



UNIVERSITAT OBERTA DE CATALUNYA (UOC)
MÁSTER UNIVERSITARIO EN CIENCIA DE DATOS (*Data Science*)

TRABAJO FINAL DE MÁSTER

ÁREA: DATA ANALYSIS Y BIG DATA (AREA 1)

Evaluación de algoritmos de aprendizaje automático para la determinación de parámetros de calidad del agua mediante teledetección

Autor: Vanesa Navarro Oronoz

Tutor: Anna Muñoz Bolas

Profesor: Albert Solé Ribalta

Barcelona, 15 de enero de 2023



Esta obra está sujeta a una licencia de Reconocimiento - NoComercial - SinObraDerivada
3.0 España de CreativeCommons.

FICHA DEL TRABAJO FINAL

Título del trabajo:	Evaluación de algoritmos de aprendizaje automático para la determinación de parámetros de calidad del agua mediante teledetección
Nombre del autor:	Vanesa Navarro Oronoz
Nombre del colaborador/a docente:	Anna Muñoz Bollas
Nombre del PRA:	Albert Solé Ribalta
Fecha de entrega (mm/aaaa):	01/2023
Titulación o programa:	Máster Universitario en Ciencia de Datos
Área del Trabajo Final:	Data Analysis y Big Data (Area 1)
Idioma del trabajo:	Castellano
Palabras clave	Calidad de agua, teledetección, aprendizaje automático

Resumen

El seguimiento de la calidad del agua es una prioridad para las necesidades de abastecimiento y consumo, y también para el buen funcionamiento de los ecosistemas naturales. En este trabajo se estudia el potencial del monitoreo de parámetros indicativos de la calidad del agua empleando técnicas avanzadas de teledetección y algoritmos de aprendizaje automático en el embalse El Val (Aragón), ya que está catalogado como -Zona Sensible- y se encuentra en situación de vulnerabilidad y riesgo medioambiental.

Este trabajo presenta los resultados de la validación de imágenes térmicas del sensor TIRS a bordo de las misiones Landsat para obtener la temperatura superficial del agua, y la validación de imágenes ópticas capturadas por el sensor MSI a bordo de la misión Sentinel para obtener la turbidez y concentración de clorofila-a entre enero de 2018 y diciembre de 2022.

En la estimación de la temperatura, las imágenes se han calibrado con los datos de campo obteniendo un modelo de regresión lineal ($R^2=0.98$) con error de 1 °C. Para la estimación de clorofila-a y turbidez se han evaluado los algoritmos Decision Tree, Random Forest y SVM con búsqueda automática de la mejor combinación de hiperparámetros y validación cruzada con 5 particiones estratificadas. En la estimación de clorofila-a el desempeño de Decision Tree y Random Forest ha sido similar explicando un 88 % de la varianza del modelo y con errores entre los 10-15 mg/l. Los modelos evaluados para estimar la turbidez no han sabido representar el comportamiento de la variable de forma adecuada.

Palabras clave: Calidad de agua, teledetección, aprendizaje automático

Abstract

Monitoring water quality is a priority for supply and consumption needs, and also for the proper functioning of natural ecosystems. This paper studies the potential of monitoring indicative parameters of water quality using advanced remote sensing techniques and machine learning algorithms in the El Val reservoir (Aragón), since it is classified as a -Sensitive Zone- and it is in a vulnerable and environmental risk situation.

This paper presents the results of the validation of thermal images of the TIRS sensor on board the Landsat missions to obtain the surface temperature of the water, and the validation of optical images captured by the MSI sensor on board the Sentinel mission to obtain the turbidity and chlorophyll-a concentration between January 2018 and December 2022.

In estimating temperature, the images have been calibrated with field data, obtaining a linear regression model ($R^2=0.98$) with an error of 1 °C. For the estimation of chlorophyll-a and turbidity, the Decision Tree, Random Forest and SVM algorithms have been evaluated with automatic search for the best combination of hyperparameters and cross validation with 5 stratified partitions. In the estimation of chlorophyll-a, the performance of Decision Tree and Random Forest has been similar, explaining 88 % of the variance of the model and with errors between 10-15 mg/l. The models evaluated to estimate turbidity have not been able to adequately represent the behavior of the variable.

Key words: Water quality, remote sensing, machine learning

Índice general

Resumen	III
Abstract	IV
Indice	V
Listado de Figuras	VI
1. Introducción	2
1.1. Contexto y justificación del trabajo	2
1.2. Motivación	5
1.3. Objetivos del trabajo	5
1.3.1. Objetivo principal (Hipótesis)	5
1.3.2. Objetivos parciales (Preguntas de investigación)	5
1.4. Impacto en sostenibilidad, ético-social y de diversidad	6
1.5. Enfoque y método seguido	8
1.6. Planificación del trabajo	9
1.6.1. Fases del proyecto	9
1.6.2. Fechas clave	11
1.6.3. Diagrama de Gantt	11
1.6.4. Recursos	12
1.7. Breve resumen de productos obtenidos	13
1.8. Breve descripción de otros capítulos de la memoria	13
2. Estado del arte	15
2.1. Marco Teórico	15
2.1.1. Conceptos básicos de teledetección	15
2.1.2. Calidad del agua mediante teledetección	17
2.2. Proyectos de referencia	21

2.2.1. Proyectos destacados a nivel nacional	21
2.2.2. Proyectos destacados a nivel internacional	23
2.3. Planteamiento del trabajo a partir del estado del arte	24
2.4. Decisiones tecnológicas	24
3. Diseño e implementación del trabajo	25
3.1. Flujo de trabajo para un programa de seguimiento de la calidad del agua mediante teledetección	25
3.2. Extracción y limpieza de los datos	26
3.2.1. Datos de la sonda instalada en El Val	26
3.2.2. Imágenes satelitales	28
3.3. Ajuste y evaluación de algoritmos	32
3.3.1. Temperatura del agua	34
3.3.2. Concentración de Clorofila-A	36
3.3.3. Turbidez	37
3.4. Resultados	39
3.4.1. Mapas de calidad del agua	39
3.4.2. Estimación del potencial ecológico del embalse	39
4. Conclusiones	41
4.1. Conclusiones generales del trabajo	41
4.1.1. Conclusiones: temperatura superficial del agua	43
4.1.2. Conclusiones: concentración de clorofila-a	44
4.1.3. Conclusiones: turbidez	45
4.2. Líneas de trabajo futuras	46
Bibliografía	47

Índice de figuras

1.1.	Imagen del embalse El Val en un estado de eutrofización.	4
1.2.	Diagrama de Gantt. Fuente: Elaboración propia.	12
2.1.	Ejemplo de un sistema de teledetección basado en satélites pasivos. Fuente: The Nature Conservancy.	16
2.2.	Espectro electromagnético. Fuente: Wikipedia.	17
2.3.	Diferencias en los colores y firmas espectrales únicas (reflectancia en función de la longitud de onda de la luz) de las aguas interiores (izquierda) a las oceánicas (derecha).	19
2.4.	Firmas espectrales de aguas pura (azul) y con presencia mayoritaria de CDOM (gris), fitoplancton (verde) y sólidos en suspensión (amarillo). Fuente: NASA-ARSET.	21
3.1.	Propuesta metodológica para el monitoreo de la calidad del agua mediante teledetección.	25
3.2.	Udicación en planta de la sonda (izquierda) y foto del equipo en detalle (derecha).	26
3.3.	Descarga de los datos públicos desde la web de la CHE mediante web-scraping con Selenium y BeautifulSoup.	27
3.4.	Función que ejecuta todas las tareas de limpieza y filtrado de datos del csv descargado original.	28
3.5.	Metadatos Landsat (izquierda) y Sentinel (derecha) con las horas exactas de toma de la imagen para la zona de estudio.	29
3.6.	Código para obtención de la lista de fechas con las imágenes Sentinel-2 disponibles en la zona de estudio en el intervalo temporal deseado mediante la interacción de los módulos geemap y el catálogo de Earth Engine.	30
3.7.	Código para obtención de los valores de temperatura superficial en el punto de evaluación.	31
3.8.	Código para obtención de los valores del índice NDCI en el punto de evaluación.	33

3.9. Estadísticas básicas (izquierda) y coeficiente de correlación de Pearson (derecha).	34
3.10. Diagrama de cajas (izquierda) y diferencia entre los datos observados vs estimados en función de la temperatura del agua (derecha).	35
3.11. Gráfico de dispersión con los datos estimados (eje X) vs datos observados (eje Y).	35
3.12. Código para crear modelo regresión lineal con datos de temperatura.	36
3.13. Gráfico de dispersión con el valor del índice de clorofila NDCI (eje X) vs datos observados (eje Y).	37
3.14. Código para crear modelo de árbol de decisión.	38
3.15. Mapa representando el estado trófico del embalse del Val en un evento de eutrofización.	39
3.16. Límites de clase de Potencial Ecológico según DMA.	40
3.17. Clasificación Potencial Ecológico embalse El Val.	40
4.1. Ubicación real de la sonda en campo y el píxel empleado para evaluación en las imágenes satelitales.	42

Siglas

ARSET Applied Remote Sensing Training. [23](#)

CCEG Competencia de Compromiso Ético y Global. [6](#)

CDOM Materia orgánica disuelta coloreada. [20](#), [21](#), [23](#)

CHE Confederación Hidrográfica del Ebro. [4](#), [8](#), [23](#), [26](#)

CRISP-DM Cross-Industry Standard Process for Data Mining. [8](#)

DMA Directiva Marco Europea del Agua. [3](#), [39](#)

ESAQS Ecological Status of Aquatic Systems with Sentinel Satellites. [22](#)

FAO Organización de las Naciones Unidas para la Alimentación y la Agricultura. [2](#)

GEE Google Earth Engine. [29](#), [31](#), [40](#)

IoT Internet of Things. [24](#)

NDCI Índice de Clorofila de Diferencia Normalizada. [32](#), [36](#), [47](#)

NDTI Índice de Turbidez de Diferencia Normalizada. [32](#), [47](#)

NTU Nephelometric Turbidity Unit o Unidades Nefelométricas de Turbidez. [38](#)

OAC Componentes Ópticamente Activos. [20](#)

ODS Objetivos de Desarrollo Sostenible 2030. [6](#), [7](#)

OMS Organización Mundial de la Salud. [2](#)

PEC Prueba Evaluación Continua. [9](#)

SAICA Sistema Automático de Información de Calidad de las Aguas. 8

SIG Sistemas de Información Geográfica. 12, 24

SVM Support Vector Machines. 33, 36

TFM Trabajo Final de Máster. 5

Capítulo 1

Introducción

1.1. Contexto y justificación del trabajo

La preservación de la calidad de los cuerpos de agua es una preocupación mundialmente reconocida. Organizaciones como la [OMS](#) consideran que el monitoreo continuo de los cuerpos de agua es uno de los prerrequisitos para evitar el deterioro de los cuerpos más vulnerables. La calidad del agua de lagos, lagunas, embalses y los ríos es fundamental para el ser humano. El agua tiene un alto valor ecológico porque es esencial para la vida y es el hábitat de un gran número de seres vivos. Aunque las masas de aguas interiores solo cubren alrededor del 1 % de la superficie de la Tierra, es el hábitat de alrededor del 40 % de las especies de peces ([FAO, 2020](#)) [1], y brinda alimentación, recreación y recursos que afectan directamente la vida y la economía de la población humana.

Por ello, el seguimiento de la calidad del agua es una prioridad, no solo en cuanto a las necesidades de abastecimiento humano, consumo agronómico e industrial, sino también para el buen funcionamiento y control de los ecosistemas naturales. La calidad del agua describe la condición física de ésta incluyendo sus características químicas, físicas y biológicas, generalmente respecto a su idoneidad para algún propósito particular, por ejemplo, el funcionamiento de los ecosistemas o salud humana. La calidad del agua está influenciada por los siguientes factores:

- Eutrofización o sobrecarga de nutrientes debido a la escorrentía agrícola y a la lixiviación séptica.
- Contaminación por fuentes industriales puntuales o deposiciones provenientes de la atmósfera.
- El cambio climático está provocando un incremento de la temperatura que estimula las floraciones de cianobacterias y una acidificación debido al CO₂ elevado.
- Cambios en la red alimentaria.

- Introducción de nuevas especies.
- Cambios en el flujo del agua ya sea por la construcción de elementos de regulación o por eventos naturales como inundaciones y sequías.

El avance en las técnicas de teledetección y el desarrollo de la tecnología en general, está facilitando y fomentando los estudios de observación terrestre mediante imágenes satelitales debido a la mayor disponibilidad de datos y a una menor inversión de recursos. Con estos métodos es posible obtener sistemáticamente grandes cantidades de información a un coste menor que con los métodos convencionales de tomas de muestras en campo. Además, el seguimiento de una masa de agua con muestreos in situ no permite tener una visión global del comportamiento del sistema.

Existen numerosos estudios de teledetección aplicados a la calidad del agua que ponen de manifiesto la capacidad que poseen los sensores remotos para controlar ciertos parámetros de calidad (aquellos que modifican directamente las propiedades ópticas o térmicas de las aguas superficiales), con los que hacer el seguimiento del estado de la masa de agua con una resolución temporal y espacial adecuada.

En el contexto legal europeo, el control del estado de los embalses es un requerimiento de la [Directiva Marco Europea del Agua](#). Esta Directiva, en su Artículo 8, obliga a los Estados miembros a establecer programas para el seguimiento de la calidad de las aguas de las masas de agua con objeto de obtener una visión general coherente y completa de su potencial ecológico en cada demarcación hidrográfica [2]. Esta obligación de la [DMA](#) se traspa al ordenamiento nacional a través del Real Decreto Legislativo 1/2001, de 20 de julio, por el que se aprueba el texto refundido de la Ley de Aguas.

Según la [DMA](#), los embalses se consideran dentro de la categoría río de naturaleza muy modificada debido a que, bien por su tamaño, por la longitud fluvial afectada o por el fuerte efecto regulador que ejercen, condicionan una modificación en el río que puede considerarse estable y duradera, llegando en ocasiones a una nueva situación de equilibrio que se estima de reversibilidad compleja y socialmente indeseada. La alteración de las condiciones naturales es tan fuerte que se considera que no pueden alcanzar el buen estado ecológico según la definición de la [DMA](#). El objetivo para estas masas es el de alcanzar el buen potencial ecológico.

El embalse El Val (Aragón) soporta las consecuencias de la eutrofización y su potencial ecológico es inferior a bueno, que es el objetivo medioambiental según la [DMA](#). De hecho, en el año 2019 se declaró “Zona Sensible” como mecanismo de protección, por el Ministerio de Transición Ecológica, lo que implica la adopción de medidas especiales para reducir la aportación de nutrientes en el área de captación y también un seguimiento específico de la calidad

de sus aguas. En los estudios más recientes realizados en el embalse por el Área de Calidad de la [Confederación Hidrográfica del Ebro](#), se ha diagnosticado su potencial ecológico como moderado, y su estado trófico como eutrófico [3]. En la siguiente imagen 1.1 se puede distinguir el color verde de las aguas del embalse debido a un proceso de eutrofización.



Figura 1.1: Imagen del embalse El Val en un estado de eutrofización.

Por estas razones, se ha considerado conveniente el refuerzo de las medidas habituales de seguimiento de la calidad del agua y control de vertidos con la instalación de una sonda automática que obtiene datos continuos en tiempo real, pero solo de un punto del embalse.

