



Universidad
del País Vasco

Euskal Herriko
Unibertsitatea

Unibertsitate Masterra
Konputazio Ingeniaritza eta Sistema Adimentsuak

Konputazio Zientziak eta Adimen Artifiziala Saila –
Departamento de Ciencias de la Computación e Inteligencia Artificial

Master Tesia

Hacia una monitorización en continua basada en
teledetección multi-temporal de infraestructuras
para el transporte de energía

*Towards NDVI-based continuous monitoring for
energy transport infrastructure maintenance*

Itziar Altuna Ruiz de Eguino

Tutorea(k)

Marco Quartulli

Vicomtech

Carmen Hernández

Konputazio Zientziak eta Adimen Artifiziala saila
Informatika Fakultatea

MDe

Master eta Doktorego Eskola
Escuela de Máster y Doctorado
Master and Doctoral School

2018ko irailan

MDe
Master eta Doktorego Eskola
Escuela de Máster y Doctorado
Master and Doctoral School

Hacia una monitorización en continua basada en
teledetección multi-temporal de infraestructuras para el
transporte de energía

*Towards NDVI-based continuous monitoring for energy transport
infrastructure maintenance*

Itziar Altuna Ruiz de Eguino

Supervisores:

Marco Quartulli
Carmen Hernández

September 27, 2018

Resumen

Este proyecto propone un acercamiento a la monitorización de zonas de interés mediante el análisis de series temporales. En concreto, se ha seleccionado un conjunto de cables de alta tensión del área de Donostia con el fin de mostrar y advertir de las líneas eléctricas que se encuentran en áreas boscosas o de alta vegetación. Para ello, se utilizan tecnologías como *Google Earth Engine* y *QuantumGIS* que son combinadas para gestionar los datos que se adquieren de ambas fuentes. También se ha realizado un estudio previo de las imágenes Sentinel-2, los diferentes tipos de nivel y el preproceso de esta clase de imágenes así como otros conceptos básicos utilizados en teledetección. Finalmente, se han analizado algunos de los índices de vegetación más populares y se han utilizado dos de ellos para realizar la experimentación así como el análisis comparativo de los resultados obtenidos con cada uno de los índices.

Agradecimientos

En primer lugar me gustaría agradecer a mis dos supervisores, Marco y Mamen, por haberme ayudado y por haber mostrado disponibilidad siempre que he tenido alguna duda o problema a lo largo de todo el proyecto.

Por otro lado, agradecer a Vicomtech la oportunidad de conocer desde dentro el funcionamiento de un Centro de Investigación y el trabajo en grupo que se hace.

Finalmente, a mis amigos y compañeros de clase, por haber compartido risas, agobios y cafés; y a mi familia por haberme dado la oportunidad de estudiar todo lo que quiero y apoyarme en todo lo que hago.

Índice

Lista de Figuras	6
Lista de Tablas	8
1 Introducción	9
1.1 Objetivos	11
1.2 Marco del proyecto de Máster	12
1.3 Elementos metodológicos y algorítmicos	12
1.4 Tipos de datos y tecnologías utilizadas	15
1.4.1 Imágenes Sentinel-2	15
1.4.2 Herramientas y estructuras de datos utilizadas	18
2 Metodología	24
2.1 Imágenes del satélite Sentinel-2	27

<i>ÍNDICE</i>	4
2.2 Índices de vegetación	29
2.3 Análisis de conglomerados	32
3 Experimentación y discusión de los resultados	33
3.1 Selección del área de estudio y descripción temporal	34
3.2 Selección y segmentación de las líneas de alta tensión	35
3.3 Obtención y preprocesado de las imágenes del satélite Sentinel-2	39
3.4 Análisis de conglomerados	39
3.4.1 Clasificación de los segmentos de las líneas de alta tensión	45
3.5 Clasificación y visualización de alarmas	51
3.6 Discusión de los resultados	55
3.7 Otros experimentos	57
4 Conclusiones y líneas futuras	60
References	62
Anexos	68
A Transformación radiométrica para imágenes Sentinel-2	69
B Preprocesado de las imágenes Sentinel-1	72

<i>ÍNDICE</i>	5
C Instalación de librería Python Earth Engine	75
D Relación estructura de datos GEE y QGIS	78
E Código	79

List of Figures

1.1	Bandas de las imágenes del satélite Sentinel-2 vs. las del satélite Landsat.	16
1.2	Consola en línea <i>Google Earth Engine</i> (GEE).	19
2.1	Estructura de trabajo.	25
2.2	Mosaico de las imágenes Sentinel-2 de nivel L1C.	28
3.1	Área exacta de estudio.	35
3.2	Cableado eléctrico de la zona del País Vasco.	36
3.3	Polilínea con un <i>offset</i> de 10 m.	37
3.4	Dos segmentos de una polilínea.	38
3.5	Cable de alta tensión subdividido en 5 segmentos.	38
3.6	Conjunto de conglomerados para la detección de zonas de alta vegetación.	40
3.7	Valores del índice NDVI en las diversas zonas que forman el conjunto de entrenamiento a lo largo del tiempo.	43

<i>LIST OF FIGURES</i>	7
3.8 Imágenes de satélite Sentinel-2 tomadas el 13 y 28 de julio respectivamente.	44
3.9 Valores de la serie temporal del índice de vegetación NDVI para un único cable de alta tensión formado por 5 segmentos: A, B, C, D, y E.	46
3.10 Cables de alta tensión del área de Donostia.	48
3.11 Valores del índice NDVI para las zonas de estudio que contienen cables de alta tensión en el área de Donostia.	49
3.12 Cuantificación de los cambios observados para imágenes captadas en fechas repetidas. La diferencia entre los valores del índice NDVI para bosque/césped se representa en rojo.	50
3.13 Mapa de los cables que se encuentran en una zona de mucha vegetación clasificada como peligrosa.	55
3.14 Correlación NDVI y GNDVI.	58
B.1 Adquisición TOPSAR	73

List of Tables

1.1	Bandas espectrales para imágenes de Sentinel-2.	17
3.1	Coordenadas del conjunto de conglomerados para la detección de zonas de alta vegetación.	41
3.2	Longitudes de los 11 cables eléctricos del área de Donostia.	42
3.3	Valores del índice de vegetación NDVI para las distintas imágenes y las distintas zonas del conjunto de entrenamiento a lo largo del tiempo.	45
3.4	Valores del índice de vegetación NDVI para las distintas imágenes y las distintas zonas del conjunto de entrenamiento a lo largo del tiempo.	46
3.5	Tabla de distancias	52
3.6	Reclasificación de cada segmento de los cables de alta tensión según el valor mínimo de la tabla de distancias.	54
3.7	Matriz de confusión usando índice NDVI	56
3.8	Matriz de confusión usando índice GNDVI	59

Chapter 1

Introducción

La gestión de proyectos suele enfocar los temas de gestión en tres temas fundamentales: la calidad del producto, los tiempos necesarios para la producción del mismo, y el coste [2]. Por ejemplo, en principio es posible obtener una calidad muy alta del producto a cambio de un aumento en el tiempo necesario para su desarrollo o a un alto coste.

El sector de la teledetección habitualmente se ha puesto al servicio de aplicaciones de seguridad y defensa. Por esta razón, los temas de calidad han dominado sobre los temas de coste y tiempo. Este enfoque tradicional está cambiando a lo largo de los últimos años con el crecimiento significativo de aplicaciones civiles. Los costes y los tiempos ya no son variables condicionadas por la alta calidad del producto a obtener. Se empieza a considerar que datos de calidad más baja pero obtenidos dinámicamente con costes limitados pueden ser más interesantes que productos de muy alta resolución y calidad recogidos anteriormente por sistemas extremadamente más caros. Este cambio de enfoque está siendo tan generalizado que ya está llegando hasta a la comunidad especializada seguridad y defensa [1].

A nivel de sistemas, la oportunidad de trabajar con datos más actuales y frecuentes se

relaciona con la posibilidad de utilizar plataformas más flexibles para la obtención de datos como, por ejemplo, constelaciones de satélites de bajo coste.

En este sentido, un ejemplo claro es el caso Planet [6] — UAVs para la observación de la Tierra y aeronaves no tripuladas del tipo de dirigibles — ya que esta empresa tiene como prioridad la obtención de datos actuales más que la calidad de los mismos.

A nivel computacional, la opción de compaginar y fusionar diferentes fuentes de datos se va orientando, por consiguiente, en la posibilidad de explotar series temporales de imágenes. Las series temporales son secuencias de datos observados en determinados instantes de tiempo y ordenados cronológicamente. Se utilizan, principalmente, para estudiar la relación de diferentes variables a través del tiempo y observar la influencia que tienen ciertas variables sobre otras. En nuestro caso, el contenido informativo de series temporales de imágenes de observación de la Tierra es significativamente más alto respecto al contenido informativo de imágenes individuales. En las secuencias de imágenes podemos observar no solo las clases de cobertura del suelo, sino incluso el desarrollo de fenómenos. Por otra parte, también se pueden clasificar diferentes zonas viendo el desarrollo de las mismas. Mientras que el análisis de imágenes individuales era capaz de, por ejemplo, identificar zonas del mundo que tuviesen un tipo de cobertura parecido, un análisis enfocado a series temporales y al desarrollo de fenómenos permite identificar de una forma similar zonas del planeta que están desarrollándose de forma semejante o paralela. Asimismo, es posible identificar fenómenos de desertificación parecidos en diferentes zonas del Mediterráneo. Otro ejemplo típico es el análisis de la dinámica de urbanización y del cambio de uso del suelo así como el desarrollo de entornos urbanos y periurbanos [33].

A nivel metodológico, todo esto conlleva una transición entre la teledetección basada en el análisis de imagen a la teledetección basada en análisis de series temporales de imágenes e, incluso, de vídeos.

1.1 Objetivos

El principal objetivo de este estudio consiste en determinar si los cables eléctricos de alta tensión de la zona de Donostia-San Sebastián (Guipúzcoa) se encuentran en una zona poblada de vegetación y, de ser así, que sirva de indicador a las empresas eléctricas para saber cuándo y dónde deben realizarse podas.

El proyecto nació con motivo de los incendios forestales ocurridos en el año 2017 en las zonas de Galicia y Portugal. Se descubrió que la principal causa de dichos incendios fue la saturación de los cortafuegos que causaron sobrecalentamientos en las líneas. Por ese motivo, dado que la forma más eficaz de evitar dichas catástrofes reside en la prevención, se van a utilizar imágenes del satélite Sentinel-2, que proporcionan datos de libre acceso que permiten realizar una monotorización de la Tierra.

Otro de los objetivos de este proyecto consiste en familiarizarse y descubrir la amplia gama de posibilidades que ofrece la plataforma *Google Earth Engine*. Esta plataforma almacena datos e imágenes de satélites - como Landstat, MODIS o Sentinel entre otros - que pueden, posteriormente, ser utilizadas por otras herramientas como sistemas de información geográfica (*Geographic Information System, GIS*).

Finalmente, y relacionado con el punto anterior, también se busca tener un conocimiento de los conceptos y herramientas más utilizadas y comunes en el campo de la teledetección. Por un lado, se planteó un estudio sobre los diferentes Índices de Vegetación para conocer las características propias y saber en qué situaciones se pueden utilizar cada uno de ellos. Por otro lado, se considera que adquirir cierta soltura a la hora de utilizar un sistema GIS como *Quantum GIS (QGIS)* es fundamental para llevar a cabo este proyecto ya que la selección de las zonas de interés se hará mediante esta herramienta.

1.2 Marco del proyecto de Máster

Este proyecto ha sido desarrollado en el centro de investigación Vicomtech, situado en el Parque Tecnológico de San Sebastián y fundado en 2001. Vicomtech pertenece a Graphicsmedia.net, una alianza estratégica que está formada por una red internacional de centros de investigación aplicada alineados con las tecnologías de *Computer Graphics* y multimedia tales como Computación Gráfica, Computación Visual y Tecnologías Multimedia, y tiene como objetivo responder a las necesidades de innovación de las empresas e instituciones. Para ello:

1. Realiza investigación aplicada y desarrolla tecnologías multimedia de interacción visual y comunicaciones.
2. Colabora estrechamente con la industria, la Universidad y otros centros tecnológicos, a quienes complementa.
3. Fomenta la movilidad y formación de sus investigadores.

Vicomtech comenzó su actividad en abril de 2001 impulsado por una red de centros tecnológicos alemanes y las Administraciones locales. Esta red de centros especializados en tecnologías de computación visual manifestó su inquietud por crear un centro similar fuera de Alemania. El Gobierno Vasco apoyó esta iniciativa junto con las Administraciones locales y, de este modo, se fundó Vicomtech en el Parque Tecnológico de San Sebastián.

1.3 Elementos metodológicos y algorítmicos

Las áreas de interés prioritario para la teledetección son, inevitablemente, aquellas con mayor grado de antropización en las que son más evidentes los efectos de las manipulaciones e influencias de la sociedad, como las áreas urbanas o las zonas de interés para las infraestructuras esenciales como las que están dedicadas al transporte de energía, agua y

telecomunicaciones.

Por otro lado, esas áreas son también las que la fenomenología observable en escenas y, por consecuencia, en imágenes y vídeos adquiridos son más complejas y relacionadas con geometrías más variadas: muros verticales, mayor presencia de oclusiones, geometrías solapadas, etc. Estos factores complican de forma evidente el estudio de la zona y la reconstrucción de las escenas con todas sus características a partir de los datos adquiridos [38].

Un primer elemento algorítmico a tener en cuenta a la hora de generar series temporales es la necesidad de preprocesar las imágenes de la secuencia para que estén calibradas y para que los efectos de ruido sean limitados y controlados [35].

Un segundo aspecto es la extracción de información desde los fotogramas individuales. Por ejemplo, en cuanto a la estimación del contenido de textura de los mismos, permite analizar y distinguir de forma eficaz diferentes tipos de vegetación en las que pueden localizarse áreas de bosque de diferentes especies [36].

Un tercer aspecto se refiere a la extracción de características primarias de las series temporales que permitan tratar ventanas de análisis de las mismas como instancias independientes en sistemas de *machine learning* o de búsqueda por contenido.

Existe una amplia gama de metodologías eficaces disponibles en la literatura dedicada al análisis de series temporales multivariadas en términos estadísticos y del procesamiento de la señal [9, 15, 27] y en ámbitos de aplicación que van desde el financiero (por ejemplo con el uso de Análisis de Componentes Independientes y *Support Vector Regressors* [23]) hasta investigaciones como la segmentación de áreas cerebrales utilizando algoritmos de *clustering* de señales de electroencefalogramas (como por ejemplo, la técnica matemática

Common spatial patterns, CSP). Específicamente significativa es, en este sentido, la amplia disponibilidad de librerías de software con alto nivel de madurez tecnológica, *Technology Readiness Level*, que implementan métodos de cálculo de distancias entre series temporales en diferentes condiciones de alineación, muestreo y periodicidad (por ejemplo, *time warping distances* [31]) lo que resulta particularmente importante en el caso de la teledetección electro-óptica por su dependencia en las condiciones de nubosidad y iluminación respecto al caso del escaneo mediante radar.

Los enfoques multi-sensor centrados en patrones deterministas y conjuntos de elementos frecuentes (*frequent itemsets*), al nivel de clases supervisadas o no supervisadas, representan una posibilidad de generar mapas multi-temporales de forma rápida y eficaz [30, 21]. La presencia de altos niveles de incertidumbre relacionados con los propios datos (ruido, datos no disponibles o *missing data*, e incertidumbre en los parámetros de adquisición) y con algunas propiedades observadas solo parcialmente en la escena, implica la posibilidad de explotar métodos probabilísticos para la clasificación, el *clustering* y la búsqueda de datos. Muchos de los métodos desarrollados desde los finales de los años 90 y basados en enfoques bayesianos se encuentran ya implementados en grandes sistemas de mapeo rápido conectados con los catálogos de imágenes de diferentes productores de datos [11]. La frontera de los desarrollos se encuentra en este caso en la integración de metodologías de aprendizaje profundo (*deep learning*) en sistemas de mapeo extendido y de consulta de imágenes por contenido [39].

Un problema concreto de los sistemas supervisados, por ejemplo para la clasificación y sucesiva vectorización de objetos en las escenas, es la necesidad de grandes conjuntos de datos etiquetados para su entrenamiento. Sistemas de etiquetado masivo [5] y de aprendizaje por transferencia (*transfer learning*) — es decir, capaces de aprovechar parte del material etiquetado para aplicaciones diferentes pero relacionadas [25, 12] — son algunas de las técnicas mas utilizadas para mitigar estos aspectos difíciles del entrenamiento.

En cuanto a trabajos relacionados con el problema que se plantea en este trabajo podemos

citar el trabajo de Matikainen [28] que, en su artículo, trata de hacer una amplia revisión bibliográfica sobre las posibilidades y limitaciones que aportan los sensores de detección remota a la hora de realizar una supervisión de los corredores de líneas eléctricas tanto a nivel de los componentes mismos de la línea así como de la vegetación circundante a las mismas. También puede encontrarse una detallada descripción de los aspectos más importantes de la monitorización de líneas de transmisión en el segundo capítulo del libro de Hu y Liu [19]. Por último, otros trabajos relacionados y realizados con distintos tipos de tecnologías tales como imágenes aéreas, datos LIDAR, imágenes de satélites de alta definición, etc. pueden encontrarse en [41, 29, 22, 32, 20, 40, 42].

1.4 Tipos de datos y tecnologías utilizadas

Como se ha mencionado en el apartado de los objetivos del proyecto, hemos trabajado con diferentes tipos de datos tales como series temporales de imágenes de satélite Sentinel-2 así como datos vectoriales que incluyen información de las líneas de alta tensión. Además, se han utilizado varias tecnologías en el tratamiento de dichas series temporales de imágenes: *Google Earth Engine* y el sistema de información geográfica *Quantum GIS*. Por ello, en esta sección realizamos una breve descripción de las imágenes Sentinel-2 así como de dichas herramientas. También presentamos algunos aspectos sobre las estructuras de datos utilizadas en ambas tecnologías así como el flujo de datos entre ambas. De igual forma, incluimos en esta sección algunas definiciones básicas pero imprescindibles para establecer una terminología apropiada.

1.4.1 Imágenes Sentinel-2

Sentinel-2 es una misión de observación terrestre que fue desarrollada por la agencia espacial europea ESA (*European Space Agency*) y que se encuentra dentro del programa Copérnico (<http://www.copernicus.eu/>). El principal motivo por el que se creó la misión Sentinel-2 es para tener un seguimiento de la evolución de los bosques, los cambios en la corteza

terrestre y la gestión de los desastres naturales mediante la observación de la Tierra. Está compuesto por dos satélites idénticos colocados en la misma órbita, en fase a 180° entre sí. Sentinel-2A fue lanzado el 23 de junio de 2015 y Sentinel-2B fue lanzado el 7 de marzo de 2017.

