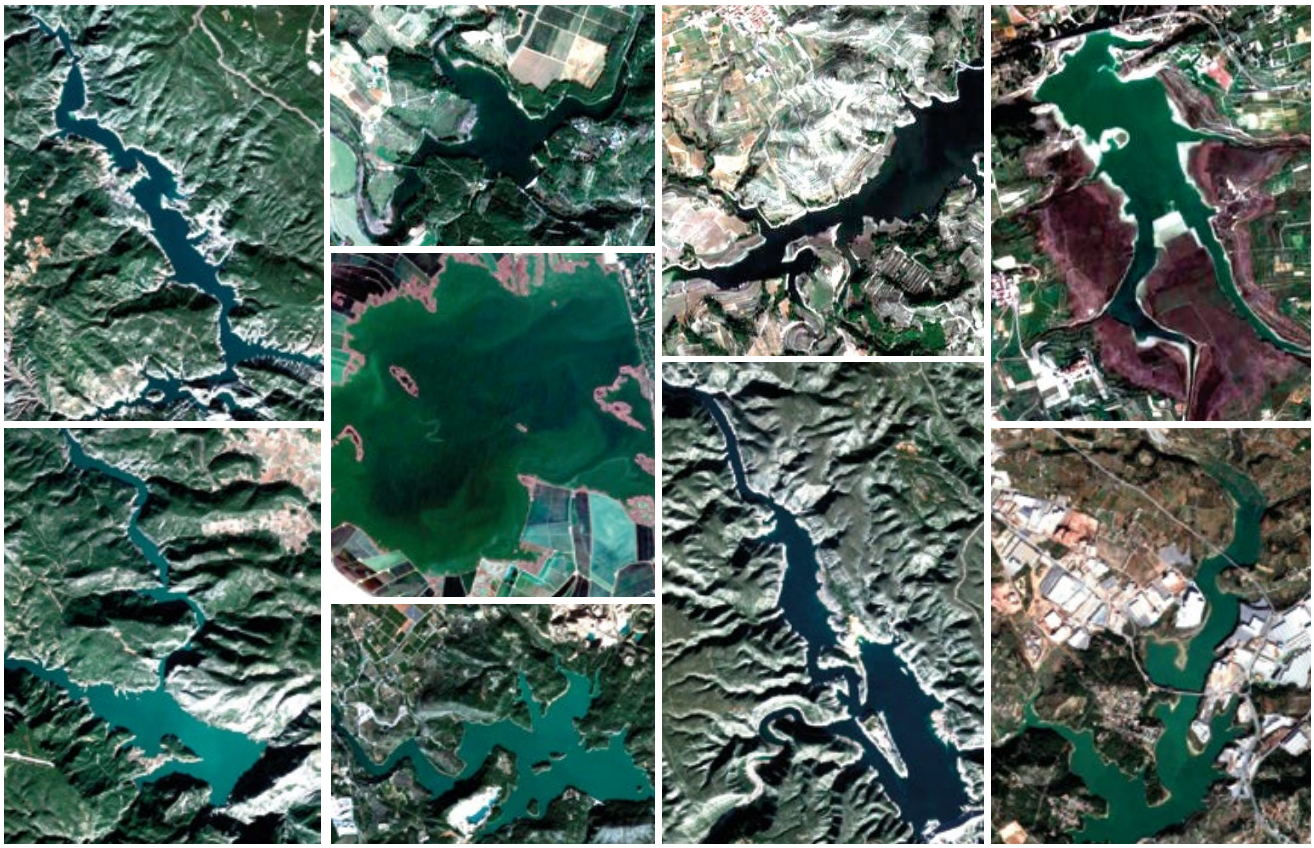
Desafíos y oportunidades de Sentinel-2 en la monitorización de la calidad de las aguas continentales

Marcela Andrea Pereira-Sandoval



Doctorado en Teledetección

Junio 2023

Directores

Ana Belén Ruescas Orient
Jesús Delegido Gómez
José Moreno Méndez

VNIVERSITAT DE VALÈNCIA

UNIVERSITAT DE VALÈNCIA

Facultat de Física
Departament de Física de la Terra i Termodinàmica

Desafíos y oportunidades de Sentinel-2 en la monitorización de la calidad de las aguas continentales

Marcela Andrea Pereira-Sandoval

Doctorado en Teledetección

Tesis Doctoral
Junio 2023

Directores

Ana Belén Ruescas Orient
Jesús Delegido Gómez
José Moreno Méndez

ANA BELÉN RUESCAS ORIENT, Profesora en el departamento de Geografía de la Facultad de Geografía e Historia, Universidad de Valencia.

JESÚS DELEGIDO GÓMEZ, Catedrático en el departamento de Física de la Tierra y Termodinámica de la Facultad de Física, Universidad de Valencia.

JOSÉ MORENO MÉNDEZ, Catedrático en el departamento de Física de la Tierra y Termodinámica de la Facultad de Física, Universidad de Valencia.

CERTIFICAN:

Que la tesis doctoral "Desafíos y oportunidades de Sentinel-2 en la monitorización de la calidad de las aguas continentales", presentada por Marcela Andrea Pereira Sandoval, ha sido realizada bajo su dirección y es favorable para optar a grado de Doctor por la Universidad de Valencia en el programa de Doctorado en Teledetección.

Y para que así conste, firman el presente certificado en Paterna, a 1 de junio de 2023.

Fdo: Ana Belén Ruescas Orient

Fdo: Jesús Delegido Gómez

Fdo: José Moreno Méndez

Financiación

Parte de esta tesis doctoral ha sido posible gracias a la financiación económica recibida por:

- La Generalitat Valenciana mediante la concesión del Programa Prometeo 2016/032 para la puesta en marcha y desarrollo de proyecto ESAQS (Estado Ecológico del los Sistemas Acuáticos con los satélites Sentinels).

Otros proyectos y contratos científicos en los que he colaborado y han financiado diversas actividades científicas como participación a congresos:

- El Contrato Framework Partnership Agreement on Copernicus User Uptake. Courses, workshops and seminars in Europe and Latin America (FPA 275/G/GRO/COPE/17/10042) financiado por la Comisión Europea.

A Juan,

Teresa, Edmundo,

Vicky, Valen, Giu y Ri.

Mis amores.

Agradecimientos

Voy a comenzar con mis directores de tesis doctoral. Ana, sin tu constante apoyo y confianza en mí, hoy yo no estaría aquí. Un millón de gracias por tus consejos, tu tiempo, por escucharme y siempre animarme a seguir cuando en más de una oportunidad creía que este momento nunca iba a llegar. Me siento muy orgullosa de ser tu primera doctoranda, significa mucho más de lo que te puedas imaginar. Ojalá existieran más Anas que tuvieran tu mismo amor por la investigación y sobre todo por tu gran humanidad. Jesús, nunca olvidaré todo lo que has hecho por mí desde que comencé este camino por la universidad. Yo, una argentina recién llegada a Valencia, con muchas ganas de aprender y vos, un profe con muchas ganas de enseñarme sobre como iba esto del mundo de la investigación con mucho cariño y sobre todo con una graaaan e infinita paciencia, un millón de gracias por haber confiado en mí cuando más lo necesitaba, eso es algo que nunca olvidaré. Pepe, tus clases magistrales en aquellas aulas del máster afianzaron mi amor por la teledetección. Gracias por haberme invitado a pasarme un día por el IPL, y quién hubiese dicho que ese día se convertiría en una estancia de aprendizaje en el maravilloso Grupo LEO. Eres un apasionado de este mundo de la investigación y esa pasión la transmites y nos la haces llegar a gente como yo que estamos orgullosos de ser tus doctorandos. Gracias a ti y a Jesús porque a pesar de no seguir junto a ustedes, siempre me han animado y apoyado para llegar hoy aquí.

Antonio y Carolina, siempre se los dije y se los seguiré diciendo “mis otros directores de tesis del corazón”. Gracias por haberme abierto las puertas de la teledetección en aguas. Casi todo lo que sé, se los debo a ustedes. Gracias por los años compartidos en la oficina, por las horas que invirtieron en mí y que dedicaron para enseñarme de qué iba todo esto de los satélites y el agua. Gracias por haberme permitido ser parte de su equipo, por la enorme paciencia que me tuvieron y por la calidez con la que me han tratado. Estaré siempre en deuda con ustedes y serán parte de mi corazón.

Sigo con el “Combo Tropical”. ¿Qué hacen juntas una colombiana, una olivense, una venezolana, una balear y una argentina en el IPL? Pues... Reír, llorar,

volver a reír y recorrer juntas el camino del Doctorado en Teledetección. Gracias por todo el tiempo compartido, por estar en el momento justo y en el lugar indicado. Carito, Nieves, Patricia y Eatidal, sin ustedes esto no hubiera sido tan divertido.

También quiero agradecer a mis otros compañeros del Grupo LEO con quienes compartí los primeros años de este camino. Quiero mencionar especialmente a MaPi y Shari. Ha sido un privilegio haberlas conocido, gracias por mostrarme otra teledetección, por compartirme sus experiencias y por sus preciados consejos. Son dos profesionales admirables, pero sobre todo dos personas maravillosas. A mis compañeros Adrián, Luca, Santi, Jochem, por permitirme compartir esta experiencia junto a ustedes y aprender tanto de lo que nada conocía. También quiero agradecer a compañeros del Grupo ISP, quienes en algún momento me han tendido un mano con una profesionalidad admirable, agradezco especialmente a Gonzalo, Jordi, Gustau y María.

Agradecer también especialmente al Grupo de Limnología, a Eduardo, Juan Miguel y Xavi. Con ustedes aprendí lo duro del trabajo de campo. La cantidad de veces que hemos cargado, hinchado y deshinchado el bote o remado cuando no llegabamos con el motor. Con ustedes aprendí a tomar muestras, a "lugolizar" y volver a cargar el bote una y otra vez. ¡Qué felicidad y cuánto esperaba las salidas de campo! Gracias por enseñarme lo bonito, lo frágil e importante que son nuestros ecosistemas acuáticos. Tampoco quiero olvidarme de ti Ramón, de quien he aprendido mucho. Gracias por preocuparte por mí en tus visitas al IPL.

No quiero olvidarme de dos personas con las que comencé este camino allá por 2014 en las aulas del Máster de Teledetección. Carolina y Guiomar. Caro, no importa en que isla estés, ni dónde estés, seguro que allí nos volveremos a ver, que para nosotros la distancia no es problema. Guio, qué alegría que después de tantos años todavía sigamos en contacto.

A mi familia de Argentina, a Teresa, Edmundo, mis padres, gracias por apoyarme siempre en todas mis decisiones y estar incondicionalmente para mí, los amo. A mis hermanas, a mis sobrinos a quienes amo infinitamente, Victoria, Valentino, Giuliana, que con sus juegos, sus ocurrencias o con una simple sonrisa me alegran los días grises. Bebu Ri, te agradezco todo lo que me estás enseñando en tan poquito tiempo, sos un ejemplo de lo que verdaderamente significa luchar en la vida. A mi ahijado Titino, a mis sobris del corazón, Tuly, Ludmi y Anita. A

mi familia de España, especialmente a Paula y Monche por animarme siempre a continuar, a mis cuñados, especialmente a Pedro por ser un ejemplo, a mis sobris Dani y Pablo. Y a mi otra familia por elección, a mis amigas de toda la vida, mis hermanas del corazón: Cintia, Lucrecia y Alejandra ¡Chicas lo logramos! Gracias por bancarme y por estar siempre conmigo en lo verdaderamente importante.

También quiero dar las gracias a los compis "visuales" que me están acompañando en el final de este camino, especialmente a María, Juanki, Ferran, Bea, Pepe, Julia y también a Lucía y Mercedes por darme la oportunidad de seguir haciendo lo que me gusta con otra mirada.

Por último, mi agradecimiento final es para una de las personas más importantes de mi vida, Juan, mi gran amor. Estuviste desde el minuto cero conmigo en esta aventura. Quién más que vos me aguantó en las subidas y bajadas de esta montaña rusa. Gracias por ser el pilar de mi vida. Te elegiría una y mil veces. Gracias amor.

Nota introductoria

Esta tesis doctoral se presenta para obtener el título de Doctorado en Teledetección de la Universidad de Valencia.

Según el reglamento vigente sobre depósito, evaluación y defensa de la tesis doctoral de la Universidad de Valencia, la tesis doctoral puede ser presentada de dos formas: 1) Tesis tradicional o 2) Tesis por compendio de publicaciones. En este caso, la tesis presentada está elaborada por compendio de publicaciones. La tesis de acuerdo con esta opción, tiene que cumplir los siguientes requisitos:

1. Se deben presentar un mínimo de tres artículos, ya publicados o aceptados en revistas indexadas en algún índice internacional como JCR (WoS) y/o SJR (Scopus) y ser el autor de la tesis, el primer firmante de cada uno de ellos.

2. La tesis debe presentar un resumen global de la temática, de los principales resultados y de las conclusiones.

3. La tesis debe incluir una copia completa de los trabajos publicados o admitidos para publicación.

La presente tesis doctoral está compuesta de las siguientes tres publicaciones:

1. **Pereira-Sandoval, M.**, Urrego, P.E., Ruiz-Verdú, A., Tenjo, C., Delegido, J., Soria-Perpinyà, X., Vicente, E., Soria, J. y Moreno, J. Calibration and validation of algorithms for the estimation of chlorophyll-a concentration and Secchi depth in inland waters with Sentinel-2. *Limnetica.* 2019, 38(1):471-487. DOI: 10.23818/limn.38.27

2. **Pereira-Sandoval, M.**, Ruescas, A., Urrego, P.E., Ruiz-Verdú, A., Delegido, J., Tenjo, C., Soria-Perpinyà, X., Vicente, E., Soria, J. y Moreno, J. Evaluation of atmospheric correction algorithms over spanish inland waters for Sentinel-2 Multispectral Imagery Data. *Remote Sens.* 2019, 11, 1469. https://doi.org/10.3390/rs11121469

3. **Pereira-Sandoval, M.**, Ruescas, A.B, García-Jimenez, J., Blix, K., Delegido, J. y Moreno, J. Supervised classification of optical water types in Spanish inland waters. *Remote Sens.* 2022, 14, 5568. https://doi.org/10.3390/rs14215568

Índice general

Acrónimos y abreviaciones

ACOLITE Atmospheric correction for Oli lite

AOP Apparent Optical Properties

BOA Bottom Of Atmosphere

C2RCC Case 2 Regional Coast Colour

C2RCCCX Case 2 Regional Coast Colour for Complex Waters

Chl-a Clorofila-a

DTC Decision trees

DMA Directiva Marco del Agua

ESA European Spatial Agency

ESAQS Ecological Status of Aquatic Systems with Sentinel

IOPs Inherent Optical Properties

iCOR Image Correction for atmospheric effects

Kd Diffuse attenuation coefficient

KNN K-Nearest Neighbor

L1C Level 1C

L2A Level 2A

linSVC Linear Support Vector Machine Classifier

NIR Near - Infrared

OAC Componentes Ópticamente Activos

OC Ocean Color

OC_490 Ocean Color 490

OWT Optical Water Types

Polymer Polynomial-based algorithm applied to MERIS

polySVC Polynomial Support Vector Machine Classifier

rbfSVC Radial basis function Support Vector Machine Classifier

RFC Random Forest Classifier

R_{rs} Remote sensing reflectance

S2-MSI Sentinel 2 - MultiSpectral Image

Sen2Cor Sentinel 2 Correction

TOA Top Of Atmosphere

TBDO Triple banda Dall' Olmo

TSM Total Suspended Matter

Z_{SD} Secchi disk depth

Resumen

La sequía, las inundaciones, el aumento del nivel del mar, así como la propia acción del hombre, tienen un impacto cada vez mayor sobre la dinámica de los ecosistemas acuáticos. En las aguas continentales, la escasez y la contaminación de este recurso está promoviendo que diversos organismos gubernamentales incluyan en sus agendas políticas, estrategias para mitigar esta situación a través de una gestión sostenible. Directivas ambientales europeas, como la Directiva Marco del Agua (DMA) establece entre sus requerimientos fundamentales, la monitorización del estado ecológico de las aguas continentales.