Debido a la disponibilidad abundante de datos en tiempo real para poder calibrar las imágenes y la situación de vulnerabilidad y riesgo medioambiental del embalse, se ha decidido que el área de estudio del presente trabajo se centre en el embalse El Val, donde el empleo de imágenes satelitales de alta resolución espacial puede suponer una fuente valiosa de información para entender el comportamiento global del cuerpo de agua y detectar de forma temprana episodios y fuentes de contaminación.

1.2. Motivación

Hace años que ejerzo profesionalmente como Analista de Sistemas de Información Geográfica y Teledetección por lo que conozco de primera mano el potencial de la información geoespacial. Durante el tiempo que he estado cursando el Máster Universitario en Ciencia de Datos, he desarrollado nuevas habilidades para el tratamiento y análisis de datos en general y la tecnología de la información geoespacial se ha optimizado, surgiendo nuevas plataformas, nuevas aplicaciones, nuevas misiones satelitales, cambiando la forma en la que accedemos y procesamos la información si lo comparamos con un proyecto de teledetección cinco años atrás.

Actualmente desarrollo mi actividad profesional en la gestión del agua de Navarra. Geográficamente el embalse El Val se sitúa en la cabecera del río Queiles cuyo recorrido transcurre por territorio navarro hasta desembocar en el río Ebro. Este río sufre una gran presión antrópica y se ve afectado directamente por la gestión y contaminación del embalse El Val. Debido al conocimiento de primera mano de esta situación y a la alta disponibilidad de datos del embalse por su situación de vulnerabilidad es que he decidido realizar este trabajo. El TFM me permite integrar y testear los conocimientos adquiridos en el máster junto con la experiencia en el área de calidad de agua.

De esta manera, la propuesta metodológica que aquí se va a estudiar puede ser extrapolable a otros cuerpos de agua vulnerables de diferentes demarcaciones hidrográficas.

1.3. Objetivos del trabajo

1.3.1. Objetivo principal (Hipótesis)

Este proyecto tiene como objetivo el monitoreo de parámetros indicativos de la calidad del agua (temperatura superficial, turbidez, y concentración de clorofila-a) empleando técnicas avanzadas de teledetección y algoritmos de aprendizaje automático en el Embalse El Val, Aragón.

1.3.2. Objetivos parciales (Preguntas de investigación)

Dentro de los objetivos parciales, se ha diferenciado entre los objetivos generales definidos en el Plan Docente para un [Trabajo Final de Máster](#), cuyo enfoque es académico, y los objetivos específicos de este trabajo cuyo carácter es 100 % técnico.

Objetivos básicos generales TFM

- Poseer y comprender conocimientos que aporten una base u oportunidad de ser originales en el desarrollo y/o aplicación de ideas, a menudo en un contexto de investigación.
- Comunicar las conclusiones del trabajo, y los conocimientos y razones últimas que las sustentan, a públicos especializados y no especializados de un modo claro y sin ambigüedades.
- Extraer, interpretar y analizar datos de diferentes entornos.
- Aplicar conocimientos de ciencia de datos con el objetivo de resolver problemas relacionados con la toma de decisiones en diferentes entornos, tanto conocidos como nuevos.
- Presentar los resultados obtenidos del análisis de los datos de forma eficiente y atractiva.
- Buscar, gestionar y usar información y recursos en el ámbito de la ciencia de datos.
- Trabajar la iniciativa, la automotivación y el trabajo de forma independiente.
- Proponer soluciones innovadoras y toma de decisiones.
- Analizar un problema de ciencia de datos y ser capaz de identificar y definir los requerimientos apropiados para resolverlo.
- Presentar y defender un trabajo en el ámbito de la ciencia de los datos ante un público experto.

Objetivos específicos del proyecto

- Obtener los datos de las diferentes fuentes de información (imágenes satelitales y sensores in situ).
- Realizar el tratamiento, limpieza y homogeneización de los datos extraídos.
- Desarrollar algoritmos basados en la correspondencia entre los datos obtenidos in situ y los valores capturados por el satélite.
- Evaluar el desempeño de diferentes algoritmos para la determinación de las variables de calidad de agua calculando su error medio absoluto y su error cuadrático medio.
- Generar cartografía de calidad del agua casi en tiempo real (cada vez que haya una imagen actualizada disponible).
- Definir el estado trófico y el potencial ecológico para el embalse El Val de forma continua, global y casi en tiempo real.

1.4. Impacto en sostenibilidad, ético-social y de diversidad

La [Competencia de Compromiso Ético y Global](#) aborda tres grandes dimensiones (sostenibilidad, comportamiento ético y responsabilidad social, y diversidad y derechos humanos) que se alinean con los [ODS](#) [4]. Este trabajo, dentro de ese marco, se encuadra dentro de los siguientes [ODS](#):

- 3 – Salud y bienestar.
- 6 – Agua limpia y saneamiento.
- 13 – Acción por el clima.
- 14 – Vida submarina.
- 15 – Vida ecosistemas terrestres.

A continuación, se mencionan las líneas de trabajo específicas de cada ODS con las que este trabajo colabora:

- **ODS3: Reducir considerablemente el número de muertes y enfermedades** causadas por productos químicos peligrosos y por la polución y **contaminación del agua**, el aire y el suelo.
- **ODS6: Mejorar la calidad del agua** reduciendo la contaminación, eliminando el vertimiento y minimizando la emisión de productos químicos y materiales peligrosos, **uno de cuyos indicadores es en concreto la proporción de masas de agua de buena calidad.**
- **ODS6: Proteger y restablecer los ecosistemas relacionados con el agua**, incluidos los bosques, las montañas, los humedales, los ríos, los acuíferos y los lagos.
- **ODS13: Fortalecer la resiliencia y la capacidad de adaptación a los riesgos relacionados con el clima y los desastres naturales** en todos los países.
- **ODS14: Prevenir y reducir significativamente la contaminación marina** de todo tipo, en particular la producida por actividades realizadas en tierra, incluidos los detritos marinos y la **polución por nutrientes.**
- **ODS15: Asegurar la conservación, el restablecimiento y el uso sostenible** de los ecosistemas terrestres y los **ecosistemas interiores de agua dulce** y sus servicios, en particular los bosques, los humedales, las montañas y las zonas áridas, en consonancia con las obligaciones contraídas en virtud de acuerdos internacionales.

La naturaleza de este trabajo se fundamenta en un recurso fundamental para la vida de cualquier especie: el agua. La garantía del suministro de agua en cantidad y en calidad suficientes, lo que implica un uso eficiente y sostenible de los recursos hídricos disponibles, es fundamental para el desarrollo de la sociedad y para la lucha contra la pobreza y las enfermedades en cualquier parte del mundo. El carácter transversal del agua hace que sea un elemento clave para el progreso sostenible a nivel económico, social y ambiental y, por ello, el desarrollo del trabajo atraviesa diferentes dimensiones y sectores de los ODS como son la salud, el clima, los ecosistemas y la vida submarina.

Por otra parte, el objetivo 6 está fuertemente relacionado con el objetivo 13 de lucha contra el cambio climático. El cambio climático afecta a todos los países en todos los continentes del

mundo. Las consecuencias del cambio climático incluyen cambios en los patrones climáticos, aumento de las temperaturas, aumento del nivel del mar y fenómenos meteorológicos más extremos. Dada la naturaleza semiárida o árida de gran parte del territorio nacional, donde se incluye el área de estudio de este trabajo, existe un riesgo de desertificación asociado al cambio climático. La gestión del agua y garantizar su calidad, es un reto cada vez más complejo en el contexto del cambio climático, y que sitúa a la gestión de los recursos hídricos en un papel destacado en el marco de las políticas públicas. Como se verá a lo largo de este trabajo, algunos de los procesos que degradan la calidad del agua están directamente relacionados, no solo con la contaminación antrópica, sino con el aumento de temperaturas.

Por todos estos motivos, este trabajo supone una herramienta para abrir nuevas líneas de trabajo y metodologías en la gestión y conocimiento de la calidad del agua que puedan ayudar en la toma de decisiones a los gestores del agua y se pueda llevar a cabo una gobernanza de los recursos hídricos de forma sostenible y respetuosa con el medio ambiente.

1.5. Enfoque y método seguido

La metodología que se va a seguir para llevar a cabo este trabajo es la **CRoss-Industry Standard Process for Data Mining (CRISP-DM)**, considerada la metodología de facto para proyectos dedicados a extraer valor de los datos. Este método se va a tomar como base para organizar el proyecto, y se personalizará o flexibilizará en función de los objetivos específicos de este trabajo.

La metodología **CRISP-DM** contempla seis fases las cuales dependen entre sí tanto en forma secuencial como cíclica. Estas fases son la comprensión del problema, la comprensión de los datos, la preparación de los datos, el modelado, la evaluación y el despliegue.

En este caso, la comprensión del problema se ha expuesto en el apartado 4.2. Actualmente el monitoreo de la calidad se realiza de forma puntual, con una sonda instalada en el embalse, pero una sola medida en un punto fijo no es representativa de la situación global del sistema y de los procesos que ocurren en la superficie total del embalse que es de 112 hectáreas, por ello se cree que el monitoreo mediante imágenes satelitales va a generar un valor añadido y convertirse en una fuente valiosa de información.

En cuanto a la comprensión y preparación de los datos, en primer lugar, los datos reales de campo se van a descargar a través del **Sistema Automático de Información de Calidad de las Aguas** de la **CHE**, que los publica de forma libre y en tiempo real. La sonda instalada realiza perfiles en toda la columna de agua obteniendo como resultado los diferentes parámetros de

calidad medidos a diferentes alturas. En el caso del presente trabajo, los resultados que interesan son los obtenidos en el primer metro de la columna de agua ya que para las imágenes satelitales la transparencia del agua limita la obtención de información y la energía que recibe el sensor satelital es la reflectancia saliente del agua. Una vez obtenidos y tratados los datos de campo, se obtienen las imágenes satelitales Landsat y Sentinel para el intervalo temporal que interese a través del servicio Google Earth Engine. Las imágenes serán procesadas de forma independiente, aplicando diferentes correcciones, ya sean imágenes térmicas u ópticas.

En la fase de modelado, donde se construye el proyecto, la metodología propuesta se fundamenta en el desarrollo de algoritmos ad-hoc basados en la correspondencia de los datos muestreados in situ por la sonda instalada en el embalse y los valores de reflectancia o temperatura capturados por los sensores instalados en los satélites para la estimación de la calidad del agua por medio de imágenes satelitales de las misiones Landsat (8 y 9) y Sentinel (2). A partir de las imágenes Landsat se obtiene el parámetro de temperatura superficial del agua y de las imágenes Sentinel se calculan las variables de turbidez y concentración de clorofila-a.

La evaluación se va a llevar a cabo midiendo el rendimiento de los diferentes algoritmos de aprendizaje automático, cuantificando su error medio absoluto y su error cuadrático medio, escogiendo el más preciso en base a los datos disponibles de campo.

El resultado o despliegue final que se persigue en este trabajo es el generar mapas temáticos para las diferentes variables de calidad del agua de estudio (temperatura superficial del agua, turbidez y concentración de clorofila-a) en el embalse El Val y definir su estado trófico y potencial ecológico casi en tiempo real.

1.6. Planificación del trabajo

La estructura, organización y planificación del presente trabajo vienen determinadas por los requerimientos de las diferentes Pruebas de Evaluación Continua (PEC) demandadas desde el Plan Docente de la asignatura. En el Plan Docente se detallan 4 PECs correspondientes a la Evaluación Continua y dos defensas del trabajo, una mediante un vídeo-presentación y otra pública ante el tribunal evaluador.

1.6.1. Fases del proyecto

Estos hitos prefijados, definen las diferentes fases del proyecto que a continuación se detallan:

Fase 1: Definición y planificación del TFM \approx PEC1

- Escoger un título que defina claramente la temática del trabajo y unas palabras clave que lo describan.
- Resumir la propuesta de la temática del trabajo.
- Contextualizar y justificar el trabajo.
- Explicar la motivación personal para la investigación de la temática.
- Establecer el objetivo general y los objetivos específicos del trabajo (alcance del TFM).
- Enmarcar el trabajo dentro de la competencia de compromiso ético y global y los objetivos de desarrollo sostenible.
- Describir la metodología a seguir y el proceso de desarrollo del trabajo.
- Planificar las fases del trabajo identificando las diferentes actividades a ejecutar su duración y la dependencia entre ellas.
- Identificar los recursos necesarios para la consecución de los objetivos del trabajo.

Fase 2: Estado del arte \approx PEC2

- Buscar bibliografía adecuada que justifique con evidencias científicas el trabajo escogido.
- Investigar la existencia de proyectos con temática y objetivos similares.
- Refinar los objetivos parciales definidos anteriormente, si procede.
- Identificar la metodología y técnicas de obtención de datos que se van a aplicar en el trabajo.
- Determinar las herramientas de desarrollo a emplear durante la implementación.

Fase 3: Diseño e implementación del trabajo \approx PEC3

- Preparar el entorno de desarrollo, investigar los conjuntos de datos disponibles, las librerías necesarias y descargar plugins para QGIS (si procede).
- Obtener el histórico de los datos de la sonda instalada en el embalse El Val y el listado de las imágenes Landsat y Sentinel disponibles en el intervalo temporal de los datos de campo.
- Procesar las imágenes satelitales (filtros, correcciones, etc..) para poder trabajarlas.
- Aplicar algoritmos para relacionar la reflectancia saliente del agua que recoge el satélite con los valores reales de la masa de agua medidos in situ.
- Calcular los errores de los diferentes algoritmos para evaluar el rendimiento de cada uno.
- Elegir el método con menor error y generar cartografía temática de la masa de agua con los diferentes parámetros de calidad analizados.
- Automatizar el proceso completo para poder generar mapas de calidad y poder definir el estado trófico y potencial ecológico de todo el embalse casi en tiempo real.

Fase 4: Redacción de la memoria \approx PEC4

- Redacción del documento de manera transversal a todas las PECs.
- Especificar las conclusiones y las lecciones aprendidas del trabajo validando los resultados obtenidos.
- Indicar las líneas de trabajo futuras.
- Repasar y corregir detalles en la redacción del documento final.
- Incorporar los comentarios e indicaciones de la tutora en la versión final del documento.

Fase 5: Presentación y defensa del proyecto \approx DEF1

- Resumir los aspectos más relevantes del trabajo y recogerlos en una presentación de diapositivas.
- Crear un vídeo exponiendo el trabajo realizado.

Fase 6: Defensa pública del TFM \approx DEF2

- Presentar y defender el trabajo frente al tribunal evaluador respondiendo a sus preguntas.

1.6.2. Fechas clave

Según la planificación que se ha detallado en el punto anterior, éstas son las fechas clave del proyecto:

- **Inicio del proyecto:** 28 de septiembre 2022
- **Entrega PEC1:** 9 de octubre 2022
- **Entrega PEC2:** 23 de octubre de 2022
- **Entrega PEC3:** 25 de diciembre de 2022
- **Entrega PEC4:** 15 enero de 2023
- **Presentación y defensa del proyecto:** 22 enero de 2023
- **Defensa pública:** fecha por determinar (final de enero/principio febrero de 2023)

El proyecto tiene una duración total aproximada de 4 meses y una carga lectiva de 12 ECTS, equivalente a 300 horas de trabajo.

1.6.3. Diagrama de Gantt

La figura 1.2 representa la planificación temporal de las diferente tareas detalladas anteriormente, mediante el uso de un diagrama de Gantt.

