



Facultad de Ciencias

VISUALIZACIÓN DE DATOS DE SATÉLITE EN 3D

(3D Visualization of Satellite Data)

Trabajo de Fin de Máster
para acceder al

MÁSTER EN CIENCIA DE DATOS

Autor: Razvan Pascal

Director/es: Fernando Aguilar Gómez,

Daniel García Díaz

Agradecimientos

Me gustaría agradecer a mis padres por ofrecerme la oportunidad de seguir formándome, un año más, por su confianza y su apoyo incondicional en todo momento.

También a los que han sido los directores de este trabajo, Fernando y Daniel, por su amabilidad y dedicación, siempre dispuestos a ayudarme y a resolver todas las dudas y problemas que surgieron durante la realización de este estudio. Sin ellos no hubiese salido adelante.

Resumen

El aumento de la cantidad y complejidad de los datos que se generan actualmente a veces supera nuestra habilidad de extraer información y conocimiento de ellos. En este trabajo se estudia la posibilidad de visualizar conjuntos de datos multidimensionales obtenidos a partir de imágenes de satélite.

La metodología propuesta cubre todo el proceso desde la obtención de los datos hasta su visualización tridimensional. Se ha combinado la visualización de modelos digitales de elevación con índices derivados de imágenes de las misiones espaciales Sentinel-2 y Landsat-8.

Los resultados obtenidos permiten explorar y establecer relaciones espaciotemporales de los datos de forma fácil e intuitiva, lográndose el objetivo establecido.

Palabras clave: datos de satélite, visualización 3D, Sentinel-2, Landsat-8, embalse, Cuerda del Pozo, índices medioambientales, teledetección.

Abstract

The increase in the amount and complexity of data currently generated, sometimes exceeds our ability to extract information and knowledge from it. In this paper we study different ways of visualizing multidimensional datasets obtained from satellite images.

The proposed methodology covers the entire process from data acquisition stage to three-dimensional visualization. A multi-method visualization technique is used, combining digital elevation models with remote sensing indices from Sentinel-2 and Landsat-8 missions.

The obtained results allow users to explore and establish spatio-temporal relationships of the data in an easy and intuitive way, achieving the established objective.

Keywords: satellite data, 3D visualization, Sentinel-2, Landsat 8, reservoir, Cuerda del Pozo, environmental index, remote sensing.

Tabla de contenido

CAPÍTULO 1. INTRODUCCIÓN.....	8
1.1 Motivación	8
1.2 Monitorización terrestre.....	11
1.2.1 Programa Copernicus	11
1.2.2 Programa Landsat	13
1.2.3 Shuttle Radar Topography Mission (SRTM)	15
CAPÍTULO 2. OBJETIVO	17
CAPÍTULO 3. CASO DE ESTUDIO	18
CAPÍTULO 4. METODOLOGÍA	19
4.1 Descarga de los datos	19
4.2 Procesado y transformación de los datos	26
4.3 Generación de las visualizaciones	29
4.4 Vistas generadas.....	33
4.4.1 MNDWI	33
4.4.2 NDVI.....	34
4.4.3 NDTI	35
4.4.4 Temperatura	35
CAPÍTULO 5. CONCLUSIONES Y TRABAJOS FUTUROS.....	37
CAPÍTULO 6. BIBLIOGRAFÍA.....	39

Listado de Figuras

Figura 1. Cronología de la serie de misiones espaciales Landsat.	14
Figura 2. Mapa de la cobertura geográfica de la misión SRTM. Fuente: [17]	15
Figura 3. Esquema de funcionamiento del sistema InSAR de la misión SRTM. Fuente: [17]	16
Figura 4. Imagen del embalse Cuerda del Pozo obtenida mediante el satélite Landsat	18
Figura 5. Proceso abstracto de generación de representaciones gráficas. Fuente: [3].	19
Figura 6. Imagen del modelo de elevaciones SRTM en el embalse Cuerda del Pozo obtenido a través de Google Earth Engine	21
Figura 7. Reflectancia de diferentes tipos de superficie. Fuente: [34]	22
Figura 8. Presencia de nubes en las ortofotos obtenidas de la misión espacial Sentinel-2.	24
Figura 9. Comparación de la aplicación de máscaras de nubes en 2D y 3D.	25
Figura 10. Presencia de nubes en las ortoimágenes obtenidas de la misión espacial Landsat-8.	26
Figura 11. Representación de la estructura dimensiona de una variable. Fuente: [37]	27
Figura 12. Proceso de mapeado de la visualización utilizado en este trabajo.....	29
Figura 13. Representación de la densidad de flujo mediante diferentes mapas de colores. Fuente: [38]	30
Figura 14. Asignación de las coordenadas UV en un modelo tridimensional. Fuente: [38]	31
Figura 15. Visualización 3D del embalse con una textura con los colores del espectro visible.....	32
Figura 16. Representación tridimensional de los datos de simulación del modelo hidrológico.....	33
Figura 17. Visualización del índice MNDWI en diferentes fechas.	34
Figura 18. Variación del volumen de agua del embalse Cuerda del Pozo según mediciones in-situ. Fuente: [44]	34
Figura 19. Visualización del índice NDVI en diferentes fechas.....	35

Figura 20. Visualización del índice NDTI en diferentes fechas..... 35
Figura 21. Visualización de la temperatura superficial en diferentes fechas. 36

Listado de tablas

Tabla 1. Tipos de productos que ofrecen SENTINEL-2.Fuente: [15] 12
Tabla 2. Características del instrumento multiespectral de los satélites Sentinel-2 en sus diferentes bandas. Fuente: [15] 13
Tabla 3. Características de los instrumentos de la misión espacial Landsat-8..... 15

Capítulo 1. Introducción

1.1 Motivación

La percepción y la cognición humana se consideran altamente correlacionadas, permitiendo que investigadores, analistas, ingenieros e incluso una audiencia sin conocimientos técnicos, obtengan información sobre estos datos de manera eficiente y efectiva al detectar características o patrones visuales [1] [2].

El cerebro humano está constantemente procesando enormes cantidades de datos e información de forma simultánea. Esto permite que seamos capaces de percibir y procesar atributos visuales como el color, tamaño, proximidad o movimiento de forma muy rápida. Si estos atributos visuales se mapean a los valores de los variables, somos capaces de detectar de forma casi instantánea cuáles son los valores mínimos y máximos, tendencias, patrones, dependencias o agrupaciones, en gran cantidad de datos, incluso diversos. Gracias a la habilidad de percepción, la visualización permite entender sistemas complejos, tomar decisiones o encontrar información que de otra forma pasaría desapercibida. [3]

La visualización de los datos es una parte fundamental dentro del “*Data science*”, sin embargo, debido a que existen muchas formas de representarlos, a veces puede ser complicado elegir la mejor opción. En la literatura se encuentran numerosos estudios que tratan de ofrecer diferentes clasificaciones de las técnicas de visualización a utilizar en función de varias características de los datos con el objetivo de ofrecer una sistematización y un marco de referencia que facilite la toma de decisiones.

Una primera clasificación se presenta en Card [4], que distingue entre *visualización científica* y *visualización de la información*. La primera trata de representar datos que tienen una correspondencia física con el objetivo de visualizar fenómenos naturales o experimentos científicos. Los datos pueden ser obtenidos a través de sensores, simulaciones o pruebas de laboratorio, y se pueden mapear algunas de sus características directamente en coordenadas geográficas o entornos tridimensionales. Se aplica en campos como la medicina para visualizar las imágenes obtenidas mediante escáneres de tomografía o de resonancia magnética; en química y biología para visualizar estructuras moleculares; en geografía y ciencias atmosféricas para la representación del terreno y los diferentes fenómenos climáticos o meteorológicos; o en la física para visualizar la formación de estrellas o el estudios de las ondas gravitacionales [5]. Por otra parte, la *visualización de la información* trata de representar datos abstractos en la naturaleza que no tienen una correspondencia espacial explícita.

Se suele aplicar en campos como las finanzas representando fluctuaciones en el mercado; el marketing para visualizar estudios de mercado, o en sociología para mapear diferentes características demográficas, entre otros [3].

Dentro de la visualización científica, los datos se pueden clasificar según la dimensionalidad asociada a cada valor de la muestra como escalares, vectores o tensores. Un campo escalar asocia un escalar a cada punto de muestra, y a menudo se usa para indicar algunas propiedades físicas como la temperatura, la presión y la densidad. Para un campo vectorial, la cantidad es especificada por un vector, dando una dirección y una magnitud, como la velocidad y el campo magnético. Los campos de tensor definen un tensor en cada punto de muestra, que puede expresarse como una matriz multidimensional, como por ejemplo las tensiones que existen en una pieza mecánica. Dependiendo de la dimensionalidad de los datos se requieren métodos de visualización concretos [6] [7] [8]:

- **Escalares:** se utilizan técnicas como mapas de altura, curvas de contorno, mapas de colores, o superficies malladas en el caso de datos 2D e se utilizan isosuperficies o renderización de volúmenes en el caso 3D.
- **Vectores:** se utilizan glifos, líneas de corriente o líneas de convolución integral.
- **Tensores:** para poder visualizar este tipo de datos se suelen descomponer en vectores y valores propios y se utilizan las técnicas asociadas a vectores y escalares.

Keim [9] y Schulz [10] propone una clasificación en función del objetivo que se pretende conseguir a través de la visualización, distinguiendo tres categorías:

- **Análisis exploratorio:** En este caso se utiliza la visualización para tratar de obtener una hipótesis a partir de un conjunto de datos desconocidos en base a patrones o relaciones entre los datos. Un ejemplo puede ser la herramienta propuesta por Ferreira [11] que permite analizar los datos de los taxis de Nueva York tratando de definir una hipótesis acerca del transporte urbano.
- **Análisis confirmatorio:** Mediante este tipo de análisis se pretende hacer uso de la visualización para confirmar o rechazar una hipótesis asociada a un determinado conjunto de datos. Un ejemplo es la aplicación de análisis visual realizada por Kumpf [12], que permite visualizar el nivel de confianza de predicciones climáticas.
- **Presentación:** En esta tercera categoría la visualización se aplica para la comunican efectiva y eficaz de los resultados obtenidos en la fase de análisis confirmatorio.

Según van Wijk [2], la visualización en la presentación de resultados es igual de importante que en la fase de análisis, ya que es una herramienta que permite la comunicación de ideas complejas de forma clara, precisa y eficiente.

En los últimos años, la cantidad de datos generados ha aumentado de forma drástica, así como su complejidad, dando lugar a que en muchas disciplinas los datos se hayan vuelto multifacéticos [8]. Según los autores las principales facetas a considerar son las siguientes:

- **Espacio-temporales:** datos que representan estructuras espaciales y/o procesos dinámicos;
- **Multivariadas:** datos de diferentes atributos como la temperatura o la presión;
- **Multimodales:** datos provenientes de diferentes fuentes de adquisición;
- **Multirun (ensembles):** datos derivados de múltiples simulaciones con diferentes configuraciones de parámetros;
- **Multimodelo:** datos resultantes de modelos de simulación acoplados que representan fenómenos de interacción física.

Dada esta nueva complejidad, las representaciones 2D tradicionales que van desde tablas de datos en bruto a diferentes tipos de gráficos o diagramas se ven limitadas en la cantidad de información o conocimiento que puedan aportar. Cuando los datos tienen tablas N-dimensionales, tienen una componente espacial tridimensional o representan un objeto en movimiento, la visualización en 3D puede ser la única forma de extraer información.

En el caso de la faceta espacio-temporal se pueden representar las relaciones de los datos en un determinado dominio geográfico y observar los cambios que se producen en el tiempo, tratando de descubrir patrones como eventos especiales o comportamientos que se repiten.

Si además los datos contienen múltiples atributos por cada localización espacio-temporal, se llega a un límite en el que es imposible poder representar todos en el mismo espacio, siendo necesario hacer uso de herramientas interactivas de visualización. Mazza [3] define tres tipos de interacción en las representaciones: **estáticas**, no permite al usuario ningún tipo de interacción; **manipulable**, permitiendo la interacción con la vista generada mediante desplazamiento, rotación o zoom; y **transformable**, pudiendo el usuario filtrar de los datos de entrada y modificando por tanto el resultado de la vista.

1.2 Monitorización terrestre

La monitorización terrestre permite recopilar información sobre los sistemas físicos, químicos y biológicos de la Tierra, pudiendo evaluar el estado y los cambios tanto en el entorno natural como en el creado por el hombre. Esta información puede dar lugar a aplicaciones como: predicción, adaptación y mitigación del cambio climático; predicción del tiempo meteorológico; monitorización y respuesta ante desastres naturales; cambios en el uso de la tierra (deforestación); etc.