Las principales características de Sentinel-2 son las siguientes:

- **Imágenes multispectrales:** Las imágenes multispectrales son aquellas que capturan datos de imágenes dentro de rangos de longitud de onda específicos a través del espectro electromagnético, y dividen la luz en un cierto número reducido de bandas. Las imágenes multispectrales de Sentinel-2 constan de 13 bandas en el espectro visible, aquel que el ojo humano puede percibir, en el infrarrojo cercano e infrarrojos de onda corta además del espectro electromagnético (Figura 1.1 y Tabla 1.1).

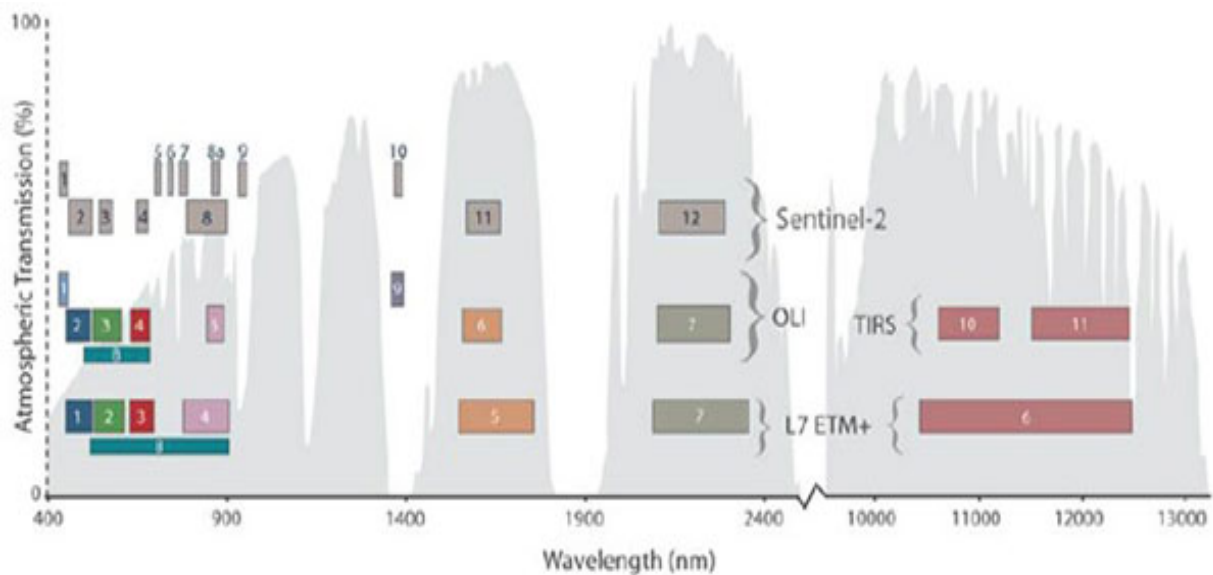


Figure 1.1: Bandas de las imágenes del satélite Sentinel-2 vs. las del satélite Landsat.

Banda	Resolución(m)	Longitud de onda (nm)	Ancho de banda (nm)	Utilización
B01	60	443	20	Detección aerosoles
B02	10	490	65	Azul
B03	10	560	35	Verde
B04	10	665	30	Rojo
B05	20	705	15	Clasif. vegetación
B06	20	740	15	Clasif. vegetación
B07	20	783	20	Clasif. vegetación
B08	10	842	115	Infrarrojo cercano (NIR)
B08A	20	865	20	Clasif. vegetación (Narrow NIR)
B09	60	945	20	Vapor de agua
B10	60	1375	30	Cirro (SWIR)
B11	20	1610	90	Nieve/Hielo/Nubes (SWIR)
B12	20	2190	180	Nieve/Hielo/Nubes (SWIR)

Table 1.1: Bandas espectrales para imágenes de Sentinel-2.

- **Cobertura global:** Sentinel-2 realiza una cobertura global sistemática de las capas de tierra de S56° a N84°, aguas costeras, y cubre todo el mar Mediterráneo.
- **Revisión periódica:** Las zonas de poca altitud se revisan cada 5 días manteniendo los mismos ángulos de visión. Las zonas de altitud más alta también se revisan cada 5 días pero con ángulos de visión diferentes.
- **Resolución espacial:** Las imágenes de Sentinel-2 cuentan con resoluciones espaciales de 10, 20, y 60 metros.
- **Campo de visión:** El campo de visión alcanza hasta 290 km.

Gracias a su cámara multiespectral, algunas de las aplicaciones de las imágenes Sentinel-2 son la vigilancia del terreno y aguas costera e interiores, y el mapeo de la cubierta terrestre. También hay que destacar el uso de imágenes Sentinel-2 para la Misión Óptica de Alta Resolución de la ESA para los Servicios Operativos de GMES, *Global Monitoring for Environment and Security*, ahora llamado programa Copérnico [13], donde los objetivos principales de la misión eran:

- Proporcionar adquisiciones globales sistemáticas de imágenes multiespectrales de alta resolución con una alta frecuencia de revisitas.
- Proporcionar una continuidad mejorada de imágenes multiespectrales proporcionadas por la serie de satélites SPOT, *Satellite Pour l'Observation de la Terre*.
- Proporcionar observaciones para la próxima generación de productos operacionales, como mapas de cobertura terrestre, mapas de detección de cambios en la tierra y variables geofísicas.

1.4.2 Herramientas y estructuras de datos utilizadas

Google Earth Engine

A finales del año 2010 y coincidiendo con la Conferencia sobre Cambio Climático que se celebró en Cancún (México), Google anunció el lanzamiento de su nuevo proyecto *Google Earth Engine* (GEE) [17]. Esta plataforma pone a disposición del usuario una gran cantidad de imágenes satelitales, tanto actuales como históricas, así como la información completa del satélite Landsat, una serie de satélites creados y puestos en órbita por Estados Unidos. Este proyecto fue creado con la intención de monitorizar y medir los cambios del medio ambiente así como dirigir recursos para actuar ante desastres futuros y evitar degradaciones forestales. Entre sus diversas aplicaciones destacan por un lado, la posibilidad de ofrecer a los científicos su estructura computacional (Google) para analizar todas la imágenes y, por otro lado, puede utilizarse para la medición, reporte y verificación (MRV)¹ las iniciativas que buscan detener la deforestación en todo el planeta.

Como se ha mencionado, GEE es proveedor de imágenes Landsat pero también tiene acceso a otros conjuntos de imágenes ráster, fotografías aéreas digitales, imágenes de satélite, imágenes digitales e incluso mapas escaneados - como por ejemplo, imágenes de los satélites Sentinel o MODIS - o imágenes de un metro de resolución proporcionadas por el Programa

¹Las siglas MRV se refieren al protocolo *Monitoring, Reporting, Verification*.

Nacional de Imágenes de Agricultura (*National Agriculture Imagery Program*, NAIP) de los Estados Unidos. GEE también ofrece la posibilidad de "subir" información ráster o vectorial propia y compartir los *scripts* diseñados.

Para utilizar los recursos de GEE es necesario crear una cuenta. Una vez dados de alta, se obtiene acceso a la consola en línea donde se pueden crear proyectos, aprender a utilizar la herramienta estudiando los ejemplos que ofrece y tener acceso a la documentación de todos los métodos que *Earth Engine* ofrece para tratar elementos de la superficie como objetos. El lenguaje de programación que se debe utilizar en la consola en línea es JavaScript aunque existe la interfaz de programación de aplicaciones API (*Application Programming Interface*) de Python para poder trabajar en este lenguaje desde un sistema GIS.

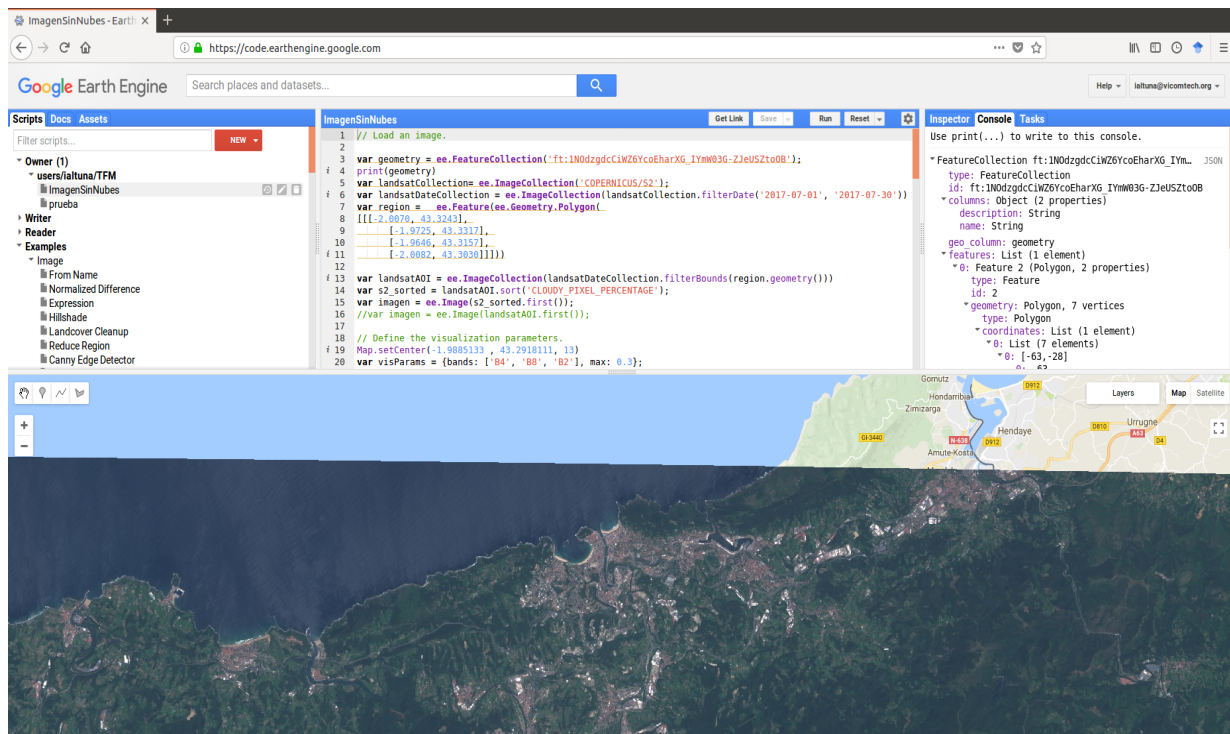


Figure 1.2: Consola en línea *Google Earth Engine* (GEE).

En este trabajo se ha utilizado la API de Python y el sistema operativo Ubuntu 16.04. Para poder instalar la librería *Python Earth Engine*, se han seguido los pasos de instalación que se encuentran en la página de desarrolladores de Google https://developers.google.com/earth-engine/python_install. Existen varias formas de realizar la instalación, pero en este caso se ha elegido la que se llama *Minimal Earth Engine Python API Installation*. Los pasos de instalación se explicarán más detenidamente en el Anexo C.

Quantum GIS

Un sistema GIS almacena y manipula información acerca del mundo como una colección de capas temáticas que pueden relacionarse entre sí. La información geográfica contiene una referencia geográfica explícita, como la latitud y la longitud, o una referencia implícita, como el domicilio y código postal, el nombre del área censal, un bosque o el nombre de una calle. Las referencias implícitas pueden derivarse de referencias explícitas utilizando un proceso automatizado llamado "geocodificación".

Las fases de un sistema GIS son las siguientes:

1. Entrada de la información en el sistema, ya sea digital o pendiente de digitalización.
2. Almacenamiento y actualización de las bases de datos geográficas, es decir, georreferenciación de la información mediante coordenadas geográficas de latitud y longitud.
3. Análisis e interpretación de los datos georreferenciados.
4. Salida de la información en forma de productos diferentes, que dependen de las necesidades del usuario.

Actualmente existen muchos sistemas de información geográfica pero el que se ha utilizado en este proyecto es *Quantum GIS* (QGIS). QGIS es un sistema de información geográfica de código libre para plataformas GNU/Linux, Unix, Mac OS y Microsoft Windows que

proporciona un amplio conjunto de herramientas para unir, almacenar y representar datos espaciales del mundo que nos rodea.

Entre sus aplicaciones, destaca su uso para estudios científicos, la arqueología, la evolución del impacto ambiental o la cartografía, entre otras. QGIS permite manejar formatos de datos ráster y vectoriales a través de bibliotecas y acceder a un gran número de bases de datos.

Estos son algunos de sus puntos "fuertes" y donde radican algunas de sus grandes potencialidades. Gracias a su flexible interfaz, permite trabajar de manera similar en cualquiera de los sistemas operativos que se han mencionado.

Otra de las ventajas de QGIS es la cantidad de complementos que se pueden instalar, muchos de ellos imprescindibles, como es el caso de MMQGIS, que es un conjunto de complementos de Python que permite manipular las capas vectoriales en QGIS y realizar efectos de animación. Por último, se debe destacar como una de sus mayores ventajas la posibilidad de integrar GRASS (*Geographic Resources Analysis Support System*) en QGIS, utilizando toda la potencialidad del análisis de GRASS en un entorno de trabajo más sencillo. GRASS es un software GIS bajo licencia GNU *General Public License* (GPL) que soporta información tanto ráster como vectorial y cuenta con herramientas de procesado digital de imágenes. Por lo tanto, la combinación de ambas herramientas es esencial para el tratamiento digital de imágenes, así como para el análisis y edición de datos vectoriales. El complemento de GRASS en *Quantum GIS* proporciona acceso a las bases de datos y funcionalidades GIS de GRASS, como la digitalización de capas vectoriales, la creación de nuevas capas vectoriales o la edición de datos GRASS en 2D y 3D.

Datos espaciales en QGIS y en *Google Earth Engine*

Una vez obtenidas e importadas las localizaciones de los cables de alta tensión en QGIS, como se explicará mas adelante en la sección 3.2, podemos empezar a extraer la información que deseamos. Para ello, tenemos que trabajar con distintas estructuras de datos de QGIS tales como:

- *qgis.utils iface.activeLayer()*: Este método de la interfaz *QgisInterface*, es el primer paso que se da para obtener un puntero a la capa activa sobre la que se trabaja. Una vez hecho esto, se puede acceder a todos los elementos que contenga esta capa.
- *selectedFeatures*: Todas las características seleccionadas por el usuario o *QgsFeature* se almacenan en una lista denominada *QgsFeatureList*.
 - *QgsFeature*: Encapsula una *feature* incluyendo su identificador, su geometría y una lista de sus atributos.
 - *QgsFeatureList*: Una lista características de tipo *QgsFeature*.

En *Google Earth Engine* un mapa(*ee.Map*) almacena el objeto que queremos visualizar (imagen, características o geometrías). Una imagen (*ee.Image*) se utiliza para mostrar un mapa y puede incluir varias bandas. Para organizar las diferentes imágenes podemos utilizar un conjunto de imágenes (*ee.ImageCollection*) al que puede aplicarse, por ejemplo, un filtrado del conjunto por alguno de los metadatos (fecha, *path&row*, ...) o una función sobre cada imagen independientemente (índice de vegetación, filtrado, ...).

En *Google Earth Engine* los datos vectoriales son de tipo Geometría (*ee.Geometry*) que incluye los siguientes subtipos:

- *Point*: Un punto es una lista de coordenadas (x, y, z) en la misma proyección.
- *LineString*: Una línea es una lista de puntos.
- *LinearRing*: Un anillo es una línea cerrada.
- *Polygon*: Un polígono es un área cerrada que representa la forma y la ubicación de entidades homogéneas.

GEE también incluye los formatos multipunto, multilínea y multipolígono²

²En un multipolígono, los polígonos son parte de la misma estructura y tienen la misma proyección y polaridad.

Después de haber visto las estructuras de datos de ambas tecnologías, existe la necesidad de combinarlas ya que, por un lado tenemos, las localizaciones de los cables de alta tensión en la estructura de datos de QGIS, mientras que, por otro lado, la colección de imágenes se obtiene gracias a GEE. Esta relación se consigue mediante la geometría extraída de cada característica *QgsFeature* con el método *exportToGeoJSON()*, que devuelve el tipo de geometría (*LineString*, *Polygon*...) y las coordenadas de ésta. Una vez hecho esto, se crea un nuevo polígono que cumpla con la estructura de datos GEE, con las coordenadas extraídas de dicha característica *QgsFeature* y ya cumple los requisitos para que se pueda usar como condición a la hora de filtrar las imágenes de Sentinel-2 y crear la colección.

Para ver cómo se ha trabajado con ambas tecnologías y cómo se han relacionado la estructura de datos GEE y QGIS, se ha realizado una pequeña explicación en el Anexo [D](#). El código completo queda recogido en el Anexo [E](#).

Chapter 2

Metodología

A fin de seguir unas pautas ordenadas, hemos dividido el proyecto en pequeñas tareas que nos han facilitado la obtención del resultado final deseado. La Figura [2.1](#) muestra claramente cuáles han sido dichos pasos.

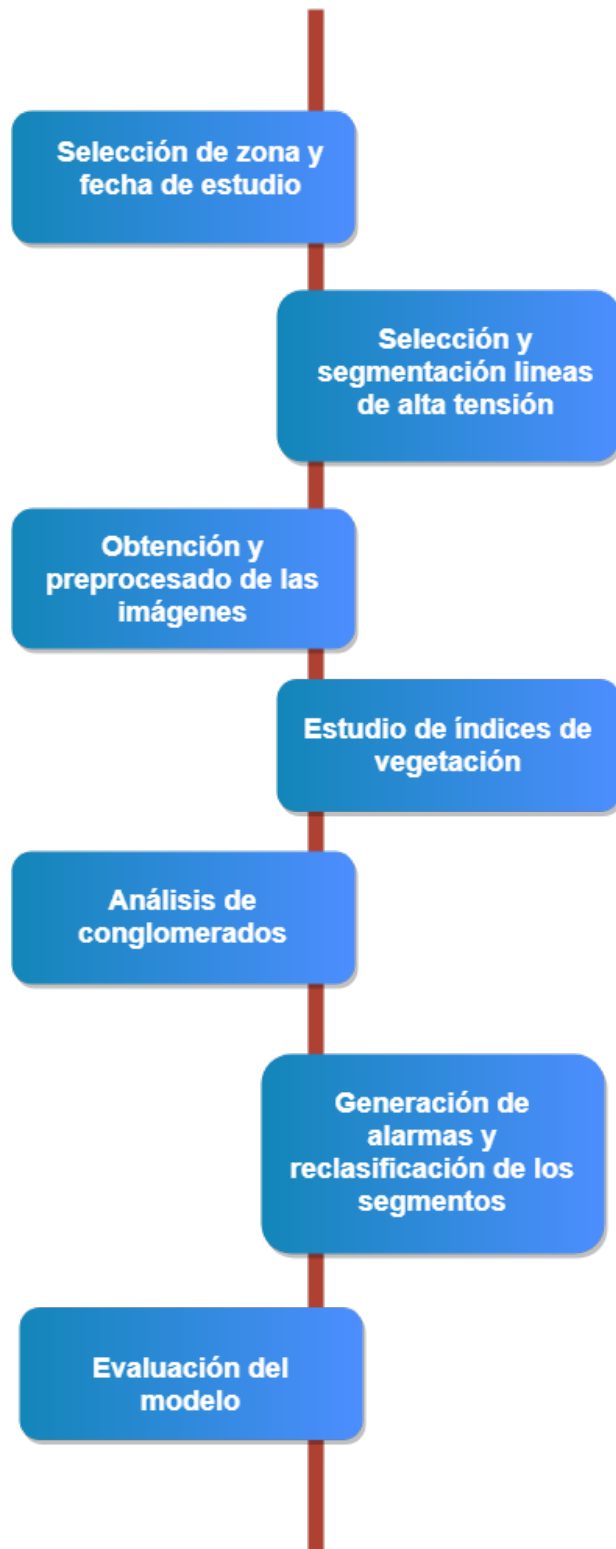


Figure 2.1: Estructura de trabajo.

