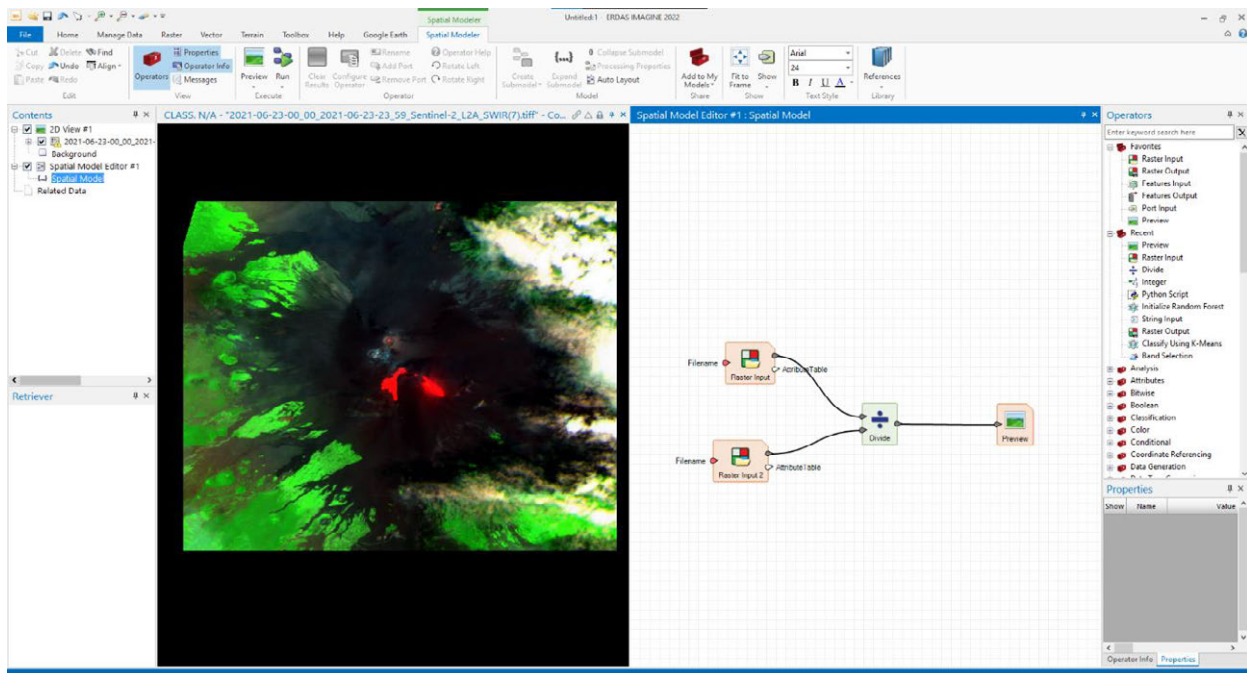


Ilustración 34: Spatial Modeler en Erdas Imagine

Una característica muy interesante de esta opción es la posibilidad de insertar un *script* de *Python* para describir el comportamiento del modelo sin necesidad de usar el tablón y arrastrar los módulos.



Ilustración 35: Módulo Python Script Erdas Imagine

En un principio, se iba a hacer uso de este módulo de *Python Script* para insertar el algoritmo DBSCAN en la propia herramienta de *Erdas Imagine* y poder así hacer uso de este algoritmo el cual no está implementado en la herramienta para clasificar imágenes satelitales. Tras una intensa investigación debida a la escasa documentación acerca de cómo funciona este módulo e innumerables intentos de insertar el algoritmo en la

herramienta, se llegó a la conclusión de la imposibilidad de la creación de nuevos módulos para *Erdas Imagine*, es decir, los *scripts* de *Python* que se pueden insertar en la herramienta son aquellos que contienen un flujo de trabajo o modelo con los módulos creados por el propio *Erdas*, por lo que no hay posibilidad de crear nuevos módulos tal y como era el propósito de este proyecto.

Debido a esta imposibilidad se exploraron otras alternativas para poder insertar el algoritmo. Después de valorar algunas opciones se decidió implementar el algoritmo en QGIS, un *software* de código abierto el cual se explica a continuación.

3.4. QGIS

QGIS es un Sistema de Información Geográfica de código abierto disponible para la gran mayoría de plataformas, las más conocidas son Windows, Linux o MacOS. Esta aplicación fue desarrollada por una organización no gubernamental cuyo objetivo es el desarrollo de programas geoespaciales de código abierto llamada OSGeo (*Open Source Geospatial Foundation*).

Es una aplicación, que, a la hora de ponerla en uso, es muy similar a *Erdas Imagine*, pero con la peculiaridad de que es de código abierto y por lo tanto permite la incorporación de módulos de código propio para utilizarlo dentro de la aplicación y así poder aplicarlo para tratar imágenes satelitales.

Esta herramienta permite la edición, visualización y análisis de imágenes satelitales a través de distintos *plugin* ya sean nativos de la aplicación o desarrollados por uno mismo. Al usarse para tratar imágenes tomadas por satélites de la superficie terrestre, los campos en los que más se usa son todos aquellos del ámbito de la teledetección como puede ser la agricultura o seguimiento de evolución de cualquier catástrofe natural. En el caso de este trabajo se utilizará QGIS 3.30.

Nada más abrir la aplicación, se presenta una interfaz con una parte central en la que aparecen las capas cargadas y un menú lateral situado a la izquierda el cual muestra distinta información sobre las capas que se están estudiando. A la derecha se muestra la caja de herramientas de procesos en la que se pueden encontrar los distintos *plugin*

que existen en la herramienta, aparte, se pueden conseguir otros *plugin* desde el centro de descargas o cargar módulos propios creados en un entorno local.

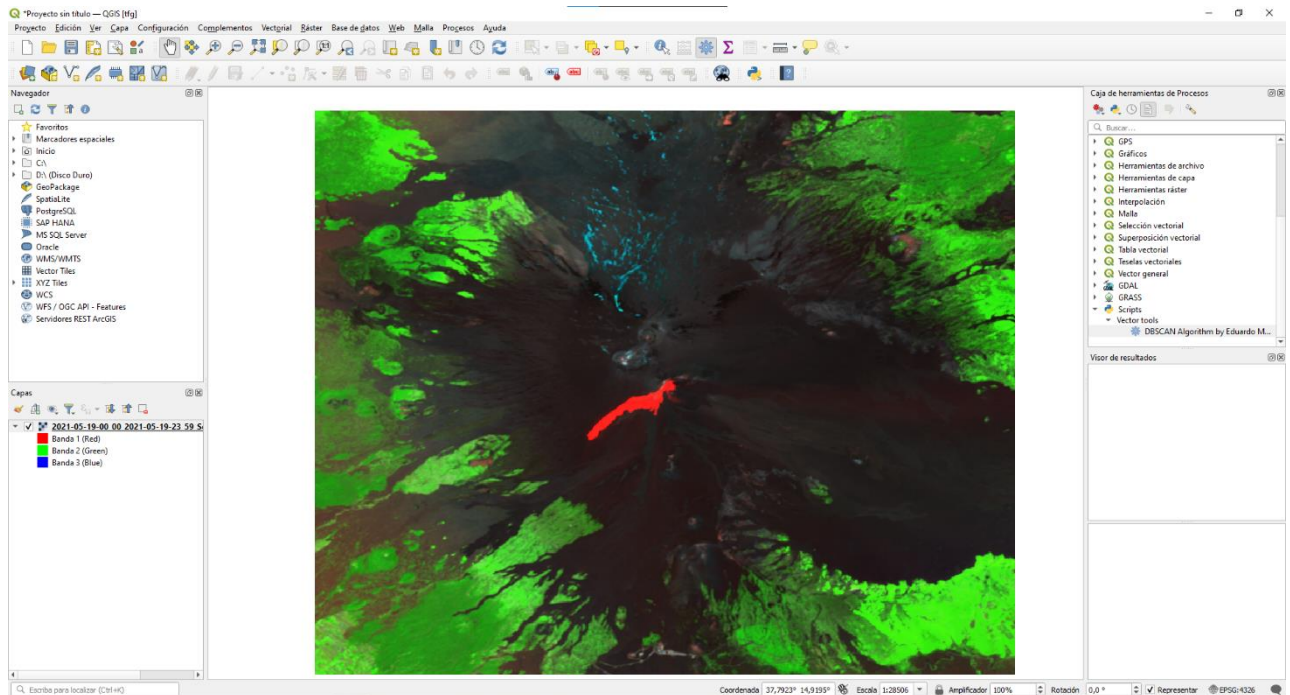


Ilustración 36: Pantalla inicio QGIS

Esta aplicación, al igual que Erdas Imagine, permite clasificar imágenes o detectar cambios entre un par de imágenes satelitales, entre muchas otras funcionalidades.

Algunos de los propósitos con los que se va a usar QGIS es, tal y como se ha comentado con anterioridad, la incorporación del algoritmo desarrollado DBSCAN dentro de la herramienta.

4. Metodología

A lo largo de este trabajo, la metodología empleada para su desarrollo está formada por una serie de etapas las cuales abarcan desde la obtención de imágenes de la zona de estudio, el desarrollo del algoritmo de clasificación DBSCAN y su posterior implantación en QGIS, la clasificación de las imágenes a estudiar y finalmente la evaluación de los resultados de la clasificación de estas imágenes para la detección de cambios.

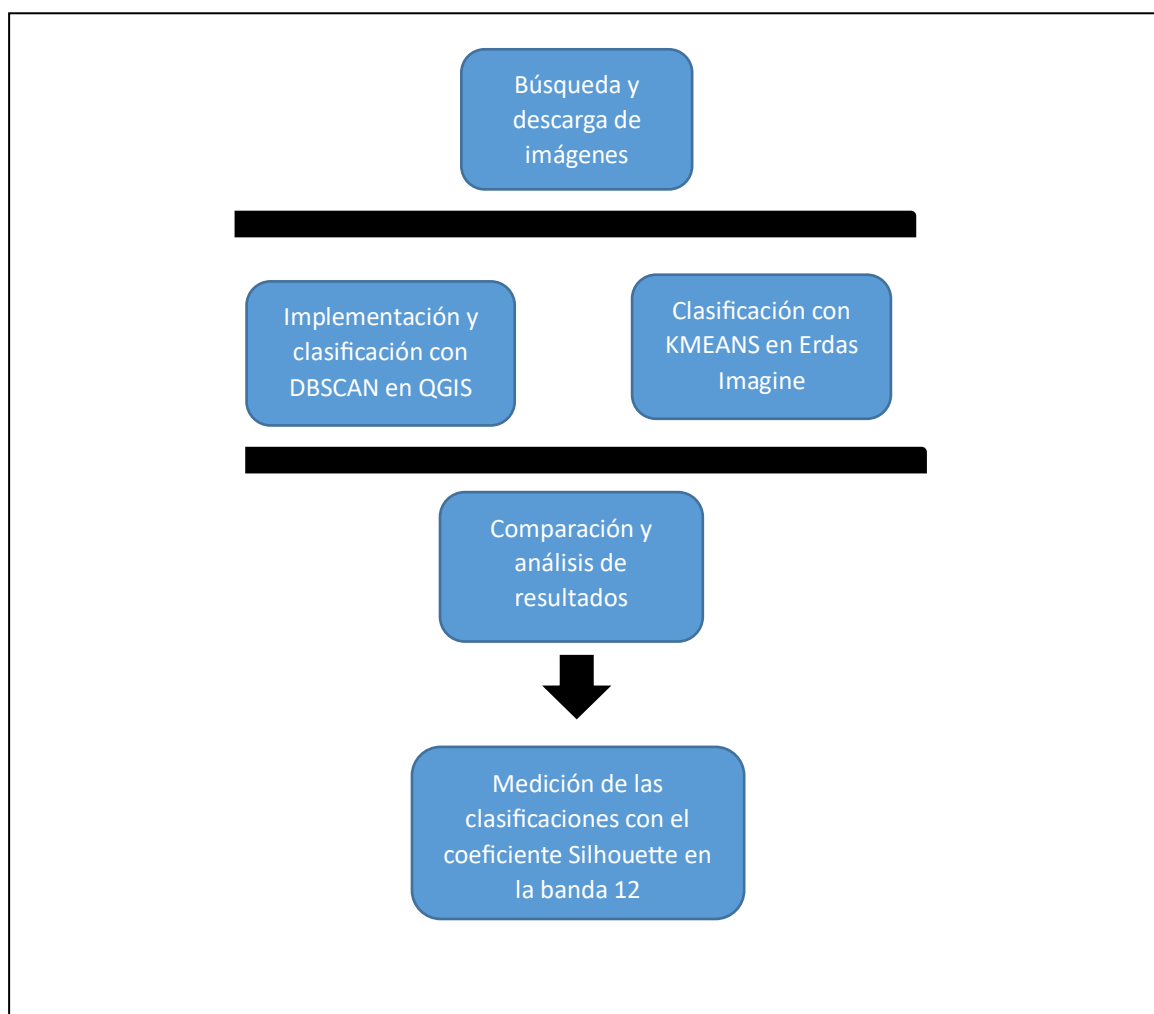
4.1. Fuente de datos

En el caso del ámbito de estudio de este trabajo, la obtención de los datos se centrará en la selección de imágenes satelitales de un terreno volcánico, para ser más concreto, se obtendrán imágenes satelitales del volcán Etna para estudiar las consecuencias y los cambios que sufre el terreno debido a sus muy frecuentes y explosivas erupciones.

La razón por la que se ha elegido estudiar el volcán Etna es por la gran explosividad que genera cada erupción volcánica y la frecuencia de estas, ya que, este volcán, es considerado el más activo de la placa Euroasiática y, por tanto, es más interesante hacer un estudio de los cambios que se producen en la superficie de este volcán antes que hacerlo que otro el cual entra en erupción una vez cada 50 años debido a la variedad de imágenes de las que se van a disponer.

Para realizar esta tarea se empleará la herramienta web llamada *Sentinel Hub EO Browser* para obtener imágenes procedentes de la misión Sentinel-2. Las instantáneas que interesan para realizar este estudio son aquellas que se hayan capturado entre los años 2017 y 2023. Un dato de relevancia para escoger las imágenes es que se debe visualizar con claridad la colada de lava causada por la erupción y, en la medida de lo posible, se debe intentar que en el área de la instantánea no haya ni masas de nieve ni nubosidad ya sea causada por la propia erupción o por otros factores. Estos requisitos a la hora de obtener los datos se imponen debido a que, para un buen resultado de la ejecución del algoritmo, se necesita que la imagen sea lo más clara posible. Ahora bien, escoger una imagen con todas estas características va a ser una tarea bastante complicada debido a que la cima del volcán se sitúa a más de 3000 metros por lo que la presencia de nieve es bastante común en los meses de otoño, invierno y primavera. Además, al estar situado en una isla en la que hay bastante viento, hay una gran presencia de nubes que impiden ver con claridad la imagen. Para finalizar, puesto que se va a estudiar un volcán en erupción, es muy fácil observar masas de cenizas y nubes de gases en las imágenes.