Estas observaciones pueden realizarse de forma remota haciendo uso de satélites, mediante estaciones de monitorización in situ, o una combinación de ambos métodos ya que se complementan. Mientras que la teledetección ofrece una vista global del fenómeno, cubriendo un dominio espacial y temporal amplio, las mediciones in situ proporcionan medidas localizadas con una gran resolución temporal [13]. Además de las componentes espaciales y temporales, los sensores pueden capturar varios parámetros a la vez, que a su vez pueden derivar nuevos atributos, dando lugar a conjuntos de datos multidimensionales, que necesitan de técnicas de visualización específicas.

1.2.1 Programa Copernicus

Uno de los programas más importantes de observación de la Tierra es el programa Copernicus [14], liderado por la Comisión Europea y la Agencia Espacial Europea. Este sistema de monitorización tiene como fuentes de obtención de datos tanto satélites de observación de la Tierra como sensores in situ (estaciones terrestres, sensores aéreos y sensores marinos). El objetivo del programa es proporcionar a los usuarios datos e información exacta, continua, de fácil acceso y de forma gratuita a través de un conjunto de servicios que abordan seis áreas temáticas: la superficie terrestre, los océanos, la atmosfera, el cambio climático, la gestión de emergencias y la seguridad.

La Agencia Espacial Europea está desarrollando cinco misiones espaciales denominadas SENTINEL con el objetivo de atender las diferentes necesidades del programa Copernicus y reemplazar los satélites de misiones anteriores que se han retirado o que lo van a hacer próximamente para de esta forma dar continuidad a los datos. Cada una de las misiones se centra en un aspecto diferente de la observación de la tierra, siendo estas: **SENTINEL-1**: proporciona imágenes radar de la superficie terrestre y oceánica en cualquier condición meteorológica, de día o de noche; **SENTINEL-2**: proporciona imágenes ópticas de alta resolución de la tierra, principalmente de la vegetación, el suelo y las zonas costeras; **SENTINEL-3**: se centra en la observación marina, concretamente en la topografía de la superficie del mar, la temperatura en la superficie del mar y la tierra, y el color de la superficie del océano y la tierra; **SENTINEL-4**,

SENTINEL-5: monitorizarán la calidad del aire atmosférico desde diferentes órbitas. Para este trabajo los se utilizarán datos de las misiones SENTINEL-2 por lo que se desarrollan más en detalle sus características. [14]

SENTINEL-2

A través de la misión SENTINEL-2, compuesta por dos satélites gemelos (2A y 2B) se proporcionan imágenes multiespectrales de gran resolución. Los satélites operan de forma simultánea en fases de 180° entre sí, en una órbita heliocéntrica a una altitud media de 786 km, permitiendo una mayor resolución. La cobertura espacial de esta misión abarca desde una latitud de 56°S hasta 84°N con una resolución temporal de 5 días en el ecuador, teniendo en cuenta los dos satélites.

Para la toma de datos se utiliza un instrumento multiespectral (MSI, por sus siglas en inglés), siendo éste un tipo de sensor pasivo capaz de recoger la luz solar reflejada en la superficie terrestre. La adquisición de los datos se realiza a medida que el satélite se mueve a lo largo de su trayectoria orbital, siendo capaz de realizar el barrido con una anchura de franja de 290 km. El haz de luz que captura el sensor es capaz de distinguir 13 bandas espectrales desde el visible e infrarrojo cercano (VNIR) hasta el infrarrojo de onda corta (SWIR). En la Tabla 2 se muestra la longitud de onda, la anchura de banda y la resolución espacial para cada una de las bandas de los dos satélites SENTINEL-2.

Como se ha mencionado anteriormente, estos datos obtenidos a través de las diferentes misiones del programa Copernicus, son accesibles de forma gratuita para cualquier usuario. La Agencia Espacial Europea proporciona ortoimágenes en proyección UTM/WGS84, de 100x100km² en el sistema de referencia Military Grid Reference System (MGRS), ofreciendo la posibilidad de aplicar correcciones atmosféricas (Nivel-2A, transformación por parte del usuario) o descargar las imágenes sin correcciones (Nivel-1C). [15]

Tabla 1. Tipos de productos que ofrecen SENTINEL-2. Fuente: [15]

Nombre	Descripción de alto nivel	Producción y Distribución	Volumen de datos
Nivel-1C	Reflectancia de la atmósfera en la geometría cartográfica.	Generación sistemática y distribución on-line.	600 MB (cada 100x100 km ²)
Nivel-2A	Reflectancia del fondo de la atmósfera en geometría cartográfica.	Generación sistemática y distribución y generación en línea en el lado del usuario (utilizando la caja de herramientas de Sentinel-2)	800 MB (cada 100x100 km ²)

Tabla 2. Características del instrumento multispectral de los satélites Sentinel-2 en sus diferentes bandas. Fuente: [15]

Bandas	SENTINEL - 2A		SENTINEL - 2B		Resolución espacial (m)
	Longitud de onda central (nm)	Ancho de banda (nm)	Longitud de onda central (nm)	Ancho de banda (nm)	
Banda 1 – Costera/Aerosoles	442.7	21	442.2	21	60
Banda 2 – Azul	492.4	66	492.1	66	10
Banda 3 – Verde	559.8	36	559.0	36	10
Banda 4 – Rojo	664.6	31	664.9	31	10
Banda 5 – Visible e infrarrojo cercano (VNIR)	704.1	15	703.8	16	20
Banda 6 – Visible e infrarrojo cercano (VNIR)	740.5	15	739.1	15	20
Banda 7 – Visible e infrarrojo cercano (VNIR)	782.8	20	779.7	20	20
Banda 8 – Visible e infrarrojo cercano (VNIR)	832.8	106	832.9	106	10
Banda 8A – Visible e infrarrojo cercano (VNIR)	864.7	21	864.0	22	20
Banda 9 – Vapor de agua	945.1	20	943.2	21	60
Banda 10 - Infrarrojo de onda corta (SWIR) – Cirrus	1373.5	31	1376.9	30	60
Banda 11 – Infrarrojo de onda corta (SWIR)	1613.7	91	1610.4	94	20
Banda 12 – Infrarrojo de onda corta (SWIR)	2202.4	175	2185.7	185	20

1.2.2 Programa Landsat

El proyecto Landsat de la Agencia Espacial Norteamericana (NASA) y el Servicio Geológico de los Estados Unidos (USGS) es la serie de misiones espaciales de observación terrestre más longeva. El primer satélite de este programa fue lanzando en 1972, actualmente se encuentra operativo el octavo satélite (Landsat-8) y se prevé el

lanzamiento de uno nuevo en el año 2020/2021 para dar continuidad a la adquisición de datos. En la Figura 1 se muestra el periodo de actividad de las diferentes misiones espaciales de la serie Landsat.

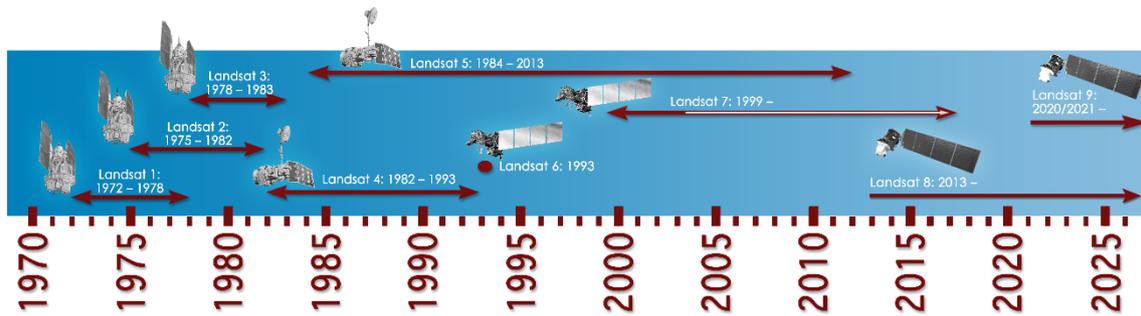


Figura 1. Cronología de la serie de misiones espaciales Landsat.

El satélite de la misión espacial Landsat-8 orbita de forma síncrona con el Sol a una altitud nominal de 705 km, cubriendo de forma completa la superficie terrestre cada 16 días. Para la adquisición de los datos, la captura se realiza en modo barrido a lo largo de la trayectoria de la nave con una anchura de franja de 185 km, reduciendo de esta forma el nivel de distorsión radiométrica en comparación con las misiones anteriores.

El satélite está compuesto por dos instrumentos: un radiómetro de barrido multicanal (Operational Land Imager, OLI) y un radiómetro infrarrojo de dos canales (Thermal Infrared Sensor, TIRS), que proporcionan imágenes a una resolución espacial de 30 metros (visible, NIR, SWIR); 100 metros (térmico); y 15 metros (pancromático). El sensor OLI opera en nueve bandas espectrales con una precisión radiométrica mejorada con respecto a los instrumentos utilizados en las misiones precedentes, proporcionando a la vez consistencia y compatibilidad con los datos históricos en siete de estas bandas. Las otras dos están diseñadas para detectar nubes cirrus y para observaciones de zonas costeras. Por su parte, el sensor TIRS, realiza observaciones en dos combinaciones de longitud de onda infrarroja térmicas, permitiendo obtener la temperatura de la superficie y estudiar el proceso de transferencia de calor y humedad en la gestión del agua o la agricultura.

Los datos son ofrecidos de forma abierta y gratuita en imágenes con un tamaño de escena de 170 x 185 km en el sistema de referencia WRS-2 (Worldwide Reference System), pudiéndose identificar cada imagen de forma unívoca por su franja vertical y fila. Los datos se ofrecen en diferentes categorías según el método de procesamiento: Reflectancia en la superficie, Reflectancia en la parte superior de la atmósfera o Imágenes en bruto; y en diferentes niveles: Nivel 1 (T1): datos que cumplen con los

requisitos de calidad geométrica y radiométrica; Nivel 2 (T2): datos que no cumplen con los requisitos del Nivel 1; Tiempo real (RT): datos que aún no se han evaluado.

Tabla 3. Características de los instrumentos de la misión espacial Landsat-8.

Bandas	Longitud de onda (micrómetros)	Resolución (metros)	Sensor
Banda 1 – Costera/Aerosoles	0.43-0.45	30	OLI
Banda 2 – Azul	0.45-0.51	30	OLI
Banda 3 – Verde	0.53-0.59	30	OLI
Banda 4 – Roja	0.64-0.67	30	OLI
Banda 5 – Infrarrojo cercano (NIR)	0.85-0.88	30	OLI
Banda 6 – Infrarrojo de onda corta 1 (SWIR 1)	1.57-1.65	30	OLI
Banda 7 – Infrarrojo de onda corta 2 (SWIR 2)	2.11-2.29	30	OLI
Banda 8 – Pancromática	0.50-0.68	15	OLI
Banda 9 – Cirros	1.36-1.38	30	OLI
Banda 10 – Infrarrojo térmico 1	10.6-11.19	100	TIRS
Banda 11 – Infrarrojo térmico 2	11.50-12.51	100	TIRS

1.2.3 Shuttle Radar Topography Mission (SRTM)

La misión topográfica SRTM llevada a cabo por la NASA y la Agencia Nacional de Inteligencia Geoespacial (NGA) junto con las agencias espaciales alemanas e italianas en febrero del año 2000, dio lugar a uno de los modelos digitales de elevación (MDE) más completo y de mayor resolución de la Tierra [16]. La misión tuvo una duración de 11 días, recopilando datos sobre la superficie terrestre que se encuentra entre los 60º de latitud norte y 54º de latitud sur, que corresponde al 80% del terreno de la Tierra.

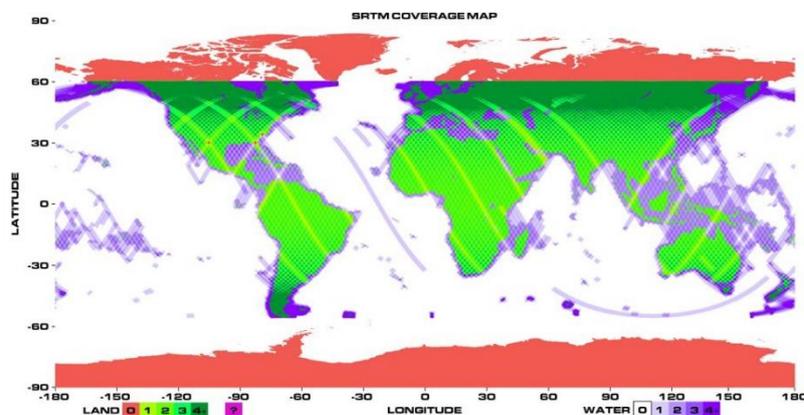


Figura 2. Mapa de la cobertura geográfica de la misión SRTM. Fuente: [17]

Para la obtención de la altimetría se utilizó una técnica denominada interferometría radar de apertura sintética (InSAR). Se equipó al transbordador espacial Endeavour con dos antenas de radar separadas mediante un mástil de 60 metros. La antena principal, situada en el propio transbordador emite un haz de ondas de radar, que al alcanzar la superficie terrestre se dispersan y son recogidas por las dos antenas. De esta forma, conociendo la distancia de separación entre las dos antenas, su posición absoluta, el rango hasta la Tierra y la diferencia de fases entre las señales de onda reflejadas; es posible calcular de forma precisa la elevación de la superficie de la Tierra. El motivo por el cual se utilizaron instrumentos radar se debe a que puede funcionar tanto de día como de noche, y las mediciones no se ven afectadas por las nubes.