1. **Selección de la zona y la/s fecha/s de estudio:** En esta fase hemos seleccionado la zona de interés y las fechas del estudio a realizar.
2. **Selección y segmentación de las líneas de alta tensión:** Hemos seleccionado diferentes secciones de las líneas de alta tensión de la zona y, en el caso de ser necesario, hemos segmentado la línea.
3. **Obtención y preprocesado de las imágenes:** Las imágenes deben ser preprocesadas antes de extraer la información más relevante para nuestro estudio. Para ello, inicialmente se realiza una calibración y un correregistro de las mismas, y también se aplican filtros para la reducción del posible ruido.
4. **Estudio de los índices de vegetación:** Dependiendo de las características del área seleccionada, resulta interesante investigar qué índices de vegetación son los apropiados para las propiedades de interés a observar. Tanto las propiedades a las que se hace referencia, como los diferentes índices que pueden utilizarse, serán comentados más adelante en la Sección [2.2](#).
5. **Análisis de conglomerados:**
 - **Selección de los conglomerados:** Se han seleccionado 5 zonas muy diferentes entre sí para crear los grupos que posteriormente han sido utilizados como *clusters* o conglomerados.
 - **Clasificación de los segmentos de las líneas de alta tensión:** Una vez seleccionados los diferentes índices de vegetación, hemos clasificado los segmentos de las líneas de alta tensión a partir de la distancia mínima a los grupos o conglomerados. No obstante, hemos realizado un experimento previo de la clasificación de un único cable de alta tensión que consta de 5 segmentos.
6. **Generación de alarmas y posible reclasificación de segmentos de las líneas de alta tensión:** A partir de una matriz de distancias de los valores de los índices de vegetación seleccionados a lo largo del tiempo a los valores de los mismos índices de vegetación de los conglomerados, puede que se establezcan "alarmas" para algunos segmentos que no habían sido considerados como sensibles al ser monitorizados.

7. **Evaluación del modelo:** Una vez obtenida la clasificación final de los segmentos de las líneas de alta tensión, hemos evaluado los resultados del sistema mediante una matriz de confusión a fin de determinar la calidad de la clasificación realizada.

2.1 Imágenes del satélite Sentinel-2

Desde mayo de 2017, la agencia ESA comenzó a proporcionar productos Sentinel-2 de nivel L2A para Europa y han reprocesado archivos de ese año. Desde mediados de marzo de 2018, los productos de nivel L2A son ya un producto operativo con respecto a la cobertura de la región euro-mediterránea y se espera una cobertura sistemática a nivel mundial para este verano de 2018. No obstante, es posible descargar productos de nivel L1C y personalizar la corrección atmosférica utilizando diferentes algoritmos (6S, 6SV, arcsi) o DEM (*Digital Elevation Model*) diferentes.

Si queremos utilizar imágenes Sentinel-2 de nivel L2A a partir de imágenes de nivel L1C, debemos realizar un preprocesamiento previo como se muestra en el Anexo A en el cual se describe la transformación radiométrica de TOA (*Top-Of-Atmosphere*) a BOA (*Bottom-Of-Atmosphere*)¹.

Las imágenes que han utilizado en este trabajo son imágenes del satélite Sentinel-2 de nivel 1C (L1C) ya que el estudio temporal ha sido desarrollado para el año 2017.

Un producto de nivel L1C está compuesto de mosaicos (orto-imágenes en proyección UTM/WGS84). Estas imágenes se obtienen como resultado de la proyección de la imagen en geometría cartográfica a partir de un modelo de elevación digital (DEM). Las mediciones radiométricas por píxel se proporcionan en forma de reflectancias en el límite superior de

¹La corrección de los niveles de reflectividad BOA o nivel de superficie permite generar imágenes corregidas atmosféricamente bajo productos de nivel 2A. Entre otras funciones, el algoritmo de corrección trabaja sobre la banda 10 *cirrus* corrigiendo la presencia de nubes y sobre las bandas del azul, rojo y SWIR para realizar una medición de la transparencia atmosférica mediante el factor AOT (*Aerosol Optical Thickness*). La corrección transforma los datos de radiancia y reflectividad pasando de TOA a BOA y generando una réplica de bandas corregidas del nivel 1C, en resoluciones de 10, 20 y 60 metros.

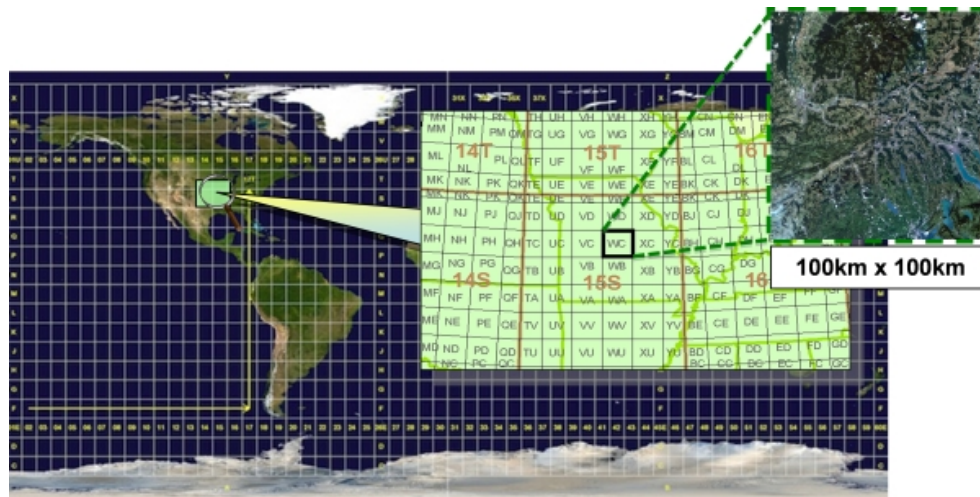


Figure 2.2: Mosaico de las imágenes Sentinel-2 de nivel L1C.

la atmósfera (*Top-of-atmosphere*, TOA) junto con los parámetros para transformarlas en radiancias. Los productos de nivel 1C se remuestrean a partir de la distancia constante (10, 20 y 60m) entre dos centros de píxeles en la imagen (*Ground Sampling Distance*, GSD) dependiendo de la resolución original de las diferentes bandas espectrales. Además, en los productos de nivel L1C, las coordenadas del píxel se refieren a la esquina superior izquierda del píxel. Estos productos también incluyen máscaras de nubes y de tierra/agua.

Por lo tanto, los productos de nivel L1C proporcionan la reflectancia ortorrectificada en el límite superior de la atmósfera (*Top-Of-Atmosphere*, TOA) con registro multispectral a nivel de sub-píxel. Es decir, no es necesario realizar ningún preprocesado de las imágenes cuando se trabaja con *Google Earth Engine* GEE ya que las imágenes adquiridas han sido corregidas radiométrica y geoméricamente incluyendo la ortorrectificación y el registro espacial de las mismas.

Si se desean utilizar imágenes Sentinel-2 de nivel L2A no hay una gran variación en el esquema de trabajo ya que simplemente debemos realizar el preprocesado requerido para imágenes de este nivel y obtener los índices de vegetación deseados.

En el caso de las imágenes Sentinel-1, el máximo nivel de procesamiento (nivel 2), las

imágenes están geolocalizadas pero no ortorrectificadas y, por tanto, requieren de un procesamiento específico el cual se encuentra especificado en el Anexo B donde podemos encontrar los pasos necesarios para el preproceso de las imágenes Sentinel-1 mediante algoritmos de *Google Earth Engine*.

Sin embargo, en este caso, no se puede calcular un índice de vegetación tal como el índice NDVI (ver siguiente 2.2). Pero existe un gran número de estudios que muestran la correlación o la relación empírica existente entre la variable esencial climática (*Essential Climate Variable*, CSV) denominada índice de área foliar² (*Leaf Area Index*, LAI) [37] y el índice de vegetación NDVI [34]. Por lo cual, en caso de utilizar imágenes Sentinel-1, puede reemplazarse el índice de vegetación por la variable ECV conocida como LAI [14, 24, 8, 7].

2.2 Índices de vegetación

Un índice de vegetación (*Vegetation Index*, VI) es una transformación espectral de dos o más bandas donde se busca mejorar la contribución de las propiedades de la vegetación y permitir comparaciones espaciales y temporales de la actividad fotosintética de la superficie terrestre. Han sido herramientas muy utilizadas en la detección de cambios para el estudio de la deforestación [10, 26]. Proporcionan un indicador de la abundancia relativa de un parámetro de interés relacionado con la vegetación, minimizando la influencia de la cubierta de suelo y de las condiciones atmosféricas. Por lo tanto, para su cálculo se requieren imágenes multiespectrales ya corregidas y en unidades de reflectancia.

Actualmente existen muchos índices de vegetación y se dividen en dos grupos según el número de bandas utilizadas: los índices de vegetación multiespectrales y los índices de vegetación hiperespectrales. A continuación, vemos algunos de los índices de vegetación más utilizados en la literatura:

²El índice de área foliar, también conocido como LAI, se define como la cantidad de área foliar (hoja verde) por unidad de área de superficie de tierra que abarca ese vegetal. Este parámetro es uno de los más utilizados para caracterizar la vegetación o los doseles o canopias de las plantas.

- **NDVI:** El índice de vegetación de diferencia normalizada (*Normalized Difference Vegetation Index*, NDVI) es uno de los índices más conocidos. Se utiliza para estimar la cantidad, calidad y desarrollo de la vegetación con base a la medición de la intensidad de la radiación de las bandas roja e infrarroja cercana del espectro electromagnético que la vegetación emite o refleja. Se calcula de la siguiente manera:

$$NDVI = \frac{\varphi_{NIR} - \varphi_{RED}}{\varphi_{NIR} + \varphi_{RED}}$$

donde NIR corresponde a la reflectividad espectral del canal infrarrojo cercano, y RED representa la reflectividad en el canal rojo visible.

- **GNDVI:** El índice de vegetación de diferencia normalizada verde (*Green Normalized Difference Vegetation Index*, GNDVI) es uno de los índices más utilizados en teledetección para la elaboración de mapa de cultivos. Al igual que el NDVI, es un indicador de la actividad fotosintética de la planta, pero a diferencia de éste, el GNDVI reemplaza la banda correspondiente al color rojo por la banda de la luz en el rango verde. Su fórmula es la siguiente:

$$GNDVI = \frac{\varphi_{NIR} - \varphi_{GREEN}}{\varphi_{NIR} + \varphi_{GREEN}}$$

- **ARVI:** El índice de vegetación *Atmospherically Resistent Vegetation Index* es un índice de vegetación basado en el índice NDVI. Podría decirse que es una versión mejorada del de la diferencia normalizada ya que toma en cuenta los efectos de la atmósfera y sustituye en la fórmula del NDVI el canal rojo por uno rojo-azul. Este índice se calcula a partir de la siguiente fórmula:

$$ARVI = \frac{\varphi_{NIR} - (\varphi_{RED} - 1 * (\varphi_{RED} - \varphi_{BLUE}))}{\varphi_{NIR} + (\varphi_{RED} - 1 * (\varphi_{RED} - \varphi_{BLUE}))}$$

- **SR:** El índice de vegetación simple NIR-rojo *Simple Ratio NIR/Red* es un índice muy conocido que se describe como la relación entre la luz dispersa que se absorbe en el rango infrarrojo cercano y la que se absorbe en el rango rojo. Se calcula de la siguiente manera:

$$SR = \frac{\varphi_{NIR}}{\varphi_{RED}}$$

- **RG:** El índice de vegetación simple rojo-verde (*Simple Ratio Red/Green* o *Red-Green Ratio*, RG) muestra la relación entre las bandas roja y verde con el índice de vegetación normalizado NDVI, y calcula un índice que excluya la banda de infrarrojo cercano como:

$$RG = \frac{\varphi_{RED}}{\varphi_{GREEN}}$$

- **SAVI:** *Soil Adjusted Vegetation Index* es un índice que pretende minimizar la influencia del suelo en la cuantificación de vegetación mediante la inclusión de un factor de ajuste L. Este factor normalmente toma un valor de 0.5, pero si la vegetación cubre altamente la zona, el valor de L es 0 ó 0.25, mientras que si la cobertura de la vegetación es baja, el valor de L es -1. A continuación, vemos la fórmula para calcular dicho índice:

$$SAVI = \frac{(1 + L)\varphi_{NIR} - \varphi_{RED}}{\varphi_{NIR} + \varphi_{RED} + L}$$

- **EVI:** *Enhanced Vegetation Index* es otra mejora del índice NDVI utilizada en áreas de dosel arbóreo denso donde el índice del área foliar es alto. Aprovechando la información en la longitud de onda azul, pueden mejorarse los valores obtenidos con NDVI ya que la información en esta parte del espectro puede ayudar a corregir las señales de fondo del suelo y las influencias atmosféricas. A la fórmula se le añaden los coeficientes C1 y L, que se utilizan para corregir la condición atmosférica. Para imágenes Sentinel 2 L1C, los valores que se utilizan para los coeficientes son de G (*gain factor*)=2.5, C1=2.4 y L=10000.

$$EVI = G \frac{\varphi_{NIR} - \varphi_{RED}}{\varphi_{NIR} + C1 * \varphi_{RED} + L}$$

2.3 Análisis de conglomerados

El análisis de conglomerados es una técnica estadística multivariante que busca agrupar objetos (en nuestro caso, las *features* seleccionadas en la zona de interés) con el fin de obtener la máxima homogeneidad en cada grupo y la mayor diferencia entre los grupos. Es una técnica eminentemente exploratoria ya que, la mayoría de las veces, no utiliza ningún tipo de modelo estadístico para llevar a cabo el proceso de clasificación. Por tanto, el agrupamiento como tal no es una tarea automática, sino un proceso iterativo de minería de datos exploratoria o interactivo de optimización multi-objetivo y es una técnica común en el análisis de datos estadísticos [18]. No obstante, puede ser vista como una técnica de aprendizaje no supervisado, es decir, una técnica muy adecuada para extraer información de un conjunto de datos sin imponer restricciones previas en forma de modelos estadísticos, al menos de forma explícita y, por ello, puede llegar a ser muy útil como una herramienta de elaboración de hipótesis acerca del problema a resolver sin imponer patrones o teorías previamente establecidas. Se utiliza en muchos campos como el aprendizaje automático, el reconocimiento de patrones, el análisis de imágenes, la búsqueda y recuperación de información, la bioinformática, la compresión de datos y la computación gráfica.

En nuestro caso, los elementos representativos de cada grupo han sido seleccionados en función de sus características espectrales de forma que mantengan una homogeneidad espectral dentro de cada grupo y sean muy heterogéneos espectralmente entre ellos.

Una vez establecidos los grupos y los objetos a clasificar, se define una medida de proximidad o de distancia que cuantifique el grado de similitud entre cada par de objetos.

Chapter 3

Experimentación y discusión de los resultados

En este capítulo presentamos las fases del estudio y selección de la zona de interés, las fechas de estudio designadas, la localización de las líneas de alta tensión y su posible segmentación.

También incluimos en este capítulo la obtención y el tratamiento de las imágenes Sentinel-2 y la selección de los índices de vegetación para este estudio. Asimismo, presentamos la clasificación realizada mediante el análisis de conglomerados y la evaluación final del sistema.

Por último, presentamos una discusión y una comparación de los resultados obtenidos en este proyecto.

3.1 Selección del área de estudio y descripción temporal

Las localizaciones de las zonas de interés, en este caso los cables de alta tensión de Donostia, se han representado mediante polígonos en QGIS. Se determinan unas fechas concretas para el estudio las cuales sirven como filtro a la hora de crear la colección de imágenes Sentinel-2 en GEE. La extracción de las localizaciones de los cables de alta tensión se explica más adelante en la Subsección 3.2 y el modo de trabajar con la estructura de datos de QGIS y de GEE ha quedado explicado en el capítulo anterior en la Sección 1.4.2 y en el Anéxo D.

En este proyecto se ha seleccionado como zona de interés la ciudad guipuzcoana de Donostia-San Sebastián. Donostia, es una pequeña ciudad costera de unos 61 km² ubicada al Norte de España. Su localización geográfica se corresponde con una longitud de W01°59'18.64" y una latitud de N43°17'30.52". Se han seleccionado varias zonas de la ciudad para realizar la extracción de los cables eléctricos de alta tensión. Para ello, se han utilizado polígonos delimitadores (*bounding box*, BB, BBox). Las coordenadas de los puntos del polígono delimitador que se ha utilizado son las siguientes: [(43.34540, -2.04620), (43.27771, -2.04054)], [(43.28833, -1.93737), (43.33517, -1.95454)].

Además de la selección de la zona de interés, debemos especificar las fechas de captura de las imágenes. Como fecha de estudio se ha elegido la época de verano de 2017 que comprende las fechas del 21 de junio al 23 de septiembre.