4.2. Organigrama



Esquema 1: Organigrama del trabajo

4.3. Metodología

4.3.1. Búsqueda y descarga de imágenes

La obtención de datos es la primera de todas las etapas que comprenden este proyecto. Tal y como se ha comentado con anterioridad, las imágenes que se buscarán estarán comprendidas entre el año 2017 y 2023.

La plataforma con la que se obtendrán estas imágenes es *Sentinel Hub*. Esta herramienta va a permitir buscar las imágenes del área deseada, en las fechas deseadas y, además, permitirá descargar tanto cada banda de manera individual como cualquier combinación de estas. En el caso de este trabajo se descargarán las imágenes de tal manera que se combinen las bandas 12, 8A y 4 para conseguir destacar las zonas que interesan para el estudio.

Para tener un contexto general de la apariencia de la interfaz de la aplicación se puede fijar en la siguiente figura.

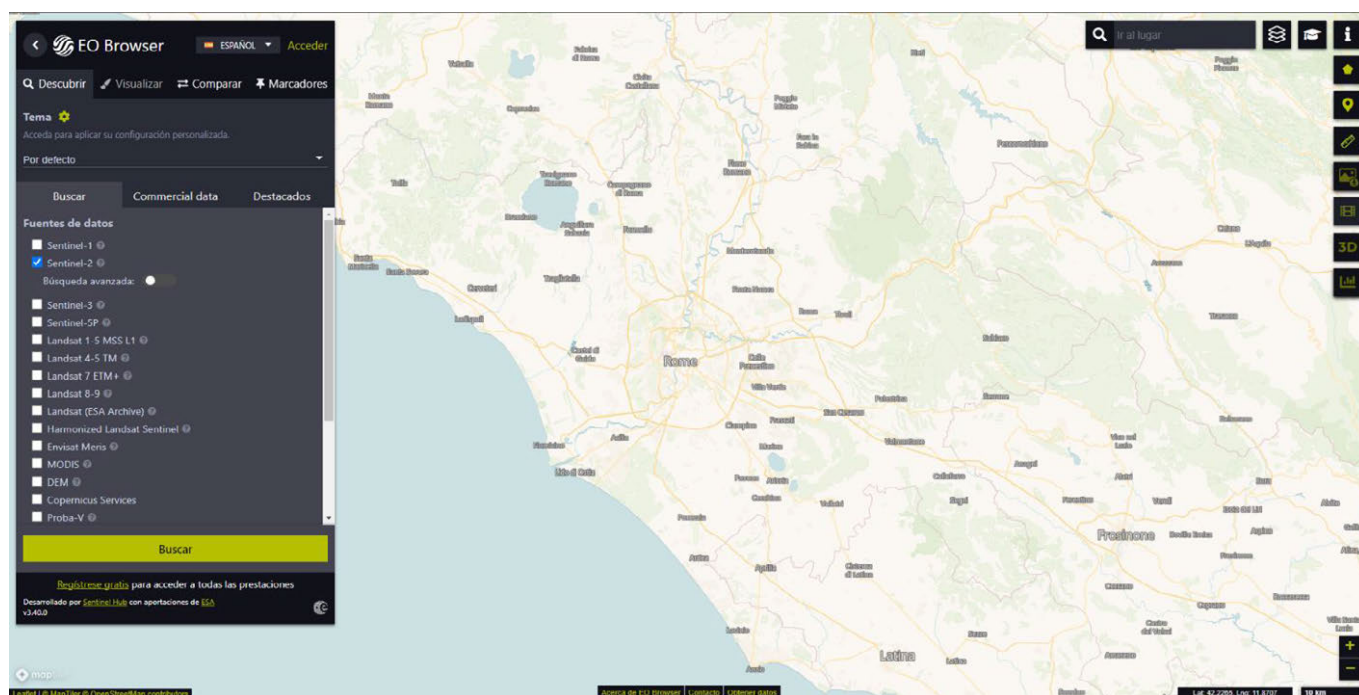


Ilustración 37: Pantalla inicio Sentinel Hub

Para empezar, se deberá seleccionar el área en la que se quiere realizar el estudio, es muy importante tener en cuenta que todas las imágenes a descargar deberán pertenecer exactamente a la misma área. En este caso el área será un cuadrado cuyos vértices se encuentran en estas coordenadas [latitud, longitud] (Sistema Geodésico Mundial 1984 (WGS 84))

[14.938316,37.695503],[15.061226,37.695503],
[15.061226,37.794593],[14.938316,37.794593]

Para establecer el área hay que escoger la opción en el menú lateral situado a la derecha de la interfaz y dibujarla.

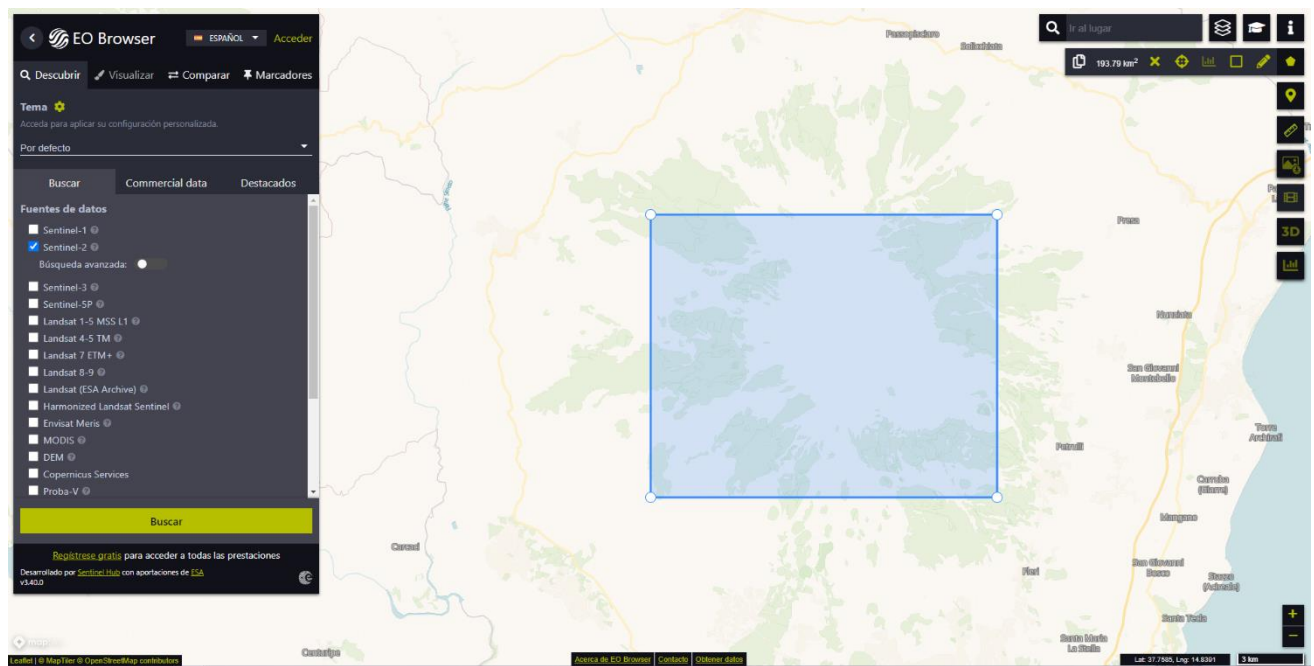


Ilustración 38: Selección área Sentinel Hub

Una vez seleccionada el área, se debe escoger la fuente de datos a partir la cual se va a obtener las imágenes, en el caso de este proyecto, se hará uso de la misión Sentinel-2. Adicionalmente, se seleccionará los satélites L2A los cuales tienen mejor calidad. Los satélites L2A también son conocidos como satélites de nivel 2A. Estos satélites ofrecen imágenes de gran calidad en las que se han corregido los efectos de la atmósfera sobre la luz reflejada por la superficie de la Tierra.

En este paso también se debe seleccionar la máxima cobertura nubosa que se quiere que presenten las imágenes, en este caso un 20%.

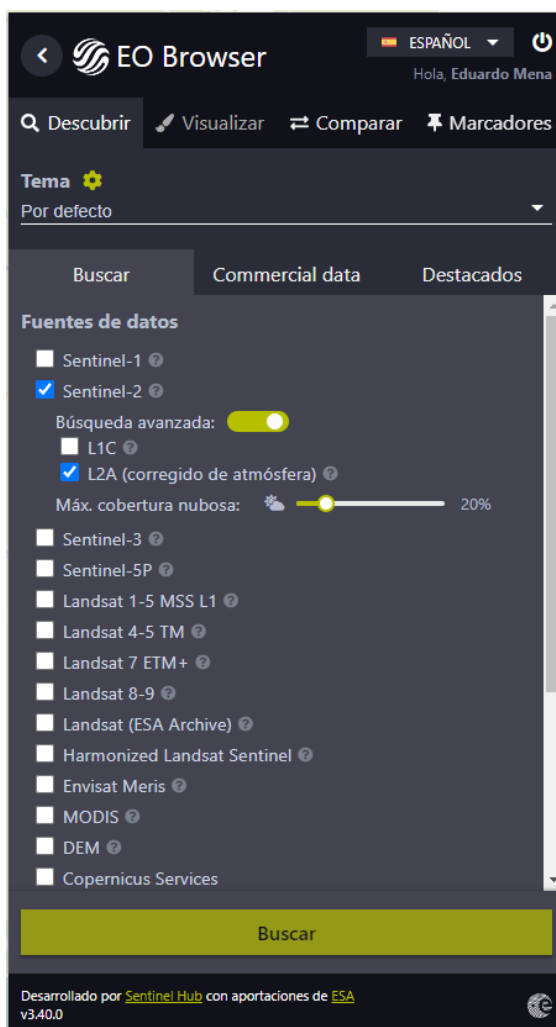


Ilustración 39: Selección origen de imágenes Sentinel Hub

Una vez realizado esto es hora de seleccionar el rango de fechas en las que queremos que se encuentren las imágenes de la búsqueda. Para esto seleccionaremos fechas entre el 1 de enero de 2017 y abril de 2023.



Ilustración 40: Selección rango de fechas Sentinel Hub

A continuación, se notará que en el calendario mostrará ciertas fechas sombreadas. Estas son las fechas que cumplen con las condiciones de nubosidad establecidas con anterioridad. Ahora solo queda navegar por ellas para conseguir las idóneas para el estudio del volcán.

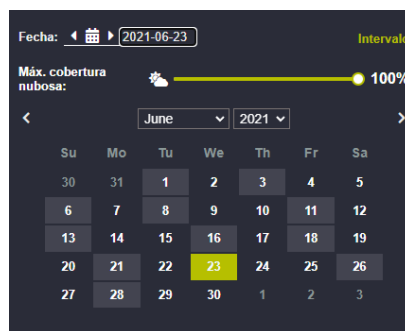


Ilustración 41: Selección fecha Sentinel Hub

Una vez seleccionada una fecha hay que seleccionar qué composición se necesita para destacar las áreas de interés, en este caso, la banda que interesa es la *SWIR*. Para seleccionar esta opción hay que hacerlo desde el menú lateral izquierdo. Una vez seleccionada, se verá una imagen tal y como se muestra en la siguiente figura.

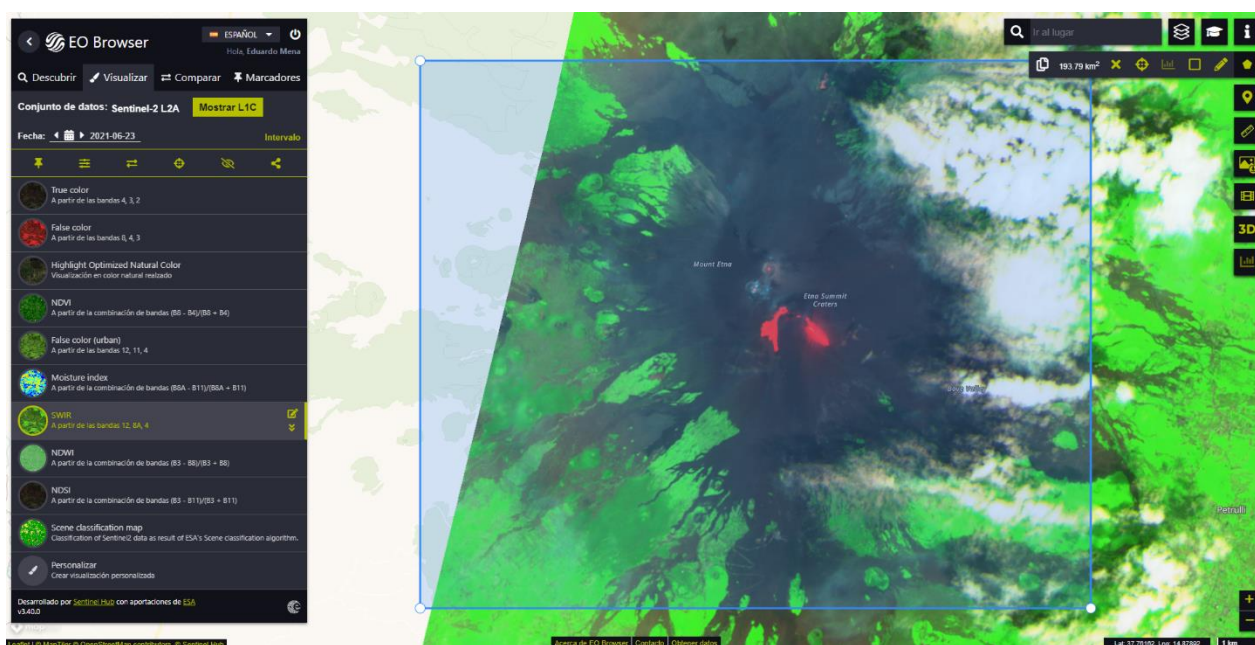


Ilustración 42: Visualización imagen Sentinel Hub

Ahora, para descargar esta imagen habrá que seleccionar el icono de descarga situado en el menú lateral derecho. Una vez dentro, habrá que seleccionar la ventana “Analítico” y ahí, dar a la opción de elegir el formato de la imagen, la resolución, el sistema de coordenadas y las capas a descargar. Para este trabajo se descargarán en un formato “*TIFF (8-bits)*”, con una resolución de 15x15 m/px y la capa *SWIR*.