Uno de los métodos tradicionales utilizados en la monitorización y evaluación de la calidad de las aguas continentales se basa en las mediciones *in situ* de muestras de agua. Se trata de un método empírico muy preciso, pero que tiene la desventaja de ser muy costoso en tiempo y dinero. Además los resultados se circunscriben a un momento determinado y a una extensión espacial muy reducida. Teniendo en cuenta que las aguas continentales presentan dinámicas influenciadas por factores externos (luz solar, viento) y por factores internos (flora acuática, elementos biológicos, físico-químicos) altamente variables en tipo y concentración, según el momento y el lugar, los métodos tradicionales resultan muy limitados por lo que es necesario aplicar herramientas y técnicas que simplifiquen las tareas de monitorización, que sean recurrentes en el tiempo y que cubran mayores superficies territoriales.

Las imágenes satelitales tienen el potencial de ofrecer una visión sinóptica, objetiva y continua, a partir de la cual es posible derivar métricas que permitan estimar el estado ecológico de las masas de agua. Esas métricas son un complemento a los tradicionales muestreos *in situ*, ya que permiten incrementar la cobertura espacial y la periodicidad en el tiempo de monitorización. Sentinel-2, con su sensor

MultiSpectral Instrument (S2-MSI), es un satélite del Programa Copernicus perteneciente a la Agencia Espacial Europea (ESA), destinado a monitorizar la vegetación de la superficie terrestre. El sensor MSI incorpora las ventajas de SPOT y Landsat, siendo la primera misión en su clase en añadir tres bandas espectrales en la región límite del rojo (*red-edge* - 690 a 790 nm). S2-MSI es una constelación de dos satélites: Sentinel-2A y Sentinel-2B, con una resolución espacial (10, 20 y 60 m) beneficiosa para las masas de aguas de medianas y pequeñas dimensiones, bandas espectrales en la región del visible, infrarrojo e infrarrojo de onda corta (desde los 443 nm a 2190 nm) y un período de revisita cada 5 días, características que permiten que sea utilizado por diferentes programas europeos para la monitorización de las aguas continentales. La plataforma de la ESA, *Copernicus Open Access Hub* provee imágenes S2-MSI de niveles L1C (reflectividad en *Top Of Atmosphere - TOA*) y L2A (reflectividad en *Bottom Of Atmosphere - BOA*) listas para ser procesadas por el usuario. Sin embargo, el producto L2A estándar es corregido de la contribución del efecto atmosférico con el algoritmo oficial de la ESA, Sentinel 2 Correction (Sen2Cor), cuyo método está orientado a eliminar el efecto atmosférico de la superficie terrestre y así obtener la reflectividad de la vegetación.

La monitorización de las aguas continentales mediante imágenes S2-MSI, presenta varios desafíos. Por una parte, la dinámica variable en tiempo y espacio de la concentración de los diversos componentes de las masas de agua, dificulta trabajar con una única metodología para la obtención de los parámetros biofísicos que determinen su calidad. Además, que S2-MSI sea una misión dirigida a monitorizar la vegetación, hace imprescindible que las reflectividades obtenidas del agua tengan un nivel de precisión aceptable para derivar formas y magnitudes asociadas a la concentración de sus diversos componentes. De esta manera, se presentan nuevas oportunidades, ya que analizando las formas y magnitudes de las reflectividades del agua es posible identificar patrones en las formas espectrales que pueden derivar en categorizaciones. En la presente tesis doctoral se derivaron mediante métodos empíricos parámetros biofísicos claves para estimar la calidad del agua, se analizó el nivel de precisión de las reflectividades del agua derivadas de un conjunto de procesadores de corrección atmosférica para aguas continentales frente a radiometría medida in situ, se establecieron categorizaciones basadas en las formas espectrales de las reflectividades del procesador que obtuvo el mejor rendimiento, y finalmente se aplicó una clasificación supervisada basada en algoritmos de aprendizaje automático para tener un previo conocimiento de la dinámica de las aguas y así optimizar la asignación de algoritmos según la forma

espectral de las reflectividades de las aguas del área de estudio.

En el primer estudio (Pereira-Sandoval et al., 2019a) incluido en esta tesis, mediante imágenes S2-MSI, se estimaron dos parámetros biofísicos asociados con la evaluación de la calidad de las aguas: la detección de los pigmentos de clorofila-a (Chl-a) y la transparencia del agua mediante la medida de la profundidad del disco de Secchi (Z_{SD}). Se analizaron, calibraron y validaron diversos algoritmos para un amplio número de embalses y lagunas, con diversas características de la Comunidad Valenciana. Se utilizaron datos de Chl-a y Z_{SD} medidas *in situ* obtenidas en el proyecto *Ecological Status of Aquatic Systems with Sentinel* (ESAQS - Programa Prometeo 2016/032); datos simulados obtenidos a partir del modelo de transferencia radiativa Hydrolight; y las bandas espectrales de S2-MSI. Para la calibración de la Chl-a, los datos simulados se dividieron según el nivel de Chl-a y se trabajó con dos algoritmos. El algoritmo de triple banda de Dall'Olmo (*Dall'Olmo three band* - TBDO), se validó en aguas eutróficas-hipertróficas, con Chl-a >10 mg/m^3, como las de la laguna de la Albufera de Valencia y los embalses de Bellús y Beniarrés, obteniéndose un R^2=0.8 y un MAE=23 mg/m^3. En cambio, en aguas con Chl-a <10 mg/m^3, los algoritmos *Ocean Color* (OC) proporcionaron mejores resultados, particularmente el OC2_490, alcanzando en el proceso de validación un R^2=0.59 y un MAE=0.89 mg/m^3. En el caso de la Z_{SD}, el algoritmo R_{490}/R_{705} se calibró con la base de datos de la Z_{SD} medida *in situ*. El 75% de los datos se utilizó para calibrar y el 25% para validar el algoritmo, obteniéndose un MAE=0.88 m. Por último, los dos algoritmos validados, TBDO y R_{490}/R_{705}, se aplicaron con óptimos resultados, de acuerdo a las medidas *in situ* de referencia, sobre una imágen S2-MSI de la Albufera de Valencia del día 07/03/2018, corregida atmosféricamente con el algoritmo oficial de la ESA Sen2Cor. De este estudio se concluye que para la estimación de la Chl-a es necesario aplicar algoritmos según el nivel de Chl-a de los embalses y lagunas (Chl-a <10 mg/m^3 y Chl-a >10 mg/m^3). En el caso de la estimación de la Z_{SD}, el mejor rendimiento se obtuvo con el algoritmo R_{490}/R_{705}. Los óptimos resultados de los algoritmos sobre una imagen S2-MSI corregida con Sen2Cor, particularmente en aguas hipertróficas, requieren indagar sobre otros métodos de corrección amosférica para el resto de embalses del área de estudio que tengan en cuenta la variabilidad de niveles tróficos existentes.

En el segundo estudio (Pereira-Sandoval et al., 2019b), a partir de una de las conclusiones derivadas del primer artículo de Pereira-Sandoval., et al 2019a, sobre la necesidad de investigar otros métodos de corrección atmosférica alternativos a Sen2Cor para imágenes S2-MSI sobre aguas continentales, se evaluó la precisión de las reflectividades derivadas de las imágenes S2-MSI procesadas con los algoritmos de corrección atmosférica para aguas continentales *Atmospheric correction for OLI 'lite'* (ACOLITE), *Case 2 Regional Coast Colour* (C2RCC), *Case 2 Regional Coast Colour for Complex waters* (C2RCCCX), *Image correction for atmospheric effects* (iCOR), *Polynomial-based algorithm applied to MERIS* (Polymer) y Sen2Cor, tras los óptimos resultados obtenidos en aguas hipertróficas. Los datos utilizados en la evaluación corresponden a las reflectividades medidas *in situ* (reflectividad IS) y las reflectividades obtenidas de cada procesador (reflectividad OBS) de diversos embalses y lagunas de la Comunidad Valenciana. Los sitios en los que se tomaron las medidas de radiometría *in situ* fueron georreferenciados y esas coordenadas geográficas se localizaron sobre las imágenes para delimitar áreas de 3 por 3 píxeles y así obtener macropíxeles. Sobre esos macropíxeles se realizaron diversos controles de calidad, para asegurar la fiabilidad de los píxeles. Finalmente los macropíxeles fueron clasificados en tres tipos según rangos de Chl-a y Z_{SD} medida *in situ*: tipo 1 (ultraoligotrófico a oligotrófico), tipo 2 (mesotrófico a eutrófico) y tipo 3 (hipertrófico). En total, se obtuvieron aproximadamente 50 macropíxeles por procesador, exceptuando C2RCCCX del que solo se obtuvieron 37, por lo que fue excluido. Se realizaron tres evaluaciones principales, en los que se aplicaron diversos análisis estadísticos para estimar la precisión de la reflectividad OBS, de cada uno de los seis procesadores, frente a la reflectividad IS. En la primera evaluación se trabajó con el total de los macropíxeles, siendo Polymer y C2RCC los algoritmos que obtuvieron los mejores resultados con un R^2 entre 0.81 y 0.83 y un MAE entre 0.009 y 0.011. En la segunda evaluación los macropíxeles se agruparon según los tres tipos de agua. Polymer y C2RCC obtuvieron en los tipos 2 y 3 rangos de R^2, entre 0.77 y 0.93, aunque también los mayores errores (MAE entre 0.007 y 0.020) que el obtenido en los otros dos tipos. En cambio en el tipo 1, la correlación fue levemente menor (R^2 entre 0.81 y 0.73), pero con un MAE≈0.007. Finalmente, en la tercera evaluación, los macropíxeles fueron analizados por tipos de aguas y bandas espectrales. En las aguas de tipo 1, la mejor correlación se obtuvo en la región del *red-edge*, con un R^2 entre 0.64 y 0.89 y un MAE<0.003, seguida por la región de visible, con un R^2 entre 0.58 y 0.89, aunque con un error más elevado, MAE<0.02. En las aguas de tipo 2, la mejor correlación se dio en la región del *red-edge*, con un R^2 entre 0.92 y 0.98 y un MAE<0.006, siendo que en la región del visible, la correlación tuvo un rango más

amplio, con un R^2 entre 0.61 y 0.99 y un error más elevado, MAE entre 0.005 y 0.0018. Y en las aguas de tipo 3, Polymer y C2RCC, obtuvieron resultados similares, tanto en las correlaciones, con R^2 entre 0.33 y 0.93, como en los errores, MAE entre 0.007 y 0.04. El resultado de estos análisis permitió diferenciar dos grupos: procesadores con rendimientos óptimos C2RCC y Polymer y procesadores con rendimientos menos favorables ACOLITE, iCOR y Sen2Cor. Finalmente se determinó que Polymer y C2RCC eran los procesadores con las correlaciones más elevadas y los menores errores, en todas las evaluaciones de precisión realizadas entre las reflectividades IS y las reflectividades OBS obtenidas de las imágenes S2-MSI de los embalses y lagunas del área de estudio.