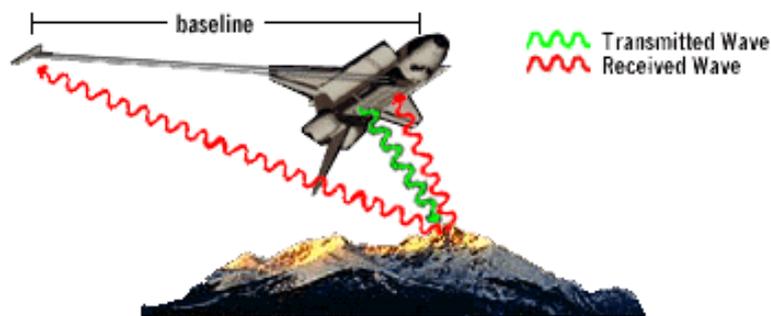


Figura 3. Esquema de funcionamiento del sistema InSAR de la misión SRTM. Fuente: [17]

El modelo digital de elevación tiene una resolución espacial de 30 metros con un error vertical de 9 metros, mejorando el objetivo inicial de la misión de 16 metros [16]. Actualmente los datos son accesibles a todo el público en su máxima resolución, sin embargo, hasta septiembre de 2014 solamente estaban disponibles en una resolución de 90 metros para las regiones fuera de los Estados Unidos.

Capítulo 2. Objetivo

El objetivo de este trabajo es explorar las aplicaciones de la representación tridimensional e interactiva de índices medioambientales multidimensionales, multivariantes y multimodales, aplicado a masas de agua en embalses. Concretamente, se combinarán modelos digitales de elevación con índices medioambientales de teledetección (cobertura de vegetación, identificación de masas de agua, turbidez, temperatura).

Como objetivo secundario también se tratará de visualizar datos generados a partir de modelos tridimensionales de simulación hidrodinámica, debido al carácter complementario que tienen las mediciones obtenidas in situ con respecto a la monitorización remota.

La metodología propuesta y las herramientas que se van utilizar son las siguientes:

- Plataforma de análisis geoespacial Google Earth Engine:
 1. Búsqueda y filtrado de imágenes de diferentes misiones espaciales de monitorización terrestre como SENTINEL-2 y Landsat-8, y de topografía como SRTM.
 2. Cálculo de los índices medioambientales mediante operaciones matemáticas con las bandas de las imágenes multiespectrales capturadas por los satélites.
 3. Descarga de los datos para un determinado rango temporal.
- Biblioteca GDAL:
 4. Conversión de las imágenes de satélite en formato .tif al formato .nc para su procesado.
- Librería UTM:
 5. Conversión del sistema de coordenadas geográficas a coordenadas UTM para tener uniformidad en las diferentes fuentes de datos.
- Librería NetCDF:
 6. Almacenamiento de los datos provenientes de diferentes fuentes en un único fichero para una mejor gestión de los mismos.
- Librería ipyvolume:
 7. Exploración de las diferentes opciones de visualización tridimensional que ofrece la librería.

Capítulo 3. Caso de estudio

Las diferentes misiones espaciales de observación terrestre a través de satélites, dada su amplia cobertura espacial, permiten el estudio de prácticamente cualquier región de la Tierra. En este trabajo, la zona de interés es el embalse de La Cuerda del Pozo, que ya ha sido objeto de estudio en otras publicaciones como [18] [19] [20] [21], entre otras.

Este embalse se encuentra en el municipio de Vinuesa en la provincia de Soria (coordenadas UTM: X: 524663, Y: 4636207, Z: 1086, Huso 30), regulando el río Duero en su cabecera junto con el embalse Los Rábanos. Su extensión de 2176 ha y su capacidad de 229 hm³ lo sitúan como uno de los seis más grandes dentro de la cuenca geográfica, abasteciendo de agua potable a Soria y parte de Valladolid, y usándose para regar 26000 hectáreas.

En el embalse de Cuerda del Pozo se están monitorizando parámetros de calidad del agua incluyendo variables físicas, químicas, biológicas y medioambientales mediante de una serie de sensores instalados en diferentes plataformas en varias zonas del embalse. Además, los datos obtenidos a través de los sensores, junto con datos meteorológicos y análisis de laboratorio; han servido de entrada a un modelo de simulación hidrodinámica tridimensional que permita la predicción y prevención de cambios en la calidad del agua.

Para tratar de explorar la visualización conjunta de datos de satélite y de datos obtenidos mediante simulaciones, el tener acceso a los datos del modelo hidrodinámico tridimensional, ha determinado esta zona de estudio. No obstante, si se considera solamente la visualización de datos de satélite, la metodología propuesta en este trabajo se puede extrapolar a cualquier región dentro de la cobertura geográfica de las misiones espaciales que se han considerado.



Figura 4. Imagen del embalse Cuerda del Pozo obtenida mediante el satélite Landsat .

Capítulo 4. Metodología

Fuchs [7] define un proceso general para la generación de representaciones visuales siguiendo una serie de pasos bien definidos (ver Figura 5). Según este modelo primero se debe realizar un procesamiento de los datos pudiéndose generar nuevas variables derivadas y se deben aplicar estándares para el fácil uso y distribución de los nuevos conjuntos de datos generados. Posteriormente se filtran o se agregan aquellos datos que son irrelevantes y la información se mapea a una representación adecuada a sus características. En la etapa de renderizado se genera la imagen mediante un determinado algoritmo en función de la representación escogida. Y finalmente se modifica para mejorar los resultados obtenidos. Cada una de las etapas se detallarán en mayor profundidad en los siguientes apartados.

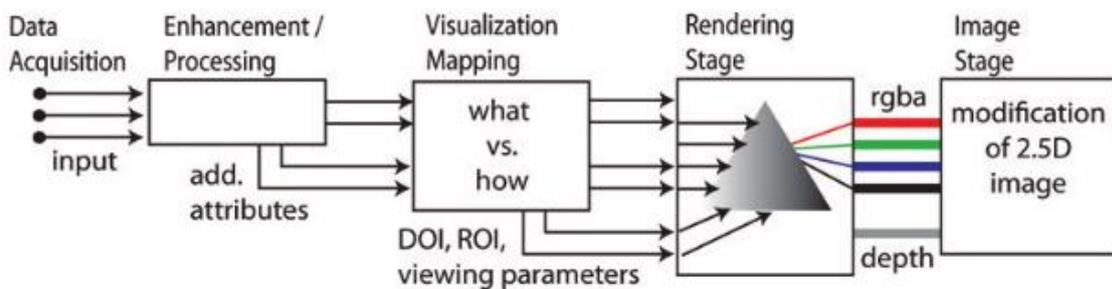


Figura 5. Proceso abstracto de generación de representaciones gráficas. Fuente: [3]

4.1 Descarga de los datos

Para este trabajo se han utilizado datos de satélite para la obtención del modelo de elevación del terreno e índices climáticos y medioambientales a nivel superficial. Y, además, se han combinado con datos provenientes de un modelo de simulación hidrológica, para la visualización de la temperatura del embalse a lo largo de su batimetría.

La descarga de los datos de satélite se ha realizado mediante la plataforma de procesamiento de datos geoespaciales en la nube Google Earth Engine. La interacción con la plataforma se puede realizar a través de las interfaces de programación de aplicaciones (API) de Javascript y Python, o mediante el entorno de desarrollo integrado (IDE) en la web.

El catálogo de datos del que dispone Earth Engine contiene varios petabytes de imágenes de satélite de las principales misiones espaciales (i.e. Landsat, SENTINEL, SRTM, MODIS, etc.); así como otros datasets geoespaciales de la cobertura terrestre,

predicciones climáticas y meteorológicas, datos socio-económicos y variables medioambientales (i.e. JRC global surface water, NCEP Climate Forecast System (CFSv2), WorldPop, etc.). Además de ser accesibles de forma gratuita por cualquier usuario, los datos han sido procesados y curados, manteniendo los metadatos con información acerca de su localización, fecha de adquisición y condiciones en las cuales las imágenes han sido tomadas o procesadas. [22]

Dentro de Earth Engine destacan dos estructuras de datos: ráster (*Image*), compuestas por diferentes bandas y un diccionario de propiedades; y vectores (*Feature*), compuestos por una determinada geometría y sus propiedades. Los conjuntos de imágenes o vectores que estén relacionados se agrupan en colecciones, por ejemplo, todas las imágenes de la misión SENTINEL-2 se encuentran en la colección *COPERNICUS/S2*. Esto, junto con la capacidad de filtrar los datos en una determinada zona de interés y/o rango temporal, facilita enormemente la obtención de los datos. La plataforma también permite realizar operaciones matemáticas con las bandas espectrales que componen una imagen, permitiendo calcular determinados índices medioambientales, sin tener que procesar las imágenes con diferente software. [23]

Datos obtenidos de Shuttle Radar Topography Mission

Para generar la visualización tridimensional se han utilizado los datos de elevaciones de la misión *Shuttle Radar Topography Mission* (SRTM) cuyas características principales se han descrito en capítulos anteriores. Dentro del catálogo de Earth Engine se encuentra la tercera versión del *dataset* (SRTM Plus), que ha sido sometido a un proceso de corrección, completando aquellas zonas en las que no existían datos de elevación mediante *datasets* topográficos en abierto como ASTER GDEM2, GMTED2010 y NED.

En la literatura se encuentran estudios en el que se combina el uso de modelos digitales de elevación (MDE) con índices de teledetección como Yacouba [24], que incorpora datos de elevación y de vegetación en la evaluación de cambios en el uso del suelo en la provincia Yunnan de China; Zhan [25], que estudia la relación entre el índice NDVI y diferentes atributos del terreno como la altitud, la pendiente y el aspecto de laderas obtenidos a través de un MDE, en el municipio Chongqing de China; o Ho [26], que utilizan un modelo de elevaciones y el índice MNDWI para generar un mapa de peligro de inundaciones.

En este estudio, dotar a la visualización de los diferentes índices medioambientales de información sobre la altitud puede servir para un mejor entendimiento del comportamiento y distribución de éstas. Por ejemplo, en zonas de gran altitud es comprensible que no exista vegetación o la temperatura sea inferior.

Para descargar los datos, se ha definido una geometría para la zona de estudio y se ha exportado como una imagen en formato .tif con una escala de 100 metros por pixel. La imagen es una matriz bidimensional con una estructura regular, en la que los índices de los ejes representan las coordenadas x e y, y los valores representan la altitud en dichos puntos. Es importante que tanto la geometría definida como la escala a la que se exporta se mantenga constante en la descarga de las demás variables, ya que las matrices deben tener la misma dimensionalidad.

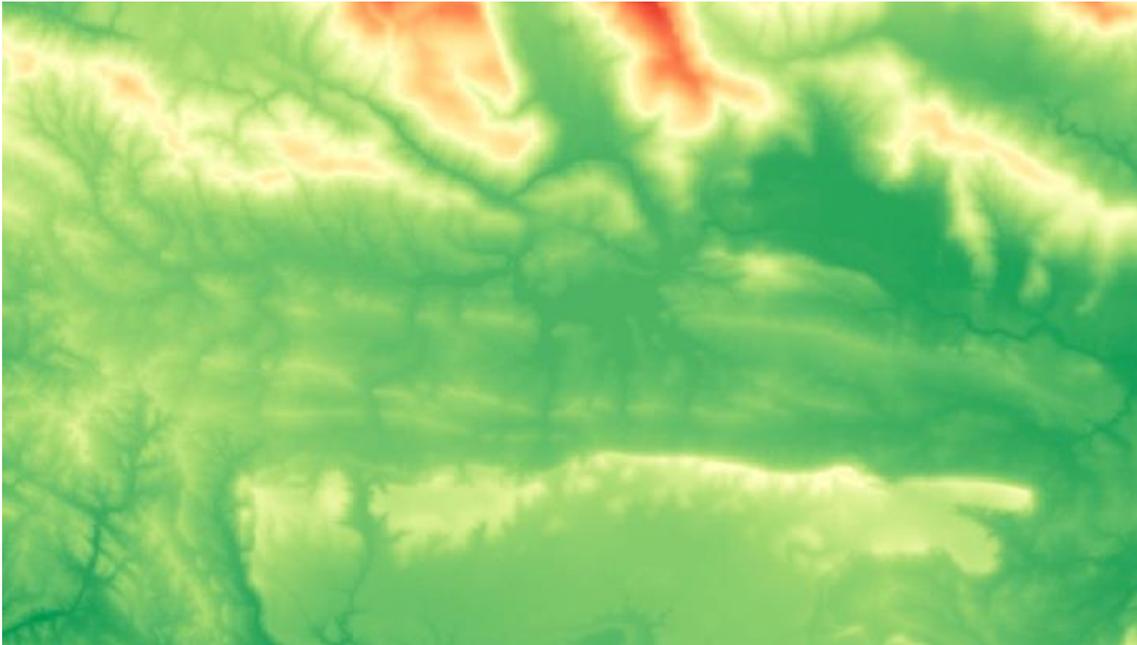


Figura 6. Imagen del modelo de elevaciones SRTM en el embalse Cuerda del Pozo obtenido a través de Google Earth Engine .