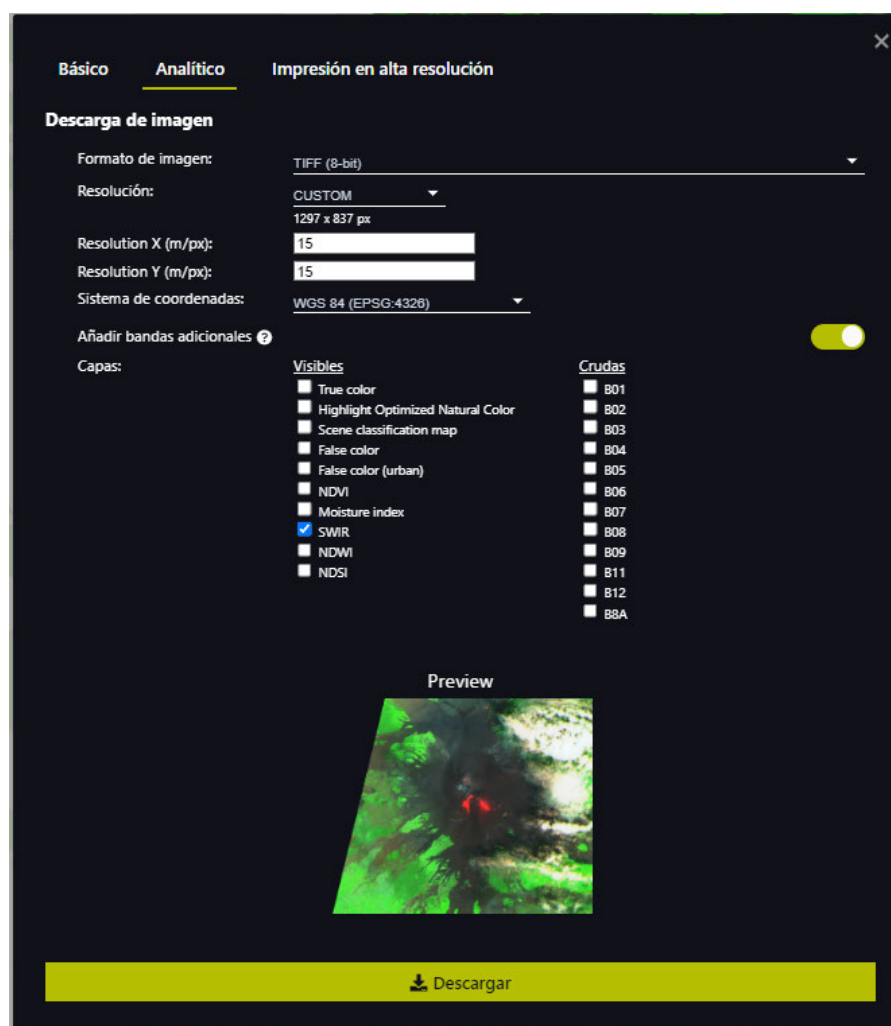
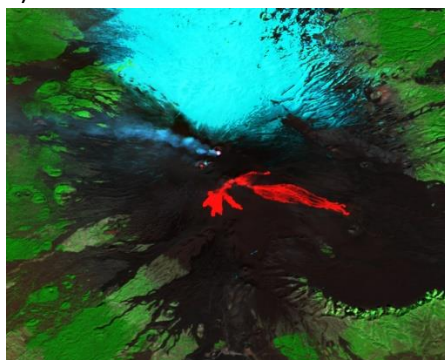


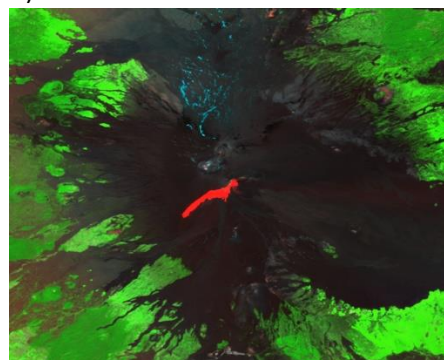
Ilustración 43: Selección parámetros de descarga Sentinel Hub

Una vez realizados todos estos pasos ya se tendrán las imágenes con las que se va a realizar el estudio. En este caso, las imágenes seleccionadas son de las siguientes fechas. 23 de junio de 2021, 23 de febrero de 2021, 19 de mayo de 2021 y 6 de junio de 2022. Se consideraron muchas otras imágenes, pero debido a la presencia de grandes masas de nieve o nubes se tuvieron que descartar. Las imágenes en cuestión son las mostradas a continuación.

a)



b)



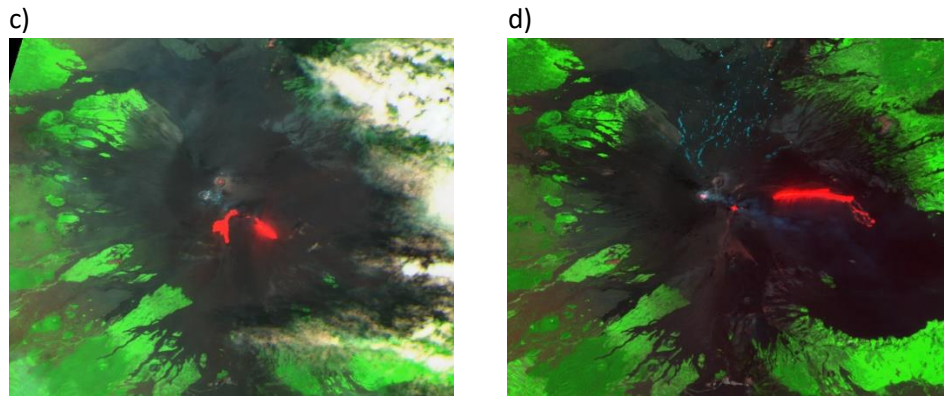


Ilustración 44: Imágenes escogidas volcán Etna a) 23 febrero 2021, b) 19 mayo 2021, c) 23 junio 2021, d) 6 junio 2022

4.3.2. Desarrollo del algoritmo DBSCAN

Tal y como se ha ido comentado a lo largo del trabajo, para realizar la tarea de detección y seguimiento de cambios del área de estudio, se va a hacer uso de un algoritmo de clasificación no supervisada denominado DBSCAN. Además, para comparar el rendimiento y los resultados de la clasificación, se va a aplicar también, a través de *Erdas Imagine*, una clasificación de las imágenes con otro algoritmo de clasificación no supervisada denominado KMEANS.

La primera tarea de todas a la hora del desarrollo del algoritmo de clasificación, antes de implementarlo en QGIS, ha sido adaptar este algoritmo para la clasificación de imágenes satelitales y la optimización de este para que se ejecute en unos tiempos no muy extensos. Toda esta adaptación fue realizada en la herramienta *Visual Studio Code* haciendo uso de una extensión para poder desarrollarlo en un *Jupyter Notebook* con una versión de *Python 3.9.7*.

En cuanto al algoritmo de DBSCAN, se ha decidido hacer uso del que está incluido en la librería de *scikit-learn* en *Python*. Un punto en contra de esta implementación incluida en *scikit-learn* es el tiempo que consume en memoria, sobre todo al aplicarse sobre matrices ya que es de complejidad n^2 . En este trabajo, cómo se va a realizar la clasificación de imágenes satelitales las cuales son matrices muy grandes y con mucho detalle y, además, la imagen va a ser de formato TIFF, siendo esta una composición de 3 bandas, la clasificación se va a realizar sobre una matriz de dimensiones *alto x ancho x 3*, haciendo que el rendimiento de este algoritmo no sea el esperado debido al tiempo y memoria que consumiría a la hora de realizar su tarea.

Debido a todo lo explicado con anterioridad, se plantearon algunas formas de optimizar el tiempo y consumo de memoria del algoritmo.

- La primera idea que se planteó fue reducir la resolución de la imagen, de manera considerable, para que así el algoritmo tuviese que tratar una matriz de menor tamaño y, por lo tanto, tardase menos y consumiese menos memoria. Esta idea fue descartada debido al propósito de la aplicación de este algoritmo. Al aplicarse a una imagen satelital para

analizar cambios producidos en esta, no tenía sentido reducir la resolución ya que en ese caso no se podría apreciar con claridad los cambios de la misma.

- Otra idea fue reducir el área de observación de la imagen, reduciéndola en varias hectáreas. Esta idea también fue descartada ya que, reducir el área de estudio de la imagen, no permitía realizar un buen seguimiento y detección de cambios.
- También se planteó realizar la clasificación de una imagen que sólo contuviese una banda, en concreto la banda 12, en vez de una compuesta por 3. Esto dio muy buenos resultados en cuanto a rendimiento en memoria y tiempo, pero en cuanto a la clasificación, si lo que queremos es hacer seguimiento de varias cubiertas terrestres (vegetación, lava, terreno volcánico), se hace imprescindible utilizar más de una banda, como puede observarse en la imagen. Aquí se puede ver la diferencia entre una imagen con la composición de bandas utilizada y cómo sería con una sola banda.

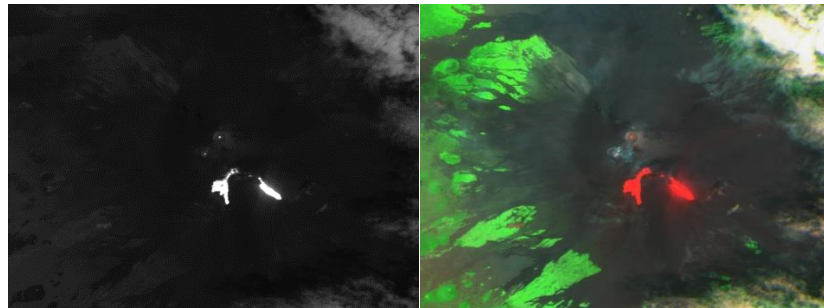


Ilustración 45: Imagen una banda y con composición de bandas 12, 8A y 4

Tal y como se puede ver, la imagen de la izquierda solo destaca la lava, mientras que la de la derecha, destaca más elementos de interés.

- También se consideró clasificar la imagen por bloques, es decir, por trozos iguales en los que sería más fácil aplicar el algoritmo, pero esto no fue una opción válida desde el principio ya que, por las características y funcionamiento del algoritmo, el resultado sería una imagen con una clasificación distinta por bloque, por lo que no tendría ningún sentido hacerlo ya que la clasificación resultante de aplicar el algoritmo de esta manera no sería coherente.
- Derivando de esta última idea, se intentó definir un solo bloque, la propia imagen, es decir, un bloque de tamaño *alto x ancho*. Aplicando esto, se observó una gran mejora en términos de optimización de tiempo y memoria. Esto fue debido a que el propio algoritmo, por las librerías que usa al leer la imagen, seleccionaba un área mucho mayor que el de la imagen, haciendo que no fuese óptimo. Es por esto por lo que se aplicó esta idea a la hora de aplicar el algoritmo y clasificar las imágenes.

Para leer la imagen se hizo uso de la librería *rasterio*, y para leer el bloque de la imagen se utilizó la librería *Window*.

Para medir el área de cada clúster también se implementó una función para que hiciese un recuento de píxeles que contenía cada uno de ellos.

4.3.3. Implementación del algoritmo en QGIS

Una vez creado el algoritmo, se procede a implementarlo en QGIS, pero para esto hay que definir una interfaz e incluir unas reglas definidas por la propia documentación de QGIS.

Pero antes de explicar lo anterior, se explicará cómo insertar un *script* propio en la herramienta.

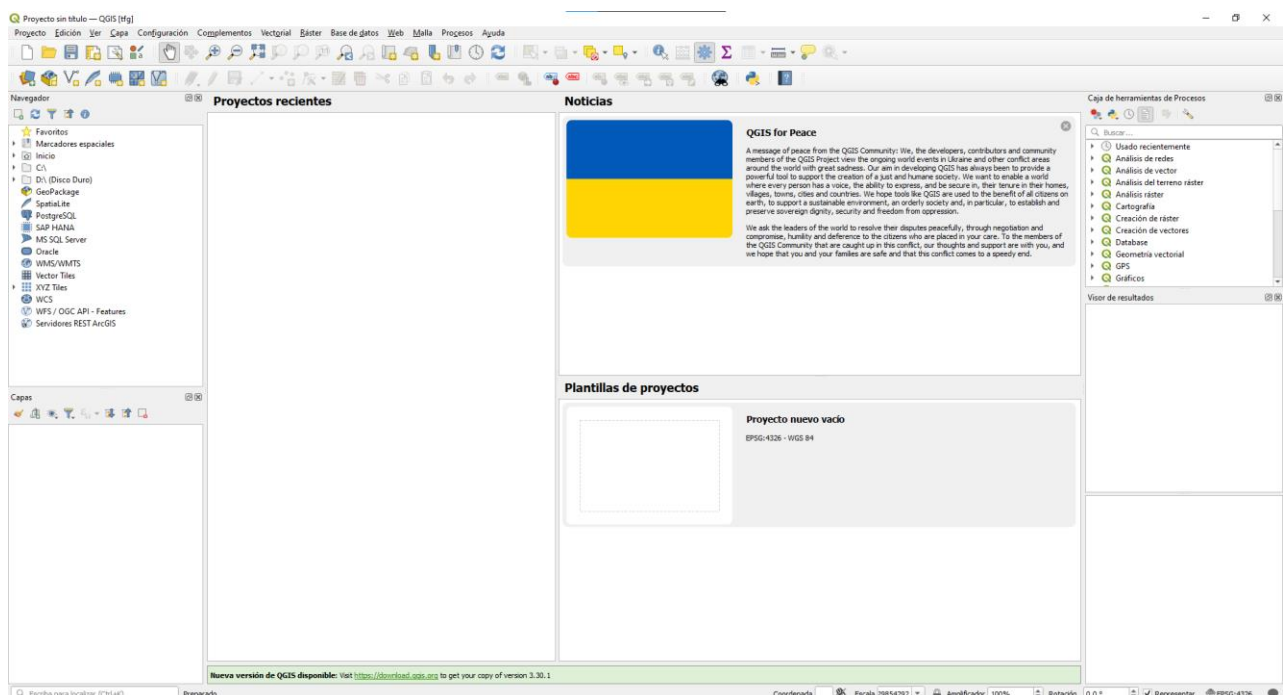


Ilustración 46: Pantalla inicio QGIS

Insertar un *script* propio a la herramienta QGIS es bastante sencillo. Simplemente hay que acceder a la caja de herramientas de procesos que, normalmente, se sitúa en un menú lateral a la derecha. Una vez en ese menú, en la parte superior aparece un símbolo con el logo de *Python*. Ahí dará la opción de crear, abrir o agregar un modelo propio.

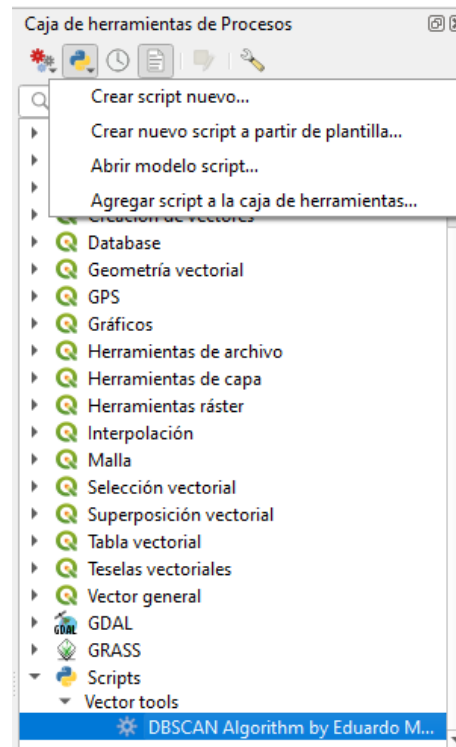


Ilustración 47: Caja herramientas de procesos QGIS

Hay que tener cuidado, pues no vale insertar un *script* tal y como se desarrollaría en un editor común, como *VSCode*, esto se debe a que, al ejecutar el algoritmo, la herramienta mostrará una interfaz en la que habrá que insertar todos los parámetros necesarios para que clasifique la imagen. Es por esto por lo que es necesario seguir una estructura y definir unas funciones para que QGIS muestre dicha interfaz. A continuación, se muestran algunas sentencias básicas para definir la interfaz tal y como requiere la herramienta.