Por último, **en el tercer estudio** (Pereira-Sandoval et al., 2022), se optó por dar un pequeño paso atrás, examinando el amplio gradiente de estados tróficos presentado en los embalses y las lagunas monitorizadas en la Comunidad Valenciana. A partir de las reflectividades de S2-MSI corregidas atmosféricamente con el procesador C2RCC, el objetivo fue aplicar una clasificación supervisada robusta para determinar *Optical Water Types* (OWTs) sobre un conjunto de embalses y lagunas del este de la Península Ibérica. Para esto, se utilizaron las reflectividades definidas como *Remote Sensing Reflectance* (R_{rs}) y diversos productos derivados del procesamiento de C2RCC sobre imágenes S2-MSI del período 2017 al 2020, como las propiedades ópticas inherentes (*Inherent Optical Properties* - IOPs), los componentes ópticamente activos (OACs) y los coeficientes de atenuación (Kds). Las coordenadas geográficas de las mediciones *in situ* fueron aplicadas sobre las imágenes para extraer el valor del píxel de una ventana de 3 por 3 píxeles con el objetivo de construir una base de datos para el análisis. De esa base, el 80% de los datos fue utilizado para entrenar y el 20% para validar los procesos de clasificación. Los datos de entrenamiento fueron definidos en cuatro categorías de OWTs: clase 1 (clara), clase 2 (moderada), clase 3 (turbia) y clase 4 (muy turbia) y sobre esos datos se aplicaron seis clasificadores de aprendizaje supervisado: 1) *K-nearest neighbor* (KNN), 2) árboles de decisión o *decision trees* (DTC), 3) bosques aleatorios (*random forest* (RFC), 4) *linear support vector machine* (linSVC), 5) *radial basis function SVC* (rbfSCV) y 6) *polynomial SVC* (polySVC). La validación del rendimiento de los clasificadores se realizó teniendo en cuenta dos situaciones: por un lado considerando únicamente las R_{rs} y por otro, las R_{rs} más las IOPs, OACs y Kds. En ambos casos, los resultados fueron muy similares, ya que se alcanzaron precisiones generales (OA) elevadas. En las las R_{rs}, los valores fueron entre 0.91 a 0.96, y en las R_{rs} más las IOPs, OACs y Kds, los valores furon entre 0.93 a 0.98. Los valores de Kappa fueron casi perfectos: en las R_{rs} estuvo entre

0.81 a 0.94 y en las R_{rs} más los IOPs, OACs y Kds, entre 0.86 a 0.96. Dado que todos los clasificadores tuvieron rendimientos semejantes, estos fueron aplicados en diferentes lagunas, embalses y situaciones. En el embalse de Contreras, los algoritmos RFC, DTC y SVCs tuvieron prácticamente el mismo rendimiento. Sin embargo, un examen más exhaustivo por áreas y añadiendo los valores de la Chl-a y *Total Suspended Matter* (TSM) derivadas de C2RCC, determinó que las clases derivadas por RFC fueron más próximas a los valores de los parámetros incorporados al análisis. De esto se determinó que era preferible trabajar con el algoritmo RFC, ya que además es un clasificador que no requiere procesamientos previos de los datos ni una exigencia computacional compleja. En un segundo análisis, RFC se aplicó sobre una serie temporal de los embalses de Sotonera, Contreras, Mediano y la Albufera de Valencia. Sotonera, normalmente es clasificado como mesotrófico, con una elevada presencia de sedimentos, RFC clasificó al embalse según zonas: clase 3 al norte y 1 al sur, variando según momentos y estados. Contreras, tiene un predominio de aguas claras y es clasificado como oligotrófico. A lo largo de la serie temporal, RFC lo clasifica como de clase 1 y 2. Mediano, tiene un comportamiento similar a Contreras, aunque en algunos momentos con cierta presencia de sedimentos, evidenciado en el cambio de color del agua. En la serie temporal, durante agosto y noviembre, pudo observarse un claro predominio de la clase 3, debido a una elevada concentración de TSM. Y la Albufera de Valencia, clasificada históricamente como hipertrófica, RFC ratifica esa clasificación en dos momentos, al clasificarla como clase 3 y 4. De este estudio se determinó que los algoritmos de clasificación supervisada basados en OWTs presentan óptimos resultados en diferentes embalses y lagunas del este de la Península Ibérica. De todos los clasificadores se optó trabajar con RFC por no requerir preprocesamiento de los datos y por tener un requerimiento computacional menos complejo en su ejecución. El análisis temporal sobre las imágenes S2-MSI, corroboró la validación en diferentes embalses y ante diversas situaciones. Obtener una clasificación basada en OWTs permite determinar cual es el tipo de agua dominate en cada momento para posteriormente realizar una asignación correcta del algoritmo a aplicar para la estimación de la calidad de las aguas. Contar con una clasificación basada en OWTs es imprescindible cuando se monitorizan aguas continentales con una amplia variedad de niveles tróficos y de composiciones complejas como los embalses y lagunas del área de estudio ya que se optimizan tiempos y recursos.

Los estudios presentados en esta tesis doctoral han demostrado que si bien S2-MSI es una misión que ha sido diseñada para monitorizar principalmente la ve-

getación, con el correcto tratamiento de sus imágenes, fue posible derivar con gran precisión las reflectividades del agua. Esto permitió identificar patrones en la dinámica de numerosas aguas continentales y detectar mediante análisis temporales eventos que alteraban esos patrones. Por otra parte, los embalses y las lagunas analizados en esta tesis doctoral presentaban dimensiones ampliamente variables, pero que gracias a la resolución espacial de S2-MSI han podido ser incluídos en los estudios. Algunas carencias de S2-MSI para el estudio de las aguas continentales fueron la resolución espectral y el ancho de bandas ya que limitó poder realizar análisis que requieren de mayor precisión como el análisis de tipos de fitoplancton y la fluorescencia de la clorofila. Se espera que los resultados de esta tesis doctoral contribuyan a potenciar el uso de S2-MSI como herramienta de evaluación de la calidad de las aguas en el contexto de la Directiva Marco del Agua Europea, así como continuar aportando conocimiento y experiencia en la evaluación del rendimiento las imágenes satelitales sobre las aguas continentales en misiones hiperespectrales.

Resumen extenso

1 Contexto

1.1 Propiedades ópticas

En la actualidad, la monitorización del estado ecológico de las aguas continentales mediante sensores remotos, permite conocer de manera más precisa la dinámica espacio-temporal de los sistemas acuáticos, así como evaluar su calidad con una periodicidad y un nivel de detalle hasta el momento sin precedentes. La Directiva Marco del Agua de la Unión Europea (DMA) [1] vigente desde el año 2000, establece un marco jurídico para proteger, regenerar el agua y garantizar su uso sostenible a largo plazo. Entre sus objetivos están el prevenir el deterioro y alcanzar un buen estado ecológico y químico de todas las superficies de agua. De esta forma, la observación mediante sensores remotos se convierte en una herramienta muy potente, a partir de la cual es posible derivar métricas que permitan cuantificar procesos bio-geoquímicos y físicos que se producen en los sistemas acuáticos, y a la vez una herramienta necesaria para avanzar hacia una gestión integral y sostenible de los ecosistemas acuáticos [2].

Entre las aguas continentales se encuentran las lagunas, embalses y las aguas costeras. A diferencia de las aguas oceánicas, las aguas continentales son mucho más complejas en composición y en dinámica. Las aguas continentales presentan particularidades que son influenciadas tanto por factores externos (luz solar, viento) como por factores internos (flora acuática, elementos biológicos, físico-químicos). En relación a la composición, las materias disueltas y particuladas (de origen orgánico e inorgánico), viva y no viva [3], son ópticamente significativas y altamente variables en tipo y concentración, según el momento y el lugar. En términos generales, esas propiedades ópticas se dividen en dos grupos: 1) Las

Propiedades Ópticas Inherentes (IOPs): cuando la luz penetra en el agua, puede ser absorbida o dispersada, por el agua en sí y por los diferentes componentes ópticamente activos, como el fitoplancton, partículas no algales, detritus, materia orgánica e inorgánica disuelta coloreada [3]. Las propiedades de absorción y dispersión de la luz en el medio acuático se especifican en términos de coeficientes de absorción y dispersión y de atenuación. Esas propiedades son las que Preisendorfer denominó como IOPs [4], siendo sus magnitudes únicamente dependientes de las materias que componen el medio acuático [5]. Cuando las IOPs varían, indican cambios en la dinámica del agua o en sus componentes ópticamente activos (OACs). 2) Las Propiedades Ópticas Aparentes (AOPs)[3][4] dependen de las IOPs, de la estructura geométrica del campo de luz ambiente y del ángulo de observación. Es importante que las AOPs presenten características suficientemente regulares y estables como para ser útiles descriptores del medio acuático. Entre las AOPs está la *Remote Sensing Reflectance* (R_{rs}) así como otras medidas radiométricas de radiancia que describen el campo de luz en el agua [3][6]. En la Figura 1 se muestran diferentes ejemplos de color del agua como resultado de la interacción de la luz solar incidente sobre la superficie acuática (absorbida y/o dispersada) y los OACs presentes en embalses y lagunas de la Comunidad Valenciana.

Figura 1: Color del agua de diferentes embalses de la Comunidad Valenciana, España.

1.2 Parámetros biofísicos

La DMA requiere identificar elementos de calidad biológica y, establecer parámetros y métricas que permitan evaluar el estado ecológico de las masas de agua. La composición, abundancia y biomasa del fitoplancton es uno de los elementos fundamentales utilizados para medir la calidad biológica y el estado trófico de los sistemas acuáticos [1]. La *Organisation for Economic Co-Operation and Development* (OECD) [7] estableció rangos de clasificación trófica de los sistemas acuáticos basados en parámetros biofísicos como la media anual del fósforo total, la media anual eufótica de clorofila-a, el valor máximo anual de clorofila-a y la profundidad media anual de visión del disco de Secchi. La concentración del pigmento de la clorofila-a (Chl-a) (mg/m^3) es una aproximación para estimar la biomasa del fitoplancton y del estado trófico de las masas de agua [8].

La profundidad del disco de Secchi (Z_{SD},m) es una medida visual que determina la claridad del agua. La profundidad de desaparición del disco es inversamente proporcional a la cantidad promedio de materia orgánica e inorgánica a lo largo del campo visual [4] [9]. Z_{SD} es en gran medida una función de la atenuación del haz de luz que depende de las IOPs siendo muy sensible a la dispersión de la luz de las partículas. Las células de las algas absorben y dispersan la luz, por lo que influyen en la visión de la profundidad de Secchi [10]. Oglesby y Schaffer [11] [10], determinaron la existencia de correlaciones entre Chl-a y Z_{SD}, sobre todo cuando la atenuación de la luz no provocada por las algas es baja o cuando covaría la Chl-a, demostraron que en el Lago Constance las lecturas de Secchi podrían utilizarse para caracterizar los estados tróficos.

1.3 Nuevos sensores remotos

Desde mediados del siglo XX, diversas misiones espaciales han incorporado diferentes tipos de sensores para el estudio de las aguas oceánicas como *OrbView2 - Sea viewing Wide Field of view Sensor* (SeaWiFS) [12], *Suommi-NPP, NOAA-20; Visible Infrared Imaging Radiometer Suite* (VIIRS) [13], *Aqua* (EOS PM) - con *Moderate Resolution Imaging Spectroradiometer* (MODIS) [14], *Environmental Satellite* (Envisat) - *Medium Resolution Imaging Spectrometer* (MERIS) [15], pero la baja resolución espacial, el tiempo de revisita, el ancho de bandas y el rango espectral, dificultaban aplicar esas imágenes satelitales en la monitorización de las aguas continentales. Desde 1972 este panorama cambió con el lanzamiento del programa Landsat de la *National Aeronautics and Space Administration - United*

States Geological Survey (NASA - USGS) y sus diferentes misiones: Landsat *MultiSpectral Scanner* (MSS) (1972) [16], Landsat-5 *Thematic Mapper* (TM) (1984) [17], Landsat-8 *Operational Land Imager* (OLI) (2013) [18] y Landsat-9 *Operational Land Imager-2* (OLI-2) [19], en los que se incrementó el período de revisita, la resolucion espacial y la resolución radiométrica.

En 2015, la Comisión Europea y la Agencia Espacial Europea (ESA) ponen en marcha el programa de observación de la Tierra Copernicus que cuenta con diversas misiones satelitales. Sentinel es una familia de satélites destinados a la observación de la Tierra. Sentinel-2 *MultiSpectral Instrument* (S2-MSI) es la misión destinada a monitorizar la superficie terrestre. La cámara multiespectral (MSI) de alta resolución se basa en las ventajas de las misiones SPOT y Landsat, siendo la primera en su clase en incorporar tres bandas espectrales en la región *red-edge* del espectro electromagnético (680 a 750 nm). S2-MSI es una constelación de dos satélites: Sentinel-2A y Sentinel-2B. Gracias a su óptima resolución espacial (10, 20 y 60 m), bandas espectrales en regiones de interés (443 a 2190 nm) y un período de revisita cada 5 días [20], ha permitido que S2-MSI sea ampliamente utilizado en diferentes programas europeos para la monitorización de las aguas continentales de tamaños pequeños a medianos como Lake Water Quality - Copernicus Global Land Service [21], Global Lakes Sentinel Services (Glass) [22], Diversity II [23] y Cyanoalert [24].

Copernicus Open Access Hub [25] provee imágenes S2-MSI de nivel L2A (bottom of atmosphere - BOA) listas para ser usadas por el usuario. Sin embargo, el producto L2A es corregido de la contribución del efecto atmosférico con el algoritmo oficial de la ESA *Sentinel 2 Correction* (Sen2Cor) [26]. El método de corrección de Sen2Cor se basa en el enfoque *dark dense vegetation* considerando una superficie lambertiana, mientras que la interfaz aire-agua tiene un reflejo especular [27][28][29] por lo que su aplicabilidad en las aguas continentales debe ser cuidadosamente evaluada.

Por otra parte, y de más reciente desarrollo, están los sensores hiperespectrales. Programas como *International Spatial Station/Hyperspectral Imager for Coastal Ocean* (ISS/HICO) (2009-2014) [30], *Earth Observing 1-Hyperion* (EO-1 Hyperion) (2000-2017) [31] fueron los primeros en incorporar espectrómetros hiperespectrales para monitorizar las aguas costeras. Actualmente programas como *International Spatial Station/DLR Earth Sensing Imaging Spectrometer Mission* (ISS/DESIS) [32] y *PRecursore IperSpettrale della Missione Applicativa* (PRISMA) [33] propor-

cionan imágenes para la monitorización de las aguas costeras y en menor medida, para las aguas continentales [34]. Además existen otros programas próximos a ponerse en marcha como *NASA Plankton, Aerosol, Cloud, ocean Ecosystem* (PACE) [35] también destinados a la monitorización de los sistemas acuáticos.

1.4 Efecto atmosférico

Un instrumento que observa la superficie del agua desde un satélite mide las radiancias ascendentes, que incluye la contribución desde la atmósfera, la superficie del agua y la columna de agua. La contribución atmosférica proviene de la radiancia solar que es dispersada una o más veces por los gases atmosféricos y los aerosoles en dirección al sensor y que supone cerca del 90% de la señal captada [36]. La radiancia que es reflejada por la superficie (brillo del sol y del cielo) es radiancia descendente que es reflejada de vuelta al sensor por la superficie del agua. La radiancia saliente del agua, el 10% restante, corresponde a la señal proveniente del agua en sí, de la interacción de la luz solar con los diferentes OACs, de la interacción de los OACs entre sí mediante procesos de absorción y dispersión (backscattering), especialmente los pigmentos de fitoplancton, sedimentos inorgánicos suspendidos y materia orgánica disuelta [37](Ver Figura 2). Sólo la radiancia saliente del agua aporta información acerca de la columna de agua, y cuando la masa de agua es poco profunda, también del fondo del mar o lago observado.