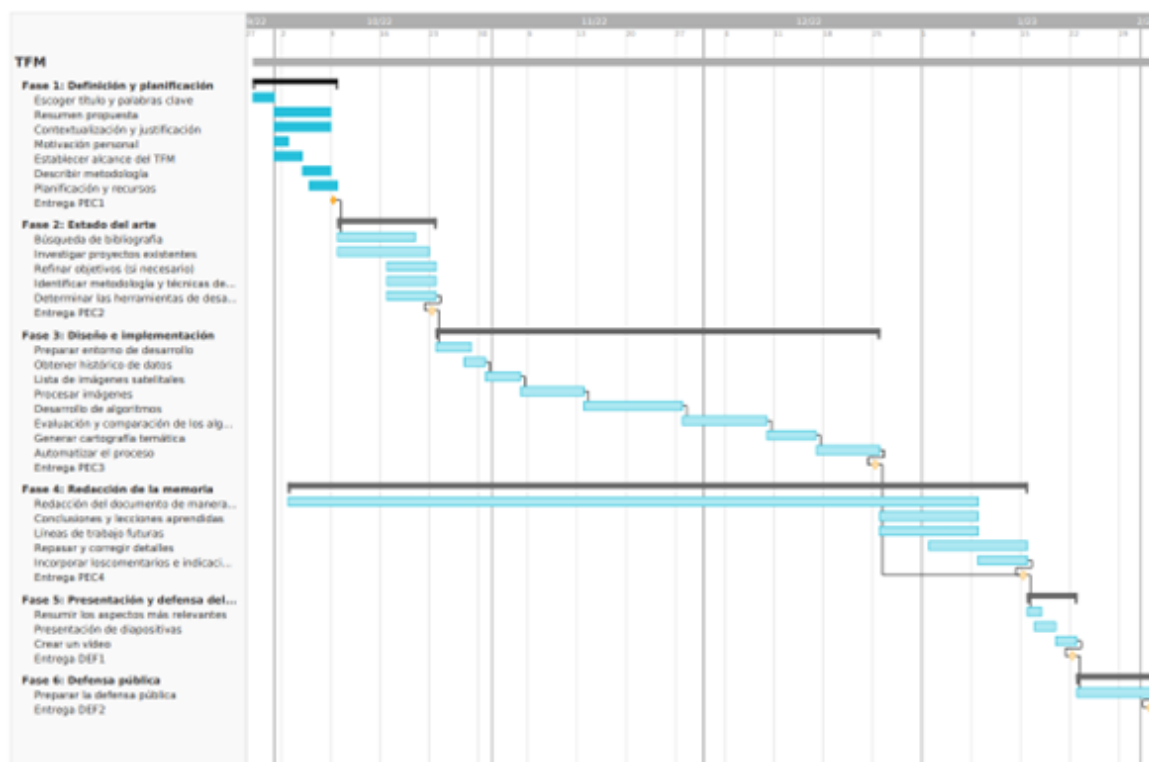


Figura 1.2: Diagrama de Gantt. Fuente: Elaboración propia.

1.6.4. Recursos

La lista de requerimientos de software y hardware necesarios para el desarrollo del proyecto es la siguiente:

Software

- **Spyder:** entorno de desarrollo integrado multiplataforma de código abierto para programación científica en el lenguaje Python.
- **Jupyter Notebook:** plataforma informática interactiva basada en la web.
- **Librerías específicas de Python** para el tratamiento de imágenes satelitales y datos geoespaciales.
- **QGIS:** SIG de software libre y de código abierto.
- **Google Earth Engine:** servicio de procesamiento geoespacial cuyo archivo públicamente disponible incluye más de cuarenta años de imágenes históricas y datasets geoespaciales que son actualizados a diario.
- **Google Chrome:** navegador web.
- **Team Gantt:** software gratuito de creación de diagramas de Gantt en línea.
- **Microsoft Word:** software de procesamiento de texto.

- **Microsoft PowerPoint:** software de presentación de diapositivas.
- **OBS Studio:** aplicación libre y de código abierto para la grabación de vídeo.
- **Github:** plataforma para alojar proyectos de desarrollos empleando el sistema de control de versiones Git.
- **Overleaf:** editor online LaTeX.

Hardware (mínimo)

PC (portátil o sobremesa) con procesador Intel(R) Core(TM) i5 @ 2.40GHz, memoria RAM de 8GB, sistema operativo de 64 bits y conexión a internet.

1.7. Breve resumen de productos obtenidos

Ver el [Repositorio Github](#) con el código de los scripts y los archivos de configuración y análisis generados.

1.8. Breve descripción de otros capítulos de la memoria

La memoria del presente trabajo se estructura en dos bloques diferenciados. El primer bloque consta de la introducción a la temática, a los objetivos del proyecto y del estado del arte en investigación de la temática escogida. Toda esta recopilación de información y análisis previo sirven de guía y referencia, sentando las bases del desarrollo del presente proyecto y estableciendo una línea metodológica. El segundo bloque constituye el desarrollo del trabajo en sí mismo, describiendo y documentando la adquisición y tratamiento de los datos necesarios para la generación de resultados y exposición de las conclusiones.

En el capítulo 2 Estado del arte, se estudian de forma resumida los procesos físicos sobre los que se basa la teledetección y se analizan brevemente las propiedades y componentes del agua que permiten la aplicación de sensores remotos para el monitoreo de aguas continentales. Al final del capítulo se enumeran algunos proyectos de referencia a nivel nacional e internacional y se explica cómo la experiencia adquirida a través del estudio de estos proyectos se aplica en el presente trabajo.

En el capítulo 3 Diseño e Implementación del trabajo, se explica de forma detallada el proceso de obtención de los datos de campo y de las imágenes satelitales, adjuntando los códigos de descarga, limpieza y homogeneización de la información para que cualquier usuario pueda reproducir el proceso. Además se estudian las relaciones estadísticas entre los datos observados en campo y los datos obtenidos por los sensores remotos, que en el caso de la temperatura superficial la medida es directa (el sensor remoto mide temperatura), pero en el

caso de la turbidez y la clorofila-a las medidas de los sensores no son directas (miden reflectancia-energía-, no los parámetros en sí), por lo que se emplean índices espectrales normalizados para relacionar los píxeles de las imágenes y los datos medidos in situ.

En el capítulo 4 Conclusiones, se explican los resultados obtenidos y el conocimiento generado gracias al presente trabajo, además de una revisión de los objetivos alcanzados. Se analiza en detalle la validación de los modelos obtenidos y se dejan abiertas líneas de investigación y experimentos futuros no explorados para lograr una mayor robustez en los modelos calibrados en este proyecto.

Capítulo 2

Estado del arte

2.1. Marco Teórico

2.1.1. Conceptos básicos de teledetección

La teledetección o percepción/detección remota (*remote sensing en inglés*) es una disciplina científica que integra un amplio conjunto de conocimientos y tecnologías utilizados para la observación, el análisis y la interpretación de fenómenos terrestres y atmosféricos. Sus principales fuentes de información son las medidas y las imágenes obtenidas con la ayuda de plataformas aéreas y espaciales.

La teledetección se conforma como una ciencia multidisciplinar basada en tres pilares fundamentales: los sensores que permiten captar la información que se desea obtener (CAPTURA), los dispositivos utilizados para el envío de la información obtenida (TRANSMISIÓN) y la forma en que se analizan dichos datos (ANÁLISIS) [5]. En la figura 2.1 se puede observar el funcionamiento de un sistema de teledetección y la interacción entre sus componentes.

Para poder obtener información a distancia es necesaria la interacción de tres elementos fundamentales: una fuente de energía, un objetivo y un dispositivo o sensor adecuado:

- La **fente de energía** es la que “ilumina” el objetivo emitiendo una onda electromagnética (flujo de fotones). También es posible medir el calor que se desprende de la superficie del objetivo (infrarrojo térmico).
- El **objetivo** es la porción de la superficie terrestre observada por la plataforma.
- El **sensor de teledetección** mide la energía solar (la radiación electromagnética) reflejada por el objetivo. El captador puede encontrarse en un satélite, avión o vehículo aéreo no tripulado.

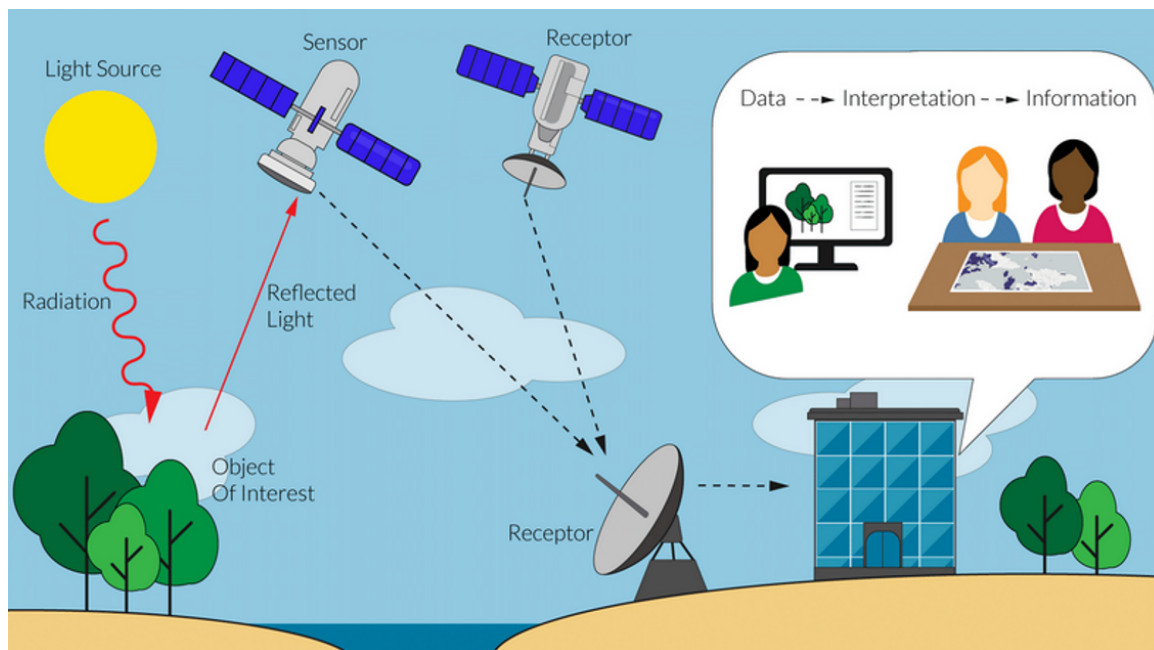


Figura 2.1: Ejemplo de un sistema de teledetección basado en satélites pasivos. Fuente: The Nature Conservancy.

Cuando la fuente de energía es el Sol, y el captador solo mide la radiación reflejada, se conoce como **teledetección pasiva**. Esta modalidad de teledetección solo es operativa durante las horas con luz solar. Por la noche, o cuando hay nubes que se interponen, no es posible realizar observaciones. Si el mismo sensor está provisto de una fuente emisora de energía que la envía hacia el objetivo (radar) y mide el eco producido se denomina **teledetección activa**. El radar atraviesa la cobertura de nubes, lo cual facilita la obtención de imágenes independientemente de las condiciones meteorológicas, tanto de día como de noche.

El **espectro electromagnético** es “el sistema que clasifica, según las longitudes de onda, toda la energía que se mueve, armónicamente, a la velocidad constante de la luz” (NASA, 2013). Dentro del espectro electromagnético se distingue una serie de regiones en función de la longitud de onda: ondas de radio, microondas, infrarrojos, visible –que percibimos como luz visible–, ultravioleta, rayos X y rayos gamma. El comportamiento de las radiaciones electromagnéticas dependerá de su longitud de onda. Cada banda del espectro permite conocer elementos y objetos específicos de la superficie terrestre.

Los sensores pasivos miden la energía en la región óptica del espectro electromagnético, que incluye el visible, infrarrojo cercano, infrarrojo medio SWIR, y el infrarrojo térmico. La figura 2.2 ilustra la clasificación de la energía según su longitud de onda, destacando la parte visible por el ojo humano.

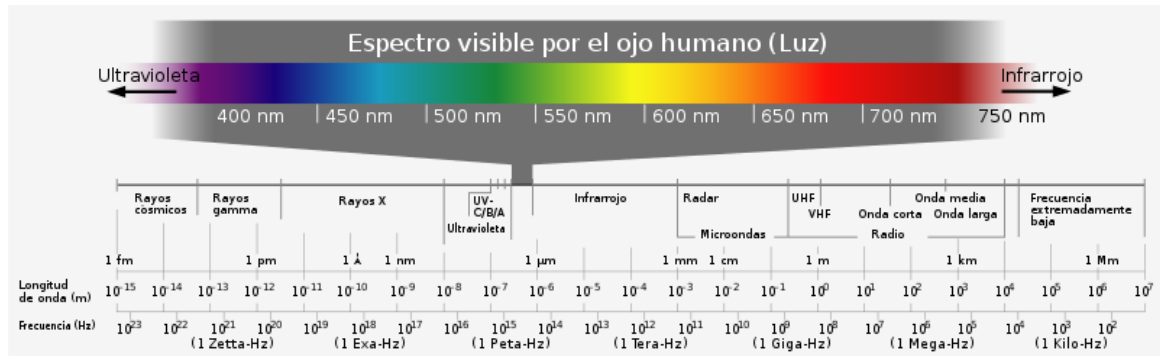


Figura 2.2: Espectro electromagnético. Fuente: Wikipedia.

La atmósfera es un factor importante al considerar los flujos de radiación entre el Sol y la superficie terrestre y entre esta y los satélites, ya que se interpone entre la superficie objeto del estudio y el sensor. Aunque es prácticamente diáfana para algunas longitudes de onda (su existencia no altera la detección de la radiación) en ciertas bandas del espectro, los componentes atmosféricos sí producen efectos que afectan a la señal teledetectada. Los gases y aerosoles que componen la atmósfera tienen un efecto triple sobre la radiación: absorción, dispersión y emisión. Existen regiones del espectro en las que no se produce absorción; son las llamadas **ventanas atmosféricas**, y se conforman como los espacios donde va a ser viable (en principio) la teledetección.

Para la identificación de objetos y procesos en la superficie terrestre se necesita conocer la **reflectividad** de estos objetos respecto a las diferentes longitudes de onda. Cada tipo de superficie interactúa con la radiación de manera diferente, absorbiendo unas longitudes de onda concretas y reflejando otras en unas proporciones determinadas. Es lo que se conoce como **firma espectral**, y para su determinación se utilizan los siguientes métodos:

- Medición directa con un espectro-radiómetro.
- Extracción de la firma desde una biblioteca puesta a punto por alguna organización.
- Simularla mediante modelos físicos.
- Extraerla de una imagen que contenga su firma espectral.

2.1.2. Calidad del agua mediante teledetección

La calidad del agua está influenciada por los siguientes factores [6]:

- Eutrofización o sobrecarga de nutrientes debido a la escorrentía agrícola y a la lixiviación séptica.
- Contaminación por fuentes industriales puntuales o deposiciones provenientes de la atmósfera.

- El cambio climático está provocando un incremento de la temperatura que estimula las floraciones de cianobacterias y una acidificación debido al CO₂ elevado.
- Cambios en la red alimentaria.
- Introducción de nuevas especies.
- Cambios en el flujo del agua ya sea por la construcción de elementos de regulación o por eventos naturales como inundaciones y sequías.

Al analizar simultáneamente el cambio de los componentes y los cambios en las propiedades de la reflectancia del agua, la abundancia o concentración de un componente particular puede estimarse cualitativamente o, a veces, cuantitativamente utilizando la teledetección. La longitud de onda óptima a utilizar depende de la sustancia medida, así como de la concentración de la sustancia medida.