Lo primero de todo es importar las librerías `qgis.core`, `qgis.analysis` y `qgis.PyQt.QTCore`. A continuación, hay que definir una serie de clases para establecer, entre otras cosas, el título, la descripción o la categoría del algoritmo.

Una vez hecho todo esto, hay que definir los parámetros de entrada y salida que tendrá nuestro algoritmo. Esto se debe definir dentro de una clase que se llame `initAlgorithm` en la que, por cada parámetro, se debe definir cómo se titula, el tipo de dato que esperará, el valor por defecto y, en el caso que tuviera, un rango de valores. A continuación, se encuentran dos ejemplos, el primero define el

parámetro *MINSAMPLES* de tipo *Double*, y el segundo define el parámetro *OUTPUT* de tipo ráster.

```
self.addParameter(QgsProcessingParameterNumber(
    self.MINSAMPLES,
    self.tr('min samples value'),
    QgsProcessingParameterNumber.Double,
    minValue=0,
    maxValue=500,
    defaultValue=15))
```

```
self.addParameter(QgsProcessingParameterFeatureSink(
    self.OUTPUT,
    self.tr("Output layer"),
    QgsProcessing.TypeVectorAnyGeometry))
```

Una vez definidos todos los parámetros se debe proceder con el desarrollo del algoritmo en sí. Para llamar a los parámetros que ya se habían definido con anterioridad, dependiendo del tipo de datos que sea, se hará de la siguiente manera.

```
eps = self.parameterAsDouble(parameters,
    self.EPS, context)

output = self.parameterAsOutputLayer(parameters,
    self.OUTPUT, context)
```

Una vez implementado todo el algoritmo, solo queda ejecutarlo para comprobar que se ha definido y que funciona tal y como se espera. En el menú desde el que se ha creado el *script*, aparecerá el algoritmo creado. Para ejecutarlo, simplemente hay que pulsar el botón derecho sobre el nombre del *plugin* creado y pulsar sobre la opción de ejecutar o bien, pulsar dos veces sobre este. Una vez ejecutado deberá aparecer una interfaz tal y como se ve en la siguiente figura.

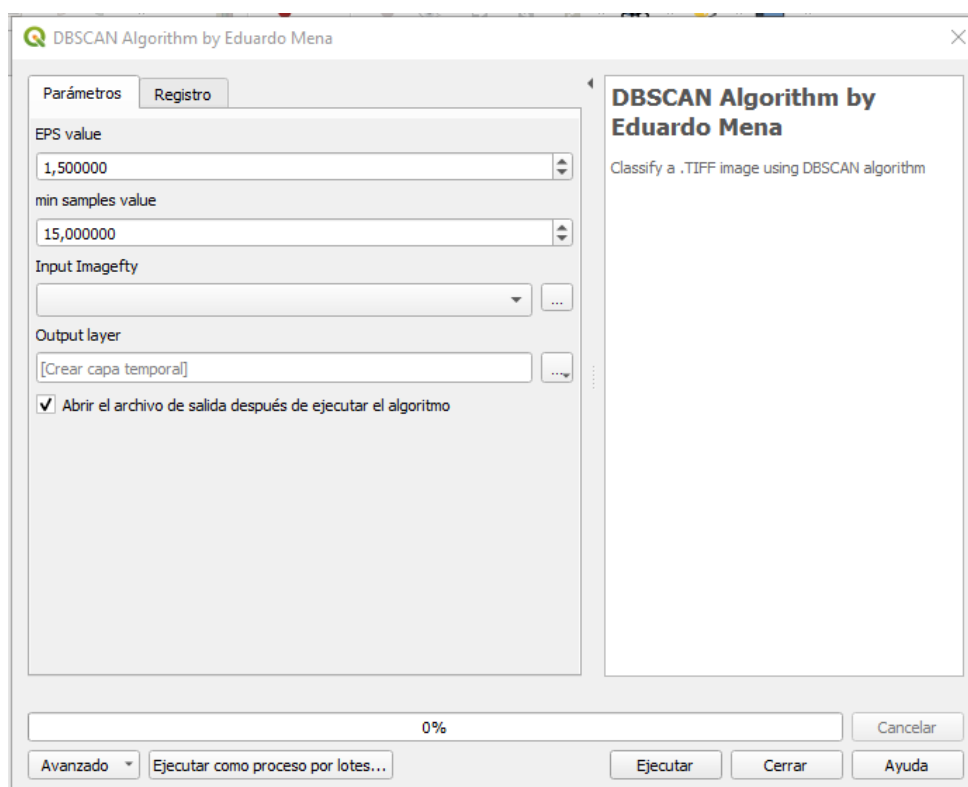


Ilustración 48: Interfaz DBSCAN QGIS

Como se puede observar aparecen distintos campos para insertar los parámetros que hemos definido. Solo queda introducirlos y pulsar sobre el botón de ejecutar. Si todo sale bien, la barra de progreso deberá mostrar el valor 100% al terminar la ejecución.

4.3.4. Evaluación de la calidad del DBSCAN

Una vez desarrollado e implementado el algoritmo en DBSCAN se ha utilizado el coeficiente Silhouette como métrica para evaluar la calidad del agrupamiento obtenido con el algoritmo. Dadas las características de dicho coeficiente, se ha utilizado una sola banda de cada imagen, en concreto, la banda número 12 correspondiente al rango espectral SWIR y que delimita el recorrido de la lava.

La implementación de este algoritmo se ha realizado haciendo uso de la librería *scikit-learn* incluida en *Python*.

Los resultados se explicarán en el capítulo siguiente.

5. Resultados y comparación del impacto

Una vez se han clasificado las imágenes satelitales escogidas mediante el algoritmo desarrollado DBSCAN y el algoritmo integrado en Erdas Imagine KMEANS solo queda comparar ambas clasificaciones para observar las posibles diferencias que puedan tener, calcular el área de los clústeres creados y finalmente evaluar el impacto del volcán Etna en el terreno.

En esta sección se mostrarán y analizarán los resultados derivados de ambas clasificaciones.

5.1. Resultados de las clasificaciones

Después de haber creado e implementado el algoritmo en QGIS y clasificar las imágenes mediante DBSCAN y KMEANS, se mostrarán resultados obtenidos a partir de ambos algoritmos.

Tal y como se ha comentado con anterioridad, dependiendo de las características únicas de cada imagen, se necesitarán unos valores distintos de los parámetros, es por esto por lo que los resultados de las clasificaciones que se muestran a continuación se han realizado con aquellos parámetros que mejor se ajustaban a cada imagen.

También hay que puntualizar, que al aplicar el algoritmo DBSCAN con las imágenes, dependiendo de los parámetros de ajuste que se utilizaban para clasificarlas, destacaba más las zonas de vegetación o las zonas de lava. Es por eso por lo que a continuación se mostrarán clasificaciones de cada imagen con distintos parámetros para luego poder decidir cuál se ajusta mejor para el estudio de este trabajo.

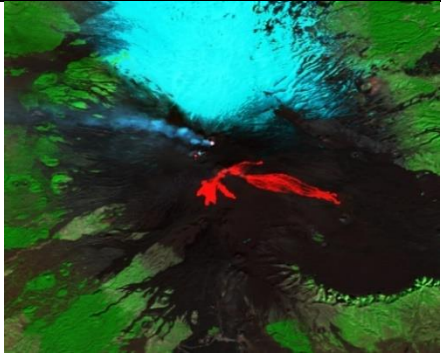
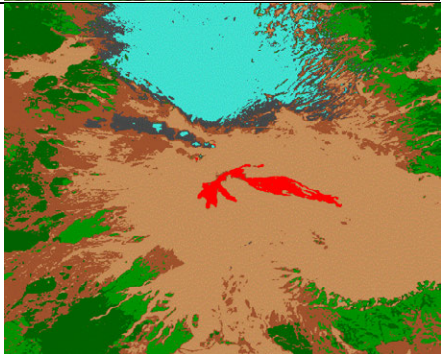
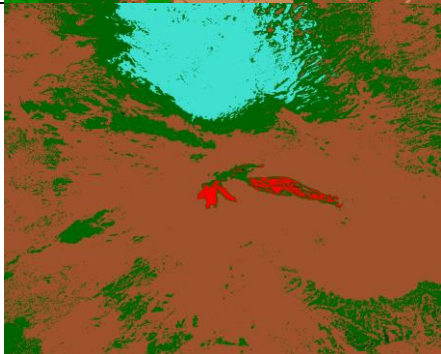
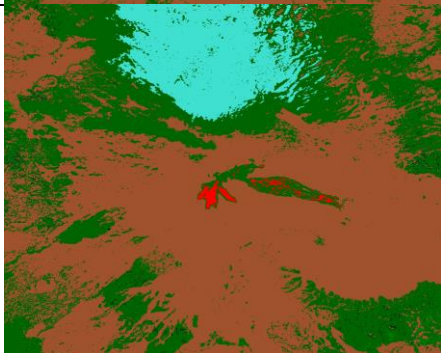
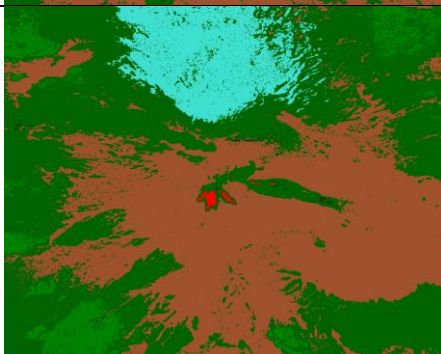
23 FEBRERO 2021		
IMAGEN ORIGINAL	Bandas 12, 8A y 4	
KMEANS	Nº Clústeres 7 Nº Iteraciones 10	
DBSCAN	eps = 2 min samples = 40	
	eps = 1.5 min samples = 40	
	eps = 1 min samples = 40	

Tabla 1 : Resultados 23 febrero 2021

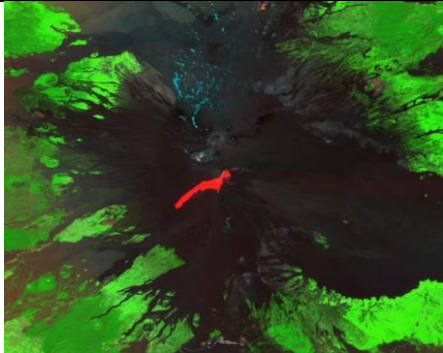
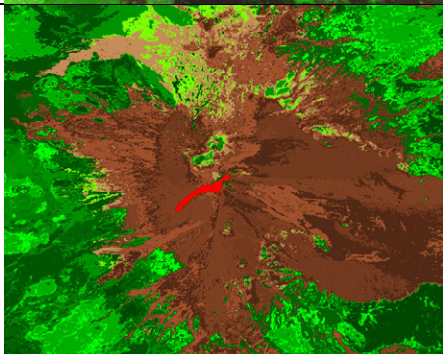
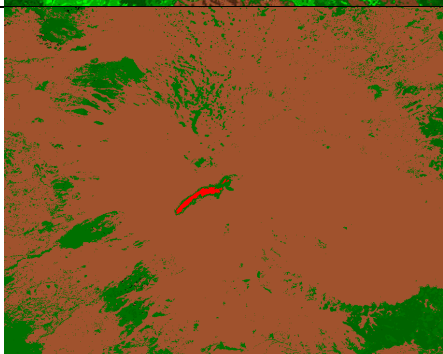
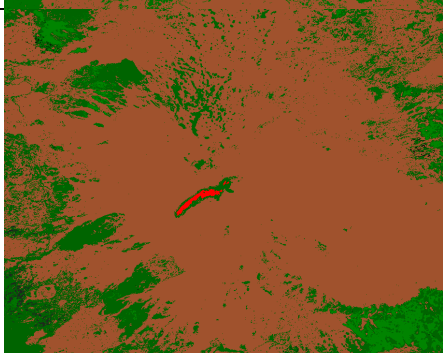
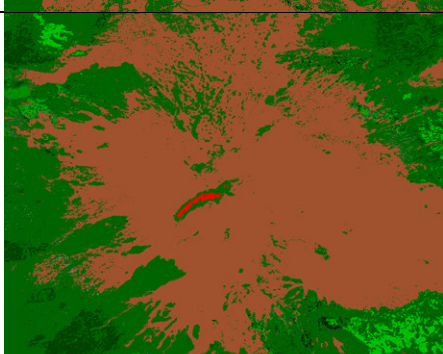
19 MAYO 2021			
IMAGEN ORIGINAL	Bandas 12, 8A y 4		
KMEANS	Nº Clústeres 20 Nº Iteraciones 20		
DBSCAN	eps = 2 min samples = 40		
	eps = 1.5 min samples = 40		
	eps = 1 min samples = 40		

Tabla 2: Resultados 19 mayo 2021

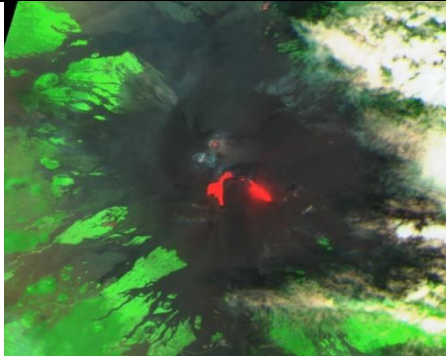
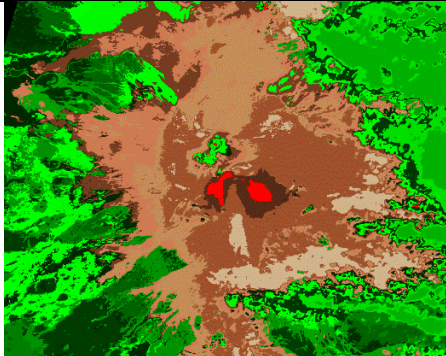
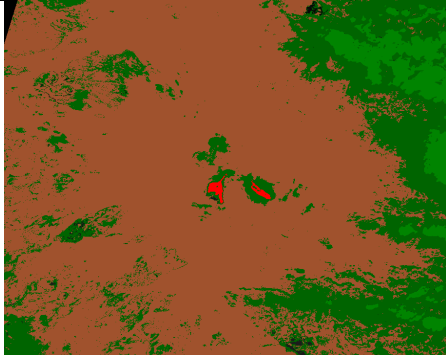
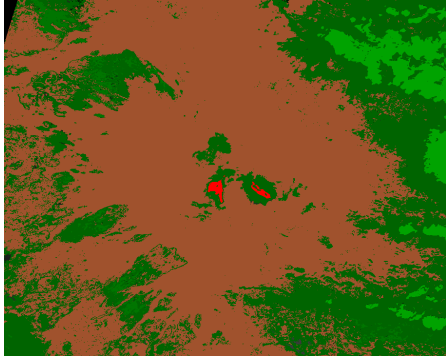
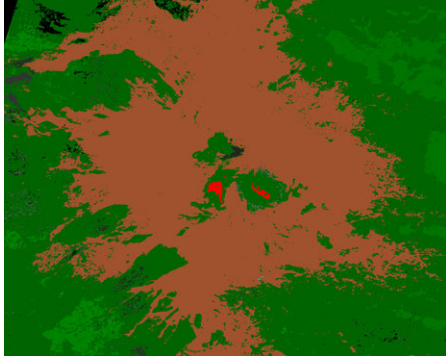
23 JUNIO 2021		
IMAGEN ORIGINAL	Bandas 12, 8A y 4	
KMEANS	Nº Clústeres 20 Nº Iteraciones 20	
DBSCAN	eps = 2 min samples = 40	
	eps = 1.5 min samples = 40	
	eps = 1 min samples = 40	