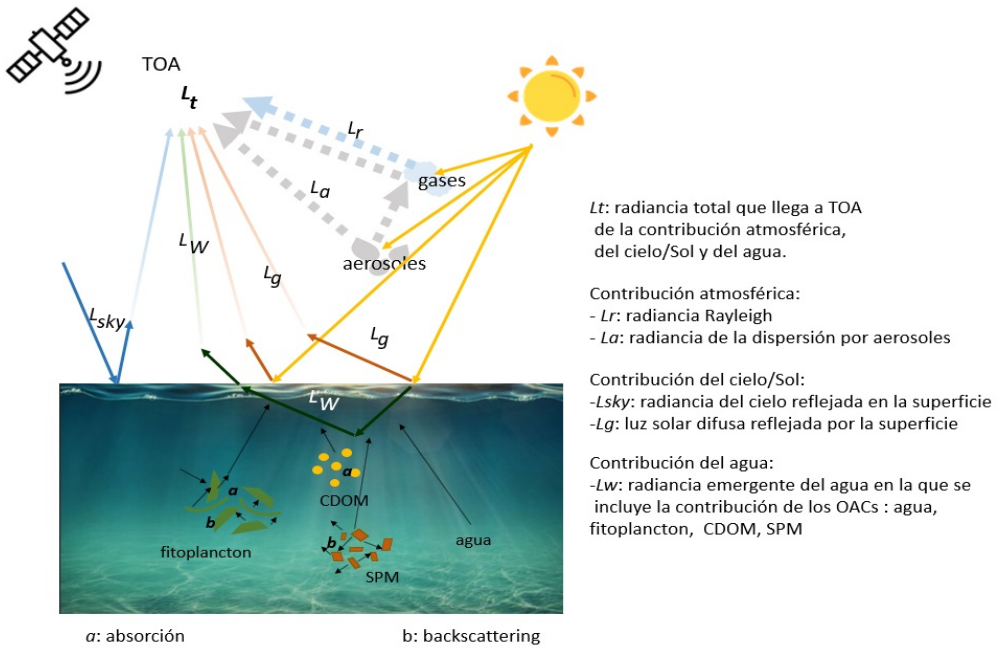


Figura 2: Contribuciones a la radiancia TOA. **Fuente:** Elaboración propia basada en: *Atmospheric correction for satellite Ocean Color radiometry*. Mobley, C.D. 2016. *NASA*.

La corrección atmosférica se refiere al proceso de eliminación de las diferentes contribuciones por parte de la dispersión atmosférica y la radiancia reflejada en la superfice de la medida total, con objeto de obtener la radiancia saliente del agua. Como puede observarse en la Figura 3, a sólo más de 30 km de altitud la radiancia aportada por la atmósfera al sensor supone ya entre el 90-95% del total.

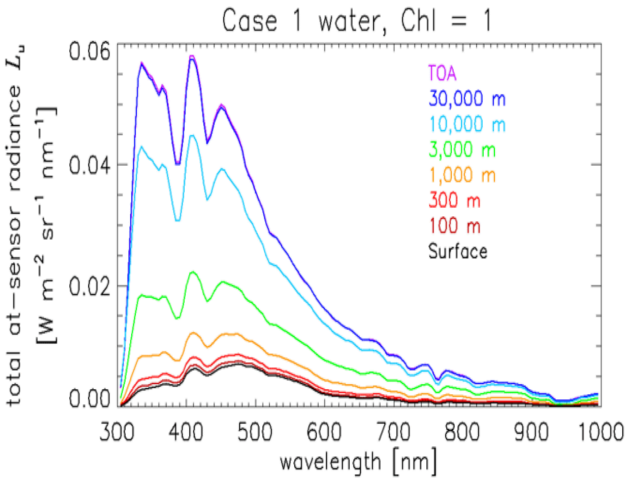


Figura 3: Radiancia TOA captada según la altitud del sensor en aguas oceánicas abiertas (Caso 1). **Fuente:** *Ocean Optics Web Book.*

El problema de la corrección atmosférica se vuelve incluso más intimidante cuando se consideran otros efectos sobre la señal como las dirección de visión del sol y del sensor, las condiciones atmosféricas y el estado de las olas en la superficie. Si consideramos aguas donde la clorofila es el único contribuyente a la radiación saliente, y además aparece en cantidades bajas (< 1 mgm^{-3}), por ejemplo en mar abierto y profundo, esta contribución atmosférica es más sencilla de corregir, ya que se asume que la reflectividad del agua en el infrarrojo cercano (*Near InfraRed*, NIR) es muy cercana a cero. En la Figura 4 se observa la fracción de contribución de los diferentes procesos en la radiancia *Top Of Atmosphere* (TOA) para un caso específico con condiciones estables de atmósfera clara, poco viento y un ángulo de visión del sensor de 30°. La radiancia saliente es de apenas el 10%, siendo la dispersión por Rayleigh la más abundante por debajo de los 500 nm. Entre los 500-1350 nm los aerosoles son los principales contribuidores. A partir de los 1350 nm la atmósfera aparece prácticamente opaca.

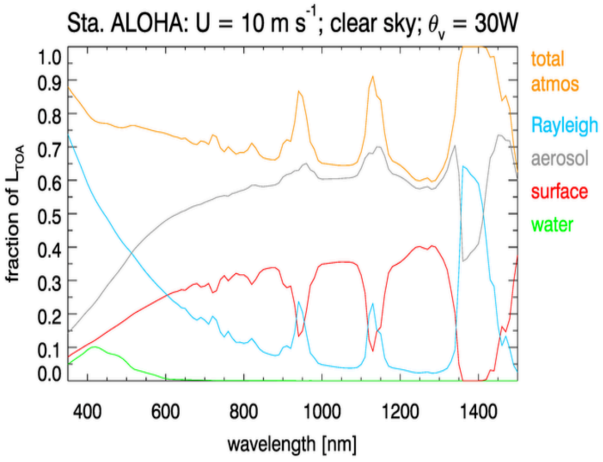


Figura 4: Fracción de radiancia que llega al sensor en aguas de Caso 1 bajo condiciones atmosféricas estables y viento por debajo de 10 ms^{-1}) . **Fuente:** *Ocean Optics Web Book.*

Sin embargo, sobre las aguas continentales la reflectividad en el NIR es mayor a cero [38] y los esquemas basados en casos de aguas de Caso 1 tienden a fallar. Un esquema del proceso detallado de corrección atmosférica es el mostrado en la Figura 5. La forma de solucionar cada uno de estos problemas depende en gran medida de los datos que se tengan, de las características de la masa de agua y del propio sensor, como se explica convenientemente en el artículo 2 que conforma el presente trabajo.

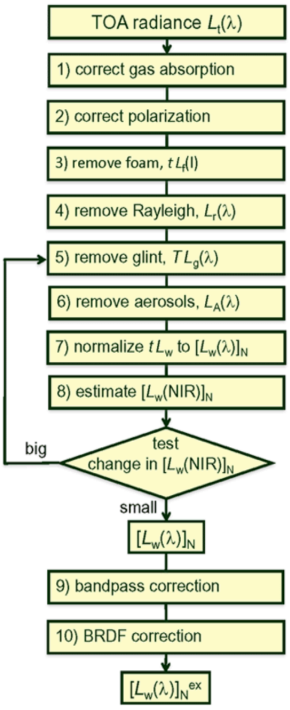


Figura 5: Esquema del proceso de corrección atmosférica sobre las imágenes de satélite
Fuente: *Ocean Optics Web Book.*

1.5 Metodologías de estimación de parámetros de calidad del agua

La reflectividad de la superficie del agua (Rrs, sr^{-1}) es la AOP a partir de la cual se derivan los OACs e IOPs [39]. Cuando la luz solar atraviesa el agua, los OACs interaccionan con esta, alterando su composición y estructura original. Cada OAC tiene un comportamiento característico que se refleja en regiones particulares del espectro electromagnético. Por ejemplo, la clorofila tiene dos picos característicos de absorción en los 440 nm y los 675 nm, siendo esa información utilizada para desarrollar algoritmos de concentración de clorofila (Chl-a). Por lo tanto, conocer las regiones del espectro electromagnético en la que cada OAC presenta su máxima influencia, permite estimar parámetros biofísicos a través de imágenes satelitales.

En este proceso, tradicionalmente se ha trabajado con métodos empíricos [40][41][42]. En estos, mediante el análisis estadístico de simples relaciones de bandas espectrales, se derivan los parámetros de interés. Este método tiene la ventaja de ser muy simple, aunque limitado a momentos y lugares concretos, requiriendo necesariamente del apoyo de datos *in situ* para su correcto ajuste. Por otra parte están los métodos basados en modelos analíticos o semianalíticos, en los que se busca "modelar" la R_{rs} en términos de las IOPs del agua a través de modelos biofísicos construidos con modelos de transferencia radiativa y datos simulados [43][44][40]. Estos métodos suelen ser muy complejos, pero al contar con una amplia variedad de datos y rangos de valores, ofrecen resultados mucho más robustos que pueden ser aplicados de manera mucho más amplia. Por último, y de más reciente desarrollo, están los algoritmos basados en el aprendizaje automático y el *deep learning* y que trabajan sobre un conjunto de datos con características conocidas. A partir de la iteración de métodos estadísticos sobre un conjunto de datos, como el total de las bandas espectrales, estos aprenden de sus características, ajustan y corrigen errores, para finalmente realizar predicciones[45][46][47][48].

2 Aportación de la tesis y organización

La monitorización mediante imágenes S2-MSI de las aguas continentales para evaluar su calidad y determinar su estado ecológico, presenta varios desafíos. Por una parte, la complejidad en la composición de las aguas como consecuencia de la presencia de diversos OACs, variables en el tiempo y en el espacio, dificulta trabajar con una única metodología para la obtención de parámetros biofísicos que determinen su calidad. Además, el hecho de que S2-MSI sea una misión diseñada para monitorizar principalmente la vegetación, hace imprescindible que el escaso porcentaje de reflectividades salientes del agua, llegue al sensor del satélite libre de la enorme contribución de los componentes atmósfericos. Por otra parte, es fundamental que ese limitado porcentaje represente una estimación precisa de la reflectividad saliente del agua, en relación a los diversos componentes, ya que solo así es posible identificar formas y magnitudes espectrales con aceptables niveles de fiabilidad. A partir del análisis de las formas y magnitudes de las reflectividades del agua es posible establecer categorías basadas por ejemplo, en tipos ópticos de agua. Contar con clasificaciones de este tipo, es fundamental cuando se monitorizan embalses y lagunas con diversos estados tróficos ya que permite determinar cual es el tipo de agua dominante en cada momento y lugar, detectar eventos que puedan estar alterando su dinámica y calidad, identificar patrones de comportamiento, así como una realizar una correcta asignación de algoritmos para la estimación de la calidad de las aguas. Esta tesis afronta cada uno de estos desafíos evaluando metodologías existentes basadas en métodos empíricos y de aprendizaje profundo, analizándolas en relación a diferentes fuentes de datos, aplicando diversas evaluaciones estadísticas y generando categorizaciones y clasificaciones.

El esquema de trabajo se resume en la Figura 6 a partir del cual se proyecta alcanzar cada uno de los objetivos específicos:

- Medición de la Chl-a y Z_{SD} y de radiometría *in situ* en diferentes embalses y lagunas del este de la Península Ibérica.
- Análisis de la Chl-a y Z_{SD} medida *in situ* en relación a las bandas espectrales de S2-MSI de nivel L2A para derivar algoritmos que permitan estimar dichos parámetros en imágenes S2-MSI.
- Evaluación estadística de la precisión de las reflectividades derivadas de diversos algoritmos de corrección atmosférica para S2-MSI y aguas continen-

tales clasificadas según niveles tróficos, frente a las radiometrías medidas *in situ* con la subsecuente elección del procesador con mejores resultados estadísticos.

- Categorización manual de tipos ópticos de agua de las reflectividades de un conjunto de imágenes S2-MSI procesadas con el algoritmo de corrección atmosférica que obtuvo los mejores resultados de precisión.
- Aplicación de algoritmos de clasificación supervisada según tipos ópticos de agua, evaluación estadística del rendimiento de los clasificadores y elección del clasificador con los mejores rendimientos en la clasificación de los embalses y lagunas según tipos ópticos de agua.

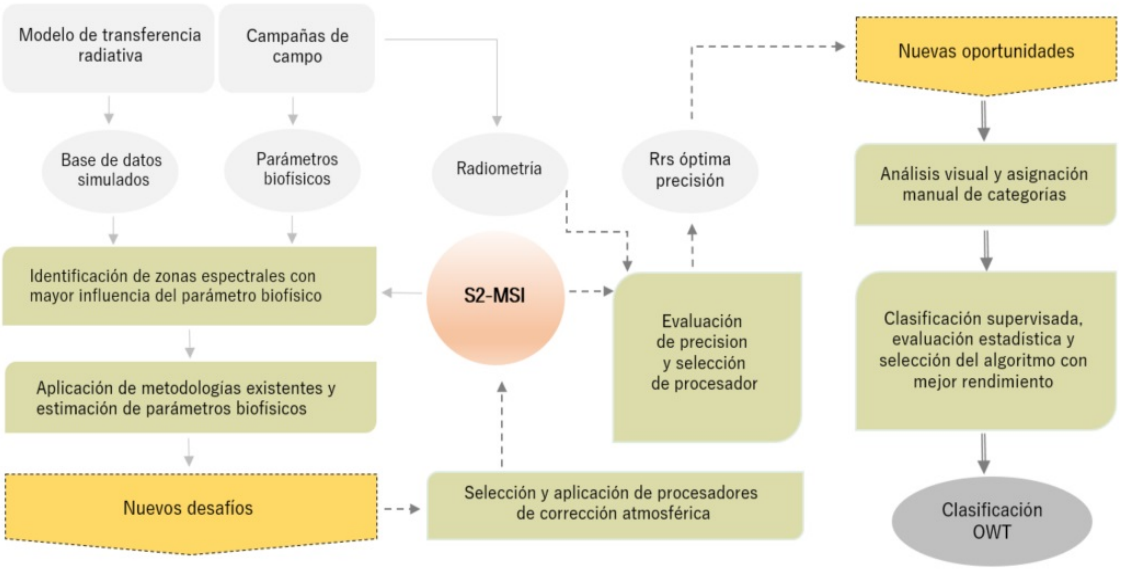


Figura 6: Esquema de trabajo.

3 Resultados

Los resultados de esta tesis se remiten a los tres artículos anexados al final de este documento. El artículo 1 (Pereira-Sandoval et al., 2019a) se basa en la calibración y validación de algoritmos para la estimación de la concentración de clorofila-a y la transparencia del agua mediante la profundidad de Secchi sobre imágenes S2-MSI. En el artículo 2 (Pereira-Sandoval et al., 2019b) se realizan evaluaciones estadísticas de la precisión de la reflectividad derivada de diversos algoritmos de corrección atmosférica para aguas continentales sobre imágenes S2-MSI. En el artículo 3 (Pereira-Sandoval et al., 2022) se aplica una clasificación supervisada a partir de una categorización basada en tipos ópticos de agua de reflectividades S2-MSI de diversas aguas continentales.

En esta sección se describen las principales aportaciones de cada uno de los tres artículos, en donde se pueden consultar las metodologías y los resultados completos en los artículos indicados.