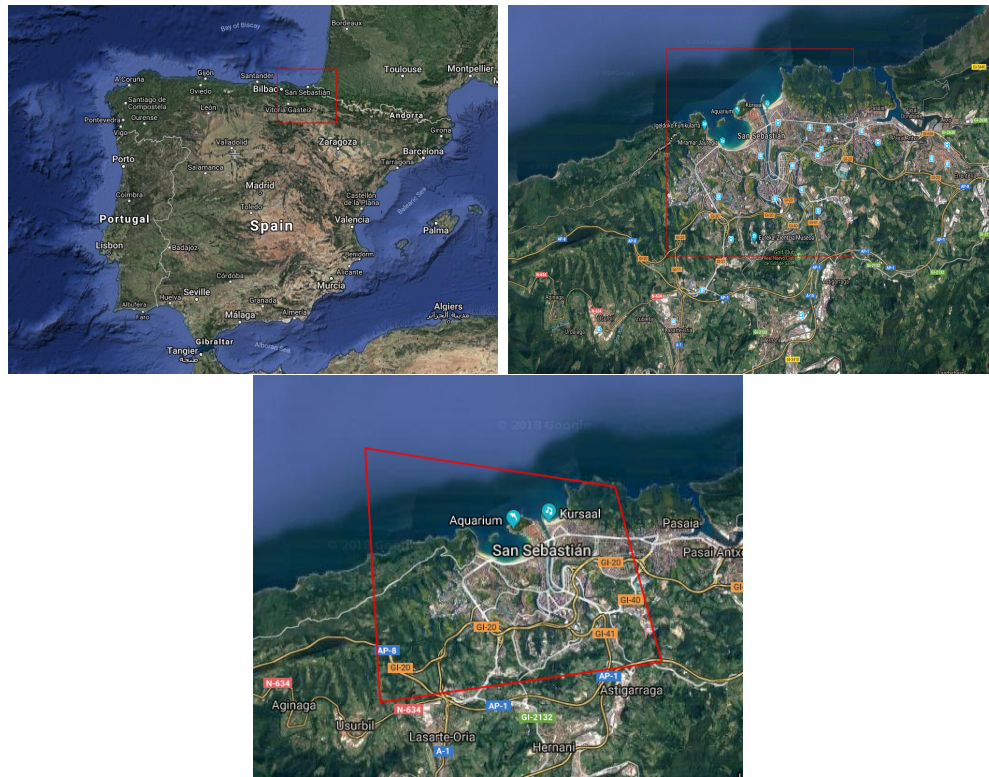


Figure 3.1: Área exacta de estudio.

3.2 Selección y segmentación de las líneas de alta tensión

Como se ha mencionado en el apartado de objetivos del proyecto (ver Subsección 1.1), el objetivo de este proyecto trata de localizar y monitorizar las zonas cercanas a cables de alta tensión y sus cortafuegos correspondientes para decidir si se encuentran en zonas con tendencias a tener una alta vegetación. Por esa razón, la primera tarea es conseguir las localizaciones de dichos puntos de interés. Mediante el portal web *“Overpass turbo”* (<http://overpass-turbo.eu/>) se pueden encontrar fácilmente la localización de las líneas eléctricas de la zona de San Sebastián y exportar los datos en diferentes formatos, como por ejemplo “.json”, “.geojson”, “.kml” o “.gpx” para, posteriormente, ser utilizados en un

sistema GIS. En nuestro caso, se utiliza el formato ".kml".

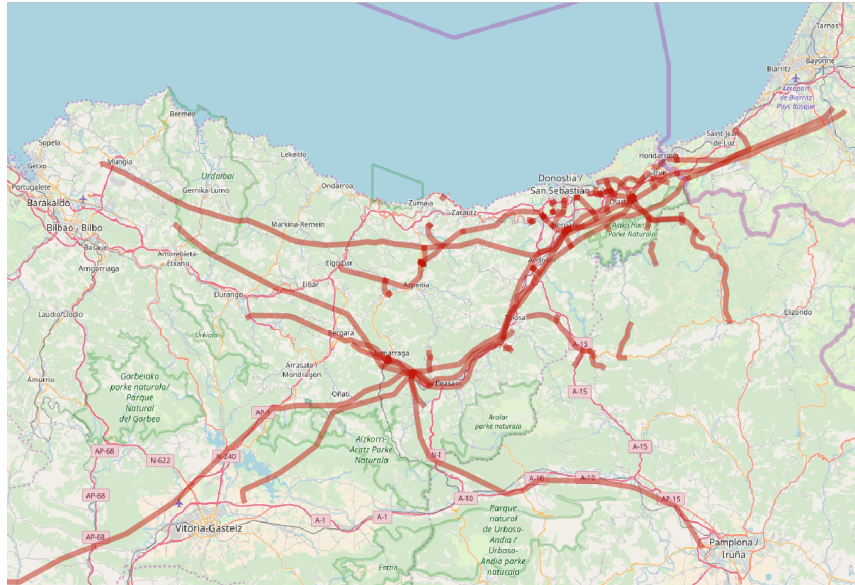


Figure 3.2: Cableado eléctrico de la zona del País Vasco.

Una vez obtenidas las localizaciones de los cables de alta tensión que queremos monitorizar, creamos polilíneas siguiendo dichos cables con un desplazamiento de 10 metros de radio alrededor de las polilíneas (*offset*).

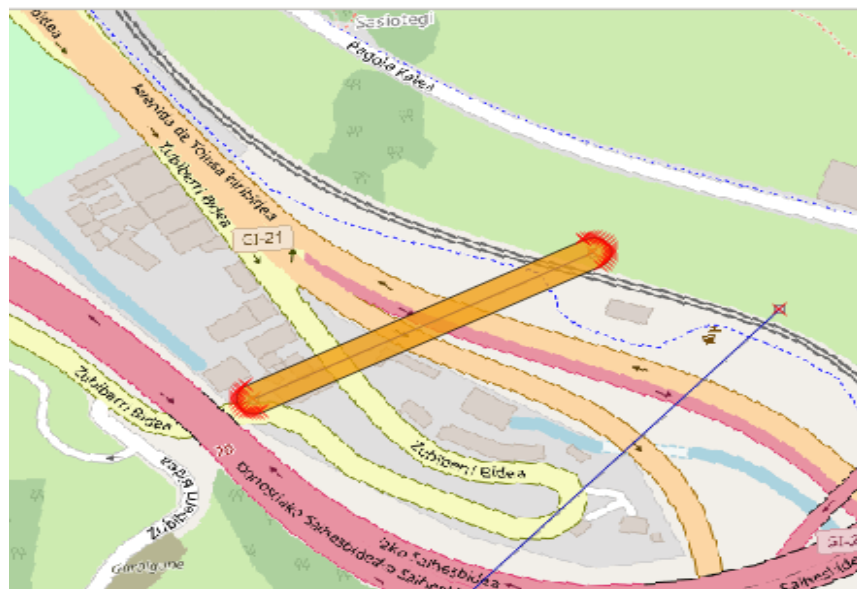


Figure 3.3: Polilínea con un *offset* de 10 m.

La utilización de este desplazamiento en los laterales de la polilínea (*offset*) es necesario para estudiar el crecimiento de la vegetación en esta franja de terreno.

Una vez establecidas tanto las polilíneas como sus franjas laterales, se segmenta cada una de ellas y se realiza un estudio particular para cada segmento de la polilínea con su *offset* correspondiente.

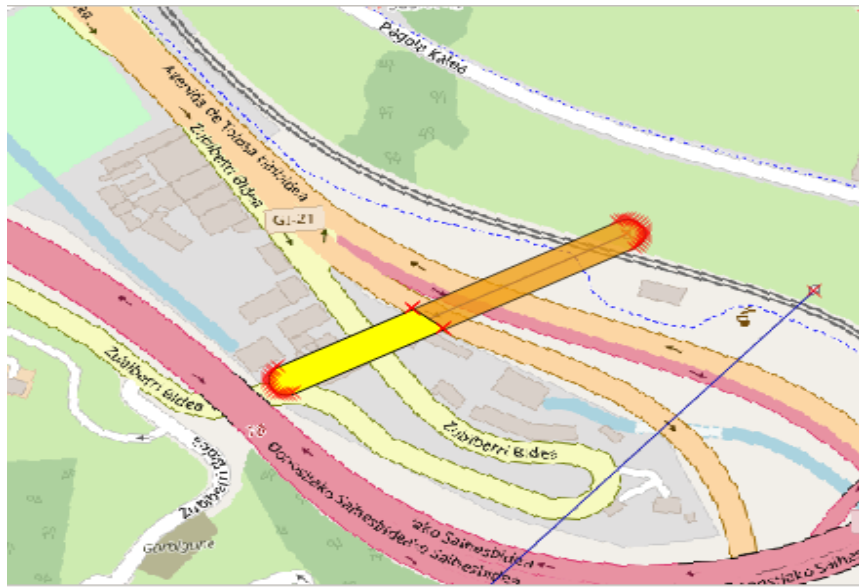


Figure 3.4: Dos segmentos de una polilínea.

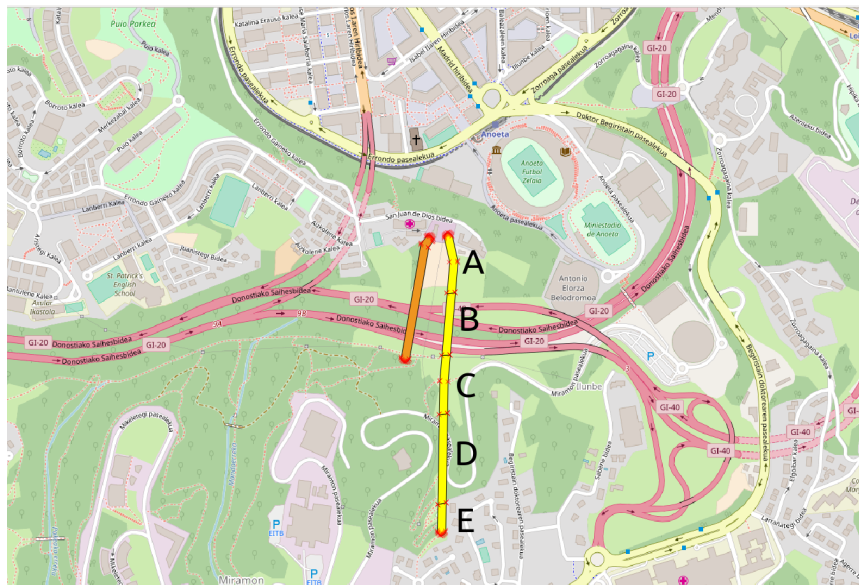


Figure 3.5: Cable de alta tensión subdividido en 5 segmentos.

3.3 Obtención y preprocesado de las imágenes del satélite Sentinel-2

Una vez seleccionada el área de estudio, hemos obtenido una serie temporal de imágenes de la zona para las fechas señaladas en la Sección 3.1. Las imágenes tomadas por el satélite Sentinel-2 se obtienen a partir de la aplicación Copernicus Open Access Hub - <http://scihub.copernicus.eu/> -, una plataforma facilitada por la ESA para la descarga de las imágenes de Sentinel-1, Sentinel-2 y Sentinel-3.

Dado que se va a trabajar con una colección de imágenes, antes de nada hay que cerciorarse de que estas son imágenes homogéneas. El paso de pre-procesamiento de las imágenes puede ser una tarea bastante ardua si se realiza manualmente. Por suerte, *Google Earth Engine* GEE cuenta con algoritmos ya implementados para realizar el preproceso de forma rápida y sencilla.

3.4 Análisis de conglomerados

Para detectar estas zonas de alta vegetación, se han seleccionado en QGIS cinco zonas con características muy diferentes como son césped, playa, mar, bosque y edificios que se han almacenado como un polígono en QGIS (ver estructuras de datos utilizadas en la Subsección 1.4.2) a fin de crear cinco *clusters* o conglomerados lo más heterogéneos posible por lo que las propiedades de las cinco zonas son bastante diferentes. En la Figura 3.6 y en la Tabla 3.1 podemos observar estas zonas y sus polígonos, así como las coordenadas de sus *bounding box*.

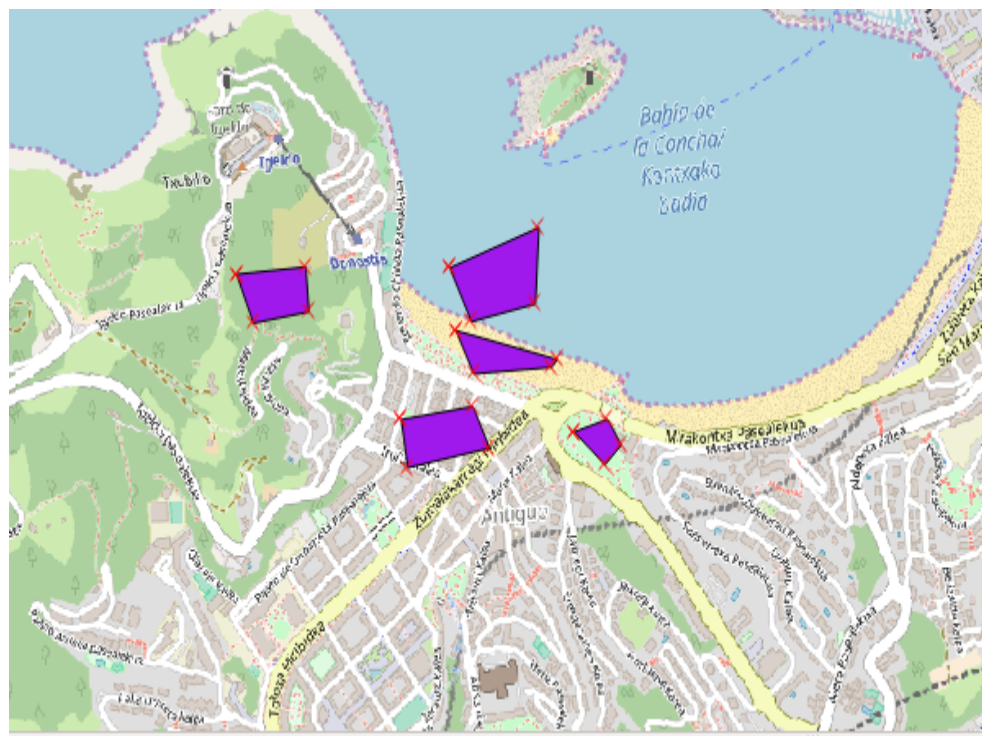


Figure 3.6: Conjunto de conglomerados para la detección de zonas de alta vegetación.

Tipo de suelo	Vértices	Longitud	Latitud
bosque	1	-2.0104976201659559	43.31867508813896706
	2	-2.00817567030345767	43.31867508813896706
	3	-2.00817567030345767	43.31777650025864546
	4	-2.01005299146888206	43.31766866881999789
	5	-2.0104976201659559	43.31867508813896706
mar	1	-2.00338356101276904	43.31835159803293323
	2	-2.00032056332181396	43.31896263345405629
	3	-2.00061698245319697	43.31745300536860555
	4	-2.00279072275000392	43.31723734115178104
	5	-2.00338356101276904	43.31835159803293323
arena	1	-2.00333415782420543	43.31709356458198812
	2	-2.0000241441904314	43.31630278736716377
	3	-2.00022175694468629	43.31615900858611212
	4	-2.00284012593856753	43.3161949533132713
	5	-2.00333415782420543	43.31709356458198812
cesped	1	-1.99962891868192094	43.31482903873404666
	2	-1.99849264534495386	43.31500876609069195
	3	-1.99809741983644296	43.31446958242609213
	4	-1.99869025809920897	43.31411012399192373
	5	-1.99962891868192094	43.31482903873404666
edificios	1	-2.00535968855532154	43.31544010957792779
	2	-2.0029389323156952	43.31554794497134253
	3	-2.00249430361862135	43.31468525646599943
	4	-2.00526088217819387	43.31446958242609213
	5	-2.00535968855532154	43.31544010957792779

Table 3.1: Coordenadas del conjunto de conglomerados para la detección de zonas de alta vegetación.

Como ya hemos dicho anteriormente, una tarea muy importante de este proyecto es la selección de los cables de alta tensión cercanos al área de estudio (en este caso, la zona de Donostia) y segmentar la zona de cableado ya que algunos cables cuentan con una longitud de varios kilómetros que pasan por zonas muy diferentes. A continuación, se muestra una tabla con las longitudes de los 11 cables eléctricos que se encuentran en la zona de Donostia (ver Tabla 3.2).

Cable	Longitud (m)	Longitud (km)
1	4260.25	4.26
2	91.05	0.09
3	717.73	0.72
4	301.25	0.30
5	2331.67	2.33
6	3766.26	3.77
7	27.67	0.03
8	45.36	0.05
9	288.66	0.29
10	37.59	0.04
11	396.00	0.40

Table 3.2: Longitudes de los 11 cables eléctricos del área de Donostia.

Por ello, es necesario realizar una buena segmentación tanto de la polilínea como de las zonas adyacentes a ésta para determinar qué partes del cableado se encuentran en las "zonas peligrosas" y cuáles no. Consideramos que las zonas "peligrosas" son aquellas zonas con mayor vegetación.

Las zonas que se clasifican como "bosque" son consideradas como zonas "peligrosas". Esto es, si una zona es etiquetada como "bosque" se debe podar la zona alrededor del cable. Por otro lado, las zonas clasificadas como "césped" no suponen un peligro tan grande como las zonas de bosque pero si requieren de una atención "especial" para evitar que la vegetación crezca en exceso.

Además de la selección de los conglomerados, debemos especificar las fechas de captura de las imágenes durante las cuales se van a observar los valores del índice de vegetación. Se han seleccionado 9 imágenes de forma aleatoria en el mes de julio de 2017 (del 1 al 30 de julio inclusive).

En la Figura 3.7 podemos observar los valores del índice NDVI para cada zona a lo largo del tiempo de estudio.

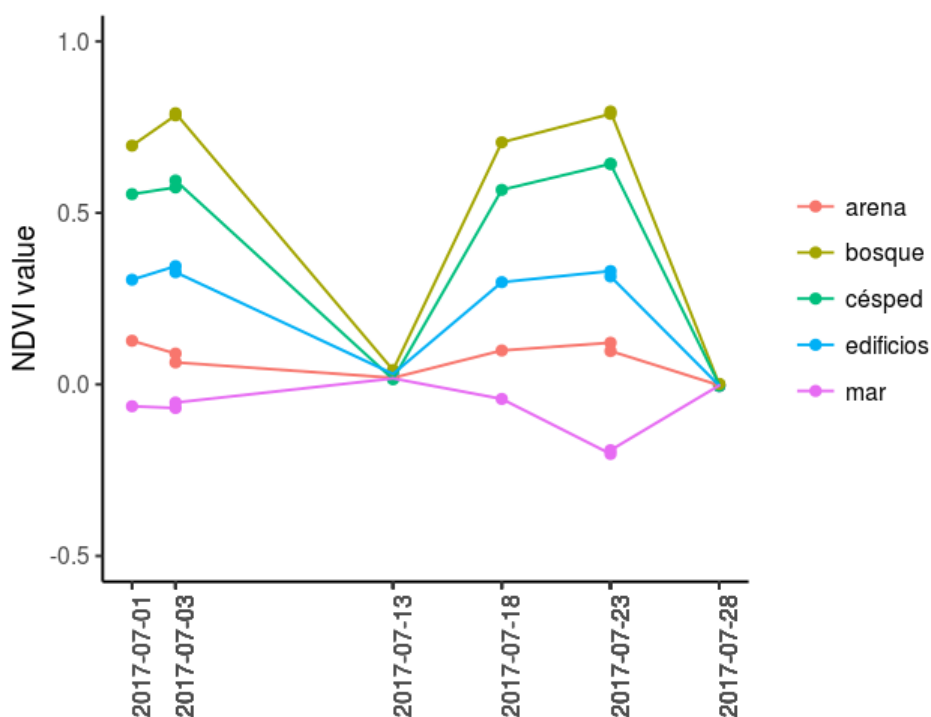


Figure 3.7: Valores del índice NDVI en las diversas zonas que forman el conjunto de entrenamiento a lo largo del tiempo.