Tabla 3: Resultados 23 junio 2021

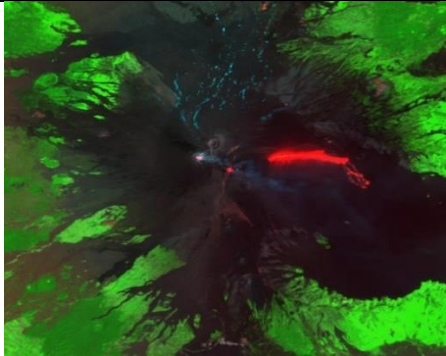
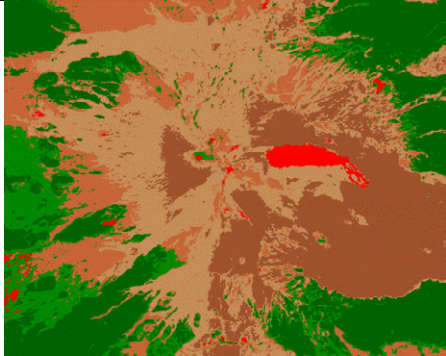
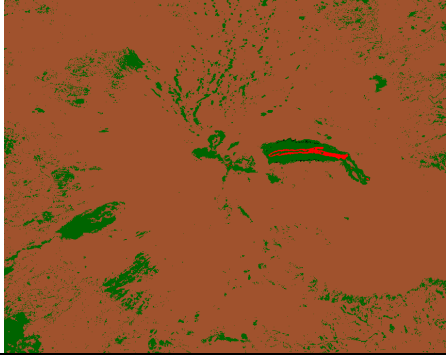
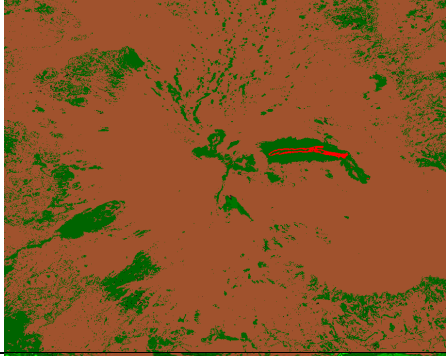
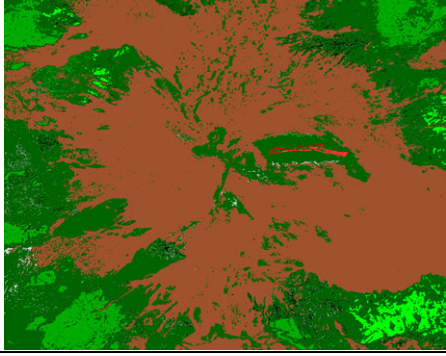
6 JUNIO 2022		
IMAGEN ORIGINAL	Bandas 12, 8A y 4	
KMEANS	Nº Clústeres 7 Nº Iteraciones 10	
DBSCAN	eps = 2 min samples = 40	
	eps = 1.5 min samples = 40	
	eps = 1 min samples = 40	

Tabla 4: Resultados 6 junio 2022

5.2. Comparación de las clasificaciones

Tal y como se puede apreciar en las imágenes de las tablas mostradas anteriormente, se puede observar como en las clasificaciones realizadas con DBSCAN, dependiendo de los valores que se les dan a los parámetros, destaca más o la superficie correspondiente a vegetación o, a la lava líquida. Por ejemplo, si utilizamos los valores de ϵ 1 y puntos mínimos 40, se aprecia muy bien la parte vegetal, pero el área de la lava no queda bien definida. Si nos vamos al ejemplo contrario, es decir con los valores de ϵ 2 y puntos mínimos 40, se puede observar que la lava queda mucho mejor definida, pero la parte de superficie de vegetación es prácticamente inexistente. Debido a todo esto y a que se pretende dar protagonismo a ambas áreas, se escogerá, como clasificación a estudiar, la realizada con los parámetros de ϵ 1,5 y puntos mínimos 40. Con estos parámetros se destaca tanto las áreas de vegetación como las áreas de lava líquida. En el caso en el que se hubiera utilizado un valor de puntos mínimos menor a 40, en este contexto, habrían salido como resultado unas clasificaciones con mucho ruido, es decir, muchos clústeres insignificantes y de un tamaño minúsculo, estropeando así la clasificación. A continuación, se puede ver el resultado de una clasificación con valor ϵ 1,2 y puntos mínimos 30.

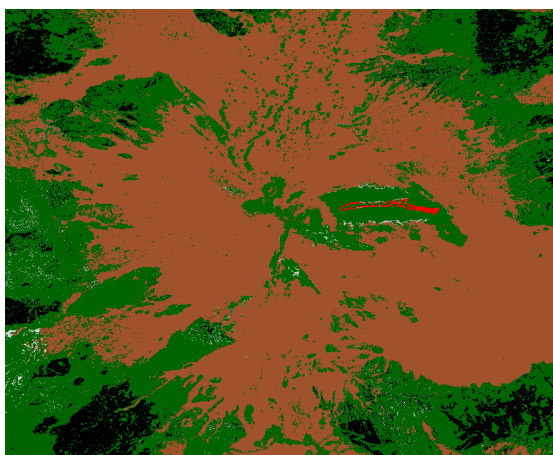


Ilustración 49: Imagen DBSCAN eps 1,2 min 30

Esta imagen, en el área de la lava no cuenta con tan solo un clúster que la define, sino que está formada por más de 3 clústeres. En el resto de la imagen se pueden apreciar un gran número de puntos blancos, los cuales son ruido, es decir, un número muy elevado de clústeres muy pequeños.

En el caso contrario, es decir, la elección de una clasificación con unos parámetros más elevados, por ejemplo, ϵ 2,5 y puntos mínimos 50, resultaría una imagen con clústeres grandes, los cuales no permitirían distinguir ninguna área de la clasificación original, es decir, sería una clasificación muy genérica. Realizar una clasificación con estos parámetros en una imagen tan grande y con tanta resolución como ésta utilizando el algoritmo DBSCAN sería totalmente ineficiente en cuanto a tiempo de ejecución y uso de memoria ya que, como se ha comentado con anterioridad, este algoritmo tiene una complejidad en memoria de n^2 en conjuntos de datos presentados en una matriz, además habría añadir que esta imagen cuenta con tres bandas por lo que la

complejidad sería $3 \cdot n^2$ lo cual es inasumible para la ejecución en unos tiempos razonables. Además, en el peor de los casos, realizar la clasificación de esta imagen con estos parámetros podría provocar el bloqueo del dispositivo que esté ejecutando el algoritmo en el caso en el que no disponga con una memoria RAM de más de 32 GB. Una solución para este problema sería la reducción de resolución de la imagen, pero se perdería mucho detalle de la misma.

Hay que puntualizar que el valor de ϵ se ha aproximado haciendo uso del método del codo, el cual consiste en realizar una clasificación del conjunto de datos con el algoritmo *K-Neighbors*, en el que tengamos como salida un gráfico que nos estime los valores de ϵ más idóneos para nuestro conjunto.

En el caso de KMEANS es más sencillo obtener una clasificación según lo deseado ya que solo depende de un parámetro llamado número de clústeres. Tal y como se explica con anterioridad, para algunas imágenes ha sido necesario crear 7 clústeres mientras que en otras ha sido necesario crear 20, esto ha ocurrido ya que, si no, había casos en los que no se clasificaba correctamente las áreas de lava líquida. A continuación, se muestra una imagen clasificada con 7 clústeres y que en la tabla ha sido clasificada con 20 clústeres, para poder observar lo mencionado con anterioridad.

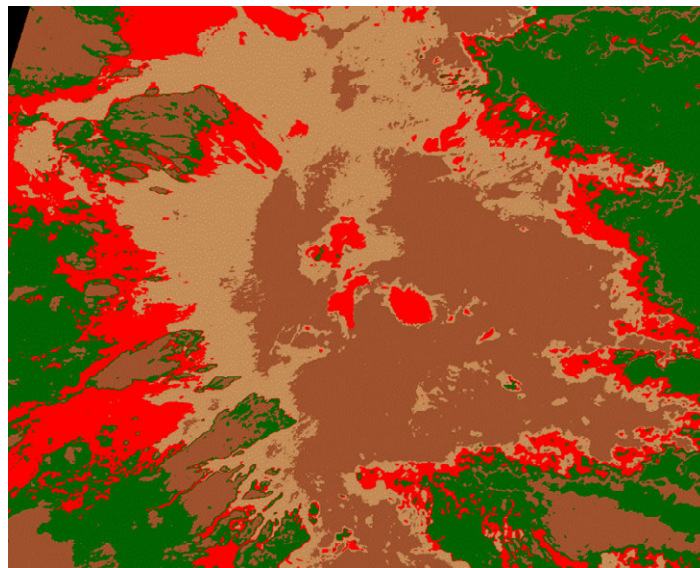


Ilustración 50: Resultado KMEANS 7 clústeres

Una vez analizados los resultados de forma rápida, se puede observar que hay bastante diferencia en el resultado de las clasificaciones dependiendo del algoritmo que se utiliza. Esto era de esperar puesto que su funcionamiento es bastante distinto.

5.3. Análisis cuantitativo de las imágenes clasificadas

Una vez realizadas todas las clasificaciones tanto con el algoritmo DBACAN como con KMEANS, hay que procesar toda la información de la imagen clasificada para poder realizar el estudio de cambios de estas calculando las áreas que ocupa cada clúster.

Para realizar el análisis cuantitativo de las imágenes clasificadas que ya se han expuesto en el apartado anterior se va a hacer uso del *plugin* ya implementado en la plataforma QGIS llamado SCP o *Semi-Automatic Classification Plugin*.

SCP es una herramienta dedicada al procesamiento de imágenes. En el caso de este trabajo se ha utilizado en la fase de postprocesamiento para realizar un análisis cuantitativo de las clasificaciones obtenidas. Como resultado de la aplicación de este algoritmo, se obtendrá una tabla que indicará el número de píxeles y porcentaje que representa cada clúster que forma la clasificación. A partir de esta tabla y, teniendo la medida de cada píxel de la imagen, se podrá calcular el área de cada clúster. En el caso de las imágenes escogidas para clasificar cuentan con píxeles con una resolución de 15 x 15 metros/píxel, es decir, cada píxel cuenta con un área de 225 m².

A continuación, se muestran unas tablas con los datos relativos al tamaño de cada imagen clasificada, en el caso de este trabajo por cada fecha, habrá una imagen clasificada con KMEANS y otra con DBSCAN. Además, solo se incluirán a la tabla los clústeres más significativos, ya que, las imágenes clasificadas con DBSCAN, tendrán como resultado un número indeterminado de clústeres creados, al contrario de lo que ocurre en KMEANS. Es por esto por lo que se va a intentar agrupar los clústeres entre los que representan la lava, la superficie de vegetación y la superficie del volcán. Para realizar esta agrupación, lo que se ha hecho es realizar la suma de todos aquellos clústeres pertenecientes a la misma área para así poder realizar una medición de los cambios y evolución de la superficie estudiada

Hay que tener en cuenta que los resultados de la clasificación con DBSCAN no va a ser tan exacta como la realizada con KMEANS. Esto es debido a que se ha intentado dar prioridad a que la clasificación contenga tanto áreas de lava como de vegetación bien definidas. Todo esto ya ha sido explicado con anterioridad.

23 FEBRERO 2021				
	Clúster	N.º Píxeles	Área total (m ²)	Área total (ha)
KMEANS	Lava	7502	1687950	1687,95
	Sup. Vegetal	292429	65796525	65796,525
	Sup. Volcán	260769	58673025	58673,025
	Nieve	107974	24294150	24294,15
DBSCAN	Lava	3185	716625	716,625
	Sup. Vegetal	213513	48040425	48040,425
	Sup. Volcán	378751	85218975	85218,975
	Nieve	66327	14923575	14923,575

Tabla 5: Análisis cuantitativo 23 febrero 2021

19 MAYO 2021				
	Clúster	N.º Píxeles	Área total (m ²)	Área total (ha)
KMEANS	Lava	1854	417150	417,15
	Sup. Vegetal	292327	65773575	65773,575
	Sup. Volcán	348568	78427800	78427,8
DBSCAN	Lava	1068	240300	240,3
	Sup. Vegetal	141353	31804425	31804,425
	Sup. Volcán	522606	117586350	117586,35

Tabla 6: Análisis cuantitativo 19 mayo 2021

23 JUNIO 2021				
	Clúster	N.º Píxeles	Área total (m ²)	Área total (ha)
KMEANS	Lava	3275	736875	736,875
	Sup. Vegetal	344035	77407875	77407,875
	Sup. Volcán	320027	72006075	72006,075
DBSCAN	Lava	1141	256752	256,752
	Sup. Vegetal	247654	55722150	55722,15
	Sup. Volcán	416151	93633975	93633,975

Tabla 7: Análisis cuantitativo 23 junio 2021

6 JUNIO 2022				
	Clúster	N.º Píxeles	Área total (m ²)	Área total (ha)
KMEANS	Lava	9791	2202975	2202,975
	Sup. Vegetal	222457	50052825	50052,825
	Sup. Volcán	436426	98195850	98195,85
DBSCAN	Lava	1249	281025	281,025
	Sup. Vegetal	93184	20966400	20966,4
	Sup. Volcán	573363	129006675	129006,675

Tabla 8: Análisis cuantitativo 6 junio 2022

Las tablas en las que se recogen los resultados derivados del análisis cuantitativo realizado con el *plugin SCP* explicado con anterioridad cuentan con un total de 4 columnas cada una. La columna **clúster**, define qué grupo se está analizando, es decir, si los datos correspondientes a esa fila pertenecen a los clústeres relativos a la lava, superficie vegetal, el resto de superficie del volcán o, en el caso en el que tenga, nieve. La columna llamada **Nº. Píxeles**, define el número total de píxeles de cada grupo de clústeres. Hay que tener en cuenta que, tal y como ya se ha comentado, en este trabajo cada píxel tiene una resolución de 15 x 15 metros/píxel, por lo que esta información será de mucha utilidad para calcular el área de la zona. Las dos últimas columnas son las relativas a las **áreas**, se ha calculado un área en metros cuadrados y otra en hectáreas para tener más perspectiva de la dimensión de las imágenes que se están estudiando.

Una vez explicada la estructura de las tablas, se da paso a la explicación e interpretación de los resultados obtenidos. Este análisis se realizará para cada tabla a continuación.

23 febrero 2021

Tal y como muestran los resultados, hay una gran diferencia entre una clasificación con KMEANS o con DBSCAN.

En cuanto a la clasificación con KMEANS define un clúster que representa la lava con un área de 1687 hectáreas, la superficie vegetal cuenta con 65796 hectáreas y el cono o superficie del volcán ocupa un área de 58673 hectáreas, además, esta imagen tiene un clúster que define un área bastante extensa de nieve siendo esta de 24294 hectáreas.