2.1.2.1. Comportamiento espectral del agua

Las características espectrales del agua hacen que ciertas propiedades de la misma sean fácilmente identificables mediante teledetección. El agua como elemento, absorbe o transmite la mayor parte de la radiación que recibe, comportándose de forma distinta en función de los elementos que se encuentren disueltos o en suspensión en ella. Así, el agua clara se comporta de forma distinta al agua turbia.

La mayor reflectividad del agua clara se produce en la región del azul, volviéndose prácticamente nula en el infrarrojo y a mayores longitudes de onda. Cuando el agua presenta turbidez, la respuesta espectral del agua obtenida por el sensor difiere de la del agua clara, viéndose modificada en una u otra región del espectro, en función del tipo de componente que ocasione esa turbidez. En la figura 2.3 se puede revisar de forma gráfica el comportamiento espectral complejo del agua.

2.1.2.2. Interacción entre luz y agua

Como ya se ha visto, la detección remota de las propiedades de una masa de agua se basa en la interacción de la radiación electromagnética con las moléculas de esta masa de agua, ya sean de agua misma, o de la materia diluida o suspendida en ella. Esta interacción, a su vez, depende de la energía de los fotones que componen la radiación electromagnética o, dicho de otro modo, de su longitud de onda, λ . Solamente les suceden dos cosas a los fotones en el agua: o son absorbidos o son dispersados.

- Los fotones son absorbidos cuando son capturados por moléculas de agua o de alguno de los constituyentes del agua.

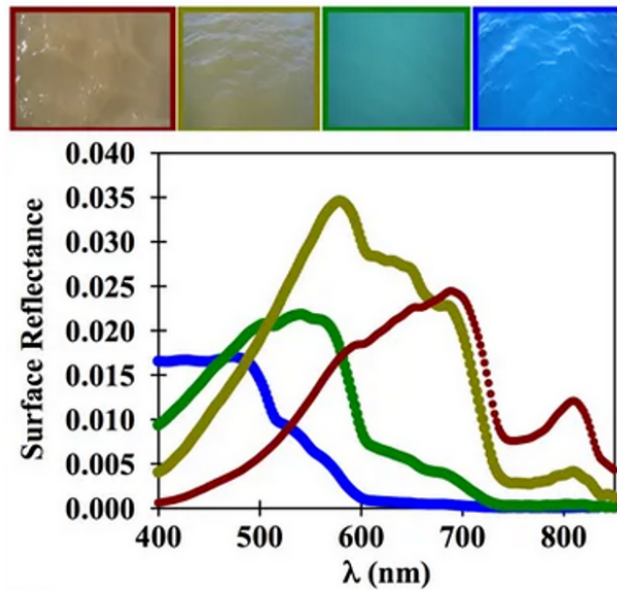


Figura 2.3: Diferencias en los colores y firmas espectrales únicas (reflectancia en función de la longitud de onda de la luz) de las aguas interiores (izquierda) a las oceánicas (derecha).

- Los fotones son dispersados cuando el agua o los componentes en agua cambian su dirección de propagación.

Propiedades ópticas inherentes

Las propiedades ópticas inherentes (IOP) son propiedades conservadoras que caracterizan el comportamiento óptico del medio y **son independientes de la luz** (su origen o intensidad), **sólo dependen de la composición agua o de las sustancias que existen en el medio acuático**.

Estas propiedades son **el coeficiente de absorción $a(\lambda)$ y el coeficiente de dispersión $b(\lambda)$** , y cuantifican la absorción y dispersión de luz de una masa de agua en cualquier longitud de onda especificada. Para una iluminación dada, las IOP determinan por completo la radiación de salida del agua (es decir, las propiedades ópticas aparentes del agua, AOP).

Propiedades ópticas aparentes

Las propiedades ópticas aparentes son aquellas propiedades que **dependen tanto del medio acuático como de la estructura direccional del campo de luz ambiental**, y que muestran suficientes características regulares y estabilidad para ser descriptores útiles de un cuerpo de agua. Las propiedades aparentes empleadas más comúnmente son [7]:

- Radiancia que sale del agua (L_w)
- Radiancia ascendente (L_u)
- Radiancia descendente (E_d)
- Reflectancia de teledetección (R_{rs})
- Coeficiente de atenuación difusa vertical

Las propiedades ópticas son el vínculo entre el sistema acuático y la información que puede obtenerse por teledetección. La composición del agua, a través de los **Componentes Ópticamente Activos**, da forma a la reflectancia de teledetección $R_{rs}(\lambda)$.

Componentes ópticamente activos del agua

Como en la gran mayoría de los ecosistemas terrestres, la luz solar constituye la entrada de energía que mantiene en funcionamiento los ecosistemas acuáticos. **El color del agua es el resultado de la interacción entre la luz y los Componentes Ópticamente Activos** que se encuentran en el agua, por lo que **la calidad del agua afecta las propiedades ópticas de ésta**. Los dos principales constituyentes que influyen en la calidad del agua son:

- **Materia orgánica disuelta coloreada (CDOM)**, la cual es una mezcla de compuestos orgánicos.
- **Materia suspendida** de origen orgánico o inorgánico, viva o muerta que incluye partículas de arcilla, minerales sin disolver, plancton y floraciones algales, entre otros.

Como ya se ha visto, los **OAC** modifican el flujo incidente de radiación solar mediante dos procesos básicos de absorción y dispersión [8]. **La reflectancia de una masa de agua depende de la concentración de sus OAC. Por tanto, si conocemos la reflectancia podemos llegar a conocer la concentración de los OAC**. Esta es la base del estudio de las aguas continentales por teledetección.

Del análisis de los espectros de los diferentes elementos que hay en una masa de agua, se llega a las siguientes conclusiones:

- La reflectancia de una masa de agua a longitudes de onda por encima de 1000 nm es muy pequeña ($R = 0$) debido a que el agua absorbe toda la energía que le llega.
- Entre 700 nm y 1000 nm sólo existe un pico de reflectancia en torno a 805 nm que identifica la cantidad de sólidos en suspensión que existen en esa masa de agua.
- Entre 400 nm y 700 nm, la reflectancia nos dará información de la existencia de todos los pigmentos fotosintéticos que están presentes y, en consecuencia, del tipo de micro algas en el agua.

- La reflectancia de una masa de agua a longitudes de onda por debajo de 400 nm depende de la cantidad de materia amarilla que haya en ella; normalmente esta respuesta también será muy pequeña.
- La capacidad de absorción selectiva de luz por parte del fitoplancton en la región visible del espectro electromagnético (400 nm – 700 nm), una región en la que las moléculas de agua apenas absorben luz, hace al fitoplancton responsable en gran medida del color del agua.
- El CDOM y los detritos absorben luz principalmente en el ultravioleta y en el azul, mientras que en las partículas minerales predomina la dispersión, que depende de la concentración y el tamaño de las partículas.
- El fitoplancton también dispersa luz en función del tamaño y forma de sus células, así como de la presencia de estructuras rígidas de naturaleza calcárea o silíceas. No obstante, su respuesta espectral depende, fundamentalmente, de la absorción debida a sus pigmentos fotosintéticos.

La figura 2.4 resume de forma gráfica las firmas espectrales del agua con diferentes componentes ópticamente activos.

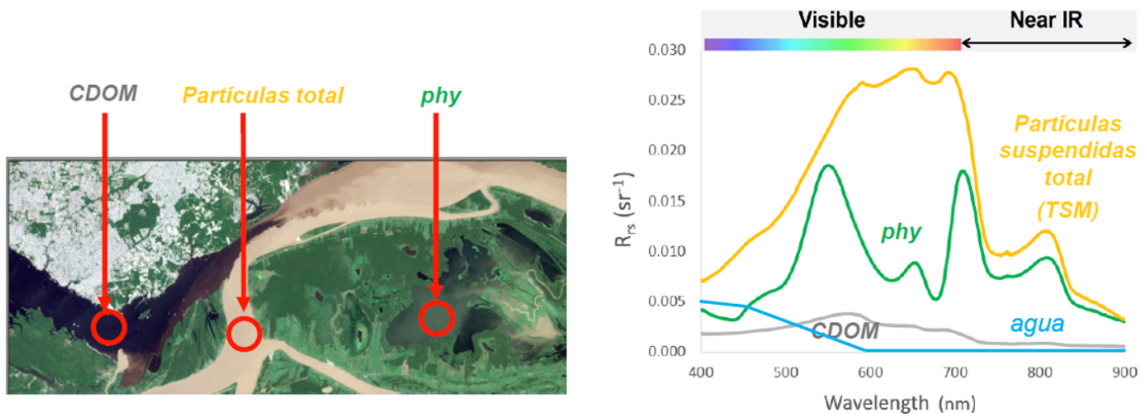


Figura 2.4: Firmas espectrales de aguas pura (azul) y con presencia mayoritaria de CDOM (gris), fitoplancton (verde) y sólidos en suspensión (amarillo). Fuente: NASA-ARSET.

2.2. Proyectos de referencia

2.2.1. Proyectos destacados a nivel nacional

En el panorama nacional destacan los trabajos realizados por diferentes grupos de investigación de la Universidad de Valencia.

En primer lugar, se sitúa el proyecto [ESAQS](#) [9]. La Generalitat Valenciana, mediante el programa PROMETEO para Grupos de Investigación de Excelencia, concedió a la Universidad de Valencia (2016/032) la subvención para la ejecución del **proyecto ESAQS (Seguimiento del estado ecológico de aguas continentales y costeras con los nuevos satélites de Observación de la Tierra)**. En el marco del proyecto [ESAQS](#) se realizaron diferentes publicaciones y apariciones en congresos presentando los resultados de la investigación cuyo objetivo principal era desarrollar y validar algoritmos para obtener indicadores ecológicos en aguas continentales y costeras con la nueva generación de satélites europeos de observación de la Tierra: Sentinel-2 y Sentinel-3.

El equipo de trabajo estaba formado por los investigadores del Laboratorio de Observación de la Tierra (Grupo LEO), del Instituto Cavanilles de Biodiversidad y Biología Evolutiva y del Departamento de Geografía de la universidad. El proyecto recopiló datos in situ durante una campaña de tres años en un conjunto diverso de masas de agua en la región de Valencia que cubren una amplia gama de estados tróficos. Estas campañas de campo medían diversos parámetros bio-físicos y espectro-radiometría, coincidiendo con el paso de la misión Sentinel-2 por la zona de estudio. Las masas de agua (8 embalses y 1 laguna costera) que se estudiaron son: Contreras, Beniarrés, Benagéber, Sitjar, María Cristina, Bellús, Regajo y Tous.

Las publicaciones que resumen los resultados y los productos generados del proyecto son las siguientes:

- Pereira-Sandoval et al. (2019) Evaluation of atmospheric correction algorithms over spanish inland waters for Sentinel-2 multispectral imagery data. Remote Sensing. <https://doi.org/10.3390/rs11121469> [10].
- Pereira-Sandoval et al. (2019) Calibration and validation of algorithms for the estimation of chlorophyll-a concentration and Secchi depth in inland waters with Sentinel-2. Limnetica. <http://www.limnetica.net/documentos/limnetica/limnetica-38-1-p-471.pdf> [11].
- Soria-Perpinyá et al. (2019) Monitoring the ecological state of a hypertrophic lake (Albufera of Valencia, Spain) using multitemporal Sentinel-2 images. Limnetica. <http://www.limnetica.net/documentos/limnetica/limnetica-38-1-p-457.pdf> [12].
- 14 – Vida submarina. Soria-Perpinyá et al. (2019) Remote sensing applications for the study of rapid flushing to remediate eutrophication in shallow lagoons (Albufera of Valencia) using multitemporal Sentinel-2 images. Hydrobiología. <https://link.springer.com/article/10.1007%2Fs10750-018-3741-6> [13].
- Soria-Perpinyá et al. (2019) Remote sensing of cyanobacterial blooms in a hypertrophic lagoon (Albufera of Valencia) using multitemporal Sentinel-2 images. Science of the Total

Environment. <https://www.sciencedirect.com/science/article/pii/S0048969719342949> [14].

Todos estos trabajos se desarrollaron en la Confederación Hidrográfica del Júcar. El área de estudio de este trabajo se sitúa dentro de la demarcación de la CHE. Por eso, en segundo lugar, se revisan los proyectos a cabo en esta demarcación.

Desde la CHE también han colaborado con la Universidad de Valencia para hacer seguimiento de algunos de sus embalses mediante teledetección. El proyecto más reciente, entregado en septiembre de 2021 que lleva por título “Seguimiento de embalses por teledetección 2020” [15], estudia los parámetros de concentración de clorofila-a, sólidos en suspensión, transparencia, turbidez, concentración de ficocianina y CDOM en 17 embalses de la cuenca.

Los resultados de este proyecto son las ecuaciones calibradas relacionando las bandas multiespectrales del satélite Sentinel-2 para estimar los diferentes parámetros de calidad de agua y mapas de calidad de cada embalse estudiado. El estudio se encuentra disponible desde el portal web de la CHE: <https://portal.chebro.es/ca/otros-estudios>.

2.2.2. Proyectos destacados a nivel internacional

A nivel internacional existen numerosas iniciativas de investigación, de entidades públicas o privadas que se centran en el estudio de la calidad de agua mediante teledetección.

Para desarrollar este trabajo, se ha revisado el proyecto de calidad de agua llevado a cabo por Copernicus, que es el programa de observación de la Tierra de la Unión Europea y los seminarios de formación del centro de Entrenamiento de Detección Remota Aplicada de la NASA (ARSET) [16] disponibles en <https://appliedsciences.nasa.gov/join-mission/training>.

Dentro del proyecto “Copernicus Global Land Services” disponible en la página web <https://land.copernicus.eu/global/> se pueden encontrar los productos “Lake Surface Water Temperature” y “Lake Water Quality”. El producto de temperatura del agua describe la temperatura de la superficie empleando datos de la misión Sentinel-3 a una resolución espacial de 1 kilómetro. El producto de calidad del agua consta de tres parámetros: turbidez, índice del estado trófico y las reflectancias de la superficie empleando para ello imágenes Sentinel-2 y generando un resultado de resolución espacial de 100 metros (el más pequeño).

Al generar resultado de resolución espacial baja, quedan excluidas la mayor parte de masas de agua del territorio nacional.

2.3. Planteamiento del trabajo a partir del estado del arte

El enfoque y la metodología planteada en el presente trabajo está basada en el aprendizaje y experiencia adquirida a partir de los proyectos llevados a cabo en territorio nacional, pero en otros embalses. Las principales diferencias que presenta este trabajo con los proyectos anteriormente revisados son:

1. En la bibliografía revisada para obtener los datos in situ se han planificado y ejecutado campañas ad-hoc coincidentes con el paso del satélite para muestrear y obtener un dato concreto en una fecha concreta. En el presente trabajo se trabajará con un sensor instalado que genera informa continua las 24 horas del día los 365 días del año, lo que nos lleva a la dimensión del Big Data y el IoT.
2. En ninguno de los proyectos revisados se han probado algoritmos de aprendizaje automático para llevar a cabo la calibración entre los datos de campo y las imágenes satelitales.

2.4. Decisiones tecnológicas

Para llevar a cabo este trabajo se emplearán herramientas que se han dominado durante el máster. En primer lugar, se ha decidido que el lenguaje de programación con el que realizar el análisis sea Python ya que es el lenguaje predominante en el sector geoespacial y para la interacción con imágenes de Google Earth Engine se emplearán librerías desarrolladas exclusivamente en este lenguaje. Por lo tanto, la captura, limpieza, tratamiento y procesado de datos, así como los modelos se desarrollarán en Python.