3.1 Artículo 1. Calibration and validation of algorithms for the estimation of chlorophyll-a concentration and Secchi depth in inland waters with Sentinel-2

Basado en los artículos: Pereira-Sandoval et al., 2019a, Pereira-Sandoval et al., 2016 y congresos: Pereira-Sandoval et al., 2018, Pereira-Sandoval et al., 2017, realizados durante el período doctoral:

- **Pereira-Sandoval, M.**, Urrego, P.E., Ruiz-Verdú, A., Tenjo, C., Delegido, J., Soria-Perpinyà, X., Vicente, E., Soria, J. y Moreno, J. Calibration and validation of algorithms for the estimation of chlorophyll-a concentration and Secchi depth in inland waters with Sentinel-2. *Limnetica.* 2019a, 38(1): 471-487. DOI: 10.23818/limn.38.27
- **Pereira-Sandoval, M.**, Soria, X., Urrego, E. P., Vicente, E., Ruiz-Verdú, A., Soria, J., Peña, R., Delegido, J., Tenjo, C., y Moreno, J. Calibration and validation of algorithms for the estimation of chlorophyll-a concentration and Secchi depth in inland waters with Sentinel-2. *International Geoscience and Remote Sensing Symposium.* July 22-27, 2018, Valencia, Spain.

- **Pereira-Sandoval, M.**, Soria, X., Urrego, E. P., Vicente, E., Ruiz-Verdú, A., Soria, J., Peña, R., Delegido, J., Tenjo, C., y Moreno, J. Objetivos y primeros resultados del proyecto ESAQS (Ecological Status of Aquatic Systems with Sentinel Satellites). *XVII Congreso de la Asociación Española de Teledetección.* 3-7 de octubre, 2017. Murcia, España.

- **Pereira-Sandoval, M.**, Ruiz-Verdú, A., Jiménez, J. C., Tenjo, C., Delegido, J., Gibaja, G., Peña, R. y Moreno, J. Ajuste de los algoritmos OC2 y OC3 de MODIS para la obtención de la concentración de la clorofila-a en lagos oligotróficos con Landdsat-8: validación en el lago Titicaca. *Revista de Topografía Azimut.*2016. (7), 45-51.
  http://revistas.udistrital.edu.co/ojs/index.php/azimut

Uno de los objetivos de esta tesis fue estimar, mediante imágenes satelitales S2-MSI), parámetros biofísicos que permitieran evaluar la calidad y determinar el estado trófico de las aguas continentales. En este estudio se trabajó con la concentración de la clorofila-a (Chl-a) y la medición de la transparencia del agua mediante la profundidad del disco de Secchi (Z_{SD}). La Chl-a es uno de los parámetros fundamentales para la detección de la calidad del agua. Es un estimador de la biomasa del fitoplancton que se aplica para detectar blooms algales y evaluar los niveles tróficos de las aguas. Z_{SD} es un parámetro ciertamente asociado al nivel de Chl-a en la columna de agua y también a otras sustancias suspendidas en la misma. La profundidad en la cual el disco desaparece es inversamente proporcional a la cantidad promedio de materiales orgánicos e inorgánicos a lo largo de campo de visión dentro de la columna de agua. Existen diversos algoritmos para estimar la Chl-a y la Z_{SD} mediante imágenes de satélite. En este estudio se analizaron, calibraron y validaron siete algoritmos para una amplia variedad de masas de aguas continentales dentro de la Comunidad Valenciana.

Para la estimación de la Chl-a, los datos usados en el procesos de calibración de los algoritmos corresponden a una base de datos simulados para aguas de Caso 2 de nivel oligotrófico-eutrófico, construida con el modelo de transferencia radiativa HydroLight. En el proceso de validación se usó la base de datos de los parámetros biofísicos medidos *in situ* en las campañas del proyecto *Ecological Status of Aquatic Systems with Sentinel* (ESAQS), con un rango de valores de Chl-a *in situ* entre 0.54 y 169 mg/m^3. Se testearon cuatro algoritmos: Ocean Color 2_443 (OC2_443), Ocean Color 2_490 (OC2_490), Ocean Color 3 (OC3) y el algoritmo de triple banda de Dall'Olmo (aquí llamado TBDO). Estos algoritmos se calibraron

mediante un ajuste polinómico entre la Chl-a de la base de datos simulados [49] y la relación de bandas espectrales de S2-MSI de los cuatro algoritmos. Del ajuste polinómico se derivaron los coeficientes que se aplicaron posteriormente en el proceso de validación. Teniendo en cuenta el óptimo rendimiento previamente obtenido por el algoritmo TBDO en aguas eutróficas e hipertróficas [50], en la validación se clasificaron las masas de agua según nivel de Chl-a *in situ* en dos grupos: Chl-a <10 mg/m^3 y Chl-a >10 mg/m^3. El algoritmo TBDO obtuvo el mejor rendimiento, con un elevado coeficiente de determinación (R^2=0.80) y un error relativo bajo (MAE=23 mg/m^3), en aguas eutróficas-hipertróficas como la Albufera de Valencia y los embalses de Bellús y Beniarrés. En el resto de embalses con Chl-a <10 mg/m^3, si bien el algoritmo OC2_490 obtuvo un coeficiente de correlación similar al de los otros algoritmos (R^2=0.59), fue el que obtuvo el menor error (MAE=0.89 mg/m^3).

Para la estimación de la Z_{SD}, se testearon tres algoritmos basados en relaciones de bandas: R_{490}/R_{560}, R_{490}/R_{705} y R_{560}/R_{705}. En la calibración y validación se utilizó la base de datos de la Z_{SD} medida *in situ* con valores entre 0.25 y 10.5 m. El 75% de los datos se usó para la calibración y el 25% para la validación. Se realizó una primera prueba aplicando la clasificación según niveles tróficos, previamente aplicado en el trabajo con la Chl-a, pero que debido a la limitada cantidad de datos (79) no se obtuvieron buenos resultados. En el proceso de calibración se aplicó un ajuste lineal entre el logaritmo neperiano de la Z_{SD} medida *in situ* y la relación de bandas espectrales de S2-MSI de cada algoritmo, siendo los algoritmos R_{490}/R_{705} y el R_{560}/R_{705} los que obtuvieron la mayor correlación, R^2=0.87 y 0.88, respectivamente. Finalmente, los coeficientes derivados del ajuste lineal se aplicaron a los datos de validación. De estos, el algoritmo con mejor rendimiento fue el R_{490}/R_{705}, con un R^2= 0.68 y con el menor error (MAE=0.88 m.)

Por último, se aplicaron los algoritmos validados, TBDO y R_{490}/R_{705}, sobre una imágen S2-MSI L2A de la Albufera de Valencia (corregida del efecto atmosférico con el algoritmo oficial de la ESA Sen2Cor) del día 07/03/2018, siendo estas aguas de tipo eutróficas-hipertróficas. La aplicación de los dos algoritmos proporcionó óptimos resultados en relación a las numerosas medidas *in situ* registradas en las campañas de campo de ESAQS. De este trabajo se concluye que para la estimación de la Chl-a es necesario aplicar algoritmos acorde a los niveles de Chl-a normalmente presentado en las aguas. Por un lado para aguas eutróficas-hipertróficas y otro para aguas ultraoligotróficas-mesotróficas. Tanto en la estimación de Chl-a como en la transparencia del agua, los algoritmos confirman su validez al ser apli-

cardos sobre imágenes S2-MSI en aguas eutróficas-hipertróficas, proporcionando una precisión suficientemente aceptable para la evaluación del estado ecológico de las aguas continentales. Sin embargo, es necesario investigar métodos de corrección atmosférica alternativos a Sen2Cor ya que los buenos resultados de los algoritmos solo se obtuvieron en aguas con esas características y no en el resto, por lo que esto puede estar asociado a la reflectividada derivada del proceso de corrección atmosférica. Por esta razón, teniendo en cuenta la amplia variedad de niveles tróficos presentes en los embalses del área de estudio es necesario realizar esta investigación.

3.2 Artículo 2. Evaluation of Atmospheric Correction Algorithms over Spanish Inland waters for Sentinel-2 Multispectral Imagery Data

Basado en los artículos: Pereira-Sandoval et al., 2019b y congresos: Pereira-Sandoval et al., 2019, Pereira-Sandoval et al., 2018 y Ruescas et al., 2016 realizados durante el período doctoral:

- **Pereira-Sandoval, M.**, Ruescas, A., Urrego, P.E., Ruiz-Verdú, A., Delegido, J., Tenjo, C., Soria-Perpinyà, X., Vicente, E., Soria, J. y Moreno, J. Evaluation of Atmospheric Correction Algorithms Over Spanish Inland Waters for Sentinel-2 Multispectral Imagery Data. *Remote Sens.* 2019b, 11, 1469. https://doi.org/10.3390/rs11121469

- **Pereira-Sandoval, M.**, Ruescas, A.B., Urrego, P., Ruiz-Verdú, A., Delegido, J., Tenjo, C., Soria-Perpinyà, X., Vicente, E., Soria, J. y Moreno, J. S2-MSI atmosperic correction evaluation over Case 2 Extreme waters: the hypertropic Albufera lagoon. *Living Planet Symposium, European Spatial Agency* 13-17 May, 2019. Milan, Italy.

- **Pereira-Sandoval, M.**, Ruescas, A. B., Urrego, E. P., Delegido, J., Ruiz-Verdú, A., Tenjo, C., Soria-Perpinyà, X., Vicente, E., Soria, J., Peña, R. y Moreno, J. Evaluación de métodos de corrección atmosférica sobre imágenes Sentinel-2 MSI en aguas continentales. *XVIII Simposio Internacional en Percepción Remota y Sistemas de Información Geográfica.* 6-9 de noviembre, 2018. La Habana, Cuba.

- Ruescas, A.B., **Pereira-Sandoval, M.**, Tenjo, C., Ruiz-Verdú, A., Steinmetz, F. and De Keukelaere, L. Sentinel-2 atmospheric correction inter-comparison over two lakes in Spain and Peru-Bolivia. *Colour and Light in the Ocean from Earth Observation, European Spatial Agency*. 6-8 September, 2016. Frascati, Rome, Italy.

Este estudio se centró en una de las conclusiones derivadas del artículo (Pereira-Sandoval et al, 2019a) sobre la necesidad de indagar en métodos de corrección atmosférica alternativos a Sen2Cor, para imágenes S2-MSI, sobre aguas continentales. Esta necesidad surgió del rendimiento particularmente notable de los algoritmos (para la estimación de la Chl-a y Z_{SD}), en aguas eutróficas-hipertróficas, como las de la laguna de la Albufera de Valencia. El resto de embalses del área de estudio presentan una amplia variedad de estados tróficos, por lo que es indispensable evaluar la precisión de las reflectividades derivadas de las imágenes S2-MSI corregidas con Sen2Cor, para comprender el rendimiento de los algoritmos aplicados. Este aspecto también ha sido objeto de análisis de diversos estudios y proyectos europeos [51][52]. La misión S2-MSI fue ideada principalmente para la monitorización de los recursos terrestres, por lo que los procesos asociados al tratamiento de estas imágenes, como la corrección de la contribución atmosférica, siguen esa línea al utilizar el enfoque *dark dense vegetation* y considerar a la superficie terrestre como lambertiana. Mientras que en la interfaz aire-agua, la superficie de las masas de agua tienen un reflejo especular [27][28][29]. En la monitorización de los sistemas acuáticos, de la señal que llega al sensor del satélite, el 90% proviene de la contribución de los gases y aerosoles de la atmósfera y el 10% restante deriva de la interacción de las radiancias de los diversos constituyentes ópticamente activos (OACs) presentes en el agua. La elevada incidencia de los componentes atmosféricos sobre la señal hace indispensable aplicar tratamientos para reducir su contribución y así obtener una correcta señal del agua y sus OACs. Es importante destacar que en las aguas de Caso 1 su contribución a la señal proviene de la región *blue-green* del espectro, mientras que en las aguas del Caso 2 la contribución tiene un origen más variable en el espectro, ya que va depender del OAC dominante.

En este estudio se utilizaron las medidas de radiometría *in situ* (reflectividad IS) y las reflectividades derivadas (reflectividad OBS) de diversos procesadores de corrección atmosférica para S2-MSI sobre aguas continentales. Las medidas de radiometría *in situ* se obtuvieron de radiómetros de amplio espectro (200 a 1100

nm) con una resolución espectral de 0.2 nm. Esas medidas fueron analizadas en laboratorio, convertidas a *Remote sensing reflectance* (R_{rs}) y ajustadas a las bandas espectrales de S2-MSI mediante procesos de convolución, aplicando la *Spectral Response Function* (SRF v2.0) de Sentinel-2. En la selección de los procesadores de corrección atmosférica se consideró la accesibilidad (de uso libre y gratuito), la complejidad de implementación y ejecución, y su aplicación en situaciones similares. Finalmente se seleccionaron seis: *Atmospheric correction for OLI 'lite'* (ACOLITE), *Case 2 Regional Coast Colour* (C2RCC), *Case 2 Regional Coast Colour for Complex waters* (C2RCCCX), *Image correction for atmospheric effects* (iCOR), *Polynomial-based algorithm applied to MERIS* (Polymer). A pesar de que Sen2Cor no es un procesador para aguas continentales, se decidió incorporarlo por los resultados obtenidos en el estudio precedente.