Los valores del NDVI negativos y cercanos a cero indican la ausencia de vegetación. Los valores por encima de 0,2 indican presencia de vegetación, aunque puede tratarse de vegetación poco densa. A partir de valores de 0,4 se presenta vegetación con mayor cobertura. Valores de 0,6 indican áreas boscosas y por encima de 0,7 hasta 1 contienen campos de cultivo de gran vigor.

Sentinel-2 revisa cada 5 días la zona y, dependiendo de si la latitud es alta, toma dos imágenes con distinto ángulo de visión. Por ese motivo, como se puede ver en la Figura 3.7, existen diferentes valores NDVI para fechas idénticas. En el mes de julio se tomaron, aleatoriamente, en cinco fechas diferentes 9 imágenes que componen nuestra serie temporal de estudio. Los valores más altos del índice NDVI corresponden a las zonas de mayor verdor

como bosques o zonas de césped. Por otro lado, se puede apreciar que hay un pico muy drástico en todas las zonas en las imágenes tomadas los días 13 y 28 de julio ya que en esos días la nubosidad era muy alta como puede observarse en la Figura 3.8.

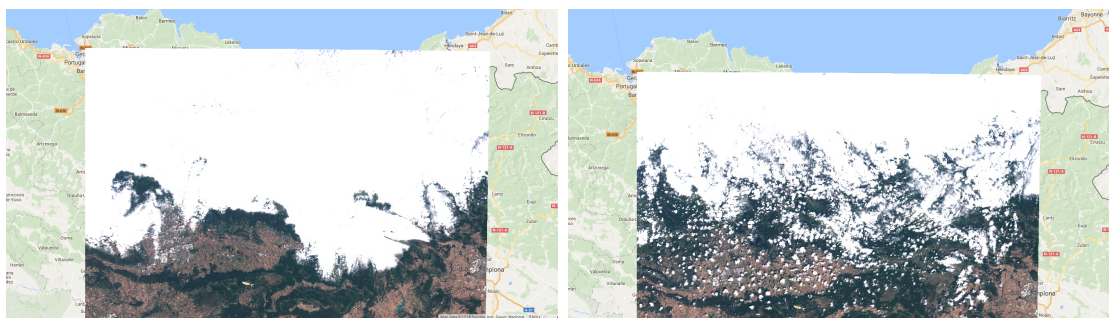


Figure 3.8: Imágenes de satélite Sentinel-2 tomadas el 13 y 28 de julio respectivamente.

Visto el impacto que pueden tener las nubes en el cálculo del índice de vegetación NDVI, hay que tomar en cuenta, para siguientes experimentos, la posibilidad de que ocurran estos contratiempos y aplicar algoritmos de corrección atmosférica que eliminen la nubosidad de las imágenes¹.

Para ver más claramente el desarrollo de las diferentes zonas en función del tiempo, la siguiente tabla recoge los valores del índice NDVI para las distintas imágenes durante el mes de julio.

¹Mediante el programa Sen2Cor podemos realizar las correcciones atmosféricas de imágenes Sentinel-2 y conseguir pasar de un nivel de producto L1C a L2A. Este programa realiza la corrección atmosférica en el entorno de SNAP replicando la estructura del archivo de bandas y permitiendo realizar un nuevo redimensionado de píxeles a 10, 20 y 60 metros de resolución espacial. Si utilizamos imágenes Sentinel-2 de nivel L1C, se toman las bandas en la parte alta de la atmósfera (TOA) que incluyen la nubosidad existente mientras que si utilizamos imágenes de nivel L2A, en estas imágenes ya han sido corregidas las imperfecciones que pueda dar la atmósfera a las bandas y, por tanto, se toman las bandas en la parte baja de la atmósfera (BOA).

NDVI	2017.07.03	2017.07.03	2017.07.13	2017.07.18	2017.07.18	2017.07.23	2017.07.23	2017.07.28	2017.07.28
mar	-0.06879	-0.05290	0.01740	-0.06367	-0.04236	-0.20262	-0.19168	-0.00282	-0.00442
arena	0.08984	0.06417	0.01931	0.12725	0.09895	0.12145	0.09731	-0.00266	-0.00314
edificios	0.34459	0.32711	0.03068	0.30515	0.29808	0.33040	0.31468	-0.00366	-0.00264
césped	0.57420	0.59433	0.01504	0.55462	0.56713	0.64284	0.64280	-0.00123	-0.00455
bosque	0.78431	0.79046	0.04198	0.69601	0.70567	0.78853	0.79562	-0.00005	0.00141

Table 3.3: Valores del índice de vegetación NDVI para las distintas imágenes y las distintas zonas del conjunto de entrenamiento a lo largo del tiempo.

De esta forma podemos monitorizar automáticamente las zonas de interés ya que son zonas de vegetación que pueden resultar altamente peligrosas si no se realiza un buen mantenimiento de las mismas o pueden ser zonas de alto riesgo de incendio.

3.4.1 Clasificación de los segmentos de las líneas de alta tensión

Después de obtener la segmentación de los cables de alta tensión, se visualiza la clasificación de cada segmento para tener una visión más "limpia" de las partes del cable respecto a los *clusters* definidos.

No obstante, y antes de realizar la clasificación de todos los segmentos de todos los cables de alta tensión, hemos realizado una clasificación previa para el cable de 5 segmentos (ver Figura 3.5) a fin de analizar la idoneidad de los grupos establecidos y de la distancia adoptada.

Clasificación de los segmentos de un único cable de alta tensión

Para este estudio preliminar de la clasificación de los segmentos de un único cable de alta tensión, hemos seleccionado una línea de gran longitud que consta de 5 segmentos denotados como A, B, C, D y E y que presentamos anteriormente en la Figura 3.5.

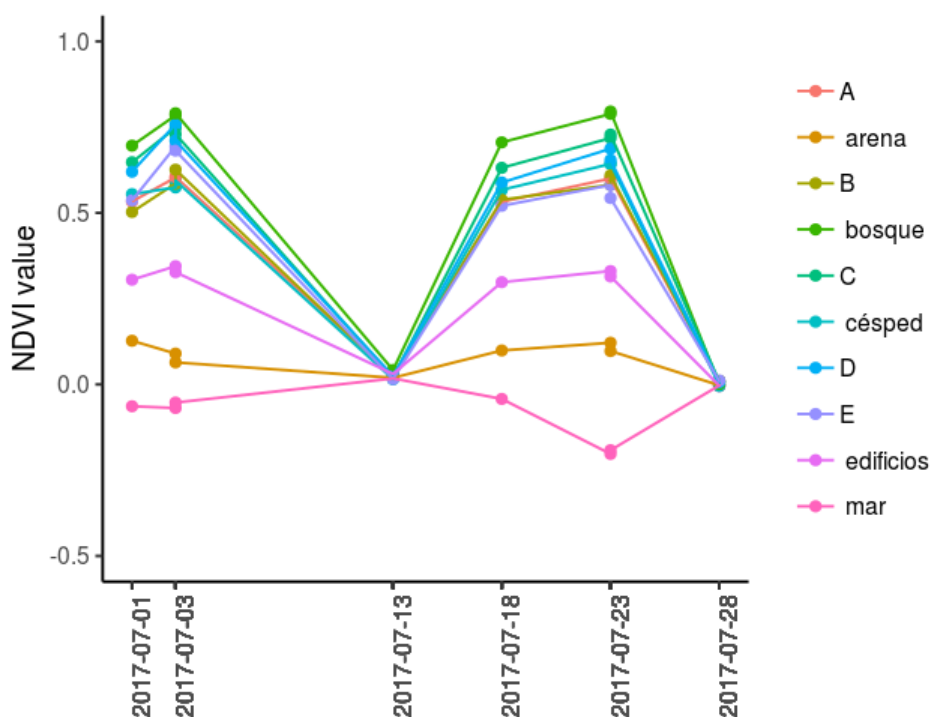


Figure 3.9: Valores de la serie temporal del índice de vegetación NDVI para un único cable de alta tensión formado por 5 segmentos: A, B, C, D, y E.

NDVI	2017.07.03	2017.07.03	2017.07.13	2017.07.18	2017.07.18	2017.07.23	2017.07.23	2017.07.28	2017.07.28
mar	-0.06879	-0.05290	0.01740	-0.06367	-0.04236	-0.20262	-0.19168	-0.00282	-0.00442
arena	0.08984	0.06417	0.01931	0.12725	0.09895	0.12145	0.09731	-0.00266	-0.00314
edificios	0.34459	0.32711	0.03068	0.30515	0.29808	0.33040	0.31468	-0.00366	-0.00264
césped	0.57420	0.59433	0.01504	0.55462	0.56713	0.64284	0.64280	-0.00123	-0.00455
bosque	0.78431	0.79046	0.04198	0.69601	0.70567	0.78853	0.79562	-0.00005	0.00141
A	0.60329	0.60485	0.02411	0.53328	0.53231	0.60014	0.60075	0.00675	0.00797
B	0.58559	0.62598	0.02070	0.50269	0.53821	0.58189	0.60936	0.00979	0.01023
C	0.74777	0.72923	0.02047	0.64766	0.63133	0.71810	0.72831	0.00759	0.00779
D	0.75829	0.71070	0.02462	0.61988	0.58853	0.68713	0.65409	0.00621	0.00686
E	0.69366	0.68152	0.01562	0.53642	0.52069	0.58119	0.54333	0.00985	0.01140

Table 3.4: Valores del índice de vegetación NDVI para las distintas imágenes y las distintas zonas del conjunto de entrenamiento a lo largo del tiempo.

Como se puede ver en la Figura 3.9, al principio, todos los segmentos del cable se encuentran

en zonas con bastante vegetación ya que han sido clasificadas como zonas de bosque y césped. A lo largo del tiempo observado, podemos apreciar cómo los valores del índice de vegetación NDVI de todos los segmentos comienzan de un valor, después descienden ligeramente para finalmente volver a subir. Las zonas C y D, se mantienen constantes en el aspecto de que comienzan siendo zonas más cercanas a el grupo "bosque" y terminan de la misma forma, mientras que el segmento E comienza el estudio siendo más cercano a "bosque" y, al final del estudio, la distancia con el grupo "cesped" es menor.

Clasificación de los segmentos de todos los cables de alta tensión

Una vez realizado el estudio de la clasificación para un cable formado para varios segmentos, hemos realizado la clasificación de todos los cables de alta tensión y de todos los segmentos que lo componen. De esta forma, realizaremos el estudio, planteado como objetivo principal, teniendo como zona de interés la ciudad de Donostia y no solamente un cable específico.

Para ello se seleccionan los cables eléctricos más cercanos a la zona de San Sebastián mediante polígonos mediante la herramienta QGIS. Como se ha mencionado anteriormente, se descartan las imágenes que contienen un alto porcentaje de nubes. Inicialmente contamos con una colección de 34 imágenes pero, después de aplicar la condición de desechar aquellas con un porcentaje de nubosidad mayor que 10, seleccionamos 11 imágenes por cada segmento de cable o *feature*.

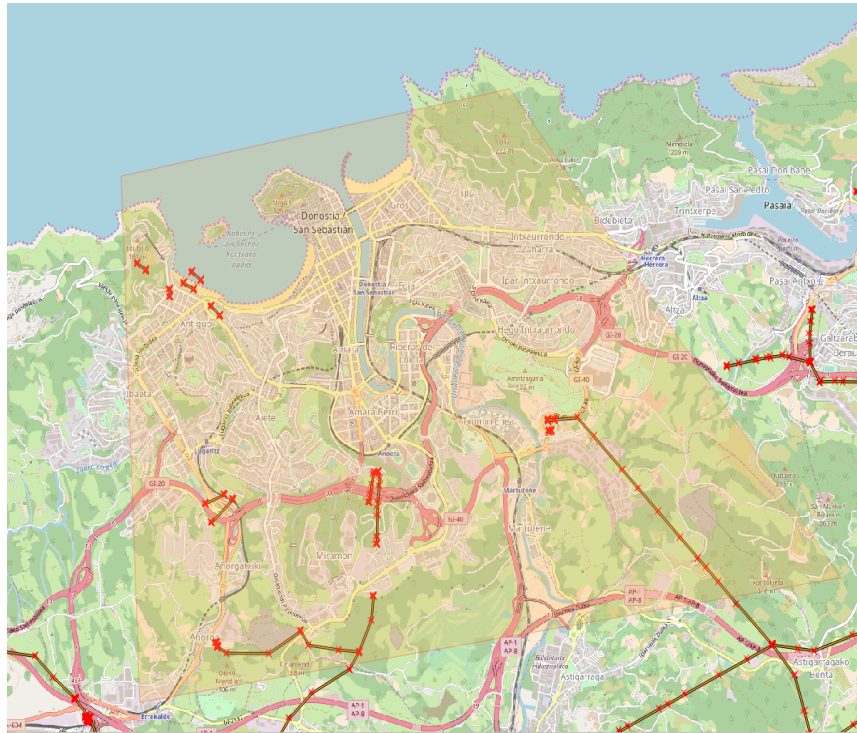


Figure 3.10: Cables de alta tensión del área de Donostia.

Después de realizar el estudio sobre los cables seleccionados, obtenemos los siguientes resultados.

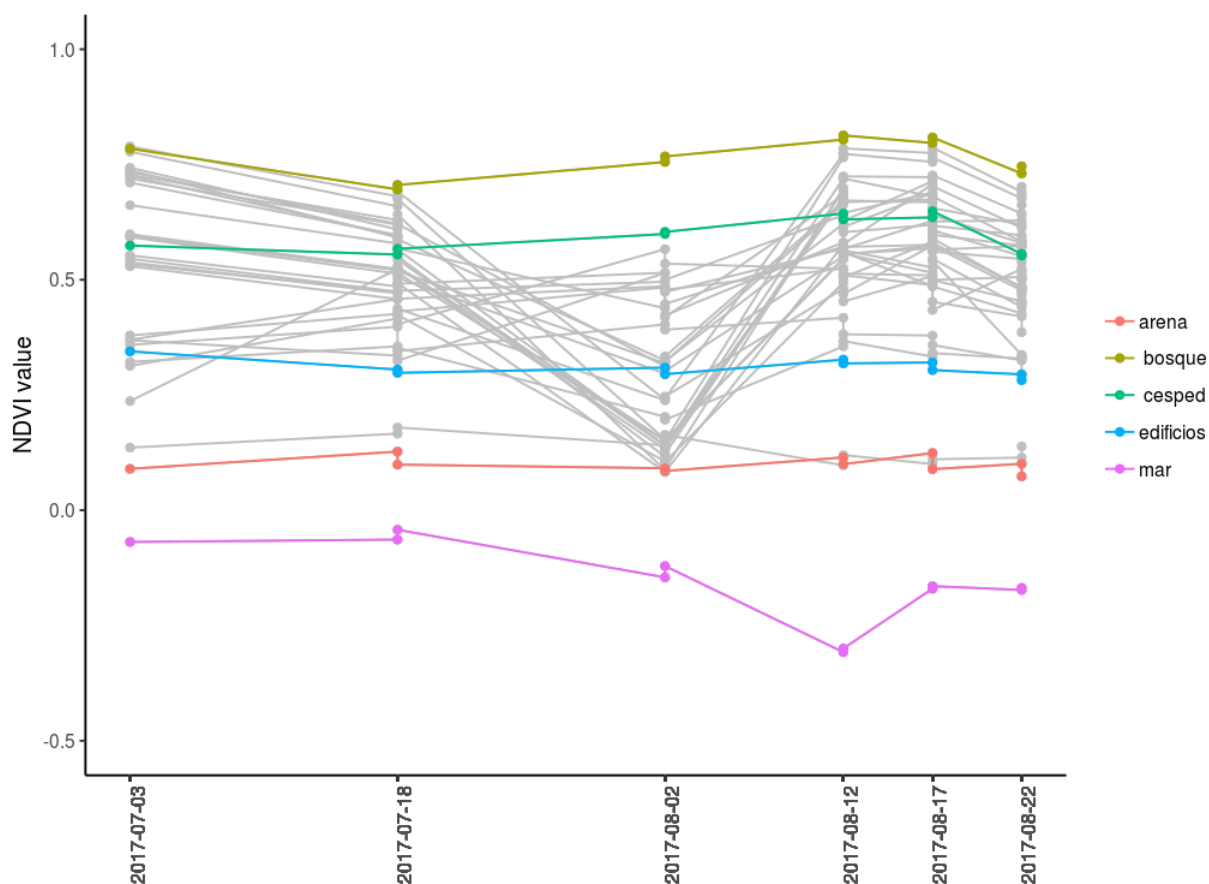


Figure 3.11: Valores del  ndice NDVI para las zonas de estudio que contienen cables de alta tensi on en el  rea de Donostia.

Como se puede ver en la Figura 3.11, existen diferentes valores para mismas fechas. Como se ha comentado en la Subsecci on 3.4 esto se debe a que se toman para el mismo d a dos fotos con diferente  ngulo de captura de imagen. Esto puede suponer un problema si la diferencia es muy alta, ya que no tiene sentido que, para las mismas  reas de estudio en las mismas fechas, se obtengan valores del  ndice NDVI muy distintos.

Para cuantificar la diferencia de dichos cambios, se establece como límite la diferencia entre los valores NDVI de los grupos "césped" y "bosque" en los días donde se producen los saltos. Si la diferencia supera el límite establecido, se debe estudiar cómo manejar este cambio mayor del esperado. Si la diferencia no supera el límite, se continúa con el estudio asumiendo que el cambio no es tan significativo como para crear grandes errores en la exactitud de los valores que se obtienen.

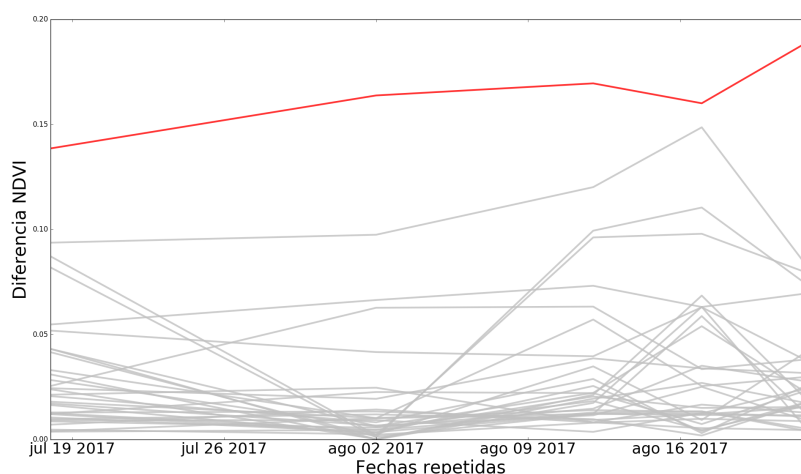


Figure 3.12: Cuantificación de los cambios observados para imágenes captadas en fechas repetidas. La diferencia entre los valores del índice NDVI para bosque/césped se representa en rojo.