Se trabajará sobre la distribución Anaconda con los entornos Spyder y Jupyter para la ejecución de los scripts. De esta forma, sobre el formato Notebook de Jupyter queda bien documentado el código y los resultados obtenidos en cada paso.

El formato de los resultados está orientado a poder consultarse en un software SIG (QGIS, por ejemplo) donde podrán ser integrados con otras capas de análisis y así realizar una gestión integral de los recursos hídricos del área de interés.

Capítulo 3

Diseño e implementación del trabajo

3.1. Flujo de trabajo para un programa de seguimiento de la calidad del agua mediante teledetección

En el siguiente diagrama 3.1, se puede observar la metodología clásica para un programa de monitoreo de los parámetros indicativos de la calidad de agua mediante técnicas de teledetección:

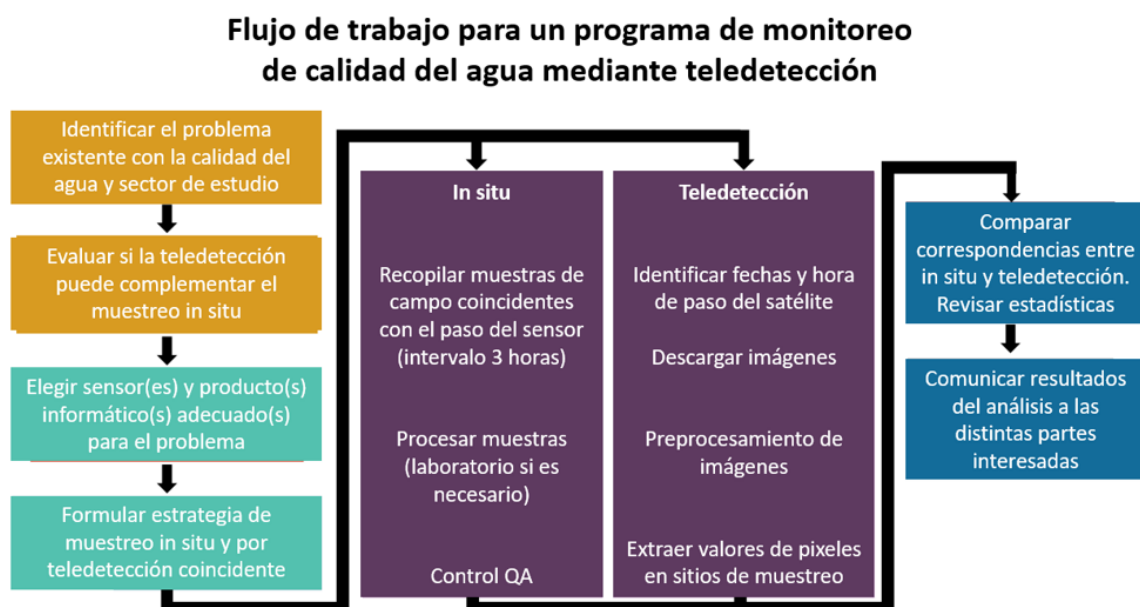


Figura 3.1: Propuesta metodológica para el monitoreo de la calidad del agua mediante teledetección.

3.2. Extracción y limpieza de los datos

3.2.1. Datos de la sonda instalada en El Val

En el embalse El Val, la [Confederación Hidrográfica del Ebro](#) ha considerado conveniente el refuerzo de las medidas habituales de seguimiento de la calidad y control de vertidos con la instalación de una sonda automática Aquadam. El sistema Aquadam está compuesto por una sonda multiparamétrica autoposicionable que mide de forma automática, autónoma y a diferentes profundidades los valores de parámetros representativos de la calidad del agua. La sonda multiparamétrica mide la temperatura, el pH, la conductividad, el redox, el oxígeno disuelto, la turbidez y la clorofila. La frecuencia de realización de los perfiles es programable, pero se suelen realizar 4 veces al día: 00:00h, 06:00h, 12:00h y 18:00h en horario GMT. La información completa generada por la sonda se puede consultar en el siguiente enlace: <https://saica.chebro.es/ElValPerfiles.php>



Figura 3.2: Udicación en planta de la sonda (izquierda) y foto del equipo en detalle (derecha).

Para descargar los datos de la pagina web, se ha empleado la técnica de **web scraping** mediante las librerías **Selenium** y **BeautifulSoup**. El uso combinado de estas librerías y la construcción de la url dinámica en función de la fecha viene ilustrada en la siguiente figura 3.3.

Los datos de la sonda están disponibles desde el día 20 de enero de 2018 hasta la actualidad y la descarga de la información se hace mediante una url dinámica en la que se va modificando el año y mes de consulta para así obtener datos del mes completo de una sola vez. Cada vez que se ejecuta el script para descargar los datos de la web, por defecto se descarga la información desde el inicio de la serie hasta la fecha actual, almacenando los datos mensuales en un archivo csv que tiene la siguiente estructura:

```

data_dirname = os.path.join(os.getcwd(), 'input_val')
# Datos disponibles desde 20-ene-18
data_ini = datetime.datetime.strptime('2018-01-20', '%Y-%m-%d').date()

today = datetime.date.today()
actual_year = today.strftime("%Y")
actual_month = today.strftime("%m")

options = webdriver.ChromeOptions()
driver = webdriver.Chrome(ChromeDriverManager().install(), chrome_options = options)
#wait = WebDriverWait(driver, 1)

for year in range(int(data_ini.strftime("%Y")), int(today.strftime("%Y"))+1):

    for month in range(int(data_ini.strftime("%m")),13):

        if month <= int(today.strftime("%m")):

            url = 'https://saica.chebro.es/ElValExport.php?anyo=' + str(year) + '&mes=' + str(month).zfill(2)
            driver.get(url)
            html = driver.page_source

            soup = BeautifulSoup(html)
            table = soup.find_all('table')

            df = pd.read_html(str(table))[0]

            filename = 'val' + str(year) + str(month).zfill(2) + '.csv'
            outpath = os.path.join(data_dirname, filename)
            df.to_csv(outpath)

```

Figura 3.3: Descarga de los datos públicos desde la web de la CHE mediante web-scraping con Selenium y BeautifulSoup.

- Fecha
- Hora
- Cota(msnm)
- Profundidad(m)
- Temp. agua($^{\circ}$ C)
- pH
- Pot. redox(mV)
- Conduct. (μ S/cm)
- Ox.dis.(mg/L)
- Turbidez(NTU)
- Clorofila(μ g/L)

De los parámetros físico-químicos obtenidos, los que interesan para el trabajo son temperatura del agua, turbidez y clorofila, puesto que son los únicos térmica u ópticamente activos que pueden ser capturados por el sensor remoto.

Los atributos fecha, hora y profundidad se emplean para filtrar los registros que son útiles para el estudio. A continuación se detalla el proceso.

- Como mediante el sensor remoto del satélite solamente se pueden reconocer valores de

la superficie, solo se han tenido en cuenta los **valores capturados por la sonda en el primer metro y medio de la columna de agua** (la sonda mide hasta los 30-40 metros de profundidad).

- Anteriormente se ha visto que la sonda realiza perfiles cada 6 horas. Para que el dato que se emplea sea coherente, **se han filtrado los perfiles por las horas coincidentes con la toma de la imagen satelital para esa zona**, así sólo se han empleado datos tomados entre las 12 y las 15 horas para asegurar la simultaneidad de los datos.

El código del proceso de limpieza y filtrado de datos originales que se ha explicado viene ilustrado en la siguiente figura 3.4.

```
def read_and_filter_file(datafile):
    # Leer archivo de datos formato csv
    df = pd.read_csv(datafile, header = 0)
    # Lista de columnas que no se necesitan para el análisis
    cols_to_delete = ['Fecha', 'Hora', 'Cota(msnm)', 'pH', 'Pot. redox(mV)', 'Conduct.(µS/cm)', 'Ox.dis.(mg/L)']
    # Transformar los cmapos de fecha y hora a datetime y obtenerlos de ahí
    df['Date'] = pd.to_datetime(df['Fecha'], format='%Y-%m-%d').dt.date
    df['Time'] = pd.to_datetime(df['Hora'], format='%H:%M').dt.time
    # Eliminar las columnas innecesarias
    df.drop(cols_to_delete, axis=1, inplace=True)
    # Renombrar columnas
    df.rename(columns={'Profundidad(m)': 'Profundidad', 'Temp. agua(°C)': 'Ta_agua',
                      'Turbidez(NTU)': 'Turbidez', 'Clorofila(µg/L)': 'Clorofila'}, inplace=True)
    # Rango de horas de los perfiles de la sonda cercanas al paso del satélite
    start_time = datetime.datetime.strptime('12:00:00', '%H:%M:%S').time()
    end_time = datetime.datetime.strptime('15:00:00', '%H:%M:%S').time()
    # Filtrar los datos de cmapo de la superficie y que estén en el rango de horas establecido
    df_superficie = df[(df['Profundidad'] < 1.5) & (df['Time'].between(start_time, end_time))]
    # Reiniciar el index
    df_superficie.reset_index(drop=True, inplace=True)
    # Borrar duplicados del mismo día. Solo nos quedamos con una toma diaria, la primera (la más cercana a la superficie)
    df_superficie = df_superficie.drop_duplicates(subset=['Date'], keep='first')
    # Eliminar la columna de la hora ya que no se necesita más
    df_superficie.drop(['Profundidad', 'Time'], axis=1, inplace=True)
    return df_superficie
```

Figura 3.4: Función que ejecuta todas las tareas de limpieza y filtrado de datos del csv descargado original.

Con estas transformaciones se ha obtenido un archivo csv un dato diario de calidad de los diferentes parámetros (fecha, temperatura agua, turbidez y clorofila-a) tomado a la hora aproximada de paso del satélite por la zona desde enero de 2018 hasta la actualidad.

3.2.2. Imágenes satelitales

En el presente trabajo se han empleado dos misiones reconocidas de observación terrestre mediante imágenes satelitales. Se han utilizado datos del sensor TIRS a bordo de la misiones Landsat (8 y 9) para las imágenes de temperatura y los datos del sensor MSI a bordo de Sentinel (2) para las imágenes ópticas.

La interacción y análisis de las imágenes satelitales se ha hecho a través de la librería **geemap**, que es un paquete de Python para mapas interactivos con **Google Earth Engine (GEE)**. **GEE** es una plataforma informática que permite a los usuarios ejecutar análisis geoespaciales en la infraestructura de Google en la nube con un catálogo de varios petabytes de imágenes satelitales y conjuntos de datos geoespaciales. **GEE** proporciona una API de JavaScript y Python para realizar solicitudes a los servidores de Earth Engine. En comparación con la documentación completa y el IDE interactivo (es decir, el editor de código de JavaScript de **GEE**) de la API de JavaScript de **GEE**, la API de Python de **GEE** tiene relativamente poca documentación y una funcionalidad limitada para visualizar los resultados de forma interactiva. El paquete **geemap** fue creado para llenar este vacío. Se basa en **ipyleaflet** e **ipywidgets**, y permite a los usuarios analizar y visualizar conjuntos de datos de Earth Engine de forma interactiva dentro de un entorno basado en Jupyter.

En ambos casos (Landsat y Sentinel) se ha buscado la información sobre el momento exacto de la adquisición de la imagen en los metadatos de ésta, y esa hora es la que se ha empleado en el filtro de los datos de campo. En la figura 3.5 se puede observar cómo viene dada esa información por los distribuidores.

Start Time	2022-12-02 10:49:08	Sensing start: 2022-12-04T11:03:29.024Z
Stop Time	2022-12-02 10:49:40	Sensing stop: 2022-12-04T11:03:29.024Z

Figura 3.5: Metadatos Landsat (izquierda) y Sentinel (derecha) con las horas exactas de toma de la imagen para la zona de estudio.

Las imágenes siempre se toman a la misma hora aproximadamente sobre la zona de estudio, que como se puede ver es sobre las 11:00 GMT en ambos casos.

Los criterios para el filtrado de imágenes han sido la delimitación geográfica y el porcentaje de cobertura nubosa que se ha limitado a 30 %. En cuanto a la distribución y extensión de las imágenes Landsat, éstas vienen dadas por la combinación de un número de ruta (path) y un número de fila (row) que identifica de forma única un centro de escena nominal. Para el caso del estudio del embalse El Val, la escena Landsat que se necesita es la **200-031** que cubre por completo el área de interés. En cuanto a los productos Sentinel, las escenas vienen divididas en gránulos, también llamados tiles. Para el caso del presente estudio solo es necesario un tile: **30TWM**.

En la figura adjunta 3.6 puede analizarse el código empleado para obtener el listado de imágenes válidas para la zona de estudio.

De todo el catálogo de productos disponibles en **GEE**, para las imágenes **Landsat** se ha

```

# Year-Month-Day Extract function
def ymdList(imgcol):
    def iter_func(image, newlist):
        date = ee.Number.parse(image.date()).format("YYYYMMdd")
        newlist = ee.List(newlist)
        return ee.List(newlist.add(date).sort())

    return imgcol.iterate(iter_func, ee.List([]))

def query_dates(date_list):

    res_list = []
    for fecha in date_list:
        string = str(fecha)
        year = string[:4]
        month = string[4:-2]
        day = string[-2:]
        date_query = year + '-' + month + '-' + day
        res_list.append(date_query)

    return res_list

geometry = ee.FeatureCollection("users/vnoronoz/punto_evaluacion_sonda") #30TWN
start = ee.Date('2018-01-21')
finish = ee.Date('2022-12-13')

#----- Sentinel 2 Collection -----#
sentinel2 = ee.ImageCollection('COPERNICUS/S2_SR_HARMONIZED') \
    .filterBounds(geometry) \
    .filterDate(start, finish) \
    .filterMetadata('CLOUDY_PIXEL_PERCENTAGE', 'less_than', 30) \
    .filterMetadata('MGRS_TILE', 'equals', '30TWM')

ymd = ymdList(sentinel2)
fechas_s2 = ymd.getInfo()
query_list_s2 = query_dates(fechas_s2)

```

Figura 3.6: Código para obtención de la lista de fechas con las imágenes Sentinel-2 disponibles en la zona de estudio en el intervalo temporal deseado mediante la interacción de los módulos geopmap y el catálogo de Earth Engine.

utilizado el producto **Collection 2** que ofrece valores de **reflectancia superficial corregida atmosféricamente y la temperatura de la superficie terrestre** derivada de los datos producidos por los sensores Landsat 8 OLI/TIRS respectivamente.

Estas imágenes contienen 5 bandas visibles e infrarrojas cercanas (VNIR) y 2 bandas infrarrojas de onda corta (SWIR) procesadas a reflectancia superficial ortorrectificada, y **una banda infrarroja térmica (TIR) procesada a temperatura superficial ortorrectificada**, que es la empleada en este trabajo.

La temperatura de la superficie de Landsat Collection 2 tiene un factor de escala de 0.00341802 y una compensación adicional de 149 por píxel. Estos coeficientes de corrección

son dados por el fabricante y se han aplicado a los valores de la banda térmica de cada imagen analizada, así como también se ha hecho la transformación a grados Celsius, ya que las unidades derivadas de la banda térmica son grados Kelvin.