Datos obtenidos de SENTINEL-2

En función de la reflectancia, cantidad de luz reflejada por una superficie expresada en porcentaje, es posible distinguir diferentes tipos de superficie según la zona del espectro electromagnético en la que se encuentre, como se puede observar en la Figura 7. Cabe destacar que la longitud de onda óptima utilizada para medir un determinado parámetro depende de la sustancia que se mide, su concentración y las características del sensor. [27]

Se han desarrollado numerosos índices que implementan diferentes combinaciones de bandas de los sensores de teledetección, que permiten identificar masas de agua [28] [29] [30] [31], cambios en la vegetación [25] [24], evaluar la calidad del agua [27] [32] [33] u otras características como mapas de inundaciones [26].

Para este trabajo se han utilizado tres índices que permiten identificar la vegetación, las masas de agua y la turbidez del agua; y que se han utilizado anteriormente en las referencias bibliográficas anteriormente mencionados. En dichos estudios, se ha

comprobado la validez de los resultados obtenidos mediante técnicas de teledetección comparándose con valores tomados in situ a través de técnicas estadísticas. A continuación se detallarán las características de cada uno de los índices utilizados en este trabajo.

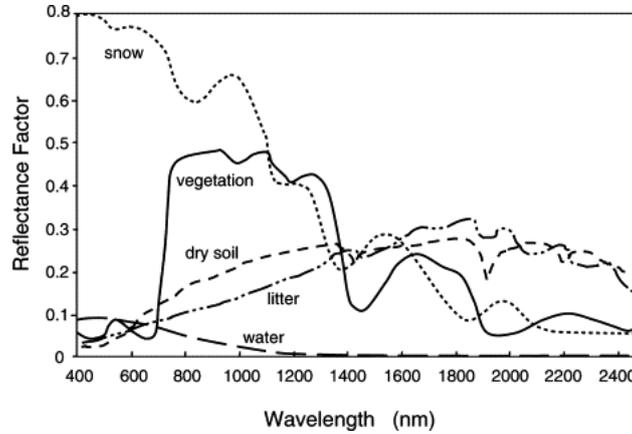


Figura 7. Reflectancia de diferentes tipos de superficie. Fuente: [34]

Modification of Normalized Difference Water Index (MNDWI)

Este índice propuesto por Xu [29] modifica el índice *Normalized Difference Water Index* (NDWI) definido anteriormente por McFeeters [28], mejorando la captación de cuerpos de agua a la vez que elimina el ruido generado por las zonas construidas que pueda haber alrededor, la vegetación y el suelo.

El índice NDWI se expresa de acuerdo a la ecuación 1, utilizando la banda verde (Banda 3, en el caso de Sentinel-2) se maximiza la reflectancia del agua, mientras que el uso del infrarrojo cercano (Banda 8) permite diferenciar el agua (baja reflectancia) de la vegetación o el suelo (alta reflectancia).

$$NDWI = \frac{B3 (GREEN) - B8(NIR)}{B3 (GREEN) + B8(NIR)} \quad (1)$$

Sin embargo, el NDWI es sensible a los terrenos urbanizados, ya que éstas tienen una reflectancia similar a la del agua tanto en la banda verde como el infrarrojo cercano, dando a menudo da como resultado cuerpos de agua sobreestimados. Por ello, Xu [29] propone cambiar el infrarrojo cercano por el infrarrojo de onda corta (Banda 11). De esta forma se intensifican los valores que obtendría el agua ya que el infrarrojo de onda corta absorbe más luz que el infrarrojo cercano, las zonas edificadas obtienen valores negativos, y la vegetación y suelo también obtienen resultados negativos al reflejar más luz en el infrarrojo de onda corta que en el cercano.

$$MNDWI = \frac{B3 (GREEN) - B11(SWIR)}{B3 (GREEN) + B11(SWIR)} \quad (2)$$

Normalized Difference Vegetation Index (NDVI)

Este es uno de los más conocidos y utilizados dentro de los índices de vegetación. Se expresa de acuerdo a la ecuación 3, utilizando el infrarrojo cercano (Banda 8) y la banda roja. En el caso de vegetación verde y densa, la mayor parte de la luz visible es absorbida mientras que el infrarrojo cercano es reflejado. Por otro lado, si la vegetación es escasa o poco saludable se refleja más luz visible y menos infrarroja cercana [35]. Según los valores obtenidos en el cálculo, se pueden determinar los siguientes tipos de suelos: los valores negativos de NDVI (valores cercanos a -1) corresponden al agua; los valores cercanos a cero (-0.1 a 0.1) generalmente corresponden a áreas áridas de roca, arena o nieve; los valores bajos y positivos representan arbustos y pastizales (0.2 - 0.4), mientras que los valores altos indican bosques templados y tropicales (valores cercanos a 1). [25] [24]

$$NDVI = \frac{B8 (NIR) - B4(RED)}{B8 (NIR) + B4(RED)} \quad (3)$$

Normalized Difference Turbidity Index (NDTI)

El índice NDTI es definido por Lacaux [32] de acuerdo a la ecuación 4. A medida que aumenta el nivel de turbidez del agua debido al aumento de partículas suspendidas en el agua, la reflectancia de la banda roja es superior a la de la banda verde. De acuerdo a los valores obtenidos se puede distinguir entre agua limpia (entre -0.2 y 0), niveles de turbidez moderados (entre 0 y 0.2), y valores superior a 0.25 para cuerpos de agua con alta turbidez.

$$NDTI = \frac{B4 (RED) - B3(GREEN)}{B4 (RED) + B3(GREEN)} \quad (4)$$

En el caso de las imágenes obtenidas del satélite Sentinel-2, se han descargado las imágenes de todo el año 2018, para poder observar algunas diferencias estacionales. Uno de los problemas de los datos obtenidos mediante teledetección es la presencia de nubes que impiden su correcta medición, por lo que es necesario realizar un filtrado según el porcentaje de píxeles afectados por nubes en cada ortoimagen. El problema de este método es que dicho porcentaje que se especifica en los metadatos es relativo a toda la ortoimagen dando lugar a dos problemas. Por una parte, es posible que se eliminen imágenes porque la cantidad de nubes que contiene la ortoimagen supera el baremo establecido, pero que no se encuentren en la zona de interés ver Figura 8.b).

Por otra parte, se puede dar el caso contrario, que la escena tenga un porcentaje bajo de nubes, pero éstas se encuentren justo en la zona de estudio (ver Figura 8.a). Como solución se ha optado por la curación manual de aquellas imágenes en las que existe presencia de nubes en la zona de interés. Esta ha sido viable dado al bajo rango temporal que se ha utilizado en este estudio, sin embargo se requiere de una metodología de análisis automático de las zonas definidas para estudios que abarquen varias décadas de imágenes de satélite.

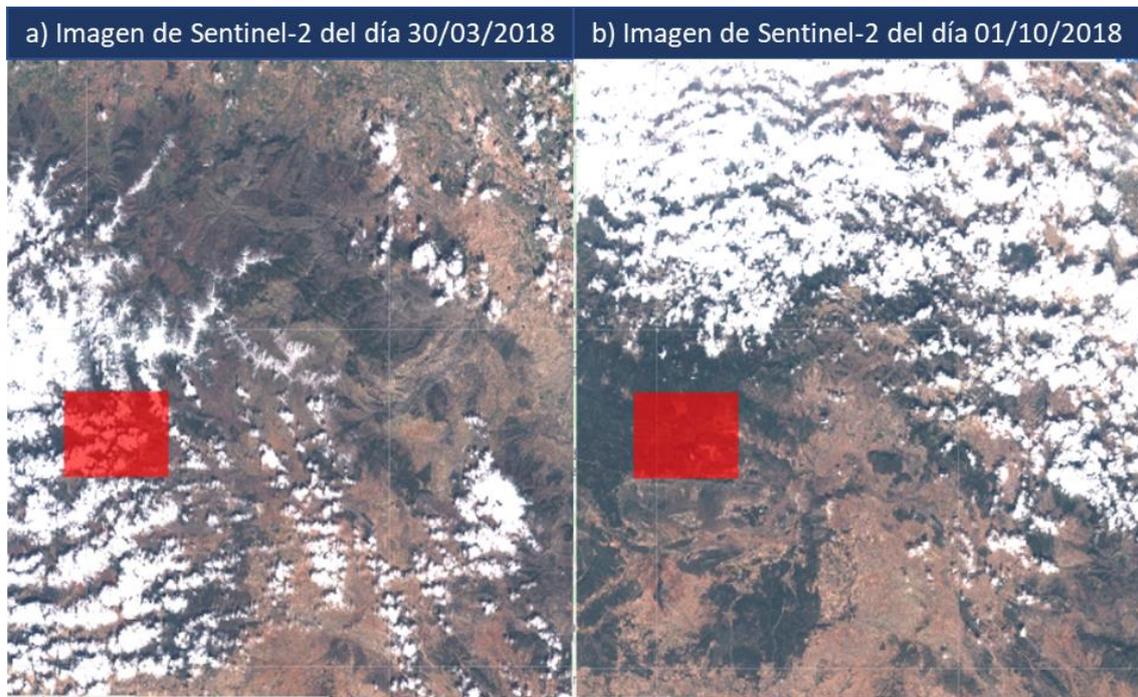
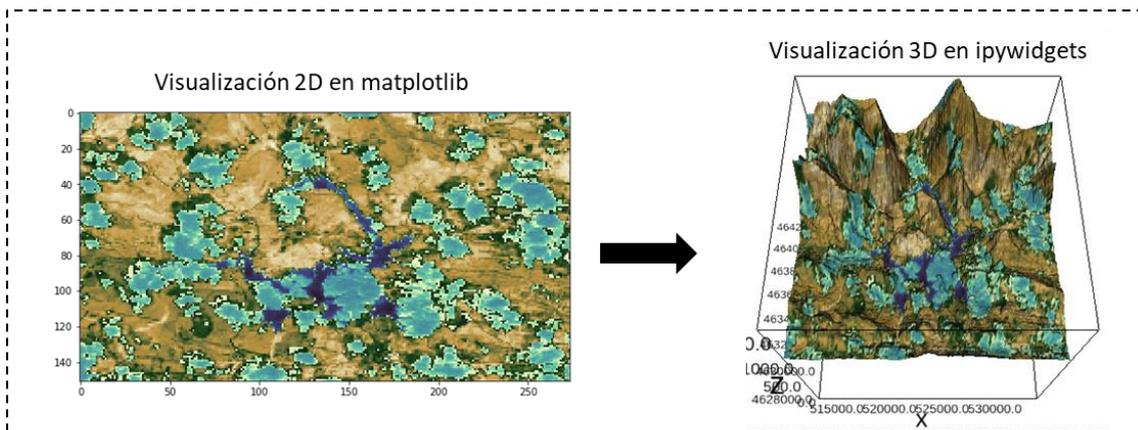


Figura 8. Presencia de nubes en las ortofotos obtenidas de la misión espacial Sentinel-2.

Debido a esta filtrado, el número total de imágenes obtenidos se ha visto reducido de 73 a 24. Dado que la herramienta Google Earth Engine permite la realización de operaciones con las bandas espectrales de las imágenes, se han calculado los diferentes índices en la plataforma y se han descargado directamente en formato .tif, al igual que en el caso del modelo digital de elevaciones, manteniendo la misma geometría y escala para la zona de estudio.

También se ha intentado aplicar máscaras que eliminen los valores de aquellos píxeles nubosos. Esta técnica es de utilidad en el caso de existan pocas nubes en la escena y hace uso, en el caso de Sentinel-2, de la banda de clasificación QA60 que indica si en el píxel existen nubes densas, cirrus o la visibilidad es perfecta. Este método puede funcionar en el caso de visualizaciones 2D en *matplotlib*, sin embargo en la representación 3D en *ipyvolume* se pierden los valores en toda la figura.