Como se puede ver en la Figura 3.12, todos los cambios se encuentran por debajo del límite establecido, por lo que se puede afirmar que la diferencia entre los valores para el par de imágenes de una misma fecha es mínima y que no supone ningún problema para el resultado final del estudio.

3.5 Clasificación y visualización de alarmas

Una vez realizada la clasificación de los segmentos de los cables de alta tensión mediante el modelo de análisis de conglomerados diseñado, y tras observar detenidamente la Figura 3.11, se puede tener una "idea general" sobre la clasificación de los segmentos de los cables de alta tensión. No obstante, un segmento de un cable puede ser reetiquetado en función del comportamiento del índice de vegetación NDVI a lo largo del tiempo. Es decir, puede suceder que un segmento no fuera etiquetado como "bosque" en el instante inicial, pero analizando las diferencias de los valores del índice de vegetación NDVI a lo largo del tiempo con respecto a los valores de dicho índice de los grupos establecidos, este segmento puede ser reclasificado como "bosque" si la distancia mínima se produce para el grupo "bosque". De esta forma, un segmento que, inicialmente, podía no ser considerado como peligroso, pasa a ser sensible de monitorización (alarma) ya que el comportamiento temporal de su índice de vegetación se encuentra más cerca del índice de vegetación correspondiente al grupo "bosque".

La Tabla 3.5 recoge las distancias de los segmentos de todos los cables de alta tensión con respecto a los 5 diferentes conjuntos de conglomerados.

línea	segmento	arena	bosque	cesped	edificios	mar
1	1	2.901797	4.412155	2.569896	0.642617	5.721632
1	2	3.926463	3.387489	1.54523	1.684933	6.746298
1	3	4.297861	3.016091	1.173832	1.993131	7.117696
1	4	4.248165	3.065787	1.223528	1.943435	7.068
1	5	0.454213	6.941906	5.099646	1.932683	3.191881
1	6	3.949917	3.364035	1.521776	1.860734	6.769752
2	1	4.679837	2.634115	0.826235	2.375107	7.499672
3	1	5.540241	1.773711	0.761069	3.235511	8.360076
3	2	4.721653	2.592299	0.791335	2.416923	7.541488
3	3	5.519715	1.794237	1.186022	3.214985	8.33955
4	1	4.37363	2.940322	1.098063	2.0689	7.193465
4	2	5.235204	2.078748	0.852865	2.930474	8.055039
5	1	5.743824	1.570128	2.069864	4.039588	8.563659
5	2	4.952817	2.361135	1.428796	3.300849	7.772652
5	3	3.900074	3.433988	1.640872	2.451761	6.699798
5	4	3.262702	4.05125	2.208991	1.745149	6.082536
5	5	4.826551	2.487401	1.379448	3.254355	7.646385
5	6	4.684753	2.629199	1.482646	3.059482	7.504588
5	7	4.789432	2.52452	1.289679	3.128979	7.609267
5	8	5.267591	2.054676	1.839251	3.808975	8.07911
6	1	6.062321	1.261474	2.03106	4.000784	8.882155
7	1	4.37859	2.935362	1.129474	2.099755	7.198424
8	1	3.96369	3.350262	1.508003	2.261076	6.783525
9	1	2.433783	4.880169	3.03791	0.539747	5.253618
10	1	3.862058	3.451894	1.609635	2.16299	6.681893
11	1	3.660102	3.65385	1.811591	1.59365	6.479937

Table 3.5: Tabla de distancias

Después de ver los valores que toma cada segmento de los cables respecto a los diferentes tipos de conglomerados, se puede decir que dichos valores guardan relación con la figura vista anteriormente donde se ve la representación gráfica de las distancias (ver Figura 3.11). Sabiendo que las zonas de máximo peligro corresponden a las zonas de bosque principalmente, se puede ver que solamente el segmento 1 de la línea eléctrica 5 y el segmento 1 de la línea eléctrica 6 se encuentran en zonas peligrosas.

Por otro lado, aunque tenemos cables de alta tensión que se encuentran en zonas alejadas del bosque, ya que están situados en zonas de césped, debemos mantener un registro de estas zonas y evitar que crezcan en exceso.

Finalmente, una vez obtenida la tabla de distancias y se reclasifica cada segmento de cada cable de alta tensión en un conglomerado en función de la distancia mínima.

línea	segmento	arena	bosque	cesped	edificios	mar	Clase
1	1	2.901797	4.412155	2.569896	0.642617	5.721632	edificios
1	2	3.926463	3.387489	1.54523	1.684933	6.746298	cesped
1	3	4.297861	3.016091	1.173832	1.993131	7.117696	cesped
1	4	4.248165	3.065787	1.223528	1.943435	7.068	cesped
1	5	0.454213	6.941906	5.099646	1.932683	3.191881	arena
1	6	3.949917	3.364035	1.521776	1.860734	6.769752	cesped
2	1	4.679837	2.634115	0.826235	2.375107	7.499672	cesped
3	1	5.540241	1.773711	0.761069	3.235511	8.360076	cesped
3	2	4.721653	2.592299	0.791335	2.416923	7.541488	cesped
3	3	5.519715	1.794237	1.186022	3.214985	8.33955	cesped
4	1	4.37363	2.940322	1.098063	2.0689	7.193465	cesped
4	2	5.235204	2.078748	0.852865	2.930474	8.055039	cesped
5	1	5.743824	1.570128	2.069864	4.039588	8.563659	bosque
5	2	4.952817	2.361135	1.428796	3.300849	7.772652	cesped
5	3	3.900074	3.433988	1.640872	2.451761	6.699798	cesped
5	4	3.262702	4.05125	2.208991	1.745149	6.082536	edificios
5	5	4.826551	2.487401	1.379448	3.254355	7.646385	cesped
5	6	4.684753	2.629199	1.482646	3.059482	7.504588	cesped
5	7	4.789432	2.52452	1.289679	3.128979	7.609267	cesped
5	8	5.267591	2.054676	1.839251	3.808975	8.07911	cesped
6	1	6.062321	1.261474	2.03106	4.000784	8.882155	bosque
7	1	4.37859	2.935362	1.129474	2.099755	7.198424	cesped
8	1	3.96369	3.350262	1.508003	2.261076	6.783525	cesped
9	1	2.433783	4.880169	3.03791	0.539747	5.253618	edificios
10	1	3.862058	3.451894	1.609635	2.16299	6.681893	cesped
11	1	3.660102	3.65385	1.811591	1.59365	6.479937	edificios

Table 3.6: Reclasificación de cada segmento de los cables de alta tensión según el valor mínimo de la tabla de distancias.

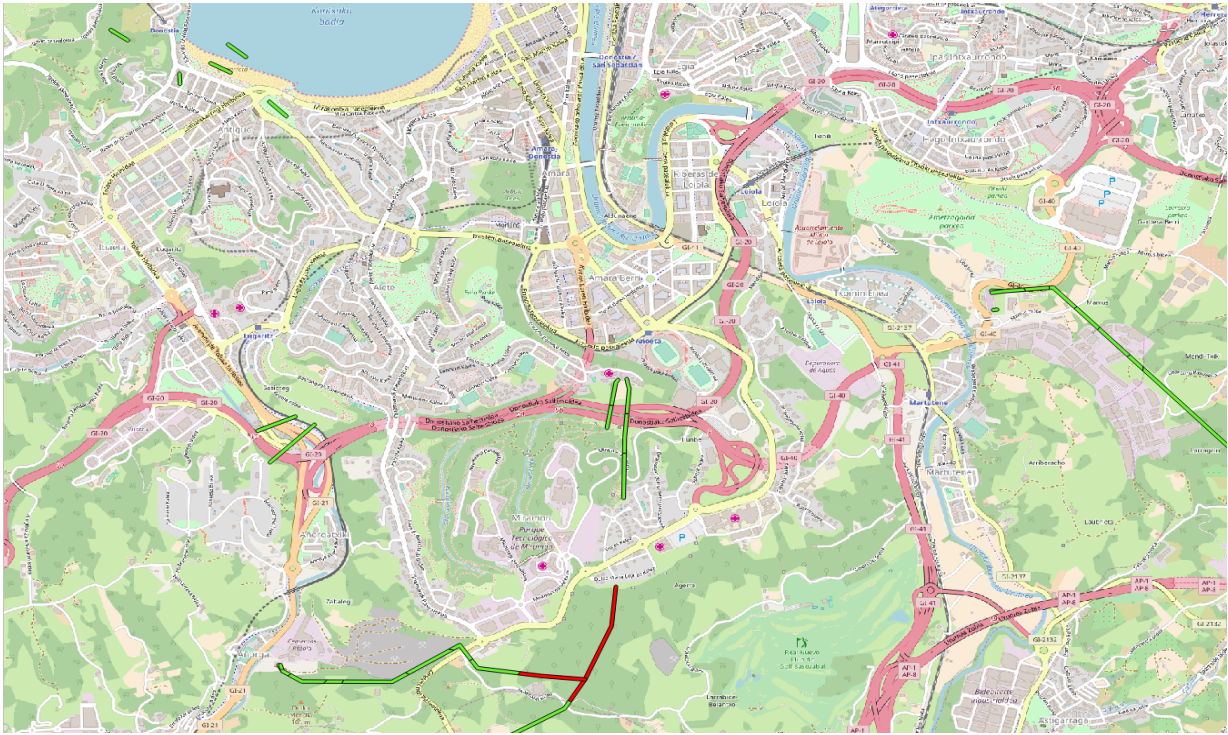


Figure 3.13: Mapa de los cables que se encuentran en una zona de mucha vegetación clasificada como peligrosa.

3.6 Discusión de los resultados

Para evaluar la clasificación realizada hemos utilizado la matriz de confusión ya que tenemos una categorización de las grupos seleccionados ("bosque", "edificios", ...) dada por los expertos. Si analizamos la matriz de confusión (ver Tabla 3.7), se pueden ver por un lado las clases "reales" (dadas por los expertos) y las clases que el modelo de clustering ha predicho. La fila *recall*, también llamado sensibilidad o *True Positive rate* (TP) que es la tasa de verdaderos positivos, mide la proporción de positivos "reales" que se identifican como tales.

		Predicción				classification overall	precision
		arena	bosque	cesped	edificios		
Clases "reales"	arena	0	0	0	0	0	
	bosque	0	2	2	0	4	50%
	cesped	0	0	17	0	17	100%
	edificios	1	0	0	4	5	80%
	truth overall	1	2	19	4	26	
	recall		100%	89.47%	100%		

Table 3.7: Matriz de confusión usando índice NDVI

Como se puede ver, en los casos de las clases "bosque" y "edificios", el valor obtenido es de un 100%, por lo que el modelo detecta y clasifica correctamente todos los segmentos que se encuentran en una zona de bosque o de edificios.

Por otro lado, se puede observar que se cometen dos clasificaciones "erróneas" ya que clasifica dos segmentos que deberían de pertenecer al grupo de "bosque" como zonas de césped. Esto puede deberse al hecho de que el índice de vegetación NDVI mide el verdor de la zona y, al ser un estudio realizado para la estación de verano, puede que esa zona se encuentre un poco seca y, por consiguiente, no tenga un color muy verde. Otro posible motivo de esta clasificación podría deberse a que durante los 3 meses de estudio, se hayan realizado podas en dicha zona. El porcentaje de acierto para la clase "césped" es de un 89.47. Este resultado se calcula teniendo en cuenta los segmentos etiquetados como "césped" y los que realmente lo son. Esto es:

$$recall_{cesped} = \frac{classification_overall_{cesped}}{truth_overall_{cesped}} = \frac{17}{19} = 0.8947 = 89.47\%$$

La columna *precision* explica qué porcentaje de predicciones positivas son correctas, por lo que la precisión a la hora de clasificar las zonas de bosque se calcula como:

$$precision_{bosque} = \frac{bosque}{classification_overall_{bosque}} = \frac{2}{4} = 0.5 = 50\%$$

Y la precisión al clasificar las zonas etiquetadas como edificios se calcula de la siguiente forma:

$$precision_{bosque} = \frac{edificios}{classification_overall_{edificios}} = \frac{4}{5} = 0.8 = 80\%$$

Finalmente, la exactitud total se calcula mediante la suma de todos los valores de la diagonal principal, ya que son las predicciones correctas, y se divide entre el número total de elementos evaluados:

$$precision_{total} = \frac{elementos_diagonal}{classification_overall_{total}} = \frac{23}{26} = 0.8846 = 88,46\%$$

Una vez analizadas todas las columnas de la matriz de confusión, podemos decir que los resultados del modelo de análisis de conglomerados adoptado ha permitido establecer un conjunto de alarmas eficientes de una forma muy eficaz y sencilla.

3.7 Otros experimentos

Una vez visto el trabajo desarrollado en este estudio con el índice NDVI, se ha repetido el mismo proceso pero con el índice GNDVI *Green Normalized Difference Vegetation Index* [16]. Como se ha explicado en la Sección 2.2, el índice GNDVI también es un índice muy popular en teledetección para crear mapas de cultivo y es un excelente indicador de la actividad fotosintética de la planta. La única diferencia a la hora de calcular el índice GNDVI respecto al índice NDVI reside en que en vez de utilizar la banda de color rojo se utiliza la banda de la luz en el rango verde.

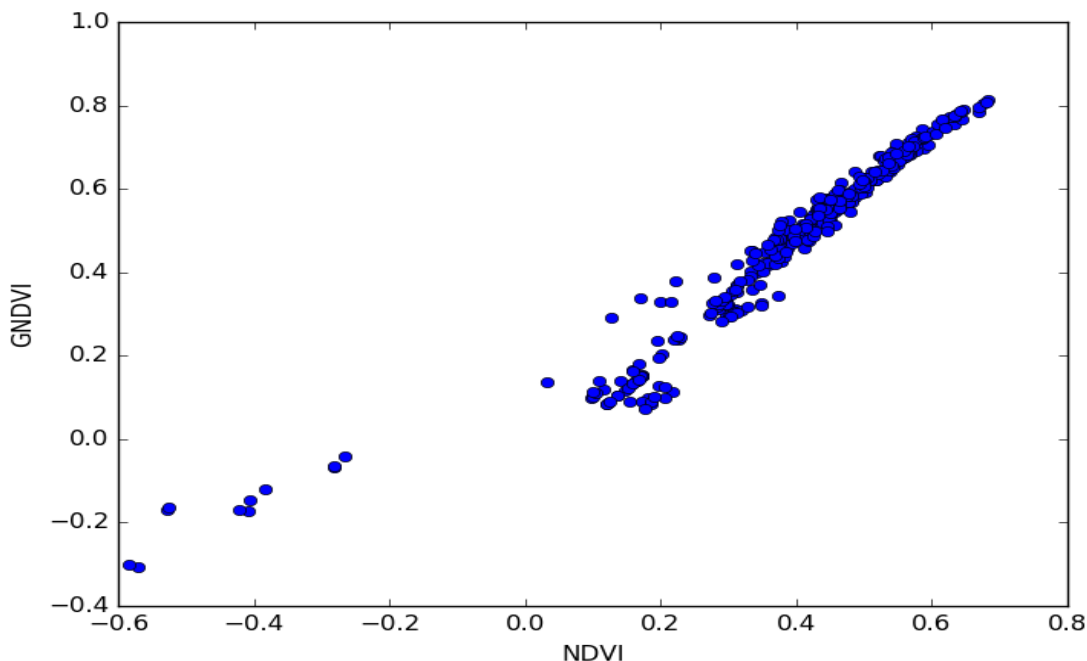


Figure 3.14: Correlación NDVI y GNDVI.

Como se puede ver en la Figura 3.14, los valores obtenidos con el índice NDVI y el GNDVI están muy relacionados.

El índice de vegetación NDVI es ideal para medir la vegetación sana y verde (vigor de la planta) en una amplia gama de condiciones. Sin embargo, puede saturarse cuando la vegetación es muy densa o cuando el área foliar es muy elevada. El índice de vegetación GNDVI resulta ser más sensible a la variación del contenido de clorofila en la vegetación que el índice NDVI y presenta un punto de saturación más alto. Se utiliza generalmente para determinar la absorción de agua y nitrógeno en las plantas, es muy útil para analizar el déficit o exceso de ambos elementos en los cultivos por lo cual permite la optimización del riego e indica cuándo se produce aislamiento del agua o varía la misma a través del suelo [16]. En áreas de vegetación escasa, ciertos estudios indican que el índice GNDVI tiene

una menor correlación que el NDVI y parece que el uso de la banda "verde" en ese tipo de áreas disminuye la sensibilidad del índice de vegetación a las variaciones de la fracción de cobertura vegetal [4].

Mediante la matriz de confusión vamos a ver la precisión de nuestro clasificador con los nuevos datos obtenidos a partir del índice de vegetación GNDVI. Los resultados obtenidos son los siguientes:

		Predicción				classification overall	precision
		arena	bosque	cesped	edificios		
Clases "reales"	arena	0	0	0	0	0	
	bosque	0	2	2	0	4	50%
	cesped	0	0	11	6	17	64.71%
	edificios	1	0	0	4	5	80%
	truth overall	1	2	13	10	26	
	recall		100%	84.615%	40%		

Table 3.8: Matriz de confusión usando índice GNDVI

Calculando de la misma manera que en la Sección 3.6 obtenemos la precisión total como:

$$precision_{total} = \frac{elementos_diagonal}{classification_overall_{total}} = \frac{17}{26} = 0.6539 = 65,39\%$$

En este caso, la precisión del modelo es de un 65.39%, por lo tanto mucho menor que para el caso utilizando el índice de vegetación NDVI. Como puede observarse en la Tabla 3.8, esta pérdida de la precisión es debida a la "errónea" clasificación de cables en zonas de "cesped" que han sido etiquetadas como "edificios". En este caso, dado que el índice GNDVI mide tanto la clorofila como la humedad o la cantidad de nitrógeno de la planta, al haber realizado el estudio en la época de verano puede tratarse de zonas de césped seco o con niveles de clorofila bajos. Por ese motivo, en este caso, las zonas donde se encuentran estos segmentos pueden haber sido clasificadas como "edificios".

Chapter 4

Conclusiones y líneas futuras

El objetivo principal de este proyecto era crear un sistema para monitorizar las zonas donde están situados los cables de alta tensión de la zona de Donostia y vigilar el crecimiento de la vegetación de dichas zonas. Hecho esto, se buscaba crear unas alarmas para los cables que se encontrasen en zonas peligrosas; esto es, zonas de alta vegetación o con tendencias a albergar mucha vegetación. A partir de un conjunto de entrenamiento de cinco zonas de características muy distintas, se ha conseguido un sistema de clasificación con una precisión de un 88% cuando se utiliza el índice de vegetación NDVI, el cual es un porcentaje de acierto muy alto y, por lo tanto, muy satisfactorio.