Transformaciones aplicadas a la Banda 10 térmica (Kelvin) de Landsat 8 y 9

- Scale: 0.00341802
- Offset: 149
- Conversión Kelvin a Celsius: -273.15

La figura 3.7 muestra el extracto de código empleado para el cálculo de la temperatura superficial en grados celsius aplicando los coeficientes indicados.

```
# Para todas las fechas de la lista extraer el valor del pixel
for f in query_list_l8:
    try:
        doi = ee.Date(f)

        image_coll = ee.ImageCollection('LANDSAT/LC08/C02/T1_L2') \
            .filter(ee.Filter.eq('WRS_PATH', 200)) \
            .filter(ee.Filter.eq('WRS_ROW', 31)) \
            .filterDate(doi, doi.advance(1, 'day'))

        # convertir el objeto coleccion a objeto imagen
        image_doi = image_coll.mosaic()

        # Seleccionar la banda termica
        thermal_band = image_doi.select('ST_B10')

        #Convert the brightness temperature to Celsius
        tempK1 = thermal_band.multiply(0.00341802)
        tempK = tempK1.add(149)
        tempC = tempK.subtract(273.15)

        out_file_Ta = 'temp' +str(f) + '.csv'
        out_path_Ta = os.path.join(results_path_Ta, out_file_Ta)

        geemap.extract_values_to_points(geometry, tempC, out_path_Ta, scale=30, crs='EPSG:4326')

    except:
        pass
```

Figura 3.7: Código para obtención de los valores de temperatura superficial en el punto de evaluación.

En el caso de las imágenes Sentinel disponibles en [GEE](#), se ha empleado el producto de **reflectancia superficial corregida atmosféricamente ortorrectificada de nivel 2A Harmonized Sentinel-2 MSI**.

Para cada imagen se han calculado dos índices robustos ampliamente estudiados y validados por la comunidad científica nacional e internacional.

- **Índice de Turbidez de Diferencia Normalizada (NDTI)**: se basa en el fenómeno de que la reflectancia electromagnética es mayor en el espectro verde que en el espectro rojo para el agua clara. Por lo tanto, con el aumento de la turbidez también aumenta la reflectancia del espectro rojo. En el estudio actual, se utilizan las bandas verde (banda 3) y roja (banda 4) de Sentinel-2.

$$NDTI = (Redband - Greenband)/(Redband + Greenband) \quad (3.1)$$

$$NDTI = (B4 - B3)/(B4 + B3) \quad (3.2)$$

- **Índice de Clorofila de Diferencia Normalizada (NDCI)**: utiliza la información del pico de reflectancia centrado en 700 nm, que es máximamente sensible a las variaciones en la concentración de clorofila-a en el agua. Además, un pico de absorción espectral amplio entre 665 nm y 675 nm se asocia generalmente a la absorción por los pigmentos de clorofila-a. La normalización a través del **NDCI** elimina las incertidumbres en la estimación de la reflectancia de detección remota, las diferencias estacionales de azimut solar y las contribuciones atmosféricas. En el estudio actual, se utilizan las bandas roja (banda 4) y de borde rojo (banda 5) de Sentinel-2.

$$NDCI = (Rededgeband - Redband)/(Rededgeband + Redband) \quad (3.3)$$

$$NDCI = (B5 - B4)/(B5 + B4) \quad (3.4)$$

El valor derivado de la reflectancia del pixel de evaluación en cada imagen para estos dos índices es el que se relacionará con los datos obtenidos in situ mediante la sonda multiparamétrica. La figura 3.8 muestra el extracto de código empleado para el cálculo del índice **NDCI** para cada imagen válida.

3.3. Ajuste y evaluación de algoritmos

En este trabajo se han tenido que tomar una serie de decisiones a priori, antes de empezar la implementación para que el desarrollo total del proyecto pudiese ajustarse a los tiempos disponibles y a los objetivos y alcance establecidos porque las opciones de ajuste y evaluación serían infinitas. Estas decisiones han acotado los experimentos del proyecto de tal forma que se llegue a definir una metodología y a implementarla de principio a fin, pudiendo realizar las pruebas

```
for f in query_list_s2:
    try:
        doi = ee.Date(f)
        image_coll = ee.ImageCollection('COPERNICUS/S2_SR_HARMONIZED') \
            .filterBounds(geometry) \
            .filterDate(doi, doi.advance(1, 'day')) \
            .filterMetadata('MGRS_TILE', 'equals', '30TWM')

        # convertir el objeto coleccion a objeto imagen
        image = image_coll.mosaic()
        ndci = image.normalizedDifference(['B5', 'B4']).rename('NDCI')
        band_C = ndci.select('NDCI')

        out_file_C = 'chla' +str(f) + '.csv'
        out_path_C = os.path.join(results_path_C, out_file_C)

        geemap.extract_values_to_points(geometry, band_C, out_path_C, scale=30, crs='EPSG:4326')

    except:
        pass
```

Figura 3.8: Código para obtención de los valores del índice NDCI en el punto de evaluación.

deseadas y llegando a medir resultados. Una de las decisiones que han delimitado el trabajo ha sido la expuesta anteriormente de emplear los índices normalizados para la calibración de las imágenes satelitales, dejando afuera otras múltiples opciones. Otra decisión preestablecida ha sido el acotar los algoritmos de aprendizaje automático que se van a comparar y cómo se van a aplicar.

De toda la selección de algoritmos posibles a evaluar se ha decidido testear los siguientes: **Decision Tree, Random Forest y Support Vector Machines (SVM)** por su facilidad de aplicación, versatilidad y resultados robustos, con lo que se cumple con el alcance del proyecto. Se ha decidido emplear estos tres algoritmos porque han sido bastante trabajados durante todo el máster, por lo que se sabe manipular sus parámetros, y en los proyectos vistos con anterioridad han ofrecido buenos resultados en predicciones numéricas. Para llevar a cabo la evaluación de algoritmos se ha empleado la librería **scikit-learn** de Python.

En todos los casos se ha dividido el conjunto de datos en **entrenamiento (80%) y test (20%)**. Para decidir cuáles son los hiperparámetros óptimos en cada algoritmo se ha utilizado la técnica de búsqueda conocida como **grid search**, con la que se entrena un modelo para cada combinación de hiperparámetros posible y también se evalúa utilizando **validación cruzada (cross validation) con 5 particiones estratificadas**. Posteriormente, se ha elegido la combinación de hiperparámetros que mejores resultados haya obtenido y es la que se ha aplicado en el trabajo para calcular los errores.

En las líneas de trabajo futuras se incluirá la evaluación de otros algoritmos diferentes a éstos y la posibilidad de emplear técnicas de ensemble learning, estrategias en las que se utiliza un grupo de modelos para resolver un problema mediante la combinación estratégica de diversos modelos de aprendizaje automático en un solo modelo predictivo.

3.3.1. Temperatura del agua

Como se ha visto anteriormente, el producto de Landsat analizado ya ofrece directamente datos de temperatura. En las siguientes tablas 3.9, se pueden analizar las estadísticas básicas de los datos insitu vs los datos de las imágenes. El grado de correlación lineal entre ambos parámetros es del 98.7%.

	count	mean	std	min	25%	50%	75%	max
ta_value	55.0	19.48	8.48	7.07	9.97	21.18	27.05	31.23
tagua_insitu	55.0	17.45	6.70	7.50	10.05	19.50	23.70	26.20

	ta_value	tagua_insitu
ta_value	1.000000	0.987019
tagua_insitu	0.987019	1.000000

Figura 3.9: Estadísticas básicas (izquierda) y coeficiente de correlación de Pearson (derecha).

En la figura 3.10, en el gráfico de la izquierda podemos ver un diagrama de cajas representando las dos variables: datos observados en campo y estimados por satélite. Si bien se observa que la distribución desde el mínimo hasta la mediana es similar, se ve que el rango desde la mediana hasta el valor máximo difiere notablemente. Los valores estimados por satélite son más altos que los observados, sobreestimando las lecturas.

Ésto puede observarse perfectamente en el gráfico de la derecha donde en el eje de abcisas se representa la temperatura estimada del agua y en el eje de ordenadas la diferencia con el valor observado real. Así podemos ver que hasta valores estimados de 20° se tienen diferencias menores a 3° mientras que para valores superiores a 20° las diferencias aumentan.

En el caso del parámetro temperatura de agua superficial se puede afirmar que la relación que existe entre los datos observados de campo y los datos estimados de la imagen satelital se ajustan a una regresión lineal. Por esta razón en el caso de la temperatura del agua se ha decidido no evaluar otros algoritmos de aprendizaje automático, ya que la regresión lineal ofrece unos resultados aceptables y claramente suficientes para el tipo de relación que existe entre los datos captados por el satélite y los medidos in situ. La figura 3.11 ilustra la correlación lineal entre las variables mediante un gráfico de dispersión.

Los resultados obtenidos mediante el modelo de regresión lineal son los siguientes:

- Coeficiente x: 0.78002794

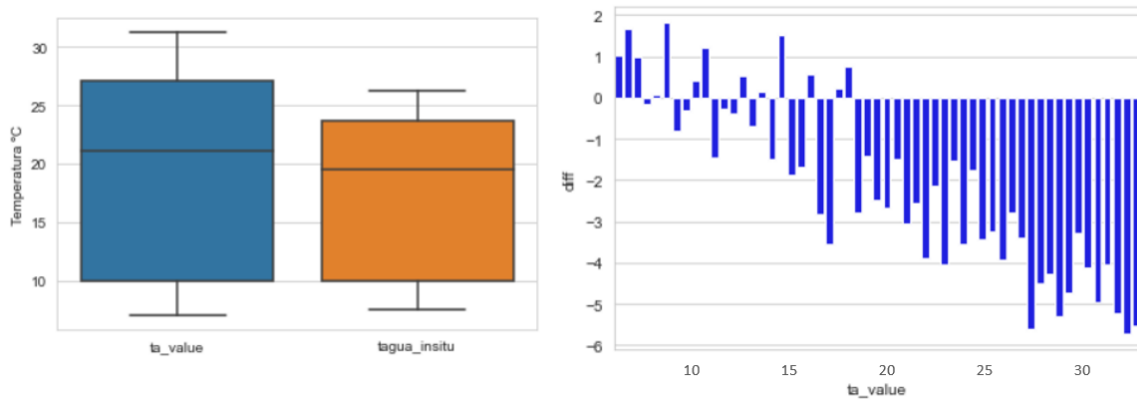


Figura 3.10: Diagrama de cajas (izquierda) y diferencia entre los datos observados vs estimados en función de la temperatura del agua (derecha).

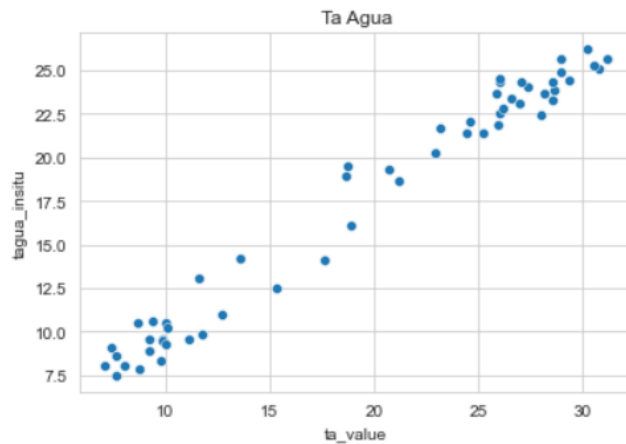


Figura 3.11: Gráfico de dispersión con los datos estimados (eje X) vs datos observados (eje Y).

- Término independiente: 2.1674049
- MAE train: 0.862, MAE test: 0.912
- RMSE train: 1.065, RMSE test: 1.088
- Variancia explicada por el modelo: train 97.4 %, test: 97.8 %

Los errores vienen dados en grados centígrados ($^{\circ}$), que es la unidad de medida de la temperatura del agua. La figura 3.12 muestra el extracto de código empleado para el ajuste de los datos.

```

def make_regression_model(df, colx, coly):

    X = colx.to_numpy()
    y = coly.to_numpy()

    X = X.reshape(-1,1)
    y = y.reshape(-1,1)

    #particion en conjuntos de entrenamiento y test (80/20)
    X_train, X_test, y_train, y_test = train_test_split(X, y, random_state=4, test_size=0.2)

    print('MODELO SKLEARN...')
    regressor = LinearRegression()
    regressor.fit(X_train, y_train)
    print('coeficiente x, termino independiente')
    print(regressor.coef_, regressor.intercept_)
    y_pred_train = regressor.predict(X_train)
    y_pred_test = regressor.predict(X_test)
    print('MSE train: %.3f, test: %.3f' % (mean_squared_error(y_train, y_pred_train),
                                         mean_squared_error(y_test, y_pred_test)))
    print('RMSE train: %.3f, test: %.3f' % (np.sqrt(mean_squared_error(y_train, y_pred_train)),
                                           np.sqrt(mean_squared_error(y_test, y_pred_test))))
    print('MAE train: %.3f, test: %.3f' % (mean_absolute_error(y_train, y_pred_train),
                                           mean_absolute_error(y_test, y_pred_test)))
    print('Exp variance train: %.3f, test: %.3f' % (explained_variance_score(y_train, y_pred_train),
                                                  explained_variance_score(y_test, y_pred_test)))

make_regression_model(df_TAGua, df_TAGua['ta_value'], df_TAGua['tagua_insitu'])

```

Figura 3.12: Código para crear modelo regresión lineal con datos de temperatura.

3.3.2. Concentración de Clorofila-A

El caso de clorofila y turbidez es diferente al de la temperatura ya que el sensor remoto no mide estos parámetros directamente. Por eso, y como se ha explicado anteriormente, se han calculado índices entre bandas para relacionar los valores de reflectancia con la concentración de los componentes. En estos casos, como la relación de los índices con la clorofila y la turbidez no es directa, se han probado los diferentes algoritmos: Decision Tree, Random Forest y SVM.

La figura 3.13 ilustra la correlación entre el índice de clorofila NDCI y los valores observados en campo, así como la correlación que existe entre ambas variables, que es del 76 %.

Los resultados obtenidos mediante el modelo con árbol de decisión (Decision Tree) son los siguientes:

- Tuning hiperparámetros: 'max-depth': 3, 'min-samples-split': 10
- MAE train: 10.674, MAE test: 11.183
- RMSE train: 17.781, RMSE test: 15.761
- Variancia explicada por el modelo: train 88.6 %, test: 85.8 %

Los resultados obtenidos mediante el modelo con Random Forest son los siguientes:

- Tuning hiperparámetros: 'max-depth': 2, 'n-estimators': 10

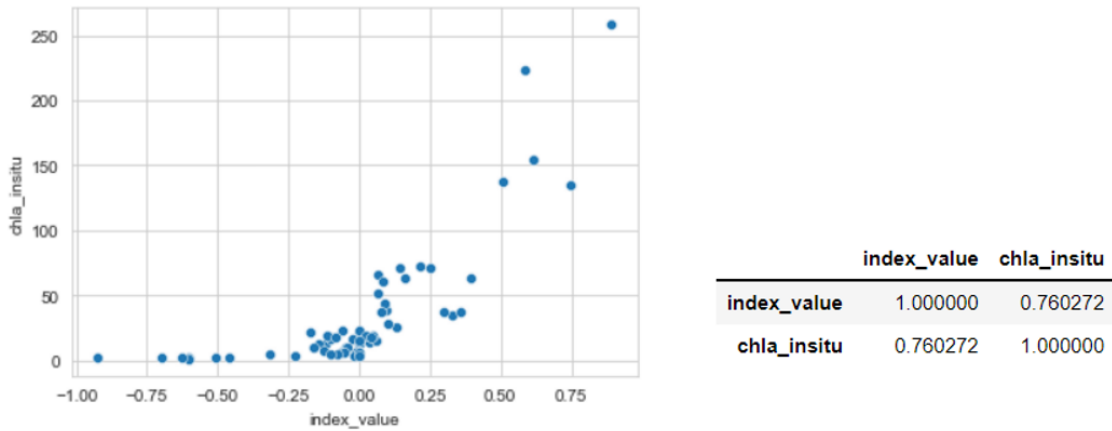


Figura 3.13: Gráfico de dispersión con el valor del índice de clorofila NDCI (eje X) vs datos observados (eje Y).