Sin máscara de nubes



Con máscara de nubes

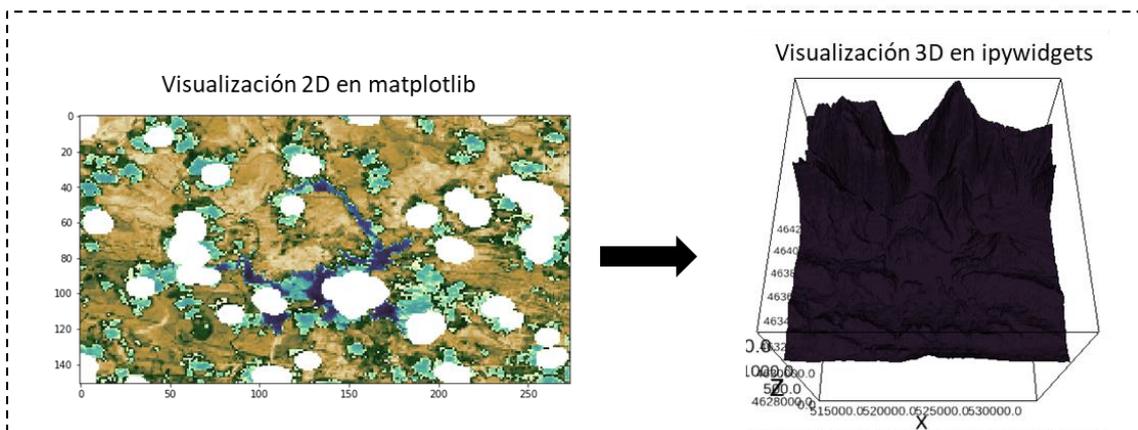


Figura 9. Comparación de la aplicación de máscaras de nubes en 2D y 3D.

Datos obtenidos de Landsat 8

A partir de los datos de la misión Landsat 8 se han obtenido los datos de temperatura superficial. Dado que este satélite lleva un sensor infrarrojo térmico de dos canales (Banda 10 y 11), para extraer la temperatura superficial simplemente se debe seleccionar una de ellas, no siendo necesario el cálculo de ningún índice como ocurría en los índices anteriores.

Al igual que en el caso de las imágenes obtenidas de la misión Sentinel-2, también se debe realizar un filtrado de las mismas según su cobertura de nubes.

El mismo problema con la filtración de las imágenes con alto porcentaje de nube se ha encontrado también para los satélites Landsat 8. En este caso, teniendo en cuenta que la resolución temporal de esta misión es de sólo 16 días, una vez realizado el filtrado, se obtienen únicamente 7 imágenes válidas para todo el año 2018.

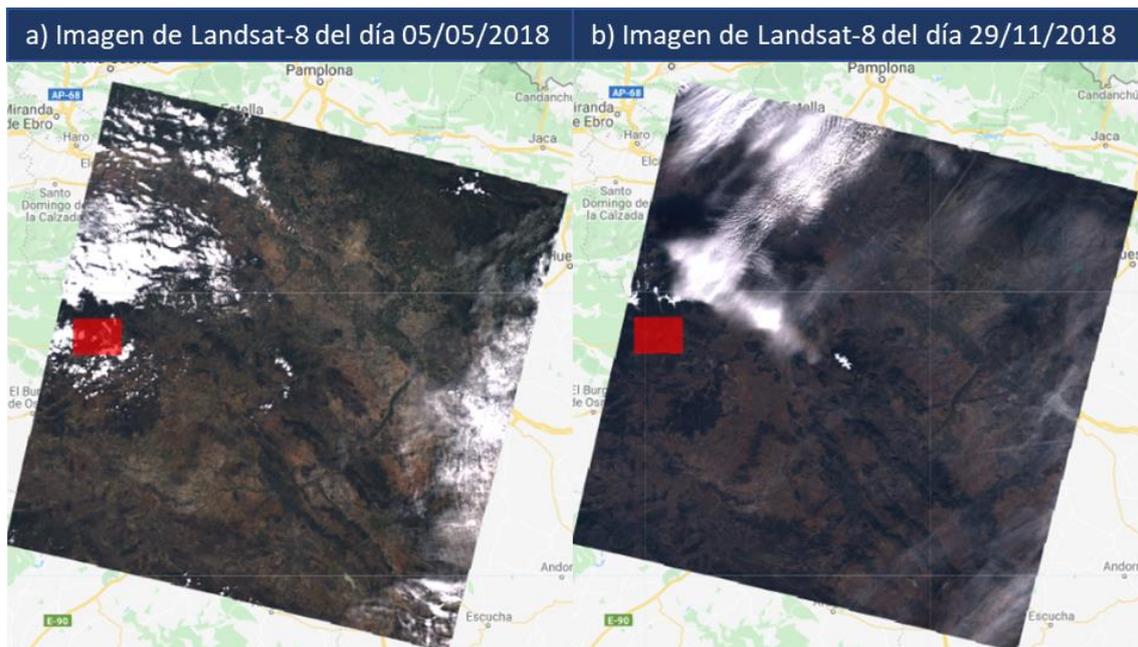


Figura 10. Presencia de nubes en las ortoimágenes obtenidas de la misión espacial Landsat-8.

4.2 Procesado y transformación de los datos

NetCDF

El formato de archivo que se ha elegido para el almacenamiento de los datos en este trabajo es NetCDF (Network Common Data Form). Este tipo de fichero permite crear, acceder y compartir datos científicos multidimensionales y multivariable, siendo comúnmente usado en áreas como la climatología, meteorología, oceanografía o sistemas de información geográfica. Entre las características de este tipo de ficheros destacan las siguientes:

- Incluyen información sobre los datos que contiene el archivo.
- Los datos se pueden georreferenciar.
- Son accesibles a través de ordenadores con diferentes formas de almacenar enteros, caracteres y números de coma flotante.
- Permiten de forma eficiente el acceso a un subconjunto pequeño de datos dentro grandes *datasets*.
- Se pueden añadir nuevos datos a un fichero sin necesidad de copiar el *dataset* o redefinir su estructura.
- Permiten el acceso de un usuario con permiso de escritura y varios usuarios con permiso de lectura de forma simultánea a un mismo archivo.
- Ofrecen compatibilidad entre versiones anteriores y futuras del formato.

Para poder definir las características de los datos y sus relaciones, los archivos netCDF disponen de tres componentes: dimensiones, variables y atributos. Combinando estas

componentes, se permite a los diferentes lenguajes o programas que soportan este formato mostrar, transformar, combinar o analizar determinados campos del *dataset*. A continuación se definen cada una de las componentes :

- **Dimensiones:** representan las dimensiones físicas (latitud, longitud, elevación y/o tiempo) u otro tipo de índices como por ejemplo ID del modelo de simulación o estación meteorológica. Se definen mediante un nombre y una longitud o tamaño. El tamaño de la dimensión puede ser un número entero positivo o tener una longitud ilimitada, actuando como un índice de registro y permitiendo a las variables expandirse a través de la misma.
- **Variables:** representan un conjunto de valores del mismo tipo y es donde se almacenan la mayor parte de los datos en un archivo NetCDF. Se definen mediante un nombre, el tipo de dato (entero, coma flotante, cadena de caracteres, etc.) y las dimensiones que las caracterizan, pudiendo no tener ninguna (variable escalar) o varias (latitud, longitud, altura y tiempo). Dentro de esta componente se encuentran también las **Variables de coordenadas**, que son unidimensionales y delimitan las coordenadas físicas correspondiente a una determinada dimensión.
- **Atributos:** definen los metadatos o datos secundarios. Se pueden describir **atributos de variables**, que proporcionan información sobre una determinada variable (nombre, unidades, sistema de referencia, etc.); y **atributos globales**, que caracterizan el propio *dataset* (titulo, descripción, autor, etc.). Existen diferentes convenciones que especifican el conjunto de atributos a definir en función del área de aplicación. [36]

Air Temperature

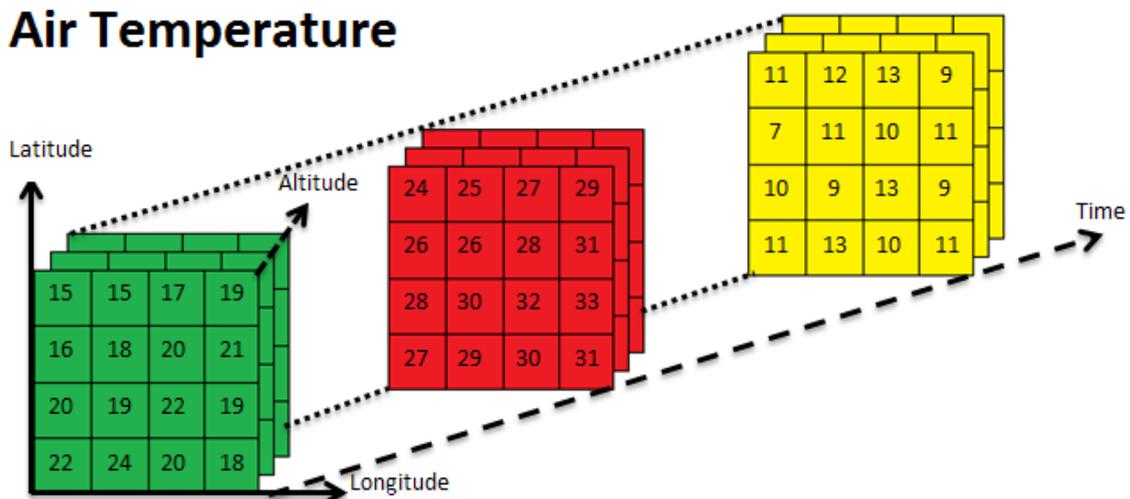


Figura 11. Representación de la estructura dimensional de una variable. Fuente: [37]

Una vez descargadas todas las imágenes, el primer paso ha sido su conversión desde el formato .tif al formato .nc para poder trabajar con los datos en el entorno Jupyter. La

primera transformación de los datos ha sido la conversión de las coordenadas del modelo digital de elevaciones desde el sistema de coordenadas geográficas (latitud/longitud) al sistema UTM (Universal Transverse Mercator). Esto es necesario para que sea coherente con el sistema utilizado en los datos de Landsat 8 y Sentinel-2.

Posteriormente, se han generado las matrices de coordenadas a partir de los vectores de coordenadas extraídos del modelo digital de elevaciones.

En el caso de las variables medioambientales y de temperatura, dado que varían a lo largo del tiempo, se ha creado para cada una de ellas un diccionario ordenado en Python con las fechas de adquisición como llave.

Finalmente, se han generado dos archivos netCDF, uno para cada misión espacial, con los datos ya preparados para su representación, así como para facilitar su distribución y reproducibilidad. Se ha optado por la generación de archivos diferentes debido a que la dimensión temporal de los datos obtenidos de los satélites Sentinel-2 es diferente de los obtenidos mediante Landsat-8, 22 imágenes frente a 9, respectivamente.

Para la creación de los archivos se han considerado tres dimensiones: coordenadas Y, coordenadas X y tiempo. En lo que respecta a las variables, se han definido: XCOR, YCOR y elevación (dependientes de las dimensiones X e Y); fechas (dependiente de la dimensión temporal); y las variables medioambientales y de temperatura (dependientes de las tres dimensiones).

En lo que respecta a los atributos, se han utilizado las 15 definiciones semánticas descriptivas que se contemplan en el modelo de metadatos Dublin Core para la descripción de los *datasets*.

Datos obtenidos del modelo de simulación

En lo que respecta a los datos de simulación, estos han sido proporcionados por los directores de este trabajo en un archivo en formato netCDF formado por 14 dimensiones y 45 variables. Los datos han sido generados por un modelo hidrológico usando el software de modelado Delft3D, a partir de diferentes parámetros obtenidos de varios sensores situados en boyas y plataformas dentro del embalse, estaciones meteorológicas, mapas batimétricos y ríos afluentes [19].

Para la visualización tridimensional se ha extraído solamente la variable temperatura, definida por su dimensión temporal, los niveles de profundidad y sus coordenadas espaciales X e Y. El objetivo de utilizar este tipo de datos era poder visualizar la estratificación térmica a lo largo de la batimetría, sin embargo, en el conjunto de datos proporcionados, la temperatura es constante en toda la masa de agua, variando únicamente a lo largo del tiempo. Esto, si bien permite la representación tridimensional

del embalse, hace que carezca de sentido el mapeado de colores según los valores de temperatura puesto que el color será el mismo en toda la figura.

4.3 Generación de las visualizaciones

Para la generación de las diferentes representaciones se ha utilizado la librería de Python *ipyvolume* que permite visualizar en Jupyter volúmenes, superficies y glifos 3D a partir de *arrays* de *numpy*. Esta herramienta, si bien cubre la mayor parte de las técnicas de visualización 3D vistas en la literatura, se encuentra todavía en fase de desarrollo, encontrándose actualmente en la versión 0.5. Esto supone que algunas características todavía no están implementadas, sin embargo, a pesar de las limitaciones del software, se ha podido hacer un análisis de las diferentes opciones de visualización que ofrece.