No obstante, existen otras formas de dar solución al problema planteado. Una de ellas, consiste en crear un heurístico que establezca un límite de los valores del índice de vegetación NDVI y que cualquier zona que supere dicho valor, quede marcada como zona que hay que podar.

Por una parte, tal y como se ha abordado el problema, se obtienen resultados muy precisos y eficaces pero también suponen un gran esfuerzo. Como se ha comentado ya en la Sección [3.4](#), los cables de alta tensión cuentan con unas longitudes de varios kilómetros y atraviesan

zonas, a veces, muy pobladas de vegetación y otras no tanto. La segmentación de dichos cables se ha realizado a mano, lo que supone una ardua tarea y resulta inviable en un área de estudio mayor. Por esta razón, como mejora futura, se propone utilizar métodos de segmentación automática.

Por otra parte, para automatizar más el proceso, se pueden utilizar métodos de clasificación no-supervisada, como K-means, para hallar los centroides de las clases de tipos de suelo.

En cuanto al tipo de imágenes que se pueden utilizar, parece interesante disponer de imágenes Sentinel-2 de nivel 2A así como de imágenes captadas por drones o de datos LIDAR que nos permitan estimar la altura de "las zonas de peligro" ya detectadas por los índices de vegetación.

Visto la clasificación "erronea" de varios segmentos tanto para el índice NDVI como GNDVI, se puede decir que el hecho de que se haya realizado el estudio en la estación de verano tiene un gran impacto en el resultado final por lo que se debería alargar el tiempo de estudio para poder observar el desarrollo de ambos índices en diferentes climas y así obtener datos más exactos.

Finalmente, en cuanto a los índices de vegetación a utilizar, pueden realizarse experimentos con más índices de vegetación o con la combinación de varios de ellos para la estimación de las zonas de alta vegetación. Concretamente, dado que varios de los cables se encuentran en zonas edificadas, la utilización del índice de vegetación TDVI o *Transformed Difference Vegetation Index* [3] ya que se centra en la detección de coberturas vegetales urbanas.

Bibliografía

- [1] Commercial Geoint Strategy. *National Geointelligence Agency*, Octubre 2015.
- [2] Roger Atkinson. Project management: cost, time and quality, two best guesses and a phenomenon, its time to accept other success criteria. *International Journal of Project Management*, 17(6):337 – 342, 1999.
- [3] A. Bannari, H. Asalhi, and P. M. Teillet. Transformed difference vegetation index (TDVI) for vegetation cover mapping. In *IEEE International Geoscience and Remote Sensing Symposium*, volume 5, pages 3053–3055 vol.5, June 2002.
- [4] Susan Barati, Behzad Rayegani, Mehdi Saati, Alireza Sharifi, and Masoud Nasri. Comparison the accuracies of different spectral indices for estimation of vegetation cover fraction in sparse vegetated areas. *The Egyptian Journal of Remote Sensing and Space Science*, 14(1):49 – 56, 2011.
- [5] Luke Barrington, Shubharoop Ghosh, Marjorie Greene, Shay Har-Noy, Jay Berger, Stuart Gill, Albert Yu-Min Lin, and Charles Huyck. Crowdsourcing earthquake damage assessment using remote sensing imagery. *Annals of Geophysics*, 54(6), 2012.
- [6] Alan S. Belward and Jon O. Skøien. Who launched what, when and why; trends in global land-cover observation capacity from civilian Earth observation satellites. *ISPRS Journal of Photogrammetry and Remote Sensing*, 103:115 – 128, 2015. Global Land Cover Mapping and Monitoring.

- [7] Safa Bousbih, Mehrez Zribi, Zohra Lili-Chabaane, Nicolas Baghdadi, Mohammad El Hajj, Qi Gao, and Bernard Mougenot. Potential of Sentinel-1 radar data for the assessment of soil and cereal cover parameters. *Sensors*, 17(11), 2017.
- [8] Manuel Campos-Taberner, Francisco Javier García-Haro, Gustau Camps-Valls, Gonçal Grau-Muedra, Francesco Nutini, Lorenzo Busetto, Dimitrios Katsantonis, Dimitris Stavrakoudis, Chara Minakou, Luca Gatti, Massimo Barbieri, Francesco Holecz, Daniela Stroppiana, and Mirco Boschetti. Exploitation of SAR and Optical Sentinel data to detect rice crop and estimate seasonal dynamics of Leaf Area Index. *Remote Sensing*, 9(3), 2017.
- [9] Chris Chatfield. *Time-series forecasting*. CRC Press, 2000.
- [10] Pol R Coppin and Marvin E Bauer. Processing of multitemporal Landsat TM imagery to optimize extraction of forest cover change features. *IEEE Transactions on Geoscience and remote Sensing*, 32(4):918–927, 1994.
- [11] Mihai Datcu, Herbert Daschiel, Andrea Pelizzari, Marco Quartulli, Annalisa Galoppo, Andrea Colapicchioni, Marco Pastori, Klaus Seidel, Pier Giorgio Marchetti, and Sergio d’Elia. Information mining in remote sensing image archives: System concepts. *IEEE Transactions on Geoscience and Remote Sensing*, 41(12):2923–2936, 2003.
- [12] Begüm Demir, Francesca Bovolo, and Lorenzo Bruzzone. Updating land-cover maps by classification of image time series: A novel change-detection-driven Transfer Learning approach. *IEEE Transactions on Geoscience and Remote Sensing*, 51(1):300–312, 2013.
- [13] M Drusch, U Del Bello, S Carlier, O Colin, V Fernandez, F Gascon, B Hoersch, C Isola, P Laberinti, P Martimort, et al. Sentinel-2: ESA’s optical high-resolution mission for GMES operational services. *Remote Sensing of Environment*, 120:25–36, 2012.
- [14] M. A. Friedl, D. S. Schimel, J. Michaelsen, F. W. Davis, and H. Walker. Estimating grassland biomass and Leaf Area Index using ground and satellite data. *International Journal of Remote Sensing*, 15(7):1401–1420, 1994.

- [15] Wayne A Fuller. *Introduction to statistical time series*, volume 428. John Wiley & Sons, 2009.
- [16] Anatoly A. Gitelson, Yoram J. Kaufman, and Mark N. Merzlyak. Use of a green channel in remote sensing of global vegetation from EOS-MODIS. *Remote Sensing of Environment*, 58(3):289 – 298, 1996.
- [17] Noel Gorelick, Matt Hancher, Mike Dixon, Simon Ilyushchenko, David Thau, and Rebecca Moore. Google Earth Engine: Planetary-scale geospatial analysis for everyone. *Remote Sensing of Environment*, 202:18–27, 2017.
- [18] Joseph F. Hair, Rolph E. Anderson, Ronald L. Tatham, and William C. Black. *Analisis Multivariante*. Prentice Hall, España:, 5a ed. edition, 1999.
- [19] Yi Hu and Kai Liu. Chapter 2 - Remote sensing and remote measurement technology of transmission lines. In Yi Hu and Kai Liu, editors, *Inspection and Monitoring Technologies of Transmission Lines with Remote Sensing*, pages 37 – 137. Academic Press, 2017.
- [20] L. Häme, J. Norppa, P. Salovaara, and J. Pylvänäinen. Power line monitoring using optical satellite data. In *CIREN Workshop 2016*, pages 1–4, June 2016.
- [21] Andreea Julea, Nicolas Méger, Philippe Bolon, Christophe Rigotti, Marie-Pierre Doin, Cécile Lasserre, Emmanuel Trouvé, and Vasile N Lazarescu. Unsupervised spatiotemporal mining of satellite image time series using grouped Frequent Sequential Patterns. *IEEE Transactions on Geoscience and Remote Sensing*, 49(4):1417–1430, 2011.
- [22] Zhengrong Li, Troy S. Bruggemann, Jason J. Ford, Luis Mejias, and Yuee Liu. Toward automated power line corridor monitoring using advanced aircraft control and multisource feature fusion. *Journal of Field Robotics*, 29(1):4–24, 2012.
- [23] Chi-Jie Lu, Tian-Shyug Lee, and Chih-Chou Chiu. Financial time series forecasting using Independent Component Analysis and Support Vector Regression. *Decision Support Systems*, 47(2):115–125, 2009.

- [24] Zbyněk Malenovský, Helmut Rott, Josef Cihlar, Michael E. Schaepman, Glenda García-Santos, Richard Fernandes, and Michael Berger. Sentinels for science: Potential of Sentinel-1, -2, and -3 missions for scientific observations of ocean, cryosphere, and land. *Remote Sensing of Environment*, 120:91 – 101, 2012. The Sentinel Missions - New Opportunities for Science.
- [25] Dimitrios Marmanis, Mihai Datcu, Thomas Esch, and Uwe Stilla. Deep learning Earth observation classification using Imagenet pretrained networks. *IEEE Geoscience and Remote Sensing Letters*, 13(1):105–109, 2016.
- [26] M Pilar Martin and E Chuvieco. Mapping and evaluation of burned land from multitemporal analysis of AVHRR NDVI images. *EARSeL Advances in Remote Sensing*, 4:7–13, 1995.
- [27] Beatriz Martínez and María Amparo Gilabert. Vegetation dynamics from NDVI time series analysis using the wavelet transform. *Remote Sensing of Environment*, 113(9):1823–1842, 2009.
- [28] Leena Matikainen, Matti Lehtomäki, Eero Ahokas, Juha Hyypä, Mika Karjalainen, Anttoni Jaakkola, Antero Kukko, and Tero Heinonen. Remote sensing methods for power line corridor surveys. *ISPRS Journal of Photogrammetry and Remote Sensing*, 119:10 – 31, 2016.
- [29] S. J. Mills, M. P. Gerardo Castro, Z. Li, J. Cai, R. Hayward, L. Mejias, and R. A. Walker. Evaluation of aerial remote sensing techniques for vegetation management in power-line corridors. *IEEE Transactions on Geoscience and Remote Sensing*, 48(9):3379–3390, Sept 2010.
- [30] S Natali, A Beccati, S D’Elia, MG Veratelli, P Campalani, M Folegani, and S Mantovani. Multitemporal data management and exploitation infrastructure. In *Analysis of Multi-temporal Remote Sensing Images (Multi-Temp)*, 2011 6th International Workshop on the, pages 217–220. IEEE, 2011.
- [31] François Petitjean, Jordi Inglada, and Pierre Gançarski. Satellite image time series analysis under Time Warping. *IEEE transactions on geoscience and remote sensing*, 50(8):3081–3095, 2012.

- [32] A. Qayyum, A. S. Malik, M. N. M. Saad, M. Iqbal, R. F. Ahmad, T. A. R. B. T. Abdullah, and A. Q. Ramli. Monitoring of vegetation near power lines based on dynamic programming using satellite stereo images. In *2014 IEEE International Conference on Smart Instrumentation, Measurement and Applications (ICSIMA)*, pages 1–6, Nov 2014.
- [33] Marco Quartulli and Igor G Olaizola. A review of EO image information mining. *ISPRS journal of Photogrammetry and Remote Sensing*, 75:11–28, 2013.
- [34] J. W. Rouse, R. H. Haas, Schell, Deering J. A., D. W., and J. C. Harlan. Monitoring the vernal advancements and retrogradation of natural vegetation. Technical report, NASA/GSFC, Final Report, Greenbelt, MD, USA, 1974.
- [35] Milindkumar V Sarode and Prashant R Deshmukh. Reduction of speckle noise and image enhancement of images using filtering technique. *International Journal of Advancements in Technology*, 2(1):30–38, 2011.
- [36] Compton J Tucker, John RG Townshend, and Thomas E Goff. African land-cover classification using satellite data. *Science*, 227(4685):369–375, 1985.
- [37] David P. Turner, Warren B. Cohen, Robert E. Kennedy, Karin S. Fassnacht, and John M. Briggs. Relationships between Leaf Area Index and Landsat TM spectral vegetation indices across three temperate zone sites. *Remote Sensing of Environment*, 70(1):52 – 68, 1999.
- [38] Qihao Weng, Dale Quattrochi, and Paolo E Gamba. *Urban remote sensing*. CRC press, 2018.
- [39] Gui-Song Xia, Xin-Yi Tong, Fan Hu, Yanfei Zhong, Mihai Datcu, and Liangpei Zhang. Exploiting deep features for remote sensing image retrieval: A systematic investigation. *arXiv preprint arXiv:1707.07321*, 2017.
- [40] Xiaowei Xie, Zhengjun Liu, Caijun Xu, and Yongzhen Zhang. A multiple sensors platform method for power line inspection based on a large unmanned helicopter. *Sensors*, 17(6), 2017.

- [41] G. Yan, C. Li, G. Zhou, W. Zhang, and X. Li. Automatic extraction of power lines from aerial images. *IEEE Geoscience and Remote Sensing Letters*, 4(3):387–391, July 2007.
- [42] Yong Zhang, Xiuxiao Yuan, Wenzhuo Li, and Shiyu Chen. Automatic power line inspection using UAV images. *Remote Sensing*, 9(8), 2017.

Anexos

Appendix A

Transformación radiométrica para imágenes Sentinel-2

En nuestro proyecto no hemos necesitado ninguna transformación radiométrica ni espacial de las imágenes ya que solo hemos necesitado las reflectancias TOA. No obstante, en este anexo mostramos los pasos a seguir en el caso de que necesitemos una transformación radiométrica de las reflectancias desde el tope de la atmósfera o TOA *Top-Of-Atmosphere* a nivel de superficie o BOA *Bottom-Of-Atmosphere*.

Aunque a partir de mayo de 2017 ESA comenzó a proporcionar productos de nivel 2A para Europa y han reprocesado archivos de ese año, es posible descargar productos de nivel L1C y personalizar la corrección atmosférica utilizando diferentes algoritmos (6S, 6SV, arcsi) o DEM diferentes. Desde mediados de marzo de 2018, los productos de nivel L2A son ya un producto operativo con respecto a la cobertura de la región euromediterránea y se espera una cobertura sistemática a nivel mundial para este verano de 2018.

Por otra parte, los productos de nivel L2A proporcionan reflectancias ortorrectificadas a nivel de superficie (*Bottom-Of-Atmosphere*, BOA) con registro multispectral a nivel sub-píxel. También se incluye un mapa de clasificación de la escena (nubes, sombras de nubes, vegetación, suelo/desierto, agua, nieve, etc.).

APPENDIX A. TRANSFORMACIÓN RADIOMÉTRICA PARA IMÁGENES SENTINEL-270

El paso de reflectancia aparente a real, es decir, a nivel de la superficie (BOA) es un proceso delicado ya que requiere de ciertas variables tales como la geometría de iluminación, el contenido en vapor de agua y el espesor óptico de aerosoles en la atmósfera (AOT). Además, tenemos que incluir el efecto de los cirros o de las nubes traslúcidas. Por todo ello, la calidad de los resultados de una corrección atmosférica depende de la exactitud de los datos de entrada disponibles sobre estas variables y efectos, y del modelo con el que se aplican a la imagen de radiancia o reflectancia TOA. El programa "Sen2Cor" proporcionado por ESA (<http://step.esa.int/main/third-party-plugins-2/sen2cor/>) realiza la transformación de las imágenes L1C a L2A mediante el programa. Este programa usa el algoritmo de corrección atmosférica ATCOR (https://www.rese-apps.com/pdf/atcor3_manual.pdf) y como DEM por defecto, el STRM 90 m (<http://srtm.csi.cgiar.org/>).

1. Importamos los datos teniendo en cuenta que cada banda tiene diferente resolución.
2. Reemplazar los valores NODATA(0) por NULL en la proyección original antes de realizar la retroproyección de los datos si el programa de procesado lo requiere.
3. En los productos L2A, hay artefactos a lo largo de los bordes de las franjas, y los valores son demasiado bajos, causados por un error en el programa "Sen2Cor". Todas las celdas que bordean una celda NULL deben tomar el valor NULL. Este paso puede que sea necesario repetirse dos veces.
4. El tamaño de los archivos se puede reducir ya que los mosaicos (o *tiles*) se superponen y se puede establecer una correspondencia entre los mosaicos vecinos que tengan la misma hora de inicio de la captura de datos.
5. Antes de realizar la retroproyección, se debe configurar correctamente la región de estudio. Es decir, si usamos bandas de diferente resolución, todas ellas deben tener una resolución de 60 m. Por ejemplo, en la proyección original, cuatro celdas de 10 m caen exactamente en una celda de 20 m, y nueve celdas de 20 m caen exactamente en una celda de 60 m. Esto se puede lograr estableciendo primero la región al *bounding box*, luego alineando la región a una resolución de 60 m. Posteriormente, debemos ajustar la resolución de la región a la resolución real de la banda de entrada, en este momento sin alineación. Otra opción es crear una cuadrícula de referencia y alinear la región de estudio a la cuadrícula de referencia.

APPENDIX A. TRANSFORMACIÓN RADIOMÉTRICA PARA IMÁGENES SENTINEL-271

6. Retroproyectar los datos preferiblemente con un método de interpolación bicúbico o Lanczos para evitar "artefactos" (tales como dientes de sierra) que pueden aparecer si utilizamos el método del "vecino más cercano".
7. Corrección atmosférica: Se aplica a cada banda de forma separada el algoritmo de corrección atmosférica. No obstante, generalmente se deben especificar los parámetros necesarios para esta corrección (parámetros 6S):
 - Condiciones geométricas (para Sentinel-2A, el valor es 25, y para Sentinel-B, el valor es 26)
 - Mes, día, hora en GMT, longitud y latitud
 - Visibilidad o profundidad óptica del aerosol (*Aerosol Optical Depth*, AOD): No se mide la visibilidad para las escenas Sentinel-2 y hay que estimar la profundidad óptica del aerosol a partir de los datos de alguna estación cercana a la zona de estudio.
 - Elevación media sobre el nivel del mar.
 - Altura del sensor.
 - Selección de la banda del sensor y del modelo atmosférico

Appendix B

Preprocesado de las imágenes Sentinel-1

Los pasos a realizar son los siguientes:

- Filtrar las imágenes con una polarización dual. La polarización dual que se utiliza es VV + VH. La variable VV indica que la energía se transmite y se recibe verticalmente polarizada. Por otro lado, VH significa que transmite verticalmente pero recibe la energía horizontalmente polarizada.
- Filtrar las imágenes recogidas en modo banda ancha interferométrica. El modo de barrido Interferométrico Ancho (IW), *Interferometric Wide (IW) swath mode*, es el modo de adquisición principal sobre tierra y satisface la mayoría de los requisitos de servicio. Adquiere datos con una franja de 250km a una resolución espacial de 5m por 20m (simple vista). El modo IW captura tres franjas secundarias utilizando Terrain Observation with Progressive Scans SAR (TOPSAR) que se trata de una forma de escaneo de imágenes ScanSAR, donde los datos se adquieren en ráfagas al cambiar cíclicamente el haz, conjunto de rayos luminosos con el mismo otigen, de la antena entre múltiples franjas adyacentes. Con la técnica TOPSAR, además de dirigir el haz en el rango como en ScanSAR, éste también se dirige electrónicamente de atrás hacia

adelante en la dirección acimutal para cada ráfaga, evitando el ondeado y dando como resultado una calidad de imagen homogénea en toda la franja.