- MAE train: 11.540, MAE test: 10.512
- RMSE train: 16.868, RMSE test: 13.713
- Variancia explicada por el modelo: train 89.9%, test: 88.0%

Los resultados obtenidos mediante el modelo con SVM son los siguientes:

- Tuning hiperparámetros: 'C': 200, 'gamma': 1
- MAE train: 14.179, MAE test: 13.542
- RMSE train: 26.033, RMSE test: 18.309
- Variancia explicada por el modelo: train 77.6%, test: 81.1%

Los errores vienen dados en mg/l, que es la unidad de medida de la concentración de clorofila. La figura 3.14 muestra el extracto de código empleado para el ajuste de los datos.

3.3.3. Turbidez

La turbidez es un parámetro muy complejo de estudiar ya que intervienen todos los componentes ópticamente activos: sedimentos, materia disuelta y clorofila que están presentes en el agua.

Los resultados obtenidos mediante el modelo con árbol de decisión (Decision Tree) son los siguientes:

- Tuning hiperparámetros: 'max-depth': 2, 'min-samples-split': 10
- MAE train: 14.830, MAE test: 92.152

```

# Creamos el objeto clasificador
dt = DecisionTreeRegressor()

# Optimización de hiperparámetros
param_grid = {'max_depth': [2, 3, 4, 5, 6, 7, 8, 9, 10],
              'min_samples_split': [10, 50, 100, 200, 500]}

# Creamos el objeto gridSearch con las combinaciones de los hiperparámetros
grid_obj = GridSearchCV(dt, param_grid, cv=5)
grid_res = grid_obj.fit(X_train, y_train)
print("Best Score:" + str(grid_res.best_score_))
print("Best Parameters: " + str(grid_res.best_params_))

Best Score:0.6088917391858996
Best Parameters: {'max_depth': 3, 'min_samples_split': 10}

dt = DecisionTreeRegressor(max_depth=3, min_samples_split=10)
dt.fit(X_train, y_train)
y_pred_train = dt.predict(X_train)
y_pred_test = dt.predict(X_test)
print('MSE train: %.3f, test: %.3f' % (mean_squared_error(y_train, y_pred_train),
                                       mean_squared_error(y_test, y_pred_test)))
print('RMSE train: %.3f, test: %.3f' % (np.sqrt(mean_squared_error(y_train, y_pred_train)),
                                       np.sqrt(mean_squared_error(y_test, y_pred_test))))
print('MAE train: %.3f, test: %.3f' % (mean_absolute_error(y_train, y_pred_train),
                                       mean_absolute_error(y_test, y_pred_test)))
print('Exp variance train: %.3f, test: %.3f' % (explained_variance_score(y_train, y_pred_train),
                                               explained_variance_score(y_test, y_pred_test)))

```

Figura 3.14: Código para crear modelo de árbol de decisión.

- RMSE train: 23.730, RMSE test: 151.855
- Variancia explicada por el modelo: train 98.9%, test: 67.2%

Los resultados obtenidos mediante el modelo con Random Forest son los siguientes:

- Tuning hiperparámetros: 'max-depth': 2, 'n-estimators': 10
- MAE train: 35.108, MAE test: 38.549
- RMSE train: 91.647, RMSE test: 61.288
- Variancia explicada por el modelo: train 85.1%, test: 95.6%

Los resultados obtenidos mediante el modelo con SVM son los siguientes:

- Tuning hiperparámetros: 'C': 500, 'gamma': 10
- MAE train: 40.722, MAE test: 73.353
- RMSE train: 110.287, RMSE test: 105.162
- Variancia explicada por el modelo: train 78.9%, test: 89.4%

Los errores vienen dados en [NTU](#), que es la unidad de medida de la turbidez.

3.4. Resultados

3.4.1. Mapas de calidad del agua

Una vez desarrollados los modelos, se pueden emplear sus predicciones para generar cartografía temática sobre los parámetros de calidad de agua que se desee.

Según sea el objetivo a analizar, se pueden crear mapas de calidad para una fecha concreta, mapas promedio con el intervalo temporal que se desee (semanal, mensual, anual...). En el ejemplo que se muestra a continuación se pone de manifiesto la ventaja que presenta una visión global del sistema frente a la medida única de la sonda. En la imagen que vemos a continuación 3.15, se puede observar la detección de un proceso de eutrofización en la cola del embalse. Esta imagen sirve como base para un sistema de alerta temprana de procesos de estas características, que no son detectados insitu por la sonda hasta pasados días.

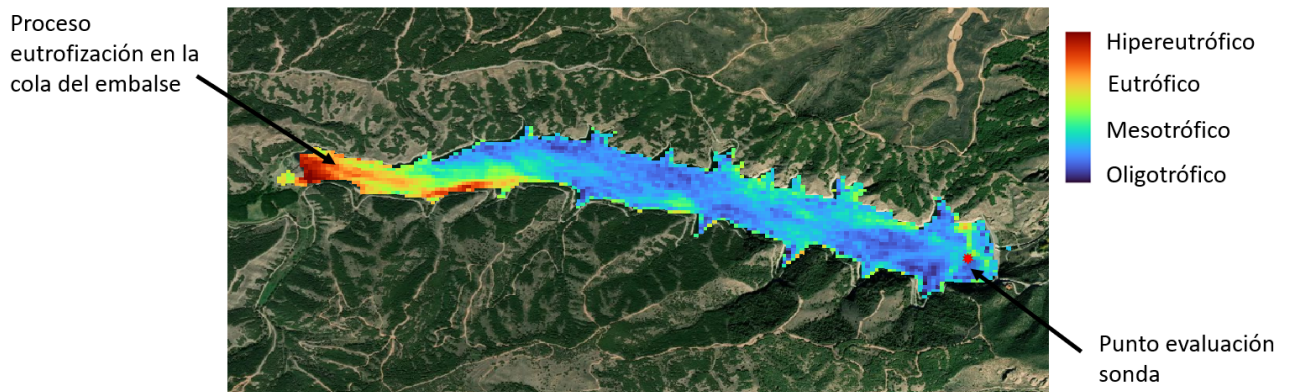


Figura 3.15: Mapa representando el estado trófico del embalse del Val en un evento de eutrofización.

3.4.2. Estimación del potencial ecológico del embalse

La DMA (Directiva 2000/60/CE) establece la necesidad de llevar a cabo diversas tareas relacionadas con la planificación y gestión de las masas de agua existentes en el territorio comunitario. Entre dichas tareas está la de realizar un seguimiento del estado de las aguas superficiales, subterráneas y de zonas protegidas. La DMA establece en su artículo 4, relativo a los objetivos medioambientales, que los Estados miembros habrán de proteger, mejorar y regenerar todas las masas de agua superficial, con objeto de alcanzar un buen estado. El estado de una masa de agua es el grado de alteración que presenta respecto a sus condiciones naturales y viene determinado por el peor valor de su estado químico y estado ecológico.

El estado ecológico de las aguas superficiales se clasifica como muy bueno, bueno, moderado, deficiente o malo. En el caso de las masas de agua muy modificadas o artificiales, que son sistemas donde las condiciones están muy alteradas y es imposible o desproporcionadamente costoso que alcancen un muy buen o buen estado ecológico (como es el **caso de los embalses**), se determina el **potencial ecológico**, que se clasifica como bueno o superior, moderado, deficiente o malo. Este concepto de “potencial” considera implícitamente la existencia y el mantenimiento de una determinada alteración en el sistema. Para el tipo de masas embalses, estas cuatro clases de calidad del potencial ecológico están definidas en la figura 3.16.

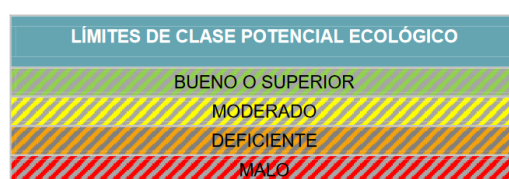


Figura 3.16: Límites de clase de Potencial Ecológico según DMA.

Los indicadores de las masas de agua que hay que monitorear vienen determinados en el **Real Decreto 817/2015**, de 11 de septiembre, por el que se establecen los criterios de seguimiento y evaluación del estado de las aguas superficiales y las normas de calidad ambiental. Éstos dependen del tipo de masa de agua y su clasificación dentro de cada tipo. En el caso del embalse El Val, este pertenece a la clase **E-T07 Monomítico, calcáreo de zonas húmedas, con temperatura media anual menor de 15 °C, pertenecientes a ríos de cabecera y tramos altos**. De los indicadores que se contemplan en el RD 817/2015: IGA, %cianobacterias, Clorofila-A y Biovolumen, el que se puede estimar mediante técnicas de teledetección es la concentración de Clorofila-A. Este indicador considera el promedio anual de Clorofila-A en el embalse para asignarle una clase del potencial ecológico según los límites publicados en el RD 817/2015. Atendiendo a estos límites, y considerando la disponibilidad de imágenes Sentinel-2 en la plataforma [GEE](#), se ha realizado la clasificación del potencial ecológico del embalse según el promedio anual de concentración de Clorofila-A, obteniendo el siguiente resultado:

Embalse	Tipo Embalse	Máximo potencial ecológico	Límites de cambio de clase de estado			MEDIA ANUAL CHLA	EQR	CLASIFICACION
			Bueno o superior/moderado	Moderado/deficiente	Deficiente/malo			
EL VAL 2018	E-T07	2.6	0.433	0.287	0.143	651	0.00	MALO
EL VAL 2019	E-T07	2.6	0.433	0.287	0.143	304	0.01	MALO
EL VAL 2020	E-T07	2.6	0.433	0.287	0.143	234	0.01	MALO
EL VAL 2021	E-T07	2.6	0.433	0.287	0.143	372	0.01	MALO
EL VAL 2022	E-T07	2.6	0.433	0.287	0.143	350	0.01	MALO

Figura 3.17: Clasificación Potencial Ecológico embalse El Val.

Capítulo 4

Conclusiones

4.1. Conclusiones generales del trabajo

Aquí se recogen las conclusiones generales del estudio y recomendaciones aplicables en cualquier programa de monitoreo de la calidad del agua mediante teledetección:

- **La viabilidad del monitoreo de calidad en masas de agua depende de las resoluciones espaciales, temporales y espectrales de las imágenes satelitales** disponibles en la zona de estudio. En el presente trabajo se han empleado imágenes Landsat para la estimación de la temperatura porque son las únicas que ofrecen una banda térmica a una resolución de 30 metros y se han empleado imágenes ópticas Sentinel-2 porque son las que mayor resolución temporal y espacial poseen del catálogo de datos abiertos.
- **La cobertura nubosa limita la técnica** del seguimiento de calidad mediante imágenes satelitales, por ello en el presente trabajo se ha aplicado un filtro del 30% a la cobertura nubosa a través de los metadatos de las imágenes.
- Es imprescindible que **los datos de campo y las imágenes satelitales sean capturados de forma simultánea**. El intervalo máximo de tiempo de muestreo simultáneo en que las condiciones de iluminación, para un día despejado, se consideran homogéneas es de 3 horas. En este trabajo no se ha planificado una campaña de muestreos en campo, sino que se han empleado los datos disponibles de los perfiles realizados por una sonda instalada in situ. Como no se tiene control sobre las horas a las que la sonda realiza la captura, se ha escogido la toma de datos más próxima al paso del satélite por la zona.
- En la calibración de este trabajo, no se ha empleado el punto exacto de medición de la sonda con su correspondiente pixel en la imagen; ya que al estar situada la sonda en la corona de la presa, el pixel que contiene la sonda in situ en la imagen es de naturaleza mixta (agua + otra cubierta) y no 100% agua. Por eso, el píxel de evaluación en la imagen ha

sido desplazado aguas adentro del embalse para garantizar que siempre se identifica como agua, añadiendo una fuente de error puesto que **la ubicación que estamos considerando en la imagen no es exactamente la misma que donde se están midiendo los datos in situ**. La figura 4.1 ilustra esta situación:



Figura 4.1: Ubicación real de la sonda en campo y el píxel empleado para evaluación en las imágenes satelitales.

- Los modelos solamente pueden ser **útiles en la zona y la temporada que fueron concebidos**, debido a que su calibración se lleva a cabo regionalmente mediante los datos de mediciones in situ para las características ópticas de los diferentes sitios. Es necesario recalibrar continuamente la relación entre las propiedades del agua y la información del imaginario satelital.
- Generalmente, la frecuencia temporal del satélite supera los límites de los procesos que se pueden observar en el agua. El tiempo de revisita es mayor que el tiempo que dura el proceso contaminante en el agua por lo que no se llega a caracterizar adecuadamente. Además, hay que tener en cuenta que **el agua (y el material que contiene) se mueve** por lo que es complejo representar mediante una toma satelital los cambios dinámicos en la calidad del agua.
- La interferencia atmosférica es significativamente alta sobre los cuerpos de agua. De la

señal medida por satélites sobre aguas continentales, alrededor del 90 % proviene de la contribución atmosférica. Esto es causado por la neblina atmosférica que dispersa la luz, especialmente en las longitudes de onda más bajas (azul, verde). Las interferencias aumentan a medida que disminuye la radiación reflejada por el agua, lo que significa que el agua con alta claridad o con altas concentraciones de algas es la más afectada. **Los procedimientos de corrección atmosférica adecuados juegan un papel importante al estimar la calidad del agua a partir de los valores de reflectancia de salida del agua, especialmente si se basan en relaciones simples o múltiples utilizando bandas del espectro de luz visible.** En este trabajo las imágenes analizadas ya vienen corregidas atmosféricamente y geoméricamente por el fabricante porque se considera fuera del alcance del proyecto evaluar diferentes modelos y técnicas de corrección atmosférica de las imágenes, aunque pudiera mejorar los resultados.

- Otra cosa a tener en cuenta sobre todo en la medición de la turbidez y la concentración de clorofila-a es que en el trabajo de campo se realizan medidas puntuales, es decir, se toman datos en puntos concretos, mientras que el píxel de la imagen es el promedio de la respuesta que da una superficie de 10x10m (100m²) o de 20x20m (400m²) dependiendo qué banda de Sentinel-2 se utilice. Esto quiere decir que, para poder hacer una comparación adecuada, se deberían tomar varias muestras en un cuadrado de esas dimensiones y de ellas, hacer el promedio. Dado que esto no es factible logísticamente, **expertos consideran que un ajuste de los algoritmos de calidad de aguas es bueno cuando la correlación entre el valor que proporciona y las medidas de campo en una población estadísticamente significativa de puntos de muestreo (distribuidos espacial y temporalmente) es aceptable, a partir de un $R^2 > 0.65$ [6].**

En base a los resultados obtenidos en este estudio y los análisis realizados en todo el proceso de desarrollo de los algoritmos para los parámetros examinados se concluye lo siguiente:

4.1.1. Conclusiones: temperatura superficial del agua

- La capacidad de monitorear la temperatura del agua es importante para evaluar los cambios en los ecosistemas fluviales. Actualmente, la recolección de datos de temperatura in situ brindan muestras puntuales (espacialmente y/o temporalmente) porque caracterizar la temperatura de todas las masas de agua a nivel global en una escala espacial regional tiene un coste muy alto. **Las imágenes infrarrojas térmicas de los satélites constituyen una alternativa y una manera complementaria de obtener datos válidos a escala regional.**
- Los datos de temperatura estimados de forma satelital, tienen un **98.7 % de correlación**

lineal con los datos medidos de campo.