Fuchs [7] propone en el caso de la representación de datos multivariable la utilización de técnicas híbridas de visualización destacando dos ventajas principalmente. Por una parte se mejora la efectividad de la visualización, ya que cada variable puede ser visualizada mediante la técnica más apropiada. Y por otra parte, permite la visualización simultánea de ambas variables a la vez que se reduce la contaminación visual. En este trabajo se han combinado la representación como superficie paramétrica para el terreno y el mapeado de colores para las variables medioambientales y de temperatura.

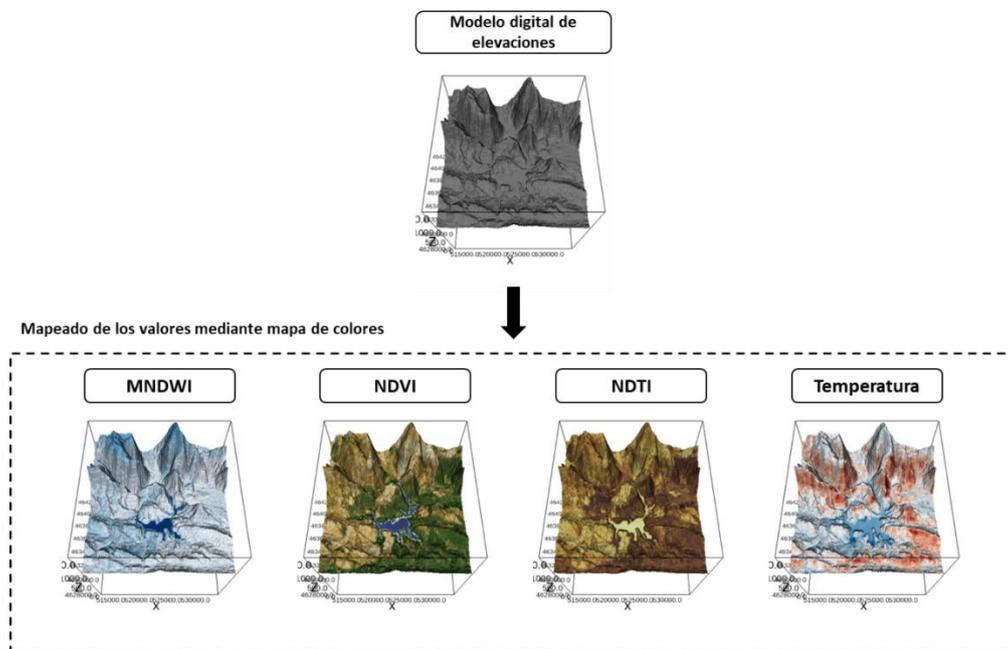


Figura 12. Proceso de mapeado de la visualización utilizado en este trabajo.

Para representar de forma tridimensional las irregularidades del terreno se ha utilizado la técnica de mallado poligonal. Este método permite generar una superficie compuesta

por vértices y caras (triángulos definidos por 3 vértices) a partir de matrices de coordenadas regularmente espaciadas de los ejes x, y, z. Las tres matrices proporcionan la posición de los vértices en el espacio y, si están definidas en una rejilla regular, la triangulación se puede inferir de forma automática a partir de las dimensiones de la matriz.

Para añadir información acerca de las variables medioambientales y temperatura se ha utilizado la técnica de mapeado de color. Este es un método común dentro de la visualización de datos escalares, que consiste en utilizar los datos como índices dentro de una tabla de consulta que contiene una serie de colores. Si bien esta técnica es unidimensional, a cada valor le corresponde un determinado color, es posible mapear los colores también a objetos bidimensionales y tridimensionales. El diseño de un mapa de colores es bastante complejo, ya que puede suponer la pérdida de información, o se desee acentuar determinadas características (Figura 13). Por ello, y dado que no existe ninguna convención para los índices de estudio, en este trabajo se han utilizado los mapas de color para oceanografía que ofrece *matplotlib*.

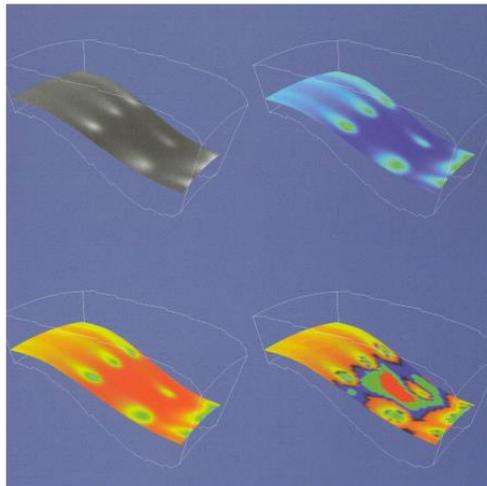


Figura 13. Representación de la densidad de flujo mediante diferentes mapas de colores. Fuente: [38]

Para poder visualizar la dimensión temporal de los datos se ha utilizado un widget interactivo que permite tanto la selección de una fecha concreta como la generación de animaciones para observar la evolución a lo largo del tiempo. Además de la generación de varias vistas para cada unidad temporal, se encuentran otras técnicas en la literatura como por ejemplo el cálculo de diferencia de los valores, uso de glifos Helix o mapeado de colores en 2 tonos [8].

Para dotar de características más avanzadas a las representaciones, y dado que es una opción que la librería *Ipyvolume* permite, se ha intentado hacer uso de la técnica mapeado de texturas. Esta técnica, según el estudio realizado por Fuchs [7], es muy versátil permitiendo combinar colores y texturas para representar múltiples valores en

un mismo punto o utilizar diferentes capas campos en la misma superficie, tratando siempre de no saturar la figura con demasiada información.

Para añadir una textura a una figura en *ipyvolume*, se debe emplear el proceso de mapeado UV. Esta técnica permite proyectar una imagen 2D en la superficie de un modelo 3D, asignando los píxeles de la imagen a los vértices que forman el mallado de la figura 3D (ver Figura 14). En los programas de modelado 3D existe la posibilidad de generar automáticamente las coordenadas UV, “aplanando” (unwrapping) el mallado poligonal que define la figura [39].

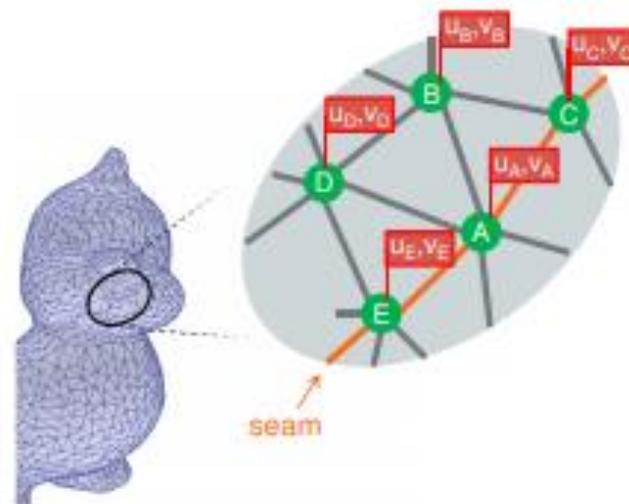


Figura 14. Asignación de las coordenadas UV en un modelo tridimensional. Fuente: [38]

En el caso de *ipyvolume*, no es posible obtener de forma automática estos valores, por lo que se han especificado unas coordenadas UV como si se tratara de una superficie plana, sin tener en cuenta la elevación. El principal objetivo fue superponer una capa con una imagen formada por los colores del espectro visible (Truecolor) generada en Google Earth Engine, a la que se le ha extraído la superficie del cuerpo de agua del embalse utilizando el *dataset Landsat Global Inland Water*. Esto permitiría tener una mejor percepción en la visualización de índices estrictamente relacionados con el embalse como el NDTI, ya que alrededor del embalse se vería la textura, mientras que en el embalse se visualizarían los valores del índice de acuerdo al mapa de colores asignado. Si bien se esperaba que la textura estuviera distorsionada por no tener en cuenta la elevación, también se ha detectado que la librería no permite visualizar la transparencia, es decir, la zona del embalse que se había extraído en la imagen 2D (ver Figura 15)

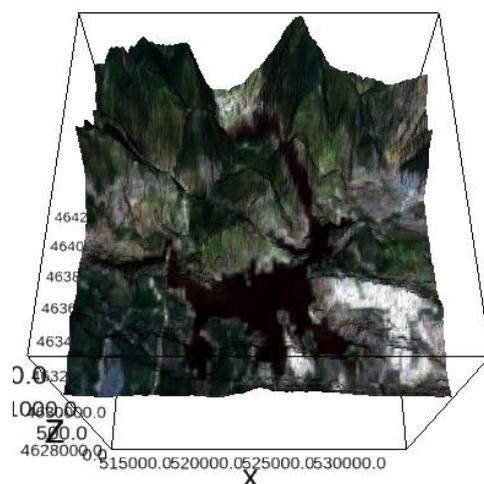


Figura 15. Visualización 3D del embalse con una textura con los colores del espectro visible.

Para la visualización de datos escalares 3D existen dos métodos principalmente, mediante isosuperficies (método indirecto); o *Direct Volume Rendering* (DVR) [38] [40] [6]. Las isosuperficies son el análogo tridimensional de las curvas de contorno, representado como una superficie geométrica sólida dentro de un volumen aquellos escalares que tengo un valor constante. Se denomina método indirecto porque para renderizar la imagen necesita generar una mallado triangular. El segundo método permite una renderización directa, aconsejable en conjuntos de datos volumétricos que son difíciles de definir mediante superficies geométricas. Esta técnica representa cada valor escalar como un *voxel* (pixel volumétrico), definiendo el color y opacidad de cada uno de ellos mediante funciones de transferencia. Las isosuperficies muestran sólo el contorno exterior de los volúmenes, útil para la representación de elementos opacos. Por otra parte, mediante DVR se consideran los datos como un medio semitransparente, pudiéndose ver la información de todo el volumen. Este tipo de técnica se utiliza en campos como la medicina, visualizando los órganos internos del cuerpo humano [41]; dinámica de fluidos computacional, renderizando la propagación de humo [42]; o en meteorología para visualizar huracanes [43]

Se han intentado aplicar estos métodos para representar los datos de simulación del modelo hidrológico. Aunque en este caso, dado que el valor de temperatura es constante en todo el volumen del embalse, la visualización no aporta gran información; puede ser de interés para visualizar las variables medidas in situ si éstas difieren.

Utilizando la técnica DVR se podría visualizar la estratificación térmica en la batimetría del embalse de forma tridimensional. Se han generado las figuras del embalse mediante las dos técnicas (ver Figura 1Figura 16) , dado que se definen espacialmente por las

dimensiones del *array* tridimensional, no ha sido posible visualizarlos de forma conjunta con la superficie generada por los datos de la superficie. En el caso de la función de *ipyvolume* que genera la visualización mediante DVR, en la documentación se especifica un parámetro “*origin*” que permite emparejar espacialmente el volumen con otras representaciones, sin embargo ese parámetro en la versión actual de la librería no está implementado. Se cree que la técnica DVR es más apropiada para la visualización volumétrica del embalse, dado que se podría hacer uso de la semitransparencia que ésta ofrece pudiendo observar fenómenos que ocurran en el interior.

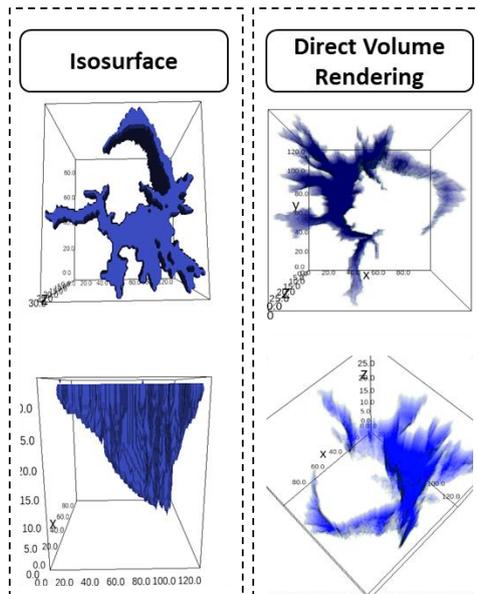


Figura 16. Representación tridimensional de los datos de simulación del modelo hidrológico.

4.4 Vistas generadas

A continuación se comentan algunas patrones que se han encontrado a la hora de examinar los datos mediante la metodología propuesta. Otra carencia que se ha encontrado en la librería *ipyvolume* es la posibilidad de añadir una leyenda con la relación entre los colores y los valores de las variables asociadas.

4.4.1 MNDWI

Para visualizar el índice MNDWI se ha utilizado un mapa de colores secuencial que varía aumenta la intensidad del color azul a medida que lo hace valor de la variable. Se ha considerado lo más oportuno dado que este índice intensifica los valores del agua. Como se puede observar en la Figura 17, la superficie del embalse se distingue claramente del terreno. En la imagen correspondiente al día 19/01/2018 se puede percibir un azul más intenso en los picos de las zonas de mayor elevación, esto puede deberse a que exista nieve o hielo.