En cambio, DBSCAN define que la lava ocupa 716 hectáreas, lo cual es un valor significativamente menor al mencionado anteriormente. Analizando este resultado y mirando la imagen clasificada, se puede observar como la clasificación con DBSCAN no tiene los clústeres tan bien definidos como los tiene KMEANS. El área de la superficie

vegetal, nieve y el resto del volcán también varía significativamente, pero no tanto como la superficie de lava.

19 mayo 2021

En esta fecha también se observa una gran disparidad en los resultados obtenidos con ambos algoritmos de clasificación, siendo la clasificación con KMEANS la que define con más precisión todas las áreas. A parte de esto, se puede observar como el flujo de lava se ha hecho más pequeña, en la fecha anterior según KMEANS, el área de lava era de 1687 hectáreas mientras que ahora es de 417 hectáreas, en cuanto a la superficie vegetal, se ha mantenido muy parecida y la nieve ha desaparecido.

23 junio 2021

Se puede observar como el flujo de lava ha aumentado ligeramente respecto a la fecha anterior. Según KMEANS, el área de lava era de 417 hectáreas mientras que ahora es de 736 hectáreas, en cuanto a la superficie vegetal, ha aumentado.

Estas tendencias también quedan reflejadas en los resultados obtenidos con el algoritmo DBSCAN.

6 junio 2022

En esta imagen, cabe destacar el hecho de que en KMEANS el área de la superficie de lava resulta notablemente mayor que con el DBSCAN. De hecho, el valor obtenido con KMEANS claramente sobredimensiona el recorrido de la lava que se observa en la imagen original.

Teniendo en cuenta todos estos datos y resultados obtenidos, se puede llegar a la conclusión de que las clasificaciones realizadas con KMEANS, en términos generales, son mejores que las realizadas con DBSCAN, es decir, los clústeres obtenidos con KMEANS tienen más sentido, en cuanto a tamaño y a lo que representan, que los obtenidos con DBSCAN. Con esto se quiere decir, que KMEANS consigue definir de mejor manera las distintas áreas que contiene la imagen ráster, mientras que DBSCAN no es capaz de clasificar e identificar correctamente y con precisión los clústeres de la imagen. Esto se ve en los datos, ya que estando ante una misma clasificación, se puede observar una gran diferencia de píxeles clasificados dependiendo del algoritmo utilizado. Eso sí, la tendencia de crecimiento o disminución del tamaño de cada clúster en cada clasificación es la misma.

A continuación, se pueden observar unos gráficos que representan las áreas (en número de píxeles) de los clústeres pertenecientes a cada grupo realizados con ambos algoritmos.

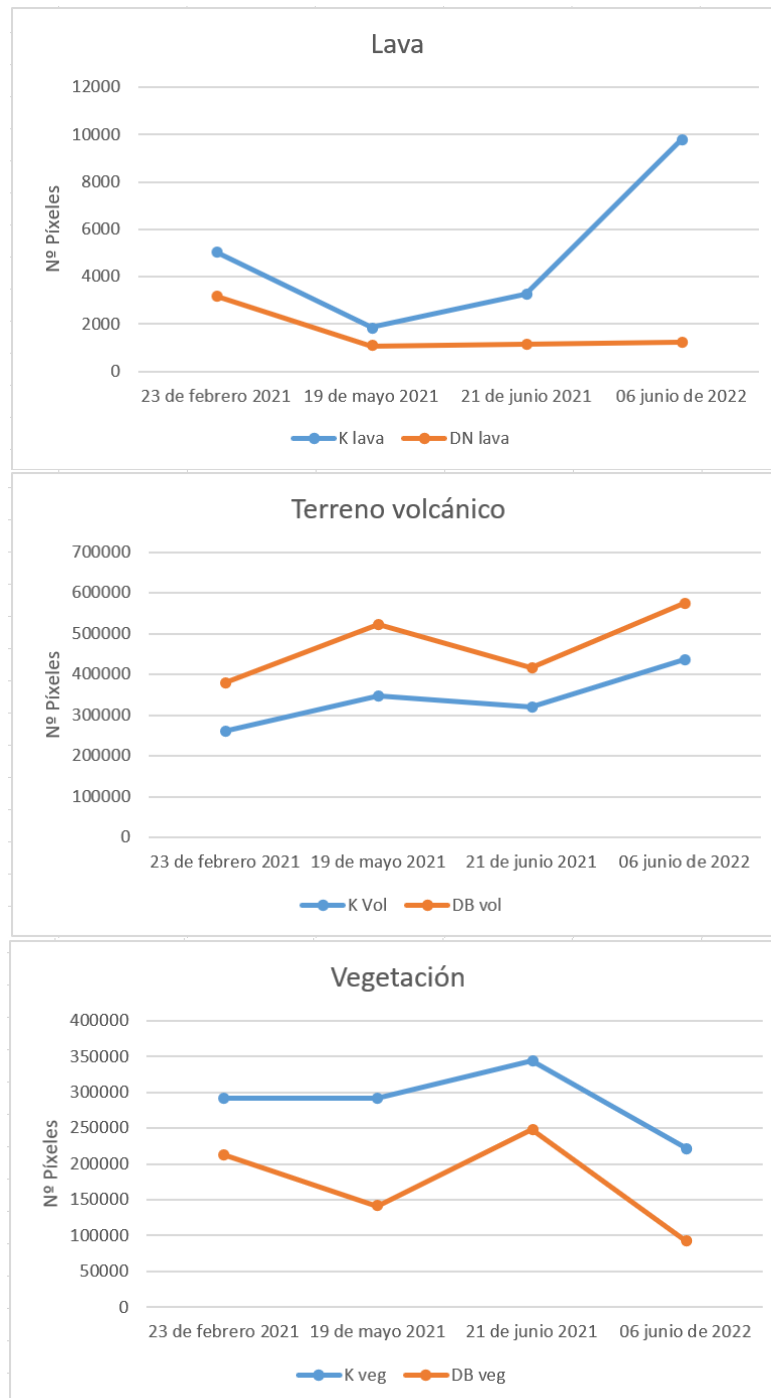


Ilustración 51: Comparación áreas clasificaciones DBSCAN vs KMEANS

Tal y como muestran estas gráficas, se puede decir que ambos algoritmos, aunque no coinciden a la hora de definir el tamaño de cada clúster, siguen las mismas tendencias, es decir, si la clasificación realizada con DBSCAN crece, la realizada con KMEANS también crece. Además, KMEANS siempre da como resultado valores por encima respecto a los obtenidos con DBSCAN. Solo el punto perteneciente al clúster de la lava del día 6 de junio de 2022, como ya comentamos con anterioridad, es anómalo y no sigue esta tendencia. Esto se puede observar visualmente en los resultados de la clasificación de ambas imágenes.

La razón por lo que ocurre esto es por la manera tan distinta que tienen ambos algoritmos a la hora de calcular los clústeres que contiene la imagen. DBSCAN maneja muy bien el ruido y los valores atípicos y puede identificar clústeres con formas irregulares, por lo que parece apropiado para la clasificación de imágenes satelitales y el seguimiento medioambiental. Sin embargo, debido al elevado número de datos a tratar y a que computacionalmente es muy costoso, finalmente los resultados con KMEANS han resultado mejores y se ajustan más a las imágenes originales. También es necesario apuntar que DBSCAN es muy sensible a los hiperparámetros (*eps* y *min samples*), y un ligero cambio de éstos puede llevar a cambios drásticos en los resultados. En este sentido y para trabajos futuros, cabría profundizar en este tema.

5.4. Aplicación del coeficiente Silhouette

El coeficiente Silhouette es una métrica para evaluar la calidad del agrupamiento obtenido con algoritmos de *clustering*.

Para poder interpretar y validar la coherencia de los clústeres creados tanto con el algoritmo de DBSCAN desarrollado e implementado con anterioridad en QGIS como con KMEANS, se ha aplicado el coeficiente Silhouette sobre la banda 12 de la imagen clasificada. La razón por la que se ha decidido hacerlo sobre esta banda es porque este coeficiente no permite su aplicación sobre imágenes con varias capas. Por ello, hemos elegido la banda 12 que es la que mejor resalta el flujo de lava emitido por el volcán, por lo que el algoritmo debería ser capaz de realizar una clasificación que la identifique correctamente.

El coeficiente Silhouette se puede definir como el resultado de la aplicación de la siguiente fórmula:

$$\frac{(b - a)}{\max(a, b)}$$

Donde a es la distancia promedio entre cada punto de un clúster y b es la distancia media entre todos los clústeres. Para entenderlo mejor, a continuación, se muestra una imagen que lo representa.

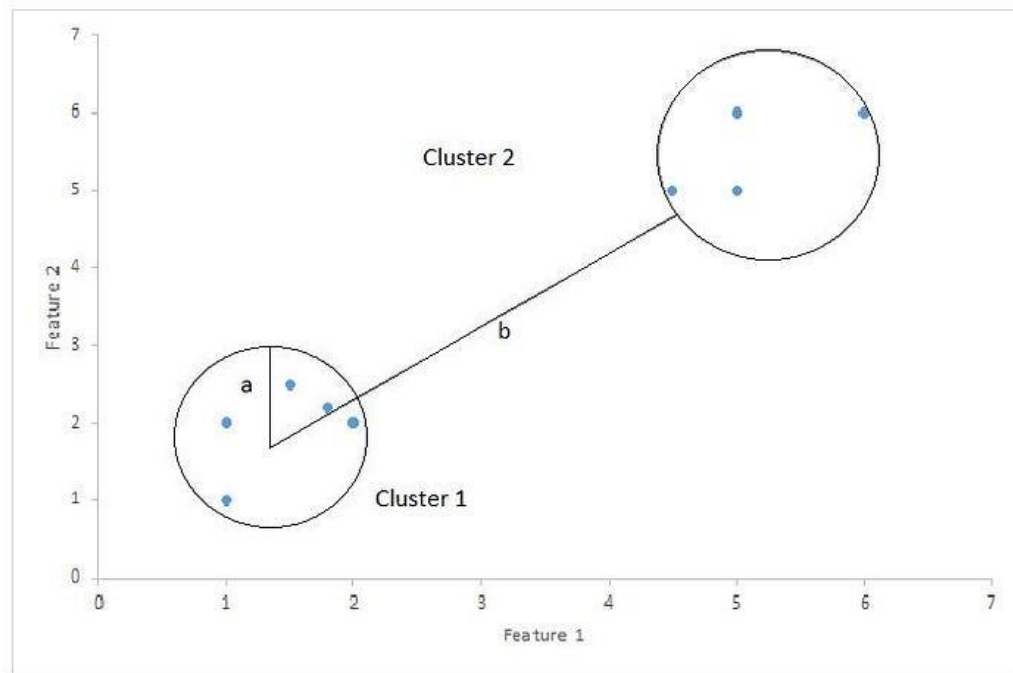


Ilustración 52: Coeficiente Silhouette [<https://towardsdatascience.com/silhouette-coefficient-validating-clustering-techniques-e976bb81d10c>]

Para aplicar el coeficiente Silhouette sobre la imagen clasificada resultante de aplicar el algoritmo DBSCAN desarrollado, se ha creado un pequeño *script* en *Python* el cual hace uso de la librería *scikit-learn*.

El resultado de la aplicación del coeficiente Silhouette sobre nuestra imagen clasificada será un valor comprendido entre -1 y 1. Para interpretar este resultado es importante saber que si nos sale un valor cercano a -1, quiere decir que el algoritmo DBSCAN desarrollado ha hecho un mal agrupamiento, es decir, los clústeres han sido asignados de manera errónea, si sale un valor cercano a 0, quiere decir que la distancia entre clústeres no es significativa, en cambio, si el resultado es cercano a 1, quiere decir que el algoritmo ha realizado un buen agrupamiento, por lo que los clústeres se pueden distinguir claramente.

A continuación, se muestran los resultados obtenidos de la aplicación del coeficiente Silhouette sobre las imágenes clasificadas con DBSCAN y KMEANS sobre la banda 12:

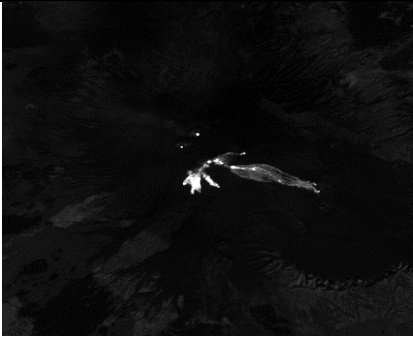
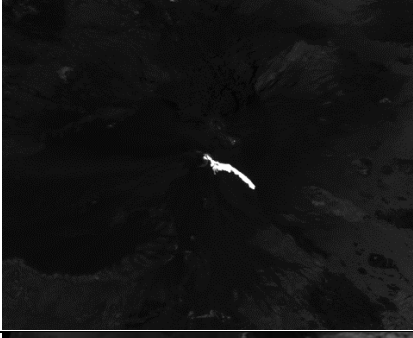
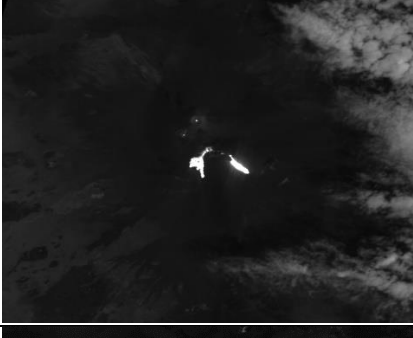
Fecha	Imagen Banda 12	Valor Coeficiente Silhouette sobre DBSCAN	Valor Coeficiente Silhouette sobre KMEANS
23 febrero 2021		0.769	0.873
19 mayo 2021		0.856	0.678
23 junio 2021		0.326	0.816
6 junio 2022		0.729	0.908

Tabla 9: Resultados coeficiente Silhouette

Tal y como se puede observar, dependiendo de la imagen analizada, se dan casos en los que DBSCAN otorga mejor clasificación que KMEANS según el coeficiente Silhouette y viceversa. En el caso de DBSCAN, otorga mejores resultados cuando la imagen presenta condiciones meteorológicas favorables, es decir, cuando no hay presencia de nubes ni de nieve y la imagen es lo más clara posible, mientras que KMEANS es más flexible en estos términos. Por ejemplo, la imagen del 23 de junio de 2021 se observa que hay una

gran diferencia entre ambos algoritmos, esto se da porque hay presencia de nubes, en cambio, en la imagen del 19 de mayo, el algoritmo DBSCAN es notablemente mejor que KMEANS ya que es capaz de clasificar mejor en condiciones óptimas.

6. Impactos sociales y ambientales

Este proyecto contribuye al conocimiento del comportamiento de los volcanes y su impacto en el medioambiente. Más allá de esto, también permite, gracias a la teledetección, realizar un seguimiento del volcán de forma periódica y observar, no solo la destrucción que puede crear, sino también cómo se regenera la zona que ha sido afectada, y todo ello con un método no invasivo y a distancia.