- Al evaluar la distribución y valores promedio de la temperatura estimada por satélite se ha observado que la media de las mediciones satelitales es superior a la obtenida en campo. Estos resultados sugieren que **las temperaturas superficiales basadas en imágenes Landsat tienen un sesgo alto**, y el grado varía en función de factores hidrológicos y meteorológicos, ya que se ha observado que **el sesgo es mayor en temperaturas altas**.
- El algoritmo empleado para ajustar los datos ha sido el de **regresión lineal** por ofrecer resultados aceptables y claramente suficientes: la variancia explicada por el modelo es de 97.4% en entrenamiento y 97.8% en test, con un **error absoluto medio de 0.9°C y un error cuadrático medio de 1°C**.
- Actualmente no existen alternativas de misiones satelitales sin coste que obtengan imágenes térmicas mejorando la resolución espacial (100 metros nativa, distribuida al público resampleada a 30 metros) que ofrece Landsat.

4.1.2. Conclusiones: concentración de clorofila-a

- El trabajo se ha orientado para **evaluar el riesgo potencial de un proceso de eutrofización** en la masa de agua a partir de los datos de las imágenes satelitales y que se convierta en una **oportunidad para implementar un tratamiento específico o practicar una intervención oportuna** en los lugares señalados.
- En aguas ópticamente complejas, como es el caso de este trabajo, la contribución de sedimentos puede resultar en una sobreestimación de clorofila a partir de las imágenes de satélite. Es posible que en el punto de calibración haya más componentes ópticamente activos en el agua además de la clorofila-a, que afecten al valor del píxel analizado. Al tratar solo con valores de reflectividad de las imágenes y no radiometría de campo, es imposible diferenciar qué componente está produciendo la respuesta espectral recibida. Por ello, es indispensable tomar datos de radiometría en los puntos donde se toman las muestras de agua de manera simultánea. Esto permite crear y alimentar una **librería espectral** con la que caracterizar las aguas y que servirá más adelante para comparar las reflectividades de las imágenes satelitales con los comportamientos espectrales de cada tipo de agua en función a los componentes que contiene.
- El **sensor MSI de Sentinel-2 no ha sido específicamente diseñado** para medir la concentración de clorofila. La misión Sentinel-3 sí que ha sido concebida con el objetivo de monitorear el estado y salud de la vegetación haciendo estimaciones de contenido de clorofila en la vegetación terrestre y ofrecer productos de cobertura mundial para el color del océano y la calidad del agua. Sus bandas están estratégicamente diseñadas para capturar los picos de absorción, reflectancia y fluorescencia de la clorofila resultando unos

productos precisos sobre pigmentos de algas, concentraciones de sólidos suspendidos y caracterización de la transparencia del agua basada en el coeficiente de atenuación difusa. Este sensor en concreto, que ha sido diseñado para los objetivos de este estudio, no es posible emplearlo en la masa de agua del embalse El Val ya que su resolución espacial es de 300 metros, habiendo sido concebido para mares y océanos.

- En el estudio de la clorofila-a se ha decidido emplear el **Índice de Clorofila de Diferencia Normalizada (NDCI)** que considera los picos de reflectancia y absorción de la clorofila en las bandas en el espectro rojo, como indicador de la relación entre la reflectancia del píxel de la imagen medida por el sensor remoto y su valor medido en campo.
- Para estimar la concentración de clorofila-a se han evaluado los algoritmos: Decision Tree, Random Forest y Support Vector Machines (SVM), obteniendo un rendimiento similar en el caso de los árboles de decisión y Random Forest, mejorando un poco su puntuación el segundo. El mejor modelo ha sido el obtenido con **Random Forest**, una profundidad máxima de 2 y 10 estimadores. De esta forma se han obtenido una variancia explicada por el modelo es de 89.9% en entrenamiento y 88% en test, con un **error absoluto medio de 11 mg/l y un error cuadrático medio de 15 mg/l**.
- Esta magnitud de los errores no suponen valores significativos cuando el embalse está en un estado de eutrofización ya que según la clasificación empleada de la OCDE un embalse está en estado eutrófico a partir de los 25 mg/l, y en El Val se alcanzan promedios anuales en torno a los 300 mg/l. En consecuencia, los modelos no son capaces de diferenciar estados donde las concentraciones de clorofila son bajas (no nocivas), lo cual no supone un problema ya que el objetivo de estos estudios es implementar un sistema de alerta temprana ante procesos que puedan suponer un perjuicio mediambiental.

4.1.3. Conclusiones: turbidez

- Como se ha revisado en la teoría, **la firma espectral del agua turbia es variable**, dependiendo de la cantidad y tipo de partículas presentes en ella, lo que hace que **la turbidez sea un parámetro muy complejo de caracterizar**. Por esta razón es recomendable realizar una campaña de espectro-radiometría en campo coincidente en tiempo y espacio con la toma de la imagen del satélite para construir una librería espectral que permita caracterizar las diferentes reflectividades recibidas por el sensor con los diferentes componentes presentes en el agua.
- La turbidez es un parámetro que está relacionado con los incrementos de caudal y las precipitaciones. De esta manera, cuanto mayor sea la magnitud del incremento de caudal en las aportaciones al embalse o mayores sean las precipitaciones, más probable es que aumente la turbidez en la masa de agua. Las relaciones no son lineales ni su coeficiente

de determinación elevado, y éstas no son las únicas variables que influyen en el aumento de la turbidez, pero sí son dos factores importantes. **Si el aumento de turbidez se da por un episodio de lluvias, la imagen satelital que capta el evento va a estar cubierta de nubes y no va a ser posible su análisis. Para cuando vuelva a tomarse una nueva imagen (revisita cada 5 días), el evento ya habrá pasado por completo.** Si no se tuviese la limitación de la cobertura nubosa en las imágenes ópticas, se podría disponer de una mayor cantidad de datos en el rango alto de turbidez.

- En el estudio de la turbidez se ha decidido emplear el **Índice de Turbidez de Diferencia Normalizada (NDTI)** que se basa en el aumento de reflectancia del espectro rojo cuando existe aumento de turbidez.
- Debido a la naturaleza compleja del parámetro, a todos los componentes que le afectan directamente y a la disponibilidad de las imágenes, en los datos se tienen **multitud de valores para medidas bajas de turbidez relacionados con una gran variedad de reflectividades** que dificultan la búsqueda efectiva de relaciones.
- **Los modelos evaluados para estimar la turbidez no han sabido representar el comportamiento de la variable de forma adecuada** por los motivos anteriormente expuestos. Los modelos son capaces de simular los aumentos grandes de turbidez de forma aceptable pero su rendimiento es mejorable cuando las variaciones de turbidez son pequeñas ya que se están cometiendo errores de magnitud significativa.

4.2. Líneas de trabajo futuras

En general hay varias líneas de trabajo que se pueden explorar y se pueden categorizar por bloques temáticos: tratamiento de las imágenes satelitales, algoritmos de aprendizaje automático en general y la relación con cada parámetro en particular.

- En cuanto a las imágenes satelitales, se puede sustituir la metodología de corrección atmosférica de las imágenes Sentinel-2 (en este proyecto se ha optado por una técnica generalista ya implementada por el distribuidor ya que este proceso queda fuera del alcance del proyecto) por una técnica más específica para cuerpos de agua y aguas ópticamente complejas: el **algoritmo C2RCC** disponible en el software libre de la ESA, SNAP, para el procesamiento de imágenes Copernicus. Entre los diferentes métodos desarrollados se encuentra el proceso Case 2 Regional Coast Colour (C2RCC), que ha sido adaptado a Sentinel-2 y es utilizado para aguas con menos turbidez como las marinas, mientras que el proceso C2X (Case-2 Extreme Cases) se emplea en aguas algo más turbias como las continentales, cuya presencia de sólidos y fitoplancton es más alta. Los productos generados directamente en el software son las propiedades ópticas inherentes (IOP), absorción y dispersión de

los diferentes constituyentes; las tres principales concentraciones ópticamente relevantes, es decir, pigmentos de fitoplancton, materia suspendida total y sustancia amarilla; y sus incertidumbres relacionadas. Es compatible con varios sensores; desde Sentinel-3 OLCI sobre Sentinel-2 MSI hasta Landsat-8 OLI. Un proyecto interesante sería investigar si se puede obtener de forma libre el código de estos algoritmos de corrección para aplicarlos sobre las imágenes del proyecto, ya que actualmente este algoritmo C2RCC viene como función dentro del software SNAP y se aplica a cada imagen de forma independiente.

- Las líneas de trabajo referentes a los algoritmos de aprendizaje automático son incontables: en este trabajo el alcance se ha acotado a la evaluación de tres algoritmos, por lo que la primera línea de trabajo sería evaluar un número mayor de algoritmos y de diferentes naturalezas. Es evidente que el alcance de este trabajo no podía incluir una revisión exhaustiva de todas las opciones. Dentro de los tres algoritmos evaluados, se han hecho la búsqueda óptima de los dos hiperparámetros más característicos de cada uno, pero también se puede revisar incluir más hiperparámetros en la fórmula. Por último, otra decisión prefijada en el proyecto ha sido la de emplear validación cruzada con 5 particiones estratificadas, variable que también podría ser investigada. Estas últimas opciones se entiende que podrían mejorar sutilmente el desempeño de los modelos, la opción de probar nuevos algoritmos es la que podría hacer una diferencia sustancial.
- Otra línea interesante para explorar sería la de evaluar si la aplicación de técnicas de ensemble learning mejora la precisión y fiabilidad de los modelos predictivos.
- En el caso de la turbidez las opciones de trabajo para alcanzar un resultado aceptable son múltiples: en vez de la correlación de los datos de campo con el índice [NDTI](#), que ha sido el explorado en este trabajo, se puede buscar otra ecuación que relacione las reflectividades capturadas por el sensor con las medidas in situ. Existen multitud de ecuaciones y relaciones entre bandas en la bibliografía que pueden ser consideradas por los algoritmos para obtener mejores precisiones. También se puede realizar un filtrado exhaustivo de los datos medidos por la sonda, seleccionando solamente aquellos que a criterio de experto se consideren significativos de algún evento que pueda ser detectado de forma remota por el satélite, ya que ahora es posible que los modelos no generen un resultado óptimo debido a que la masa de agua está muy influenciada por los procesos de eutrofización superficiales.
- Aunque con la estimación de la concentración de clorofila-a se han obtenido resultados aceptables, otra línea de trabajo puede ser sustituir el índice [NDCI](#) por una ecuación calibrada ad-hoc estudiando las diferentes respuestas de las bandas espectrales que detallen de una forma más precisa las pequeñas variaciones de clorofila-a.
- En cualquier caso, se recomienda la continuidad de recolección de datos in situ y satelitales que alimenten los algoritmos y permitan realizar ajustes más robustos a futuro.

Bibliografía

- [1] Organización de las Naciones Unidas para la Alimentación y la Agricultura (FAO). El estado mundial de la pesca y la acuicultura. <https://www.fao.org/3/ca9229es/online/ca9229es.html>, 2020. [Online; visitado 01-Octubre-2022].
- [2] Dirección General de Medio Ambiente Comisión Europea. La directiva marco del agua de la UE. <https://data.europa.eu/doi/10.2779/75335>, 2014. [Online; visitado 08-Octubre-2022].
- [3] Confederación Hidrográfica del Ebro. Estado trófico del embalse de El Val y programa de medidas. https://www.chebro.es/documents/20121/55149/EC16013_SENSIBLES_VAL_IF_vD.pdf, 2017. [Online; visitado 01-Octubre-2022].
- [4] Naciones Unidas. La Agenda para el Desarrollo Sostenible. <https://www.un.org/sustainabledevelopment/es/development-agenda/>. [Online; visitado 01-Octubre-2022].
- [5] Universidad Nacional de Quilmes. Introducción a la teledetección. <https://static.uvq.edu.ar/mdm/teledeteccion/unidad-1.html>. [Online; visitado 29-Septiembre-2022].
- [6] Domínguez et al. *Estudio de aguas continentales mediante teledetección*. UNED, 2011.
- [7] A.; Gurlin D Palacios, S.; Mehta. Integrating Remote Sensing into a Water Quality Monitoring Program. NASA Applied Remote Sensing Training Program (ARSET). <https://appliedsciences.nasa.gov/join-mission/training/english/arset-integrating-remote-sensing-water-quality-monitoring-program>, 2019. [Online; visitado 15-Octubre-2022].
- [8] J.; McCartney S. Mehta, A.; Torres-Pérez. Monitoring Coastal and Estuarine Water Quality: Transitioning from MODIS to VIIRS. NASA Applied Remote Sensing Training Program (ARSET). <https://appliedsciences.nasa.gov/join-mission/training/english/>

- [arset-monitoring-coastal-and-estuarine-water-quality-transitioning](#), 2021. [Online; visitado 15-October-2022].
- [9] Universidad de Valencia. Ecological Status of Aquatic Systems with Sentinel Satellites. <https://leoipl.uv.es/esaqs/index.php>, 2016. [Online; visitado 15-October-2022].
- [10] Pereira-Sandoval et al. Evaluation of atmospheric correction algorithms over spanish inland waters for Sentinel-2 multispectral imagery data. <https://doi.org/10.3390/rs11121469>, 2019. [Online; visitado 15-October-2022].
- [11] Pereira-Sandoval et al. Calibration and validation of algorithms for the estimation of chlorophyll-a concentration and Secchi depth in inland waters with Sentinel-2. <http://www.limnetica.net/documentos/limnetica/limnetica-38-1-p-471.pdf>, 2019. [Online; visitado 15-October-2022].
- [12] Soria-Perpinyá et al. Monitoring the ecological state of a hypertrophic lake (Albufera of Valencia, Spain) using multitemporal Sentinel-2 images. <http://www.limnetica.net/documentos/limnetica/limnetica-38-1-p-457.pdf>, 2019. [Online; visitado 15-October-2022].
- [13] Soria-Perpinyá et al. Remote sensing applications for the study of rapid flushing to remediate eutrophication in shallow lagoons (Albufera of Valencia) using multitemporal Sentinel-2 images. <https://link.springer.com/article/10.1007%2Fs10750-018-3741-6>, 2019. [Online; visitado 15-October-2022].
- [14] Soria-Perpinyá et al. Remote sensing of cyanobacterial blooms in a hypertrophic lagoon (Albufera of Valencia) using multitemporal Sentinel-2 images. <https://www.sciencedirect.com/science/article/pii/S0048969719342949>, 2019. [Online; visitado 15-October-2022].
- [15] Confederación Hidrográfica del Ebro. Seguimiento de embalses por teledetección. https://portal.chebro.es/documents/20121/55149/S_112_21_INFORME_CHE_Rev9.pdf/a24d3d49-df0e-7aa3-8570-c93e9471c05f?t=1646987319554, 2020. [Online; visitado 15-October-2022].
- [16] A Mehta, A.; Flores. Processing Satellite Imagery for Monitoring Water Quality. NASA Applied Remote Sensing Training Program (ARSET). <https://appliedsciences.nasa.gov/join-mission/training/english/arset-processing-satellite-imagery-monitoring-water-quality>, 2018. [Online; visitado 15-October-2022].