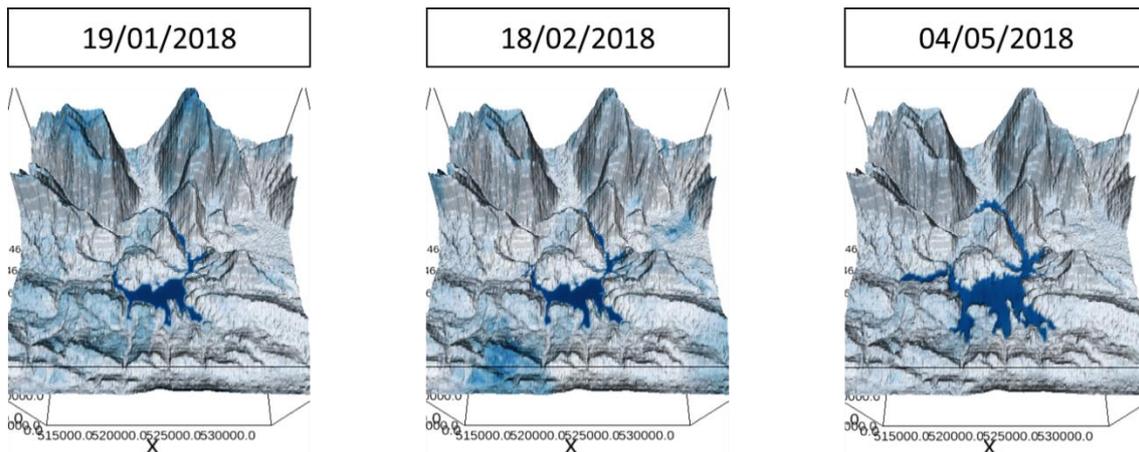


Figura 17. Visualización del índice MNDWI en diferentes fechas.

También se percibe un aumento del volumen de agua durante el mes de mayo con respecto a las imágenes correspondientes a enero y febrero. Esto se debe a que en esos primeros meses del año todavía no se ha producido el deshielo. Este cambio de volumen se confirma con los datos de medición in situ que se muestran en la Figura 18. Variación del volumen de agua del embalse Cuerda del Pozo según mediciones in-situ. Fuente:Figura 18.

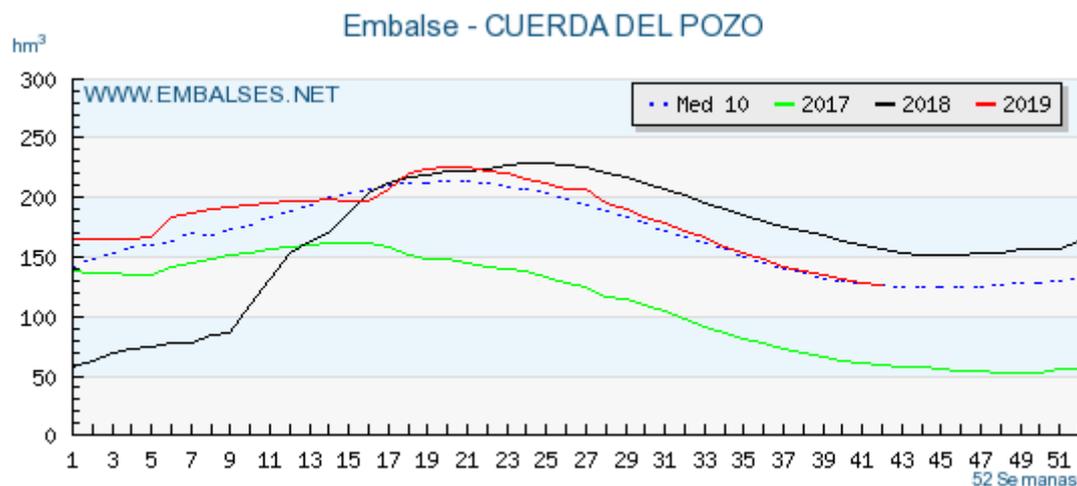


Figura 18. Variación del volumen de agua del embalse Cuerda del Pozo según mediciones in-situ. Fuente: [44]

4.4.2 NDVI

En el caso del índice de vegetación NDVI, se ha utilizado un mapa de colores divergente de la librería *cmocean*, concretamente el mapa topo. En este caso, se asocia el color azul a valores negativos (agua), blanco a valores cercanos a cero (zonas áridas) y amarillos y verdes a valores positivos (vegetación escasa/seca y vegetación densa/sana). En la primera imagen de la Figura 19 se puede observar como en el perímetro del embalse, dado su bajo nivel existen zonas áridas, que posteriormente desaparecen al aumentar el volumen y la superficie del agua. También destaca que en las zonas elevadas la vegetación es escasa. En la tercera imagen, correspondiente a un

día de verano, la vegetación se empieza a secar en comparación con las fechas anteriores correspondientes a días de primavera. Las manchas que se perciben en la segunda y tercera imagen se deben a perturbaciones debido a la presencia de nubes.

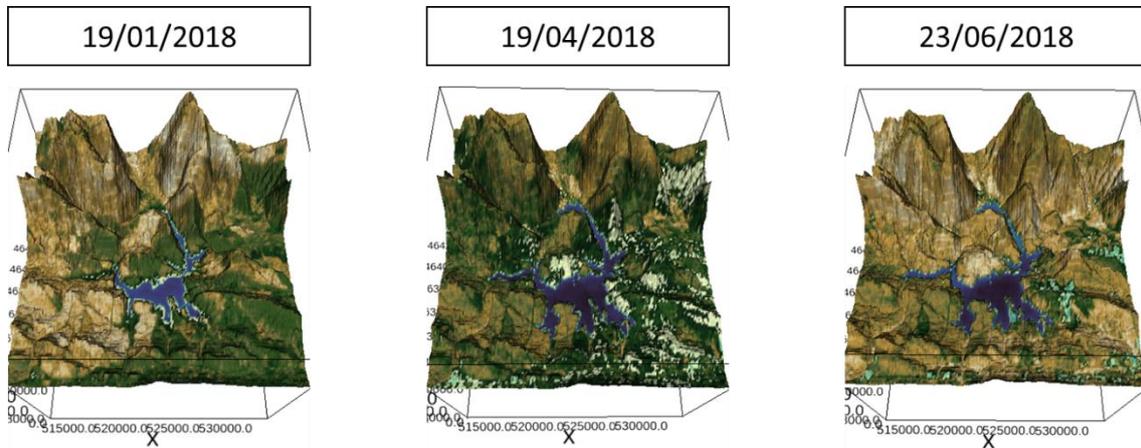


Figura 19. Visualización del índice NDVI en diferentes fechas.

4.4.3 NDTI

Para el índice de turbidez, se ha utilizado otro mapa de colores de la librería *cmocean*, denominado *turbid*. Utiliza colores marrones claros y oscuros a medida que el valor de la variable aumenta, trata de representar el incremento de sedimentos en el agua. En este caso, al tratarse de un índice de calidad del agua, la información que aporta el modelo de elevaciones es nula. Se observa como en los días 19 de enero y 23 de julio el embalse tiene una turbidez inferior al día 9 de mayo, siendo una posible causa de ello, los sedimentos que arrastran las lluvias con el agua de escorrentía.

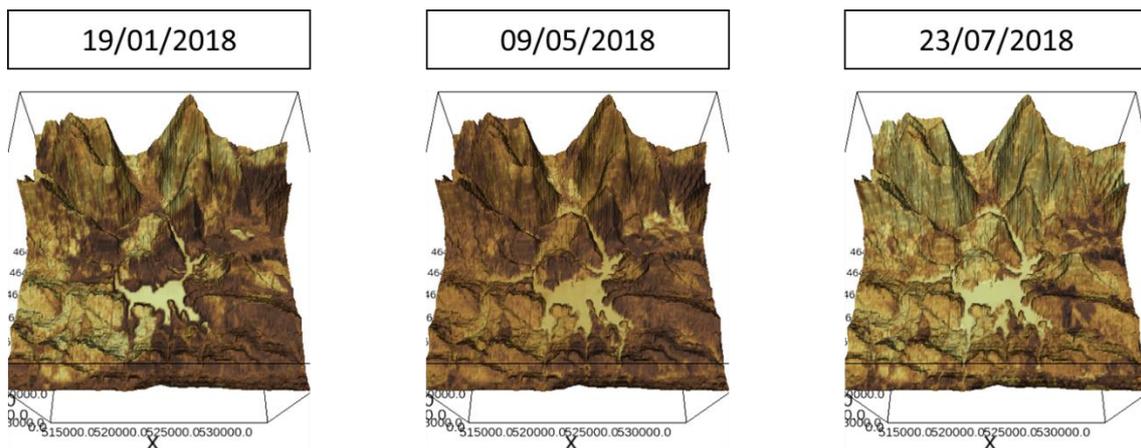


Figura 20. Visualización del índice NDTI en diferentes fechas.

4.4.4 Temperatura

En el caso de la temperatura, se ha utilizado el mapa de colores divergente *coolwarm* de *matplotlib*, que asocia el azul a valores bajos, el blanco a valores intermedios y el rojo a valores altos. La temperatura registrada por el satélite en todo el

dataset varía entre los 0°C y los 39°C. Gracias a la información que aporta el modelo de elevaciones se puede observar como en las zonas de mayor altitud la temperatura suele ser inferior con respecto a las zonas llanas, siendo este fenómeno constante en las tres fechas escogidas. También se aprecia un aumento de temperatura en la superficie del embalse en los días de junio y septiembre con respecto al 19 de abril.

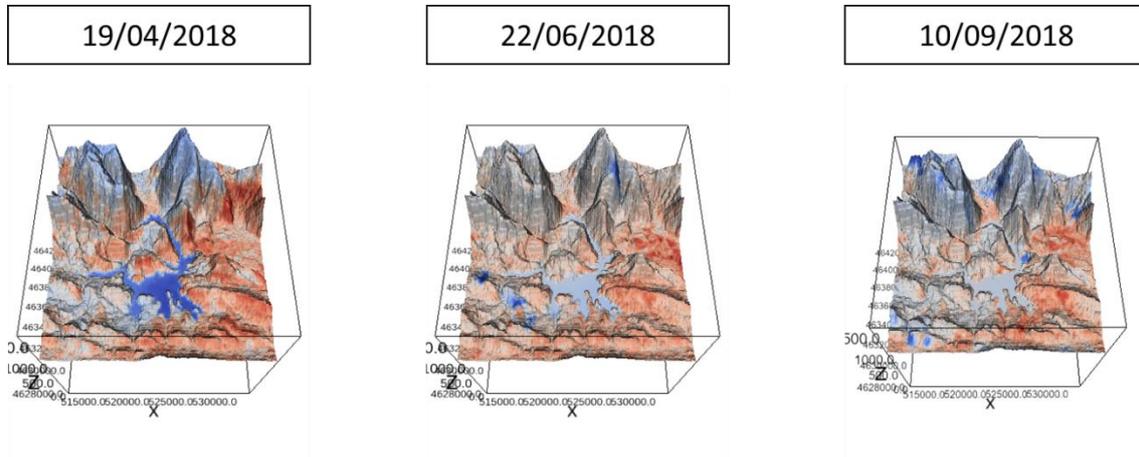


Figura 21. Visualización de la temperatura superficial en diferentes fechas.

Capítulo 5. Conclusiones y trabajos futuros

En este trabajo se ha propuesto una metodología para la visualización tridimensional a partir de técnicas de “Data Science” y de diversas librerías de Python. Además de utilizar distintos paquetes de software libre, se han analizado datos en abiertos, lo cual demuestra el enorme potencial para la extracción de información de este tipo de datos. Se han cubierto prácticamente todas las fases del ciclo de vida, desde la descarga de los datos y su curación, su procesado y análisis y finalmente la visualización y el almacenamiento.

En vista de los resultados obtenidos se puede concluir que los objetivos propuestos han sido alcanzados, obteniendo unas representaciones gráficas que permiten abstraer una gran cantidad de información. La interactividad de las gráficas permite una fácil exploración e interpretación de los datos y su relación espacial y temporal, tanto para investigadores como usuarios sin conocimientos en la materia. Esto puede conseguirse no sólo con el uso de gráficas en tres dimensiones, sino con animaciones que proporcionan una dimensión temporal y aportan información de la evolución de los parámetros en el análisis.

Durante la realización de este estudio se han encontrado algunos problemas como la perturbación de los resultados debido a la presencia de nubes, la utilización de texturas o la visualización de los datos de simulación, que no se han podido resolver de forma óptima.