Cada sitio en el que se tomaron las medidas de radiometría *in situ* fue georreferenciado, siendo esas coordenadas geográficas utilizadas para delimitar ventanas de 3 por 3 píxeles (macropíxel) sobre las imágenes. En cada macropíxel se realizó un examen exhaustivo para eliminar píxeles dudosos. De los píxeles válidos, se calculó el promedio, la desviación estándar y el coeficiente de variación que se limitó hasta un 15% para asegurar la homogeneidad del conjunto. Una última evaluación fue aplicada para garantizar la fiabilidad de los datos; se hizo un recuento de la cantidad de píxeles válidos en el macropíxel y si el número era mayor al promedio (por ejemplo, 5 sobre 9 píxeles), el macropíxel se considera representativo para el análisis. Aproximadamente, se obtuvieron 50 macropíxeles por cada procesador, aunque de C2RCCCX solo se obtuvieron 37, lo que condicionó su utilización. Por último, los macropíxeles se clasificaron en tres tipos, según rangos de Chl-a y Z_{SD} medida *in situ*: tipo 1 (ultraoligotrófico a oligotrófico), tipo 2 (mesotrófico a eutrófico) y tipo 3 (hipertrófico).
Para analizar la precisión de la reflectividad derivada de las imágenes S2-MSI corregidas con cada procesador, sobre los embalses y lagunas de nuestra área de estudio, se evaluó la reflectividad OBS frente a la reflectividad IS mediante diversas estadísticas: coeficiente de determinación (R^2), error medio cuadrático (RMSE), bias, error absoluto medio (MAE) y ratio. A partir de los macropíxeles válidos, se realizaron tres análisis. En el primero se evaluaron todos los macropíxeles, sin distinción, y se determinó que era posible diferenciar los procesadores según su rendimiento en óptimos y con rendimientos menos favorables. En el primer grupo se encuentran C2RCC, C2RCCCX y Polymer, al presentar una baja dispersión de los datos, muy alineados a la línea 1:1 (ratio entre 1.13 y 1.71), R^2 entre 0.81 y 0.83 y MAE entre 0.009 y 0.011. En el segundo grupo, algoritmos con

rendimientos menos favorables están ACOLITE, iCOR y Sen2Cor, al presentar una elevada dispersión de los datos sobre la linea 1:1 (ratio entre 0.4 y 0.55), R^2 entre 0.02 y 0.34, y MAE superior, entre 0.022 y 0.038.
En el segundo análisis, se evaluaron los macropíxeles teniendo en cuenta el tipo de agua. Aquí, se mantuvo la diferencia de algoritmos según su rendimiento. Por un lado, con rendimientos óptimos, Polymer y C2RCC para el tipo 1, con R^2 de 0.81 y 0.73, ratios de 0.85 y 1.35, respectivamente, y MAE≈0.007. Para el tipo 2, Polymer y C2RCC tuvieron resultados similares, R^2≈0.93, ratios de 0.8 y 1.86, respectivamente y MAE≈0.007. Y para el tipo 3, Polymer mostró el mejor ajuste (R^2=0.93) y el menor error (MAE=0.015), frente a C2RCC (R^2=0.77), con MAE=0.020. Si bien C2RCCCX también obtuvo buenos resultados, este fue descartado por el escaso número de macropíxeles disponibles. Los procesadores con rendimientos menos favorable fueron ACOLITE, iCOR y Sen2Cor. Sin embargo, ACOLITE y Sen2Cor mostraron una leve mejoría en sus rendimientos en las aguas de tipo 2 y 3, con un R^2 entre 0.40 y 0.55 y MAE entre 0.020 y 0.013.
Por último, en el tercer análisis, los macropíxeles fueron evaluados por tipo de agua y bandas espectrales, manteniéndose la diferencia de algoritmos según rendimientos. En las aguas de tipo 1, entre las bandas 443 y 665 nm, Polymer y C2RCC obtuvieron los mejores rendimientos, con R^2 entre 0.58 y 0.89 (a excepcion de la banda 443 de C2RCC que fue menor) y MAE<0.02 (a excepción de la banda 490 de C2RCC que fue relativamente mayor); en el *red-edge* el R^2 fue también muy similar, entre 0.64 y 0.86 y un MAE<0.003 (menor que en el visible); y en el NIR, en la banda 865, el rendimiento de ambos procesadores también fue similar, con un R^2 muy bajo, entre 0.06 y 0.12. En las aguas de tipo 2, en el rango visible, entre las bandas 443 y 665, Polymer y C2RCC tuvieron similares resultados: R^2 entre 0.61 y 0.99 y MAE entre 0.005 y 0.018; en el *red-edge* el R^2 estuvo entre 0.92 y 0.98 y el MAE<0.006; y en el NIR, en la banda 865, el R^2 entre 0.72 y 0.8. Y finalmente en las aguas de tipo 3, Polymer y C2RCC mantuvieron resultados similares en el espectro visible, con un R^2 entre 0.47 y 0.93, MAE entre 0.009 y 0.032; en el *red-edge* el R^2 estuvo entre 0.33 y 0.89 y el MAE entre 0.007 y 0.04; y en el NIR, en la banda 865 Polymer obtuvo la mejor correlación R^2=0.82. Finalmente, los procesadores con un rendimiento menos favorable continuaron siendo ACOLITE, iCOR y Sen2Cor. Sin embargo, ACOLITE y Sen2Cor mejoraron sus rendimientos en las aguas de tipo 2 y 3, en las bandas del visible, aunque con un R^2 ampliamente variable entre 0.1 y 0.87.

De este trabajo, se concluye que es posible evaluar la precisión de la reflectividad de los embalses y lagunas del área de estudio, mediante analisis estadísticos entre

la reflectividad derivada de los seis procesadores de corrección atmosférica frente a la reflectividad medida *in situ*. De esa evaluación se determinó que Polymer y C2RCC fueron los procesadores que obtuvieron los coeficientes de correlación más elevados y los errores más bajos en todos los análisis. En ambos procesadores, la precisión mejora al aplicar una clasificación trófica por tipos de agua. También fue posible detectar ciertas fortalezas y debilidades en el rendimiento de los procesadores en algunas bandas o regiones espectrales. C2RCC es un algoritmo en constante evolución por lo que se esperan mejoras en su rendimiento, particularmente en las bandas del visible.

3.3 Artículo 3. Supervised Classification of Optical Water Types in Spanish Inland Waters

Basado en los artículos Pereira-Sandoval et al., 2022 y Ruescas et al., 2021, realizados durante el período doctoral:

- **Pereira-Sandoval, M.**, Ruescas, A.B., García-Jimenez, J., Blix, K., Delegido, J. y Moreno, J. Supervised Classifications of Optical Water Types in Spanish Inland Waters. *Remote Sens.* 2022, 14, 55568. https://doi.org/10.3390/rs14215568
- Ruescas, A.B., **Pereira-Sandoval, M.**, Perez-Suay, A. Exploring DESIS for inland water quality in Spanish reservoirs. The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences, Volume XLVI-1/W1-2021. 1st DESIS User Workshop – Imaging Spectrometer Space Mission, Calibration and Validation, Applications, Methods, 28 Sept.–1 Oct. 2021, virtual.

Este estudio se centró en la heterogeneidad presentada en los embalses y lagunas monitorizadas en el proyecto ESAQS. La diversidad de componentes, las características físicas, las condiciones climáticas, son factores que determinan que embalses y lagunas presenten dinámicas particulares, condicionadas a su vez por la localización y el momento en el que se realiza la monitorización. Estos factores determinan que las aguas continentales sean más complejas, en relación a la composición, que las aguas oceánicas. Las aguas continentales pueden ser clasificadas según la riqueza de sus nutrientes, en niveles que van desde oligotróficas a hipertróficas, y modificar esa categoría ante eventos repentinos. Se asume por lo tanto

que las propiedades ópticas, asociadas al nivel trófico, pueden variar en el tiempo tanto de manera natural como intencionada. La clasificación óptica basada en el tipo de agua (*Optical Water Types* - OWTs) analiza la forma y/o magnitud espectral de la reflectividad del agua. Diversos proyectos europeos han desarrollado y validado herramientas de clasificación basadas en esta metodología. El software *SeNtinel Application Platform* (SNAP) de la Agencia Espacial Europea incorpora esta clasificación mediante la herramienta *OWTs classification*, pero que solo puede ser ejecutada en imágenes Sentinel-3 OLCI y ENVISAT MERIS. En este trabajo se aplicó una clasificación supervisada basada en OWTs de un conjunto de embalses y lagunas, a partir de las reflectividades de imágenes S2-MSI corregidas atmosféricamente con el procesador Case 2 Regional Coast Colour (C2RCC).

Se utilizaron los datos de diecisiete imágenes S2-MSI procesadas entre 2017 y 2020 de diversos embalses y lagunas de la Comunidad Valenciana y de las provincias de Huesca y Cuenca. Se trabajó con las bandas espectrales R_{443}, R_{490}, R_{560}, R_{665}, R_{705}, R_{740}, R_{783} y R_{865}, y con diversos productos derivados del proceso de corrección de C2RCC: IOPs relacionadas a la dispersión de la luz, absorción de detritos, absorción total, absorción de pigmentos de clorofila, absorción de materia orgánica disuelta coloreada, coeficientes de atenuación, la profundidad de la columna de agua, del cual proviene el 90% de la irradiancia emergente del agua y las concentraciones de clorofila-a y materia suspendida total. En total se trabajó con 21 tipos de datos, aquí llamados bandas. Los OACs derivados de C2RCC fueron validados previamente [53] sobre algunos embalses y lagunas del Proyecto ESAQS con óptimos resultados al evaluarlos en aguas con clorofila (Chl-a) <10 mg/m^3 y materia suspendida total (TSM) <10 mgL^{-1} frente a las medidas *in situ* de esos parámetros. Por otra parte, de esas medidas *in situ*, así como de otras coordenadas geográficas seleccionadas aleatoriamente, se tomaron las coordenadas geográficas para extraer el valor de los píxeles de una ventana de 3 por 3, con el propósito de construir una base de datos para el análisis. De esa base, una parte se usó para entrenar y otra para validar. Del 80% se analizó la forma espectral de la R_{rs} mediante interpretación visual, gracias al conocimiento y la experiencia adquirida previamente en el análisis de la radiometría medida *in situ*. Los datos de entrenamiento fueron etiquetados en categorías de acuerdo a diferentes propuestas de clasificación OWTs [54][55][56], dando como resultado cuatro clases: clase 1 (clara), clase 2 (moderada), clase 3 (turbia) y clase 4 (muy turbia).

En el entrenamiento se aplicaron seis clasificadores de aprendizaje supervisado: 1) K-nearest neighbor (KNN), 2) decision trees (DTC), 3) random forest (RFC), 4)

linear support vector machine (linSVC), 5) radial basis function SVC (rbfSCV) y 6) polynomial SVC (polySVC). El análisis de rendimiento de los diferentes clasificadores se realizó considerando dos situaciones: por un lado solo sobre las R_{rs} y por otro, sobre las R_{rs} más los productos derivados de C2RCC (IOPs, OACs y Kds). En ambos casos, el proceso de validación arrojó una precisión general (OA) elevada. En las R_{rs}, la OA fue de 0.91 a 0.96, y en las R_{rs} más los productos derivados de C2RCC, la OA fue de 0.93 a 0.98. Los valores de Kappa fueron casi perfectos, en las R_{rs} fue de 0.81 a 0.94, y en las R_{rs} más los productos derivados de C2RCC fue de 0.86 a 0.96.
Una de las ventajas de la clasficación con RFC y SVC es que es posible extraer estadísticas acerca de la importancia de los diferentes insumos utilizados para clasificar. Esto resulta de vital importancia porque se podría utilizar para seleccionar las variables que están efectivamente ayudando a la clasificación y eliminar las que no aportan nada, pudiendo reducir así la dimensionalidad en los modelos. Esta reducción no es tan importante en el presente caso, pero cuando se trabaja con grandes masas de datos, puede reducir los tiempo de procesamiento. También supone una manera gráfica, visual de conocer que bandas son más importante para la discriminación de las clases buscadas, lo que aporta información a las variables y las relaciones que el modelo en cuestión está teniendo en cuenta, resultando así más explicativo, y dejando atrás el concepto de caja negra tan asociado al aprendizaje máquina.

Con estos resultados, se aplicaron los clasificadores en diferentes embalses y situaciones. En un primer análisis, en el embalse de Contreras, en la imagen del 11/05/19, se aplicaron todos los clasificadores. De acuerdo a los parámetros biofísicos medidos *in situ*, Contreras normalmente es clasificado como ultraoligotrófico-oligotrófico, aunque en ciertos momentos del año se vuelve mesotrófico a eutrófico. RFC, DTC y SVCs tuvieron rendimientos muy similares, mostrando diferencias de clases por áreas. Por esto, el embalse se dividió en cuatro áreas y cada una fue analizada según los valores promedio de Chl-a (2 a 7 mg/m^3) y de TSM (4 a 15 mg/L) derivados del procesamiento de C2RCC. Al analizar estos datos con las categorías presentadas por los clasisficadores, las clases derivadas de RFC fueron las más próximas a estos valores. RFC tiene la ventaja de no requerir de un preprocesamiento de los datos para su ejecución (estandarización o normalización). Si bien es computacionalmente complejo, no lo es tanto como los kernels (SVCs). Por estas razones se procedió a seleccionar este clasificador en las siguientes evaluaciones. En un segundo análisis, RFC se aplicó a análisis temporales de los embalses de Sotonera, Contreras, Mediano y la Albufera de Valencia. Sotonera suele

presentar una elevada presencia de sedimentos, siendo clasificado como mesotrófico. En este embalse, se analizaron cuatro fechas: en la primera, el 03/02/19, el embalse es clasificado casi por completo como clase 3. Tres meses después, el 26/05/19 en el embalse predominan dos clases: al norte la clase 3 y al sur la clase 1. El 15/07/19, la clase 3 se extiende más hacia el sur, coincidiendo con un alto contenido de Chl-a en ese momento. Y en la última fecha, el 26/10/19, tras un verano seco y con un bajo nivel del agua, el embalse se clasifica como clase 3. Contreras, es uno de los embalses más estables, con predominio de aguas claras debido a su mayormente estado oligotrófico (promedio de Chl-a <2.5 mg/m^3), aunque ocasionalmente presenta entrada de sedimentos proveniente de las precipitaciones. La clasificación RFC en cuatro fechas: 09/01/2019, 11/05/2019, 19/08/2019 y 12/11/2019, ratifica esta clasificación de este embalse como de clase 1 y 2. Mediano tiene un comportamiento similar a Contreras, aunque la presencia de sedimentos es más frecuente y evidente sobre todo por el cambio del color del agua. En las cuatro fechas analizadas en este embalse, 11/01/2019, 27/03/2019, 04/08/2019 y 12/11/2019, se observa que en agosto y noviembre hubo un claro predominio de la clase 3 por sobre la 1 y 2. Según los bajos valores de Chl-a, esto se debió a una gran concentración de TSM. Finalmente, la Albufera de Valencia es un caso especial. Desde 1980, el promedio anual de Chl-a es de ≈150 mg/m^3, siendo en numerosas oportunidades la cianobacteria la clase dominante del fitoplancton. En esta laguna se analizaron dos fechas, 11/01/2019 y 25/07/2019. RFC clasifica la imagen de enero en clase 3 y 4, mientras que en la imagen de verano, la clase 4 es prácticamente inexistente con un claro predominio de la clase 3. En verano existe un descenso de la ficocianina motivado por el agotamiento de nutrientes que ocurre pasado el pico de producción de la primavera [57], lo que puede estar provocando el aumento de la clase 3 observado en la imagen de verano.