En septiembre de 2015 los líderes y representantes de la ONU llegaron a un acuerdo para dar inicio a la llamada Agenda 2030, la cual tenía varios objetivos, como, por ejemplo, erradicar el hambre en el mundo o la acción por el clima. Estos objetivos son conocidos como los Objetivos de Desarrollo Sostenible (ODS). Están compuestos por 17 objetivos los cuales, cada uno de ellos, tienen unas metas específicas que se deben de cumplir antes del 2030. Estos objetivos no solo se tienen que llevar a cabo por los gobiernos de los diferentes países, sino que también deben cumplirlos empresas o la sociedad en general.



Ilustración 53: Objetivos de desarrollo sostenible

[https://www.guardiacivil.es/es/institucional/Responsabilidad_Social_Corporativa/rsc]

Una vez explicado todo esto, la relación de este trabajo con el cumplimiento de estos objetivos es bastante estrecha. En cuanto a los volcanes, estos tienen una gran relación con la salud y bienestar (ODS 3), acción por el clima (ODS 13) y vida de ecosistemas terrestres (ODS 15).

Tal y como se ha explicado a lo largo de todo el trabajo, el volcán Etna es el volcán más activo de Europa por lo que sufre erupciones constantemente, las cuales suelen ser bastante explosivas. Además, el volcán está rodeado de varios pueblos y ciudades, entre ellos está la ciudad de Catania, la cual vive en una constante amenaza. A parte de esto, las erupciones

volcánicas resultan en una gran cantidad de emisiones de Dióxido de Carbono y de distintos gases que pueden ser nocivos para la salud.

La teledetección, en este caso, a través de imágenes satelitales, permite analizar y evaluar los daños producidos por los efectos de las distintas erupciones. Por lo que se puede decir que es una herramienta muy útil para, detectar las zonas que suelen estar más afectadas por las erupciones y actuar en consecuencia para prevenir el daño en futuras ocasiones, también es útil para el seguimiento día a día de la erupción y así poder prever en qué sentido se mueven los ríos de lava, además, también permite el análisis post-erupción para evaluar los daños y modificaciones del terreno causados. Finalmente, la teledetección se puede usar para medir la temperatura de la lava o cantidad y dirección de emisiones que produce.

7. Conclusiones

En este último capítulo del proyecto, se recogerán las conclusiones obtenidas tras haber desarrollado y analizado todos los resultados ya mencionados en secciones anteriores.

Gracias a este trabajo, he adquirido nuevos conocimientos en varias áreas en las que no había trabajado con anterioridad. Una de estas áreas es la teledetección, muy útil hoy en día en diversos campos, para obtener información relevante acerca de lo que ocurre en la superficie terrestre. También he podido profundizar en el conocimiento de los satélites, aprendiendo mucho acerca de su funcionamiento y características fundamentales. Por otro lado, he conocido nuevas herramientas como pueden ser *Erdas Imagine* o *QGIS*. Para finalizar, aunque ya lo conocía, he ahondado en el funcionamiento del algoritmo DBSCAN, su comportamiento y sus límites a la hora de clasificar imágenes.

A continuación, se analizarán los objetivos que se plantearon al inicio del proyecto para determinar cuáles se han conseguido y cuáles no y el porqué.

Objetivo 1: *“Obtención de imágenes en las que se aprecie la lava del volcán Etna. Estas imágenes deben ser lo más claras posibles, es decir, sin nubes, con la menor cantidad de nieve posible y con una buena resolución a partir la cual realizar el análisis.”*

Este objetivo se consiguió con éxito, aunque no fue nada fácil ya que, el volcán Etna, al estar situado a más de 3000 metros de altura y en una isla, cuenta con varios meses en los que la cima está llena de nieve, además, es muy complicado obtener una imagen con el cielo despejado en esa zona. Por otro lado, como se querían imágenes que mostrasen el río de lava, era muy complicado encontrar una imagen sin una columna de cenizas. Dicho esto, y tras un largo proceso de investigación, se ha conseguido obtener las imágenes deseadas.

Objetivo 2: *“Desarrollo del algoritmo DBSCAN y su posterior implementación en la herramienta Erdas Imagine/QGIS.”*

Comenzando por el desarrollo del algoritmo DBSCAN, se tuvieron diversos problemas, no a la hora de desarrollar el algoritmo, sino en optimizar al máximo el procesamiento ya que, al tratarse de una imagen de dimensiones y resolución muy grandes y teniendo en cuenta que la complejidad en memoria de DBSCAN para conjuntos de datos representados en matrices es de n^2 , y las imágenes escogidas tienen 3 dimensiones (3 bandas, la 12, 8A y 4), es decir, $3n^2$, el algoritmo desarrollado no era capaz de ejecutarse ya que bloqueaba el ordenador en el que estaba alojado. A pesar de todo esto, se consiguió optimizar el algoritmo y hacer que se ejecutase en tiempos de menos de un minuto.

En cuanto a la implementación del algoritmo desarrollado en la herramienta *Erdas Imagine*, se pensaba que esta herramienta permitía la incorporación de código externo para crear un módulo y que este se pudiera utilizar dentro de un modelo creado con la opción *Spatial Modeler* de la propia herramienta. Tras un par de semanas de investigación y de múltiples pruebas, se determinó que *Erdas Imagine* no permite la introducción de módulos con código externo, es decir, solo se pueden desarrollar modelos con los módulos que ya están incorporados en la herramienta. Tras descubrir esto, se decidió hacer uso de la herramienta *QGIS* de código abierto e implementar ahí el algoritmo previamente desarrollado con total éxito, por lo que se da este objetivo por completado.

Objetivo 3: *“Clasificación de las imágenes escogidas para realizar el estudio con DBSCAN, previamente desarrollado e implementado y con KMEANS.”*

A la hora de realizar la clasificación se ha tenido que hacer un estudio de los parámetros que mejor se ajustaban a las características de cada imagen. Esto se ha tenido que hacer tanto con DBSCAN como con KMEANS. Para realizar la clasificación mediante DBSCAN se ha hecho uso del algoritmo desarrollado en *Python* como parte de este trabajo, mientras que para clasificar con KMEANS se ha hecho uso de la herramienta Erdas Imagine.

Objetivo 4: *“Estudio de la evolución de la lava emitida e impacto sobre las zonas más cercanas al volcán (superficie vegetal...) y comparación de las clasificaciones realizadas tanto por DBSCAN como con KMEANS.”*

Una vez se tienen las clasificaciones de cada imagen, se ha realizado un análisis de manera visual de las diferencias y peculiaridades que tienen los resultados con cada algoritmo utilizado.

Objetivo 5: *“Análisis cuantitativo, a partir del plugin SCP, de los resultados de la clasificación realizada tanto con DBSCAN como con KMEANS.”*

Se ha hecho uso del plugin SCP integrado en la aplicación QGIS a partir el cual se han introducido las imágenes clasificadas tanto por KMEANS como con DBSCAN, y como resultado se han obtenido distintas tablas con el número de píxeles por clúster. A partir de estos datos y sabiendo la resolución de cada píxel se ha podido realizar un cálculo de áreas de cada grupo de clústeres para poder realizar un estudio cuantitativo de cambios.

Objetivo 6: *“Evaluación de resultados de la clasificación realizada con el algoritmo de DBSCAN desarrollado y KMEANS en imágenes con una sola banda haciendo uso del coeficiente Silhouette, desarrollado en Python.”*

Se ha desarrollado un pequeño *script* capaz de calcular el valor del coeficiente Silhouette para realizar una evaluación de la clasificación realizada por el algoritmo DBSCAN sobre imágenes con una sola banda, en concreto, la banda 12. La media de resultados obtenidos es de un valor aproximado a 0,7 para las clasificaciones realizadas con DBSCAN, mientras que las realizadas con KMEANS es de 0.8.

Una vez realizada la verificación de ambos algoritmos con el coeficiente Silhouette, se puede llegar a la conclusión de que el algoritmo desarrollado DBSCAN es mejor en aquellos casos en los que la imagen está compuesta por una sola banda y presenta clústeres irregulares. Además, una vez analizadas y verificadas todas las clasificaciones realizadas con ambos algoritmos se puede decir que KMEANS tiende a sobrevalorar o generalizar sus clústeres mientras que DBSCAN tiende a infravalorarlas.

Una vez evaluados todos los objetivos, se han llegado a unas conclusiones generales. En primer lugar, DBSCAN maneja muy bien el ruido y los valores atípicos y puede identificar clústeres con formas irregulares, por lo que parece apropiado para la clasificación de

imágenes satelitales y el seguimiento medioambiental. Sin embargo, computacionalmente es muy costoso, por lo que muestra limitaciones a la hora de clasificar imágenes multibanda. Por otro lado, se ha observado que KMEANS es más propenso a generalizar en sus clasificaciones, es decir, tiende a hacer los clústeres más grandes o genéricos de lo que en realidad son. Esto se puede observar en el escenario de las imágenes del 6 de junio de 2022, en las que KMEANS define el clúster de lava más grande de lo que en realidad es, en cambio si analizamos únicamente la banda 12 con DBSCAN, se observa que el resultado se ajusta más con la imagen original.

7.1. Trabajo futuro

Dado que DBSCAN es muy sensible a los hiperparámetros (*eps* y *min samples*), y un ligero cambio de éstos puede llevar a cambios drásticos en los resultados, en este sentido y para trabajos futuros, cabría profundizar en este tema.

Por otro lado, y dentro del seguimiento medioambiental mediante imágenes de satélite, el algoritmo DBSCAN podría ser utilizado para realizar clasificación de la temperatura de la superficie terrestre obtenida a partir de las bandas 10 y 11 de Landsat-8, lo cual permitiría actuar sobre una única banda y poder así obtener resultados de seguimiento de la actividad volcánica con menor coste computacional. Otra opción podría ser actuar sobre índices extraídos de imágenes de teledetección como el NDVI (*Normal Difference Vegetation Index*) o el NBR (*Normalized Burn Area*).

De forma más general, también se podría realizar una automatización del proceso, es decir, automatizar todos los pasos para llegar al resultado final, desde la obtención de imágenes hasta realizar el análisis cuantitativo mediante SCP. Esto permitiría realizar un seguimiento periódico del volcán Etna y crear alertas según la evolución de la erupción.

También, debido a la complejidad tanto en ejecución como en memoria, sería interesante el desarrollo de un trabajo para la optimización del algoritmo DBSCAN.

Finalmente, se podría extender el trabajo a realizar una validación más exhaustiva de los resultados obtenidos mediante una matriz de confusión y métricas adecuadas.

8. Bibliografía

Clustering. (s. f.). scikit-learn. <https://scikit-learn.org/stable/modules/clustering.html#dbscan>

About Volcanoes | U.S. Geological Survey. (s. f.). <https://www.usgs.gov/programs/VHP/about-volcanoes>

Acharya, J. (2019, 18 marzo). *What is Remote Sensing?* -. <https://www.igismap.com/what-is-remote-sensing/>

Admin, & Admin. (2022). ¿Qué tipos de teledetección existen? *Geomatas*. https://geomatas.com/que-tipos-de-teledeteccion-existen/#Tipos_de_teledetecci%C3%B3n:Instrumentos_de_teledetecci%C3%B3n_activa.

Alexandre. (2022, 29 noviembre). *Machine Learning & Clustering: el algoritmo DBSCAN*. Formation Data Science | DataScientest.com. <https://datascientest.com/es/machine-learning-clustering-dbscan>

Bandas Del Espectro Electromagnético - Espectro Electromagnético. (s. f.). <https://sites.google.com/site/wikiespectroelectromagnetico/bandas-del-espectro-electromagnetico>

Bordino, J. (2022, 30 junio). Tipos de volcanes. *ecologiaverde.com*. <https://www.ecologiaverde.com/tipos-de-volcanes-3584.html>

Borotkanych, N. (2023, 26 abril). ¿Qué Es La Resolución Espacial En Una Imagen Satelital? *EOS Data Analytics*. <https://eos.com/es/blog/resolucion-espacial/>

CNICE. (s. f.). *Resolución de las imágenes satélite*. http://concurso.cnice.mec.es/cnice2006/material121/unidad1/i_resolucion.htm

colaboradores de Wikipedia. (2020). Sentinel (satélite). *Wikipedia, la enciclopedia libre*. [https://es.wikipedia.org/wiki/Sentinel_\(sat%C3%A9lite\)](https://es.wikipedia.org/wiki/Sentinel_(sat%C3%A9lite))

colaboradores de Wikipedia. (2023). Espectro electromagnético. *Wikipedia, la enciclopedia libre*. https://es.wikipedia.org/wiki/Espectro_electromagn%C3%A9tico#Radiofrecuencia

Copernicus: Sentinel-4. (2014, 12 febrero). <https://www.eoportal.org/satellite-missions/copernicus-sentinel-4>

CS221. (s. f.). <https://stanford.edu/~cpiech/cs221/handouts/kmeans.html>

Descubre QGIS. (s. f.). <https://www.qgis.org/es/site/about/index.html>

EO Browser - eo science for society. (2023, 18 enero). eo science for society. <https://eo4society.esa.int/resources/eo-browser/>

ERDAS IMAGINE. (s. f.). Hexagon. <https://hexagon.com/products/erdas-imagine>

Estos son nuestros satélites | Copernicus. (s. f.). <https://www.copernicus.eu/es/sobre-copernicus/infraestructura/estos-son-nuestros-satelites>

Etna. (s. f.). <https://www.volcanodiscovery.com/es/etna.html>

European Space Agency. (2009a, diciembre 9). *ESA - Eduspace ES - Inicio - ¿Qué es la teledetección?* https://www.esa.int/SPECIALS/Eduspace_ES/SEMO1U3FEXF_0.html