- En casos reales como este se hace imprescindible la curación de los datos, una etapa esencial para poder llegar a buenos resultados, ya que es necesario filtrar, seleccionar, transformar y preparar los datos. Nuevos métodos para lidiar con la presencia de nubes se deberían investigar en futuros trabajos.
- Los problemas encontrados en la superposición de texturas y mapa de colores se deben en parte a que la librería utilizada para la visualización de los datos se encuentra aún en desarrollo. Sin embargo, en versiones futuras, quizás se puedan solucionar de forma fácil. También se podría recurrir a otras herramientas de visualización más maduras como pueden ser Mayavi 3D, ParaView o Blender.
- En lo que respecta a la visualización de datos de simulación, se debería tratar en un estudio diferente ya que su carácter volumétrico requiere de técnicas de visualización específicas que difieren de las empleadas con datos de satélite.

Se considera que el estudio realizado es de utilidad y puede servir de base a trabajos futuros en la materia. Además de la implementación en otras herramientas de visualización de datos, como se ha mencionado anteriormente, se podría utilizar para complementar otras gráficas en *dashboards* de monitorización medioambiental. Destacar también que la metodología propuesta se podría aplicar también a otras localizaciones geográficas y/o índices medioambientales, modificando algunos parámetros.

Este trabajo ha servido, a nivel formativo, para reforzar los conocimientos adquiridos en el máster en lo que respecta a la programación en Python; manejo de diferentes tipos de datos, uso de los metadatos, utilidad e importancia de los datos en abierto y técnicas de visualización de datos más avanzadas y, en particular, sobre datos reales que son diariamente utilizados por científicos, investigadores y administraciones públicas. También ha permitido que conozca cosas totalmente nuevas como las misiones espaciales de la ESA y la NASA, diferentes índices que se pueden derivar de las imágenes de satélite o un primer acercamiento a otros lenguajes de programación como JavaScript.

Capítulo 6. Bibliografía

- [1] T. Vinh Phan, «Immersive data visualization and storytelling based on 3D virtual reality platform.,» 2016.
- [2] J. J. van Wijk, «The value of visualization,» de *Visualization, 2005. VIS 05. IEEE*, 2005.
- [3] R. Mazza, *Introduction to information visualization*, Springer, 2009.
- [4] S. K. Card, J. D. Mackinlay y B. Shneiderman, *Readings in Information Visualization: Using Vision to Think*, Morgan Kaufmann, 1999.
- [5] B. R. Kent, *3D scientific visualization with Blender*, San Rafael: Morgan & Claypool Publishers, 2015.
- [6] J. Hodgins y A. Safonova, «Carnegie Mellon University - Computer Graphics,» 2007. [En línea]. Available: http://www.cs.cmu.edu/~jkh/462_s07/24_visualization.pdf. [Último acceso: 3 8 2019].
- [7] R. Fuchs y H. Hauser, «Visualization of multi-variati scientific data,» *Computer Graphics*, vol. 28, nº 6, pp. 1670-1690, 2009.
- [8] J. Kehrer y H. Hauser, «Visualization and visual analysis of multicafeted scientific data: A survey,» *IEEE Transactions on Visualization and Computer Graphics*, vol. 19, nº 3, pp. 495-513, 2013.
- [9] D. Keim, J. Kohlhammer, G. Ellis y F. Mansmann, *Mastering the information age. Solving problems with visual analysis*, Goslar: Eurographics Association, 2010, pp. 154-175.
- [10] H.-J. Schulz, T. Nocke, M. Heitzler y H. Schumann, «A design space of visualization tasks,» *IEEE transactions on visualization and computer graphics*, vol. 19, nº 12, pp. 2366-2375, 2013.

- [11] N. Ferreira, J. Poco, H. T. Vo, J. Freire y C. T. Silva, «Visual exploration of big spatio-temporal urban data: A study of New York City Taxi Trips,» *IEEE Transactions on Visualization and Computer Graphics*, vol. 19, nº 12, pp. 2149-2158, 2013.
- [12] A. Kumpf, B. Tost, M. Baumgart, M. Riemer, R. Westermann y M. Rautenhaus, «Visualizing Confidence in Cluster-based Ensemble Weather Forecast Analyses,» *IEEE Transactions on Visualization and Computer Graphics*, vol. 24, nº 1, pp. 109-119, 2018.
- [13] B. Spath, O. Faust, L. Pettersson, T. Hamre, D. Vitulli, A. Allen y T. Spracklen, «Integration of In-situ and remote sensing Data for Water Risk Management,» de *Integrating Sciences and Information Technology for Environmental Assessment and Decision Making*, 2008.
- [14] Agencia Espacial Europea, «Copernicus,» [En línea]. Available: <https://www.copernicus.eu/es>.
- [15] Agencia Espacial Europea, «Sentinel Online,» [En línea]. Available: <https://sentinel.esa.int/web/sentinel/home>.
- [16] T. G. Farr, P. A. Rosen, E. Caro, R. Crippen, R. Duren, S. Hensley, M. Kobrick, M. Paller, E. Rodriguez, L. Roth, D. Seal, S. Shaffer, J. Shimada, J. Umland, M. Werner, M. Oskin y Burb, «The Shuttle Radar Topography Mission,» *Reviews of Geophysics*, vol. 45, nº 2, 2007.
- [17] NASA, «Shuttle Radar Topography Mission,» [En línea]. Available: <https://www2.jpl.nasa.gov/srtm/index.html>.
- [18] I. Martínez García, «Monitorización de la calidad del agua en embalses mediante técnicas de observación remota,» 2017.
- [19] F. Aguilar Gómez, «Gestión de Datos en Entornos Cloud: aplicación para la ESFRI LifeWatch,» Santander.
- [20] A. P. Monteoliva Herreras, «Aproximación basa en ciberinfraestructuras remotas y modelado para mejorar la gestión de la eutrofización en lagos y embalses,» Santander, 2016.

- [21] M. Castrillo, D. García, F. Aguilar, J. Marco y A. Monteoliva, «Hydrodynamic and Water Quality modelization of Cuerda del Pozo reservoir,» de *Delft Software Days (2018)*, Delft, 2018.
- [22] N. Gorelick, M. Hancher, M. Dixon, S. Ilyushchenko, D. Thau y R. Moore, «Google Earth Engine: Planetary-scale geospatial analysis for everyone,» *Remote Sensing of Environment*, vol. 202, pp. 18-27, 2017.
- [23] Google, «Google Earth Engine,» [En línea]. Available: <https://earthengine.google.com/>.
- [24] D. Yacouba, H. Guangdao y W. Xingping, «Assessment of land use cover change using NDVI and DEM in Puer and Simao counties, Yunnan Province, China,» *World Rural Observations*, vol. 2, nº 1, pp. 1-11, 2009.
- [25] Z.-Z. Zhan, H.-B. Liu, H.-M. Li, W. Wu y B. Zhong, «The relationship between NDVI and Terrain Factors - A case study of Chongqing,» de *International Conference on Environmental Science and Engineering*, 2011.
- [26] L. Ho, M. Umitsu y Y. Yamaguchi, «Flood hazard mapping by satellite images and SRTM DEM in the Vu Gia - Thu Bon alluvial plain, Central Vietnam,» de *International Archives of the Photogrammetry, Remote Sensing and Spatial Information Science*, Kyoto, 2010.
- [27] J. C. Ritchie, P. V. Zimba y J. H. Everitt, «Remote Sensing Techniques to Assess Water Quality,» *Photogrammetric Engineering & Remote Sensing*, vol. 69, nº 6, pp. 695-704, 2003.
- [28] S. K. McFeeters, «The use of the Normalized Difference Water Index (NDWI) in the delineation of open water features,» *International Journal of Remote Sensing*, vol. 17, nº 7, pp. 1425-1432, 1996.
- [29] X. Hanqiu, «Modification of normalised difference water index (NDWI) to enhance open water features in remotely sensed imagery,» *International Journal of Remote Sensing*, vol. 27, nº 14, pp. 3025-3033, 2006.
- [30] K. Rokni, A. Ahmad, A. Selamat y S. Hazini, «Water feature extraction and change detection using multitemporal landsat imagery,» *Remote sensing*, vol. 6, nº 5, pp. 4173-4189, 2014.

- [31] E. Özelkan, «Water body detection analysis using NDWI indices derived from Landsat-8 OLI,» *Polish Journal of Environmental Studies*, vol. 29, nº 2, pp. 1-11, 2019.
- [32] J. Lacaux, Y. Tourre, C. Vignolles, J. Ndione y M. Lafaye, «Clasification of ponds from high-spatial resolution remote sensing: Application to Rift Valley Fever Epidemics in Senegal,» *Remote Sensing of Environment*, vol. 106, pp. 66--74, 2007.
- [33] M. Elhag, I. Gitas, A. Othman, J. Bahrawi y P. Gikas, «Assessment of Water Quality Parameters Using Temporal Remote Sensing Spectral Reflectance in Arid Environments, Saudi Arabia,» *Water*, vol. 11, nº 3, 2019.
- [34] A. R. Huete, «Remote sensing for environmental monitoring,» de *Environmental Monitoring and Characterization*, Academic Press, 2004, pp. 183-206.
- [35] Earth Observatory, «Measuring Vegetation (NDVI & EVI),» NASA, 30 Agosto 2000. [En línea]. Available: https://earthobservatory.nasa.gov/features/MeasuringVegetation/measuring_vegetation_2.php. [Último acceso: 28 Julio 2019].
- [36] University Corporation for Atmospheric Research, «Network Common Data Form (NetCDF),» [En línea]. Available: <https://www.unidata.ucar.edu/software/netcdf/>.
- [37] GeoSolutions, «NetCDF-family serving basics,» [En línea]. Available: https://geoserver.geo-solutions.it/edu/en/multidim/netcdf/netcdf_basics.html.
- [38] C. D. Hansen y C. R. Johnson, *The visualization handbook*, Elsevier, 2005.
- [39] C. Yuksel, M. Tarini y S. Lefebvre, «Rethinking Texture Mapping,» de *SIGGRAPH*, Los Angeles, 2017.
- [40] T. McReynolds y D. Blythe, *Advanced Graphics Programming using OpenGL*, San Francisco: Morgan Kaufmann, 2005.
- [41] H. Hauser, G. I. Bischì y E. Gröller, «Two-Level Volume Rendering,» *IEEE Transactions on Visualization and Computer Graphics*, vol. 7, pp. 242-252, 2001.

- [42] O. Staubli, C. Sigg, R. Peikert, D. Gubler y M. Gross, «Volume rendering of smoke propagation CFD data,» de *VIS 05. IEEE Visualization*, Minneapolis, 2005.
- [43] T. Elboth y A. Helgeland, «Hurricane Visualization Using Anisotropic Diffusion and Volume Rendering».
- [44] «Embalses.net,» [En línea]. Available: <https://www.embalses.net/pantano-125-cuerda-del-pozo.html>. [Último acceso: 10 octubre 2019].
- [45] I. Milovanovic, D. Foures y G. Vettigli, *Python Data Visualizatin Cookbook - Second Edition*, Packt, 2015.
- [46] S. Soler López, «Análisis de una especie amenazada utilizando técnicas de observación de la tierra: una aplicación de data science,» 2017.
- [47] C. Tominski, P. Schulze-Wollgast y H. Schumann, «3D information visualization for time dependent data on maps,» de *Ninth International Conference on Information Visualisation (IV'05)*, Londres, 2005.
- [48] S. Hsiang, R. Kopp, A. Jina, J. Rising y M. Delgado, «Estimating economic damage from climate change in the United States,» *Science*, vol. 356, pp. 1362-1369, 2017.
- [49] W. Aigner, S. Miksch, W. Müller, H. Schumann y C. Tominski, «Visualizing time-oriented data - A systematic view,» *Computers & Graphics*, vol. 31, nº 3, pp. 401-409, 2007.
- [50] S. Dübel, M. Röhlig, H. Schumann y M. Trapp, «2D and 3D presentation of spatial data: A systematic Review,» de *IEEE VIS International Workshop on 3DVis*, 2014.
- [51] J. Soo Yi, Y. ah Kang y J. T. Stasko, «Toward a Deeper Understanding of the Role of Interaction in Information Visualization,» *IEEE Transactions on Visualization and Computer Graphics*, vol. 13, nº 6, pp. 1224 - 1231, 2007.
- [52] Ö. Gürsoy y R. Atun, «Investigating surface water pollution by integrated remotely sensed and field spectral measurement data: A case study,» *Polish Journal of Environmental Studies*, vol. 28, nº 4, pp. 2027-2040, 2019.
- [53] A. Bock, *Tailoring Visualization Applications for Tasks and Users*, Linköping: LiU-Tryck, 2018.

- [54] W. Burkhard , «Chapter 5 - Visualization techniques for scalar fields,» 2006. [En línea]. Available: https://www.cs.auckland.ac.nz/courses/compsci716s2t/lectures/716Handout5_ScalarIcons_4up.pdf. [Último acceso: 2018 08 20].
- [55] J. Tierny, «Introduction to volume rendering,» [En línea]. Available: <https://pdfs.semanticscholar.org/90a9/04994dcfd1b8266f859260c75fe6804049d2.pdf>. [Último acceso: 2018 08 20].