De este artículo se concluye que es posible aplicar algoritmos de clasificación supervisada basados en OWTs con resultados estadísticos óptimos en lagunas y embalses, con diversas dinámicas y complejidades, del este de la Península Ibérica. Si bien los resultados estadísticos de los diferentes clasificadores fueron muy similares, se optó por trabajar con el algoritmo RFC por no requerir procesamiento previo de los datos y tener un requerimiento computacional menos complejo que el resto de los procesadores. El análisis temporal de la clasificación OWTs de los embalses y las lagunas del área de estudio sobre imágenes S2-MSI en diferentes momentos del año, corroboró los resultados de la validación. Fue de gran valor tener conocimiento previo sobre la dinámica de estas masas de agua, gracias a la constante monitorización *in situ* que se realiza sobre estos embalses. Obtener

una clasificación basada en OWTs permite determinar cual es el tipo de agua dominate en cada momento para posteriormente realizar una asignación correcta del algoritmo a aplicar para la estimación de la calidad de las aguas. Contar con una clasificación basada en OWTs es imprescindible cuando se monitorizan aguas continentales con una amplia variedad de niveles tróficos y composiciones como los presentados en los embalses y lagunas del área de estudio ya que se optimizan tiempos y recursos. La información proporcionada por la clasificación OWTs se puede utilizar como producto final, para detectar diferentes tipos de floraciones de algas o también como producto intermedio para profundizar en el conocimiento de la dinámica del agua.

4 Conclusiones

Las principales conclusiones obtenidas en los diferentes estudios llevados a cabo en esta tesis doctoral sobre los desafíos y oportunidades presentados por S2-MSI en la monitorización de las aguas continentales se resumen a continuación:

1. En las aguas continentales no es posible aplicar un único algoritmo para estimar la Chl-a. Es necesario hacer un análisis previo para diferenciar rangos según niveles de concentración de Chl-a.

   (a) Para aguas con elevado contenido de Chl-a (>10 mg/m^3), el algoritmo de triple banda de Dall'Olmo (TBDO) obtuvo un óptimo rendimiento con un R^2=0.99 en la calibración y un R^2=0.8 en la validación y un error muy bajo (MAE=23 mg/m^3) en relación al rango de Chl-a medida *in situ*.

   (b) Para aguas con bajo contenido de Chl-a (<10 mg/m^3), el algoritmo OC2_490, si bien obtuvo un rendimiento similar al resto (R^2=0.59), fue el algoritmo que obtuvo el menor error (MAE=0.89 mg/m^3).

2. Para la estimación de la Z_{SD}, en aguas con una Z_{SD} medida *in situ* <10 m, el algoritmo R_{490}/R_{705} obtuvo el mejor rendimiento con un R^2=0.68 y un MAE=0.88 m.

3. Del análisis estadísticos realizado entre la reflectividad observada, derivada de los procesadores C2RCC, C2RCCCX, Polymer, ACOLITE, iCOR y Sen2Cor, y las reflectividades provenientes de las medidas de radiometría *in situ* se determina que:

   (a) A nivel general, Polymer y C2RCC son los procesadores que presentan las mejores correlaciones (R^2 entre 0.81 y 0.83) y los menores errores (MAE entre 0.0089 y 0.01).

   (b) La precisión estadística de Polymer y C2RCC se incrementa si previamente se aplica una clasificación basada en el nivel de concentración de Chl-a y Z_{SD}.

   (c) Polymer y C2RCC son los procesadores con los mejores rendimiento siendo sus curvas espectrales relativamente semejantes ya que muestran un mejor rendimiento pues la forma de sus curvas espectrales son las más parecidas a las obtenidas en las radiometrías medidas *in situ*.

4. Obtener reflectividades del agua y de sus componentes con un elevado nivel de precisión y sin la contribución atmosférica permite derivar formas y magnitudes espectrales que pueden ser asociados a los componentes ópticamente activos presentes en el agua mediante clasificaciones basadas en tipos ópticos de agua (*Optical Water Types*, OWTs).

5. La aplicación de algoritmos de clasificación supervisada basada en tipos ópticos ha dado óptimos resultados en embalses con diversas dinámicas y niveles de complejidad.

6. Si bien los diferentes algoritmos de aprendizaje supervisadao utilizados tuvieron un óptimo rendimiento, se seleccionó el RFC por su menor complejidad en el procesamiento de las imágenes.

7. El análisis de series temporales basado en tipos ópticos sobre varios embalses proporcionó óptimos resultados en relación a los parámetros biofísicos medidos *in situ* validados gracias al conocimiento previamente adquirido sobre la dinámica de estos embalses en las numerosas campañas de campo.

8. En resumen, los estudios presentados en esta tesis doctoral demuestran que el uso de imágenes S2-MSI sobre aguas continentales:

   (a) es adecuado para la monitorización del estado ecológico de las aguas continentales con diferentes dinámicas y de dimensiones espaciales ya que sus bandas espectrales permiten derivar parámetros biofísicos claves como la Chl-a y la Z_{SD}, exigidos por la Directiva Marco del Agua de la Unión Europea.

   (b) permite la aplicación de metodologías basadas en clasificaciones supervisadas para obtener tipos ópticos de agua permitiendo evolucionar en el procesamiento de los datos a métodos más complejos.

Bibliografía

[1] European Parliament 2000. https://ec.europa.eu/environment/water/water-framework/index_en.html.

[2] Papathanaopoulou, E., Simis, S. et al. Satellite-assisted monitoring of water quality to support the implementation of the Water Framework Directive. EOMORES white paper. 2019. 28pp. DOI: 10.5281/zenodo.3463051

[3] Mobley, C.D. Light and Water: Radiative Transfer in Natural Waters. 1994. Academic, San Diego.

[4] Preisendorfer, R. W. Application of Radiative Transfer theory to Light Measurements in the Sea. Union of Geodetic Geophysical Institute Monograph. 1961. Vol 10. pp. 11-30.

[5] Kirk, J. T. O. Light and Photosynthesis in Aquatic Ecosystems. 2011. 3rd Edition, Cambridge University Press, Cambridge, England.

[6] Ministerio de Medio Ambiente. Metodología para el establecimiento del estado ecológico según la Directiva Marco del Agua en la Confederación Hidrográfica del Ebro. 2005. https://www.miteco.gob.es/es/agua/publicaciones/Protocolos_muestreo_biologico_con_portada_tcm30-214764.pdf.

[7] Caspers, H. OECD: Eutrophication of Waters. Monitoring, Assessment and Control. -154 pp. Paris: Organisation for Economic Co-Operation and Development 1982 (Publié en français sous le titre» Eutrophication des Eaux. Méthodes de Surveillance, d'Evaluation et de Lutte«). 1984. International Review of Hydrobiology, 69(2): 200-200.

[8] Ruiz-Verdú, A., Simis, S. G. H., De Hoyos, C., Gons, H. J. and Peña-Martínez, R. An evaluation of algorithms for the remote sensing of cyanobacterial

biomass. Remote Sensing of Environment. 2008. Volume 112. Issue 11. Pages 3996-4008. ISSN 0034-4257. https://doi.org/10.1016/j.rse.2007.11.019

[9] Preisendorfer, R. W. Secchi disk science: Visual optics of natural waters. Limnol. Oceanogr. 1986. 31(5). 909-926. https://doi.org/10.1016/j.rse.2007.11.019

[10] Tilzer, M. M. Secchi disk-chlorophyll relationships in a lake with highly variable phytoplankton biomass. Hydrobiologia. 1988. 162. 163–171. https://doi.org/10.1007/BF00014539

[11] Oglesby, R. T. & Schaffer, W. R. The response of Lakes to phosphorus. In K. S. Porter (ed.) J., Nitrogen and phosphorus-food production, Waste and the environment. 1975. Ann Arbor Sci.: 23–57.

[12] SeaWiFS. https://oceancolor.gsfc.nasa.gov/data/seawifs/

[13] VIIRS. https://ladsweb.modaps.eosdis.nasa.gov/missions-and-measurements/viirs/

[14] Aqua. https://oceancolor.gsfc.nasa.gov/data/aqua/

[15] MERIS. https://earth.esa.int/eogateway/instruments/meris

[16] Landsat MSS. https://landsat.gsfc.nasa.gov/multispectral-scanner-system/

[17] Landsat TM. https://www.usgs.gov/landsat-missions/landsat-5

[18] Landsat 8. https://www.usgs.gov/landsat-missions/landsat-8

[19] Landsat 9. https://landsat.gsfc.nasa.gov/satellites/landsat-9

[20] Sentinel-2. https://sentinel.esa.int/web/sentinel/missions/sentinel-2

[21] Copernicus Global Land Service. https://land.copernicus.eu/global/products/lwq

[22] Global Lakes Sentinel Services. https://un-spider.org/es/links-and-resources/gis-rs-software/glass-global-lakes-sentinel-services

[23] Diversity II. http://www.diversity2.info/

[24] CyanoAlert. https://www.cyanoalert.com/

[25] Copernicus Open Access Hub. https://scihub.copernicus.eu/

[26] Sen2Cor. http://step.esa.int/main/third-party-plugins-2/sen2cor/

[27] Kaufman, Y. and Sendra, C. Algorithm for automatic atmospheric corrections to visible and near-IR satellite imagery. Int. J. Remote Sens.1988, 9, 1357-1381.

[28] Ouaidrari, H. and Vermote, E. Operational atmospheric correction of Landsat TM data. Remote Sens. Environ. 1997, 70, 4-15.

[29] Gao, B.C., Montes, M.J. Davis, C.O. and Goetz, A.F. Atmospheric correction algorithms for hyper-spectral remote sensing data of land and ocean. Remote Sens. Environ. 2009, 113, 17-24.

[30] HICO. https://oceancolor.gsfc.nasa.gov/data/hico/

[31] Hyperion. https://www.usgs.gov/centers/eros/science/usgs-eros-archive-earth-observing-one-eo-1-hyperion

[32] DESIS. https://www.dlr.de/eoc/desktopdefault.aspx/tabid-13614

[33] PRISMA. https://https://www.eoportal.org/satellite-missions/prisma-hyperspectral#summary

[34] Exploring DESIS inland water quality in Spanish reservoirs. : https://isprs-archives.copernicus.org/articles/XLVI-1-W1-2021/65/2022/isprs-archives-XLVI-1-W1-2021-65-2022.pdf

[35] PACE. https://pace.oceansciences.org/oci.htm

[36] IOCCG. Atmospheric Correction for Remotely-Sensed Ocean Colour Products. Wang, M. Technical Report. 2010. IOCCG. Dartmouth, Canada.

[37] IOCCG. Remote Sensing of Ocean Colour in Coastal and Other Optically-Complex Waters. Sathyendranath, S.(ed). Reports of the International Ocean-Colour Coordinating Group, No. 3. 2000. IOCCG. Dartmouth. Canada.

[38] Wang, M. Atmospheric correction of Ocean Color RS observations. In Proceedings of the IOCCG Summer Lecture Series. Villefranche-sur-Mer, France, 21 July–2 August 2014. pp. 1–58.

[39] IOCCG. Remote Sensing of Inherent Optical Properties: Fundamentals Test of Algorithms, and Applications. Lee, Z.-P (ed). Reports of the International Ocean-Colour Coordinating Group, No. 5. 2006. IOCCG. Dartmouth. Canada.

[40] Matthhews, M. W. A current review of empirical procedures of remote sensing in inland and near-coastal transitional waters. International Journal of Remote Sensing. 2011, 23-21, 6855-6899.

[41] Delegido, J., Tenjo, C., Ruiz-Verdú, A. y Peña, R. Empirical model for chlorophyll-a determination in inland waters from the forthcoming Sentinel-2 and 3. Validation from HICO images. Revista de teledetección. 2014, 41, 37-47.

[42] O'Reilly, E.O. and Werdell, J. Chlorophyll algorithms for ocean color sensors-OC4, OC5 & OC6. Remote Sens. Environ. 2019, 229, 32-47.

[43] Doerfffer, R. and Schiller, H. MERIS Lake Water Algorithm for BEAM Algorithm Theoretical Basis Document (ATDB). 2008. No. 1.0, pp.1–1 (Geesthacht:GKKSS Forschungszentrum)

[44] Doerfffer, R. and Schiller, H. MERIS Regional Coastal an Lake Case 2 Water Project – Atmospheric Correction Algorithm Theoretical Basis Document (ATDB). 2008. No.1.0, pp1–42 (Geesthacht:GKKSS Forschungszentrum)

[45] Bourel, M., Crisci, C. and Martinez, A. Consensus methods based on machine learning techniques for marine phytoplankton presence-absence prediction. Ecological Informatics. 2017, 42, 46-54. https://doi.org/10.1016/j.ecoinf.2017.09.004

[46] Chou, J-S., Ho, C-C. and Hoang, H-S. Determining quality of water in reservoir using machine learning. Ecological Informatics. 2018, 44, 57-75. https://doi.org/10.1016/j.ecoinf.2018.01.005

[47] Watanabe, F. S. Y., Miyoshi, G. T., Rodrigues, T. W. P., Bernardo, N. M. R., Rotta, L. H. S., Alcântara, E. and Imai, N. N. Inland water's trophic status classification based on machine learning and remote sensing data. Remote Sensing Applications: Society and Environment. 2020, 19, 100326. https://doi.org/10.1016/j.rsase.2020.100326

[48] Grendaite, D. and Stonevicius, E. Machine Learning Algorithms for Biophysical Classification of Lithuanian Lakes Based on Remote Sensing Data. Water. 2022, 14, 11, 1732. https://doi.org/10.3390/w14111732

[49] Ruiz-Verdú, A., Jiménez, J.C., Lazzaro, X., Tenjo, C., Delegido, J., Pereira, M., Sobrino, J. and Moreno, J. Comparison of MODIS and LANDSAT-8 retrievals of clorophyll-a and water temperature over Lake Titicaca. XVIII Iberian Congress of Limnology. IRTA. 2016. July 4-8. Tortosa, Spain.

[50] Guibaja, G., Ruiz-Verdú, A., Romo, S., Soria, J.M., Tenjo, C., Pereira-Sandoval, M., Delegido, J., Peña, R. and Moreno, J. Mapping water quality in the Albufera of Valencia Lake with the new Sentinel-2 Earth Observation satellite. XVIII Iberian Congress of Limnology. IRTA. 2016. July 4-8. Tortosa, Spain.

[51] Doxani, G., Vermote, E., Roger, J.C., Gascon, F., Adriensen, S., Frantz, D., Hagolle, O., Hollstein, A., Kirches, G., Li, F., Louis, J., Mangin, A., Pahlevan, N., Pflug, B., and Vanellemont, Q. Atmospheric Correction Inter-Comparison Exercise. Remote Sens. 2018, 10, 352.