- European Space Agency. (2009b, diciembre 10). *ESA - Eduspace ES - Inicio - Interferencia atmosférica*. https://www.esa.int/SPECIALS/Eduspace_ES/SEMPYD3GXF_0.html
- European Space Agency. (2012, 2 julio). *ESA - Eduspace ES - Inicio - Teledetección, a fondo*. https://www.esa.int/SPECIALS/Eduspace_ES/SEMZ3YD3GXF_0.html
- European Space Agency. (2014, 14 enero). *ESA - Eduspace ES - Desastres naturales - Etna*. https://www.esa.int/SPECIALS/Eduspace_Disasters_ES/SEMY3QD1XOG_0.html
- FAQ: What is remote sensing? (2021, 7 septiembre). Ursa Space Systems. <https://ursaspace.com/blog/faq-what-is-remote-sensing/>
- Fig.4. DBSCAN: core, border, and noise points. (s. f.). ResearchGate. https://www.researchgate.net/figure/DBSCAN-core-border-and-noise-points_fig1_258442676
- Franzpc. (2021). Lista de índices espectrales en Sentinel 2 y Landsat. *El blog de franz*. <https://acolita.com/lista-de-indices-espectrales-en-sentinel-2-y-landsat/>
- GeeksforGeeks. (2023). DBSCAN Clustering in ML Density based clustering. *GeeksforGeeks*. <https://www.geeksforgeeks.org/dbscan-clustering-in-ml-density-based-clustering/>
- GISGeography. (2022a). Image Classification Techniques in Remote Sensing [Infographic]. *GIS Geography*. <https://gisgeography.com/image-classification-techniques-remote-sensing/>
- GISGeography. (2022b). Image Classification Techniques in Remote Sensing [Infographic]. *GIS Geography*. <https://gisgeography.com/image-classification-techniques-remote-sensing/#2-supervised-classification>
- González, L. (2021). DBSCAN Teoría. *Aprende IA*. <https://aprendeia.com/dbscan-teoria/>
- Ingeoexpert. (2020a). ¿Qué es la teledetección y qué sistemas existen? *Ingeoexpert*. <https://ingeoexpert.com/2018/07/06/que-es-la-teledeteccion/>
- Ingeoexpert. (2020b, noviembre 13). *Clasificaciones de imágenes de satélite - Ingeoexpert*. <https://ingeoexpert.com/articulo/clasificaciones-de-imagenes-de-satelite/>
- Jauregui, A. F. (2022). DBSCAN en Python: aprende cómo funciona. *Ander Fernández*. <https://anderfernandez.com/blog/dbscan-python/>
- Kogut, P. (2023, 13 febrero). Teledetección Satelital: Tipos, Usos Y Aplicaciones. *EOS Data Analytics*. <https://eos.com/es/blog/teledeteccion/>
- Las misiones Sentinel*. (s. f.). https://www.esa.int/Space_in_Member_States/Spain/Las_misiones_Sentinel
- Living Textbook | Atmospheric absorption and transmission | By ITC, University of Twente*. (s. f.). <https://ltb.itc.utwente.nl/491/concept/79556>
- López, I. G. (2020, 1 septiembre). *Técnicas de Clasificación de Imágenes en la Teledetección*. <https://ecoscript.org/tecnicasclasificacionimagenes/#:~:text=En%20general%2C%20para%20la%20clasificaci%C3%B3n,de%20im%C3%A1genes%20basado%20en%20objetos>
- Low Earth orbit*. (s. f.). https://www.esa.int/ESA_Multimedia/Images/2020/03/Low_Earth_orbit
- Manzanas, J. (2022, 29 abril). Qué tipos de volcanes existen y cómo se clasifican. *okdiario.com*. <https://okdiario.com/curiosidades/tipos-volcanes-909078>

- Mehmood, M., Shahzad, A., Zafar, B., Shabbir, A., & Ali, N. (2022). Remote Sensing Image Classification: A Comprehensive Review and Applications. *Mathematical Problems in Engineering*, 2022, 1-24. <https://doi.org/10.1155/2022/5880959>
- Missions - Sentinel Online - Sentinel Online. (s. f.-a). Sentinel Online. <https://sentinels.copernicus.eu/web/sentinel/missions>
- Missions - Sentinel Online - Sentinel Online. (s. f.-b). Sentinel Online. <https://sentinels.copernicus.eu/web/sentinel/missions>
- Nacional, I. G. (s. f.-a). Plan Nacional de Teledetección. <https://pnt.ign.es/satelites-sentinel#:~:text=Los%20dos%20sat%C3%A9lites%20de%20la,la%20Tierra%20cada%20seis%20d%C3%ADas.>
- Nacional, I. G. (s. f.-b). Volcanología. Geoportal oficial del Instituto Geográfico Nacional de España. <https://www.ign.es/web/vlc-teoria-general>
- Ochoa, F. (2023). Tipos de resolución. *El blog de franz*. <https://acolita.com/tipos-de-resolucion/>
- Park, B. K. (2016). Quality Evaluation of Poultry Carcass. En *Science Direct*. <https://doi.org/10.1016/b978-0-12-802232-0.00009-8>
- Portillo, G. (2021, 15 septiembre). Cómo se forman los volcanes. *Meteorología en Red*. https://www.meteorologiaenred.com/como-se-forman-los-volcanes.html#Como_se_forman_los_volcanes
- Prácticas de Teledetección: Clasificación. (s. f.). [https://www.um.es/geograf/sigmur/teledet/clasific1.html#:~:text=Clasificaci%C3%B3n%20por%20m%C3%ADnima%20distancia&text=Es%20este%20espacio%20de%20variables,sea%20m%C3%ADnima%20\(figura%205\).](https://www.um.es/geograf/sigmur/teledet/clasific1.html#:~:text=Clasificaci%C3%B3n%20por%20m%C3%ADnima%20distancia&text=Es%20este%20espacio%20de%20variables,sea%20m%C3%ADnima%20(figura%205).)
- Rodríguez, H. (2023, 24 febrero). Volcanes: qué son y cómo se forman. *www.nationalgeographic.com.es*. https://www.nationalgeographic.com.es/ciencia/volcanes-que-son-y-como-se-forman_18140
- Ruge.Axessnet. (2023, 18 mayo). Los satélites geoestacionarios ¿Qué son y para qué sirven ? axessnet. <https://axessnet.com/que-son-y-para-que-sirven-los-satelites-geoestacionarios/#:~:text=Los%20primeros%20son%20aquellos%20que,un%20punto%20determinado%20del%20planeta.>
- Sanz, F. (2023). Algoritmo K-Means Clustering – aplicaciones y desventajas. *The Machine Learners*. https://www.themachinelearners.com/k-means/#Algoritmo_K_Means
- Sentinel Hub. (s. f.). <https://www.sentinel-hub.com/>
- Sentinel Online - ESA - Sentinel Online. (s. f.). Sentinel Online. <https://sentinels.copernicus.eu/web/sentinel/home>
- Sharma, A. (2022). How to Master the Popular DBSCAN Clustering Algorithm for Machine Learning. *Analytics Vidhya*. <https://www.analyticsvidhya.com/blog/2020/09/how-dbscan-clustering-works/>
- Singh, S. (2021, 29 diciembre). All you need to know about the DBSCAN Algorithm - Analytics Vidhya - Medium. *Medium*. <https://medium.com/analytics-vidhya/all-you-need-to-know-about-the-dbscan-algorithm-f1a35ed8e712>

Supervised image classification using parallelepiped algorithm. (2019, 26 febrero). 50 North | GIS blog from Ukraine. <http://www.50northspatial.org/supervised-image-classification-using-parallelepiped-algorithm/>

Technologies, R. (2022). ERDAS Imagine: Software para trabajar con imagen de satélite. Geoinnova. <https://geoinnova.org/blog-territorio/erdas-imagine-software-para-trabajar-con-imagen-de-satelite/>

Teledetección | Ministerio de Transportes, Movilidad y Agenda Urbana. (s. f.). <https://www.mitma.gob.es/instituto-geografico-nacional/observacion-del-territorio/teledeteccion>

Tipos de Volcanes y Erupciones | EcoExploratorio: Museo de Ciencias de Puerto Rico. (s. f.). <https://ecoexploratorio.org/amenazas-naturales/volcanes/tipos-de-volcanes-y-erupciones/>

Types of orbits. (s. f.-a). https://www.esa.int/Enabling_Support/Space_Transportation/Types_of_orbits

Types of orbits. (s. f.-b). https://www.esa.int/Enabling_Support/Space_Transportation/Types_of_orbits#MEO

Types of orbits. (s. f.-c). https://www.esa.int/Enabling_Support/Space_Transportation/Types_of_orbits#GEO

User, S. (s. f.). LiDAR. geospatial. <https://www.geospatial.com.co/servicios/tecnologia-lidar.html>

Wikipedia. (2023). Teledetección. *Wikipedia, la enciclopedia libre.* <https://es.wikipedia.org/wiki/Teledetecci%C3%B3n>

Yıldırım, S. (2021, 14 diciembre). DBSCAN Clustering — Explained - Towards Data Science. *Medium.* <https://towardsdatascience.com/dbscan-clustering-explained-97556a2ad556>

Instituto, & Nacional, G. (s/f). Ign.es. Recuperado el 20 de mayo de 2023, de <https://www.ign.es/web/resources/docs/IGNCnig/OBS-Teledeteccion.pdf#:~:text=La%20teledetecci%C3%B3n%20es%20la%20t%C3%A9cnica,informaci%C3%B3n%20interpretable%20de%20la%20Tierra.>

Miller, F. P., Vandome, A. F., & McBrewster, J. (Eds.). (2010). *Global Volcanism Program.* Alphascript Publishing.

(S/f-a). Org.ar. Recuperado el 20 de mayo de 2023, de http://cursosihla.bdh.org.ar/SR_Exactas/2_Teor%C3%ADa/2%20Resoluciones%20en%20Teledetecci%C3%B3n.pdf

(S/f-b). Usgs.gov. Recuperado el 20 de mayo de 2023, de <https://pubs.usgs.gov/gip/7000038/report.pdf>

(S/f-c). Dnr.sc.gov. Recuperado el 20 de mayo de 2023, de <https://www.dnr.sc.gov/geology/pdfs/education/Volcanoes-pg.pdf>

(S/f-d). Mural.uv.es. Recuperado el 20 de mayo de 2023, de <http://mural.uv.es/gislo/telemedida.pdf>

Bhardwaj, A. (2021, 14 diciembre). Silhouette Coefficient - Towards Data Science. *Medium.*

<https://towardsdatascience.com/silhouette-coefficient-validating-clustering-techniques-e976bb81d10c>

sklearn.metrics.silhouette_score. (s. f.). scikit-learn. https://scikit-learn.org/stable/modules/generated/sklearn.metrics.silhouette_score.html#:~:text=The%20Silhouette%20Coefficient%20is%20calculated,is%20not%20a%20part%20of.

9. Anexos

9.1. Anexo A: Código implementado en QGIS

```

1. # qgis library imports
2. import PyQt5.QtCore
3. import qgis.PyQt.QtCore
4. from qgis.core import (QgsProcessing,
5.                         QgsProcessingAlgorithm,
6.                         QgsProcessingException,
7.                         QgsProcessingOutputNumber,
8.                         QgsProcessingParameterDistance,
9.                         QgsProcessingParameterFeatureSource,
10.                        QgsProcessingParameterVectorDestination,
11.                        QgsProcessingParameterRasterDestination,
12.                        QgsProcessingParameterFeatureSink,
13.                        QgsProcessingParameterNumber,
14.                        QgsProcessingParameterEnum,
15.                        QgsProcessingParameterRasterLayer,
16.                        QgsRasterLayer)
17. from qgis.analysis import QgsNativeAlgorithms
18. from qgis.PyQt.QtCore import QApplication
19.
20.
21. import numpy as np
22. from sklearn.cluster import DBSCAN
23. import rasterio
24. from rasterio.windows import Window
25. from matplotlib.colors import ListedColormap
26. import os
27. from osgeo import gdal
28. import rasterio.plot as plot
29. from PIL import Image
30. from matplotlib import cm
31. import matplotlib.pyplot as mplot
32.
33.
34. class DBSCANAlgorithm(QgsProcessingAlgorithm):
35.
36.     def __init__(self):
37.         super().__init__()
38.
39.     def name(self):
40.         return "dbscanalgorithm"
41.
42.     def tr(self, text):
43.         return QApplication.translate("dbscanalgorithm", text)
44.
45.     def displayName(self):
46.         return self.tr("DBSCAN Algorithm by Eduardo Mena")
47.
48.     def group(self):
49.         return self.tr("Vector tools")
50.
51.     def groupId(self):
52.         return "vectortools"
53.
54.     def shortHelpString(self):
55.         return self.tr("Classify a .TIFF image using DBSCAN algorithm")
56.
57.     def helpUrl(self):
58.         return "https://qgis.org"
59.
60.     def createInstance(self):
61.         return type(self)()
62.
63.     INPUT = 'INPUT'
64.     EPS = 'EPS'
65.     MINSAMPLES = 'MINSAMPLES'

```

```

66.     OUTPUT = 'OUTPUT'
67.     IN = 'IN'
68.
69.     def initAlgorithm(self, config=None):
70.         #self.addParameter(QgsProcessingParameterFeatureSource(
71.         #    self.INPUT,
72.         #    self.tr("Input layer"),
73.         #    [QgsProcessing.TypeVectorAnyGeometry]))
74.
75.         self.addParameter(QgsProcessingParameterNumber(
76.             self.EPS,
77.             self.tr('EPS value'),
78.             QgsProcessingParameterNumber.Double,
79.             minValue=0,
80.             maxValue=100,
81.             defaultValue=1.5))
82.
83.         self.addParameter(QgsProcessingParameterNumber(
84.             self.MINSAMPLES,
85.             self.tr('min samples value'),
86.             QgsProcessingParameterNumber.Double,
87.             minValue=0,
88.             maxValue=500,
89.             defaultValue=15))
90.
91.         self.addParameter(QgsProcessingParameterFeatureSink(
92.             self.OUTPUT,
93.             self.tr("Output layer"),
94.             QgsProcessing.TypeVectorAnyGeometry))
95.
96.         self.addParameter(
97.             QgsProcessingParameterRasterLayer(
98.                 self.IN,
99.                 self.tr('Input Image'),
100.                 defaultValue=None
101.             )
102.         )
103.
104.     def processAlgorithm(self, parameters, context, feedback):
105.
106.         eps = self.parameterAsDouble(parameters, self.EPS, context)
107.         min_samples = int(self.parameterAsDouble(parameters, self.MINSAMPLES,
108. context))
109.
110.         input_image = self.parameterAsRasterLayer(parameters, self.IN, context)
111.         # layer = QgsRasterLayer(input_image, "layer")
112.
113.         #(sink, dest_id) = self.parameterAsSink(parameters, self.OUTPUT, context,
114.         #    source.fields(), source.wkbType(),
115.         source.sourceCrs())
116.
117.         #output = self.parameterAsSink(parameters, self.OUTPUT, context,
118.         input_image.fields(), input_image.wkbType(), input_image.sourceCrs())
119.         output = self.parameterAsOutputLayer(parameters, self.OUTPUT, context)
120.
121.         # Abrir la imagen TIFF
122.         with rasterio.open(input_image.source()) as src:
123.             profile = src.profile
124.             height, width = src.height, src.width
125.             block_size = max(height, width) + 1 # tamaño de bloque en píxeles
126.
127.             # Crear un archivo TIFF para la imagen segmentada
128.             profile.update(count=1, dtype=rasterio.uint8)
129.             output_path = self.parameterAsFileOutput(parameters, self.OUTPUT,
130. context)
131.             with rasterio.open(output_path, 'w', **profile) as dst:

```

```

132.         for y in range(0, height, block_size):
133.             for x in range(0, width, block_size):
134.
135.                 # Leer el bloque de la imagen
136.                 window = Window(x, y, min(block_size, width - x),
min(block_size, height - y))
137.                 data = src.read(window=window, masked=True)
138.                 data = np.moveaxis(data, 0, -1)
139.
140.                 # Aplicar DBSCAN al bloque
141.                 dbscan = DBSCAN(eps=eps, min_samples=min_samples)
142.                 labels = dbscan.fit_predict(data.reshape((-1, data.shape[-
1])))
143.                 final = dbscan.labels_
144.                 final = final.reshape(src.shape)
145.
146.                 # Escribir el bloque etiquetado en la imagen segmentada
147.                 dst.write(final.astype(rasterio.uint8), window=window,
indexes=1)
148.
149.             # Guardar la imagen segmentada
150.             with rasterio.open(output_path) as segmented_image:
151.                 final = segmented_image.read(1)
152.                 mplot.figure(figsize=(10, 10))
153.                 mplot.imshow(final)
154.                 mplot.savefig(output, bbox_inches='tight', format='tiff', dpi=1600)
155.
156.
161.         return {self.OUTPUT: output_path}
162.

```