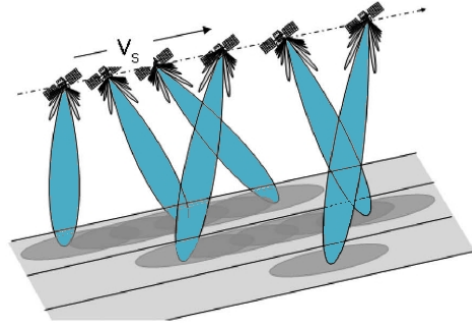


Figure B.1: Adquisición TOPSAR

- Filtrar la imagen para captarla desde diferentes ángulos.
- Crear una colección a partir de las medias de las diferentes polarizaciones y de los diferentes ángulos.

Una vez realizado este paso inicial, los pasos de preproceso que da *Google Earth Engine* para derivar el coeficiente de retrodispersión de cada píxel. El coeficiente de retrodispersión representa el área de retrodispersión del objetivo (sección transversal del radar) por área de superficie de la unidad, dicho de forma más sencilla, mide la intensidad en decibelios del eco reflejado. Para calcular dicho coeficiente se seguirán los siguientes pasos:

- Aplicar el archivo de órbita (utilizando órbitas restituidas). Actualiza los metadatos de la órbita con un archivo de órbita restituido. Dichos archivos de órbita son accesibles desde la página de Sentinel.
- Eliminar el ruido del borde GRD, *Ground Range Detected* o Detectado Rango de Tierra. Elimina el ruido de baja intensidad y los datos no válidos en los bordes de la escena. Eliminar el ruido termal. Se quita el ruido aditivo en sub-franjas para ayudar a reducir las discontinuidades entre las franjas secundarias para escenas en modos de adquisición de franjas múltiples.

- Calibración radiométrica. Se calcula la intensidad de la retrodispersión utilizando los parámetros de calibración del sensor en los metadatos de GRD.
- La corrección del terreno (ortorectificación) convierte datos de geometría de rango de tierra a coeficientes de retrodispersión.

Appendix C

Instalación de librería Python Earth Engine

Como se ha explicado en el Capítulo 1, la instalación se ha hecho sobre un sistema operativo Ubuntu 16.04 y los pasos a realizar son los siguientes:

- Comprobar que la versión de Python que tenemos es la 2.6 o 2.7 y, en caso de no ser así, instalarla mediante el sistema de gestión de paquetes *pip*.

```
sudo apt-get install python-pip
```

- Instalar *Google APIs Client Library*.

```
sudo apt-get install python-pip
```

- Asegurarse de que existe una librería criptográfica disponible. Si no se encuentra ninguna, existen varias opciones pero la que hemos utilizado en este caso es la li-

brería *pyCrypto*.

```
python -c "from oauth2client import crypt"
# Si este comando devuelve un error, ejecutar el siguiente comando.
sudo pip install pyCrypto
```

- Instalar *Earth Engine Python API* mediante *Python Package Index*.

```
sudo pip install earthengine-api
```

- Crear el archivo de credenciales para autorizar el acceso a *Earth Engine*.

```
python -c "import ee; ee.Initialize()"
# Si este comando no da error, existe el fichero de credenciales y
# no habrá que continuar con los siguientes pasos.
```

Las API de *Earth Engine* usan el protocolo *OAuth 2.0* para autenticar clientes. Para autenticarse, primero hay que configurar un archivo de credenciales en el ordenador que autorice el acceso a *Earth Engine* en nombre de su cuenta de Google. Este proceso se activa llamando al método *ee.Initialize()*. Este comando devuelve un error ya que no se dispone de dicho archivo de credenciales. Se debe copiar el comando que se muestra en el error en una terminal nueva y al ejecutarlo, se inicia la creación del archivo de credenciales. Como se ha dicho anteriormente, debemos tener una cuenta en *Google Earth Engine* y una vez registrados en la página, podremos ver un código que nos autoriza el acceso a los datos de *Earth Engine*. Copiando dicho código en la terminal donde se está creando el archivo de credenciales se finaliza la instalación.

- Comprobar que la instalación es correcta.

```
# Importar Earth Engine Python Package
import ee
```

```
# Inicializar el objeto Earth Engine usando las credenciales de
autenticación.
ee.Initialize()

# Imprimir la información de una imagen.
image = ee.Image('srtm90_v4')
print(image.getInfo())
```

Appendix D

Relación estructura de datos GEE y QGIS

```
# Selección de la capa activa (QGIS)
layer = qgis.utils iface.activeLayer()
# Guardar todas las QgsFeature seleccionadas en una QgsFeatureList (QGIS)
selected_features = layer.selectedFeatures()
# Seleccionar imagenes Sentinel 2 (GEE)
sentinelCollection = ee.ImageCollection('COPERNICUS/S2');
# Filtrar las imagenes Sentinel 2 por fecha (GEE)
sentinelDateCollection = ee.ImageCollection(sentinelCollection
    .filterDate('2017-06-21', '2017-09-23'))
# Seleccionar la primera QgsFeature de la QgsFeatureList (QGIS)
p = selected_features[0]
# Extraer la geometria de la QgsFeature seleccionada (QGIS)
g = p.geometry()
# Exportar la geometria a formato GeoJSON (QGIS)
gJSON = g.exportToGeoJSON()
o = json.loads(gJSON)
# Utilizar las coordenadas extraidas para crear un poligono en GEE (GEE)
b = ee.Geometry.Polygon(o['coordinates'])
# Filtrar la coleccion de imagenes por area (GEE)
sentinelAOI = ee.ImageCollection(sentinelDateCollection.filterBounds(b))
```


Appendix E

Código

```
import ee
import matplotlib.pyplot as plt
import matplotlib.pyplot as pl
import numpy as np
import matplotlib.cm as cm
import time
import json
import datetime
import copy
import shutil
import pandas as pd
from PyQt4.QtGui import QColor
from PyQt4.QtGui import *
from PyQt4.QtCore import *
from scipy.stats.stats import pearsonr
from datetime import datetime, timedelta
from datetime import date
from sklearn.metrics import confusion_matrix

ee.Initialize()
start_time = time.time()

layer = qgis.utils.iface.activeLayer()
```

```
print("Nombre de layer", layer.name())

#features seleccionadas
selected_features = layer.selectedFeatures()
count= len(list(selected_features))
print("Feature number", count)
#Filtrar las fotos por fecha
sentinelCollection = ee.ImageCollection('COPERNICUS/S2');
sentinelDateCollection = ee.ImageCollection(sentinelCollection.filterDate('
    2017-06-21', '2017-09-23'))

i=0
val = [None] * count
valGNDVI = [None] * count
etiquetas = []
clases= []
datak = [None] * count
fechasRep = [None] * count
dife= [None] * count
datosBosque=[]
datosCesped=[]
hayBosque= False
hayCesped= False
featuresNull = []
featuresNotNull=[]

#analizar cada feature
while i < count:

    p = selected_features[i]
    #crear un vector con el nombre de cada feature. p[1] el el atributo de
    descripcion
    etiquetas.append(p[1])

    print("ETIQUETAS", etiquetas)
    g = p.geometry()
    gJSON=g.exportToGeoJSON()
    o=json.loads(gJSON)
```

```

b = ee.Geometry.Polygon(o['coordinates'])
#filtrar la coleccion de imagenes por la geometria de la feature que se
esta analizando
sentinelAOI = ee.ImageCollection(sentinelDateCollection.filterBounds(b))
print("Imágenes filtradas por feature", sentinelAOI.size().getInfo())
imagenesConNubes = [item.get('id') for item in sentinelAOI.getInfo().get('
features')]

#crear dos arrays para separar las features sin nombre de las nombradas(
mar, arena, bosque...)
if p[1]==NULL:
    featuresNull.append(p)
else:
    featuresNotNull.append(p)
values=[]
valuesGNDVI=[]
datas=[]
diferencias=[]
fechasR = []
imagenes=[]
#crear coleccion de imagenes sin nubes
for icn in imagenesConNubes:
    icnindex= imagenesConNubes.index(icn)
    cloudPercentage=ee.Image(imagenesConNubes[icnindex]).get('
CLOUDY_PIXEL_PERCENTAGE').getInfo()
    if cloudPercentage< 10.0:
        imagenes.append(icn)
    else:
        print("IMAGEN CON MUCHAS NUBES")
print("IMÁGENES SIN NUBES", len(imagenes))
#analizar cada imagen de la coleccion
for image in imagenes:
    print 1
    #conseguir la fecha de recepcion de la imagen
    date1 = ee.Image(image).get('system:time_start')
    indexPhoto= imagenes.index(image)
    dat= date.fromtimestamp(date1.getInfo() / 1e3)
    datas.append(dat)

```

```
# calcular NDVI
nir = ee.Image(image).reduceRegion(ee.Reducer.mean(), o).getInfo().
get('B8')
red = ee.Image(image).reduceRegion(ee.Reducer.mean(), o).getInfo().
get('B4')

#en caso de que nir o red sean nulos, darle el mismo valor que el
contrario para formzar que
#NDVI sea 0 e imputar los valores posteriormente

if nir == None:
    print("NIR IS NONE")
    nir = red
if red == None:
    print("RED IS NONE")
    red = nir

if nir == 0 and red == 0:
    ndvi= 0
else:
    ndvi = ((nir - red)/(nir + red))

values.append(ndvi)

#GNDVI
green = ee.Image(image).reduceRegion(ee.Reducer.mean(), o).getInfo()
.get('B3')
if green == None:
    print("GREEN IS NONE")
    green = nir

if nir == 0 and green == 0:
    gndvi= 0
else:
    gndvi = ((nir - green)/(nir + green))

valuesGNDVI.append(gndvi)
```

```

etiquetasR= []

#Cuando hay dos fotos de la misma fecha

if indexPhoto>0 :

    a= copy.copy(datas[indexPhoto])
    b= copy.copy(datas[indexPhoto-1])

    if a == b: #si hay dos fechas iguales
        print("FECHA REPETIDA", a)
        fechasR.append(a)
        etiquetasR.append(p[1])
        if p[1]=='bosque':
            hayBosque= True
            print("Bosque index i e index i-1", values[indexPhoto],
values[indexPhoto-1])
            datosBosque.append(max([values[indexPhoto], values [
indexPhoto-1]]))
        elif p[1]=='cesped':
            hayCesped= True
            datosCesped.append(max([values[indexPhoto], values [
indexPhoto-1]]))

        dif= abs(values[indexPhoto]-values[indexPhoto-1])
        diferencias.append(dif)

#Todas las imagenes procesadas de una feature
dife[i]=diferencias

fechasRep[i]=fechasR
if i == 0:
    tamanoColumnas= datas
else:
    if len(tamanoColumnas)< len(datas):
        tamanoColumnas= datas

```

```

datak[i]=datas
val[i]=values
print ("FECHA Y DATOS" , datak[i] , val[i])

valGNDVI[i]= valuesGNDVI

# Grafica de la diferencia de valores para imagenes tomadas en la misma
fecha

h=plt.figure(1)
plt.plot(fechasRep[i][:], dife[i][:], color='silver', linewidth=3, alpha
=0.8)
pr=[str(j) for j in fechasRep[i][:]]
plt.xticks(fontsize=26)
ax = h.add_subplot(111)
ax.set_xlabel('Fechas repetidas', fontsize=30)
ax.set_ylabel('Diferencia NDVI', fontsize=30)

i+=1
#Deberia de estar seleccionado el feature de bosque y cesped para
cuantificar la diferencia
if hayBosque and hayCesped:

    print("DATOS DATOSBOSQUE ", datosBosque )
    print("DATOS DATOSCESPED", datosCesped)

    difCB= [abs(x - y) for x, y in zip(datosBosque, datosCesped)]
    print("DIFERENCIAS BOSQUE Y CESPED", difCB)
    plt.plot(fechasRep[0][:], difCB, color='red', linewidth=3, alpha=0.8)
h.show()

#Crear fichero
print("DATA Y COLUMNAS" , val[:], tamañoColumnas)

```

```

fp1= open( '/tmp/datosTimeBased.txt ', 'w')
fp2= open( '/tmp/datosSuma.txt ', 'w')
fp3= open( '/tmp/datosPrueba.txt ', 'w')

#crear un dataframe con los valores NDVI calculados, los index seran las
    etiquetas
#y las columnas seran las fechas.
fp= open( '/tmp/datos.txt ', 'w')
fechasColumnas=[str(i) for i in tamanoColumnas]
df = pd.DataFrame(data=val [:], columns=fechasColumnas)
df[ 'etiqueta ' ] = etiquetas

print( "REPLACE" )
#buscar los valores 0 y sustituirlos por n.a para hacer interpolacion linear
dfn= df.replace(0, np.nan)
dfni=dfn.interpolate()
numrows=len( dfni.index )

dfni=dfni.sort_values( by=[ 'etiqueta ' ], na_position='first ' )
dfnicopia=dfni.copy()
nulls=len( featuresNull )
nonnulls=len( featuresNotNull )
fp7= open( '/tmp/dfnnull.txt ', 'w')
fp8= open( '/tmp/dfnonnull.txt ', 'w')
fp9= open( '/tmp/dfnicopia.txt ', 'w')

dfniNullcopia=dfnicopia.head( nulls )
dfniNullcopia[ 'oldindex ' ]= dfniNullcopia.index.values
dfniNullcopia= dfniNullcopia.sort_index( axis='index ' )
dfniNullcopia = dfniNullcopia.reset_index( drop=True )

dfniNull=dfni.head( nulls )
dfniNull= dfniNull.sort_index( axis='index ' )
dfniNull = dfniNull.reset_index( drop=True )

dfniNoNullcopia=dfnicopia.tail( nonnulls )
dfniNoNullcopia[ 'oldindex ' ]= dfniNoNullcopia.index.values

```

```

dfniNoNullcopia.index = range(nulls , numrows)

dfniNoNull=dfni.tail(nonulls)
dfniNoNull.index = range(nulls , numrows)

fp7.write(dfniNull.to_string())
fp8.write(dfniNoNull.to_string())
fp7.close()
fp8.close()
frames = [dfniNull , dfniNoNull]
framescopia=[dfniNullcopia , dfniNoNullcopia]
resultdf = pd.concat(frames)
resultdfcopia = pd.concat(framescopia)
fp9.write(resultdfcopia.to_string())
fp9.close()
fp.write(resultdf.to_string())

#crear tabla de distancias
#Se crea un fichero para cada paso para ver si hace correctamente

tablaDistancias = []
#valoresresta = []

for elements in range(0, len(featuresNull)):
    clases.append(featuresNull[elements][3])
    nombreColumnas=[]
    for noNullElements in range(len(featuresNull), len(resultdf.index)):
        nombreColumnas.append(resultdf['etiqueta'].iloc[noNullElements])
        print("NombreColumnas", nombreColumnas)
        resultdfsn=resultdf.drop('etiqueta', axis=1)

    prueba1 = resultdfsn.iloc[[noNullElements]].values
    prueba2 = resultdfsn.iloc[[elements]].values
    prueba=[x - y for x, y in zip(prueba1, prueba2)]
    pruebadf = pd.DataFrame(data=prueba, columns=tamanoColumnas)

fp3.write(pruebadf.to_string())

```



```
diffTsa= pruebaadf.abs()
fp1.write(diffTsa.to_string())
r= np.sum(diffTsa.values)
print("SUMA FILA", r)
tablaDistancias.append(r)

fp.close()
fp3.close()
fp1.close()
fp2.close()
print("TABLA DISTANCIAS", tablaDistancias)
fp4= open('/tmp/distancias.txt', 'w')

print("TODOS LOS VALORES BIEN ", tablaDistancias, nombreColumnas)

splitted= np.array_split(tablaDistancias, len(featuresNull))

dist = pd.DataFrame(data=splitted, columns=nombreColumnas)

fp4.write(dist.to_string())
fp4.close()
fp5= open('/tmp/pruebaminimo.txt', 'w')
dist['Clase'] = dist.loc[:, ['arena', 'bosque', 'cesped', 'edificios', 'mar']
    ].idxmin(axis=1)
distR=dist.iloc[:, :-1]
fp5.write(dist.to_string())
fp5.close()
fpalarms= open('/tmp/alarms.txt', 'w')
alarms=dist.loc[dist['Clase'] == 'bosque']
fpalarms.write(alarms.to_string())
fpalarms.close()
if alarms.empty== False:
    indicesalarmas=alarms.index.values.tolist()
    indicesalarmasdf=resultdfcopia.iloc[indicesalarmas,:]
    numerodefeature= indicesalarmasdf['oldindex'].values.tolist()
    print(numerodefeature)
```

```
print( numerodefeature )
featurescambiar = []
for f in range (0, len( numerodefeature )):
    numact = numerodefeature [ f ]
    featurescambiar .append( selected_features [ numact ])

iface .mapCanvas() .setSelectionColor( QColor( "red" ) )
ids = [ ides .id() for ides in featurescambiar ]
layer .setSelectedFeatures( ids )
layer .triggerRepaint ()
else :
    print( "NO HAY ZONAS DE PELIGRO" )

predictedClasses = list( dist .iloc [ :, -1 ] )

g = plt .figure (2)

#Confusion Matrix
print( "Confusion Matrix" )
print( confusion_matrix( clases , predictedClasses ) )

#Grafica valores NDVI

area = (30 * np.random.rand(50))**2

for j in range (0, count):

    if selected_features [j][1] == 'mar':
        print( "ES MAR" )
        plt .plot( datak [j], val [j], color="blue", linewidth=3)
    elif selected_features [j][1] == 'bosque':
        print( "ES Bosque" )
        plt .plot( datak [j], val [j], color="darkgreen", linewidth=3)
    elif selected_features [j][1] == 'cesped':
        print( "ES cesped" )
        plt .plot( datak [j], val [j], color="olivedrab", linewidth=3)
```

```
elif selected_features[j][1]== 'edificios':
    print("ES edificios")
    plt.plot(datak[j], val[j], color="sienna", linewidth=3)
elif selected_features[j][1]== 'arena':
    print("ES arena")
    plt.plot(datak[j], val[j], color="yellow", linewidth=3)

else:
    plt.plot(datak[j], val[j], color="silver", linewidth=3, alpha=0.8)

plt.xticks(fontsize=26)
ax = g.add_subplot(111)
ax.set_xlabel('Valores verano 2017', fontsize=30)
ax.set_ylabel('NDVI value', fontsize=30)
g.show()
#h.show()

#CORRELACION

plt.figure(3)
plt.subplot(1,1,1)
plt.plot(val, valGNDVI, 'bo')
plt.xlabel('NDVI')
plt.ylabel('GNDVI')
print np.corrcoef(valGNDVI, val)
plt.show()

print("—— %s seconds ——" % (time.time() - start_time))
```