[52] GLaSS Deliverable 3.2 Global Lakes Sentinel Services, D3.2: Harmonized Atmospheric Correction Method. Technical Report. 2014. D3.2: Harmonized atmospheric correction method, EOMAP, WI, BC, CNR, SYKE, TO, VU/VUmc. www.glassproject.eu/downloads

[53] Urrego, E.P., Delegido, J., Tenjo, C., Ruiz-Verdú, A., Soriano-Gonzalez, J., Pereira-Sandoval, M., Sorià-Perpinyà, X., Vicente, E., Soria, J.M. and Moreno, J. Validation of chlorophyll-a and total suspended matter products generated by C2RCC processor using Sentinel-2 and Sentinel-3 satellites in

inland waters. In Proceedings of the XX Congress of the Iberian Association of Limnology. Murcia, Spain, 26-29 October 2020.

[54] Uudeberg, K., Ansko, I., Poru, G., Ansper, A. and Reinart, A. Using optical water types to monitor changes in optically complex inland and coastal waters. Remote Sens. 2019. 11, 2297. https://doi.org/10.3390/rs11192297

[55] Uudeberg, K., Aavaste, A, Köks, K.L., Ansper, A., Uusöe, M., Kangro, K., Ansko, I., Ligi, M., Toming, K. and Reinart, A. Optical Water Type Guided Approach to Estimate Optical Water Quality Parameters. Remote Sens. 2020, 12, 931. https://doi.org/10.3390/rs12060931

[56] Soomets, T., Uudeberg, K., Jakovels, D., Zagars, M., Reinart, A., Brauns, A. and Kutser, T. Comparison of Lake Optical Water Types Derived from Sentinel-2 and Sentinel-3. Remote Sens. 2019, 11, 2883. https://doi.org/10.3390/rs11232883.

[57] Sorià-Perpinyà, X., Vicente, E., Urrego, P., Pereira-Sandoval, M., Ruiz-Verdú, A., Delegido, J., Soria, J. M. and Moreno, J. Remote sensing of cyanobacterial blooms in a hypertrophic lagoon (Albufera of València, Eastern Iberian Península) using multitemporal Sentinel-2 images. Sci. Total. Environ. 2020, 698, 134305. https://doi.org/10.1016/j.scitotenv.2019.134305.

[58] Wu, M., Zhang, W., Wang, X. and Luo, D. Application of MODIS satellite data in monitoring water quality parameters of Chaohu Lake in China. Environmental Monitoring and Assessment. 2009, 148, 255-264.

Lista de publicaciones

Artículo 1

"Calibration and validation of algorithms for the estimation of chlorophyll-a concentration and Secchi depth in inland waters with Sentinel-2"

Este artículo fue publicado en Limnetica en 2019. Esta revista tuvo en 2019 un factor de impacto de 0.918 y la posición de 19/22 (Q4) en la categoría Limnología.

Limnetica, 38(1): 471-487 (2019). DOI: 10.23818/limn.38.27

Calibration and validation of algorithms for the estimation of chlorophyll-*a* concentration and Secchi depth in inland waters with Sentinel-2

Marcela Pereira-Sandoval[1,*], Esther Patricia Urrego[1], Antonio Ruiz-Verdú[1], Carolina Tenjo[1], Jesús Delegido[1], Xavier Soria-Perpinyà[2], Eduardo Vicente[2], Juan Soria[2] and José Moreno[1]

[1] IPL - University of Valencia. Catedrático José Beltrán, 2. 46980 Paterna, Valencia (Spain).
[2] Institut Cavanilles de Biodiversitat i Biología Evolutiva (ICBiBE). Universitat de València. C/ Catedrático José Beltrán, 2. 46980-Paterna, València, (Spain).

[*] Corresponding author: Marcela.Pereira@uv.es

Received: 09/03/18 Accepted: 12/09/18

ABSTRACT

Calibration and validation of algorithms for the estimation of chlorophyll-*a* concentration and Secchi depth in inland waters with Sentinel-2

Chlorophyll-*a* concentration and Secchi disk depth are two of the most important biophysical parameters used to assess water quality and determine the ecological state of inland waters. The *Ocean Color 2* and *Dall'Olmo three-band* algorithms were used to estimate chlorophyll-*a* concentration and the calibration of the ratio 490/705 nm was used to produce an algorithm for estimating Secchi disk depth. These algorithms have been calibrated for the Sentinel 2-Multispectral Instrument (S2-MSI) and validated using *in situ* measurements of chlorophyll-*a*, Secchi disk depth and radiometry. This data was taken in the Valencia region reservoirs as part of the project Ecological Status of Aquatic Systems with Sentinel Satellites (ESAQS). The results show that for estimating chlorophyll-*a* concentration, it is better to apply a prior classification based on their trophic status. For eutrophic and hypertrophic waters, the TBDO algorithm had an error of 23 mg/m^3 over a chlorophyll-*a* concentration range of between 10 to 169 mg/m^3. For ultraoligotrophic to mesotrophic waters, the better algorithm was OC2_490, which resulted in an error equal to 0.9 mg/m^3 over a chlorophyll-*a* concentration range of between 0.54 to 5.8 mg/m^3. For the estimation of water transparency by Secchi disk depth, we have obtained good results with the ratio 490/705 nm, with an error equal to 0.88 m over a Secchi disk depth range of between 0.26 to 8.1 m. These algorithms have been applied to S2-MSI images and satisfactory results have been obtained for different reservoirs in the Valencia region (Spain).

Key words: Ocean Color, Dall'Olmo three-band, chlorophyll-*a*, Secchi disk depth, Sentinel-2, HydroLight

RESUMEN

Calibración y validación de algoritmos para la estimación de la concentración de la clorofila-a y profundidad de Secchi en aguas continentales con Sentinel-2

La concentración de clorofila-a y la profundidad del disco de Secchi son dos de los parámetros biofísicos más importantes utilizados para evaluar la calidad del agua y determinar el estado ecológico en aguas continentales. Los algoritmos Ocean Color 2 y triple banda de Dall'Olmo fueron aplicados para la estimación de la concentración de la clorofila-a. El ratio 490/705 nm se usó para producir un algoritmo para la estimación de la profundidad del disco de Secchi. Esos algoritmos han sido calibrados para el Instrumento Multiespectral de Sentinel 2 (S2-MSI) y validados usando medidas in situ *de la clorofila-a, profundidad del disco de Secchi y radiometría. Estos datos se tomaron en los embalses de la región de Valencia en el contexto del proyecto Estado Ecológico de los Sistemas Acuáticos con Satélites Sentinel (ESAQS). Los resultados muestran que para la estimación de la concentración de la clorofila-a es mejor aplicar previamente una clasificación basada en su estado trófico. Para aguas eutróficas e hipertróficas, el algoritmo triple banda de Dall'Olmo tuvo un error de 23 mg/m^3 sobre un rango de concentración de clorofila-a entre 10 a 169 mg/m^3. Para aguas ultraoligotróficas a mesotróficas el mejor algoritmo fue el*

OC2_490, con el cual se obtuvo un error de 0.9 mg/m^3 sobre un rango de concentración de clorofila-a entre 0.54 a 5.8 mg/m^3. Para la estimación de la transparencia del agua mediante la profundidad del disco de Secchi, hemos obtenido buenos resultados con el ratio 490/705 nm, con un error igual a 0.88 m sobre un rango de profundidad de disco de Secchi entre 0.26 a 8.1 m. Estos algoritmos han sido aplicados a imágenes S2-MSI obteniendo resultados satisfactorios en diferentes embalses en la región de Valencia (España).

Palabras clave: *Ocean Color, triple banda de Dall'Olmo, clorofila-a, profundidad de disco de Secchi, Sentinel-2, HydroLight*

INTRODUCTION

Chlorophyll-*a* concentration [Chl-*a*] is a phytoplankton biomass estimator in water bodies. It is one of the fundamental parameters of water quality used to detect algal blooms and assess eutrophication levels. The anomalous productivity of phytoplankton biomass relative to a "normal situation" (or the ecological optimum, depending on the watershed characteristics and climate of a given water body) is an indicator of eutrophication (Zheng and DiGiacomo, 2017). The [Chl-*a*] derivation from satellite data relies mostly on the absorption signal of phytoplankton. The variability of the [Chl-*a*] specific absorption coefficient is spectrally minimal around the red absorption peak (670 nm), where the influence of accessory pigments is minimal, and is higher around the blue absorption peak (440 nm), where the absorption of accessory pigments and dissolved organic matter is also high (Bricaud *et al.*, 1995; Stramsky *et al.*, 2001, Zheng & DiGiacomo, 2017).

Secchi disk depth (Z_{SD}) is a measurement of water transparency. The depth at which the disk disappears into the water is inversely proportional to the average amount of organic and inorganic materials along the path of sight in the water. The sense of sight is an integral part of the measurement procedure (Preisendorfer, 1986).

The European Water Framework Directive (WFD) establishes a framework for community-wide action for assessing the ecological status, management and conservation of the different water bodies in the member states. In this context, and in view of the need for comprehensive management of the aquatic systems of the Valencia region (south-eastern Spain), the Ecological Status of Aquatic Systems with Sentinel satellites (ESAQS) project was developed. The Regional Water Authority (Confederación Hidrográfica del Júcar, CHJ) collaborates and helps this project. The main objective of ESAQS is to develop and validate algorithms for the estimation of ecological quality indicators of inland waters (i.e. chlorophyll-*a* concentration, Secchi depth, colored dissolved organic matter and suspended solids) using the data provided by the Multispectral Instrument (MSI) sensor on-board the Sentinel-2 (S2-MSI) mission. The MSI scenes have a spatial resolution of 10 m, a high temporal resolution of 5 days and an adequate spectral resolution. Table 1 summarizes the principal features of S2-MSI. These characteristics make this sensor an exceptional instrument for the retrieval of biophysical parameters in inland waters, as compared to other satellite missions designed either for ocean color applications (e.g. the Ocean and Land Colour Imager, OLCI, on Sentinel-3 or the Moderate-Resolution Imaging Spectroradiometer, MODIS-Aqua), or the land-oriented Operational Land Imager (OLI) Landsat-8, which have a lower spatial resolution (300, 1000 and 30 m, respectively). S2-MSI data is suitable for the study of small irregularly-shaped water masses like most reservoirs in the region (Fig. 1) and small area as 80 ha (Regajo reservoir). Remote sensing data adds an extra value to limnological data collected from traditional *in situ* measurements (Giardino *et al.*, 2001), thanks to the synoptic view and frequency of the satellite passes. Remotely-sensed optical imagery over water bodies has been used as a cost-effective way for monitoring water quality.

The aim of this paper is to present the results of the calibration of water quality algorithms for the estimation of [Chl-*a*] and Z_{SD} from S2-MSI spectral bands and the validation with *in situ* [Chl-*a*] and Z_{SD} measurements. These algorithms cover a wide range of trophic states, from hypertrophic to ultraoligotrophic, in line with the characteristics

Table 1. Main characteristics of S2-MSI spectral bands. *Principales características de las bandas espectrales de S2-MSI.*

Band number	Central wavelength (nm)	Bandwidth (nm)	Spatial resolution (m)
1	443	20	60
2	490	65	10
3	560	35	10
4	665	30	10
5	705	15	20
6	740	15	20
7	783	20	20
8	842	115	10
8a	865	20	20
9	945	20	60
10	1380	30	60
11	1610	90	20
12	2190	180	20

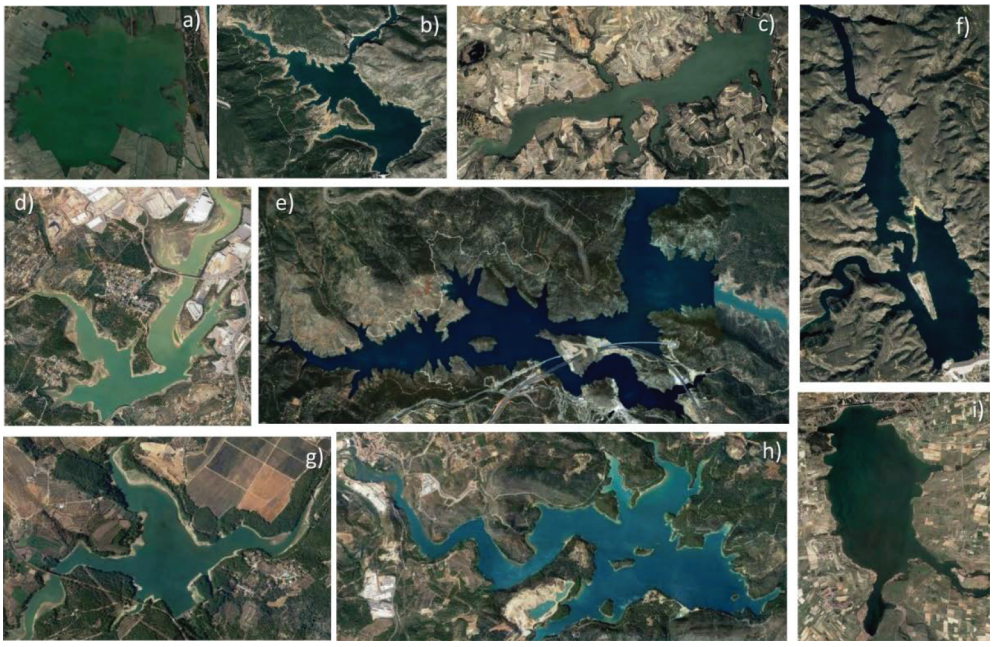


Figure 1. Reservoirs in study area. a) Albufera of Valencia, b) Benagéber, c) Beniarrés, d) María Cristina, e) Contreras, f) Tous, g) Regajo, h) Sitjar and i) Bellús. *Embalses en el área de estudio. a) Albufera de Valencia, b) Benagéber, c) Beniarrés, d) María Cristina, e) Contreras, f) Tous, g) Regajo, h) Sitjar y i) Bellús.*

of reservoirs measured in our study region. This objective fits into the main ESAQS goal of formulating new algorithms for inland water ecological state monitoring from remote sensing.

DATA AND METHODS

Study area

Since 2017, the first year of the ESAQS project, a large number of field campaigns have been carried out in the Valencia region. The study area includes the reservoirs of Benagéber, Bellús, Beniarrés, Contreras, María Cristina, Regajo, Sitjar, Tous and the Albufera of Valencia (Fig. 2), covering a wide gradient of trophic states. For each reservoir, between 1 and 4 *in situ* measuring points were taken at a suitable distance from shoreline to avoid mixed pixels (land-water mixed reflectance). The field campaigns were planned on cloud-free days on which the S2 satellites acquired images over the reservoir region.

In situ data: [Chl-*a*] and Z_{SD}

[Chl-*a*] and Z_{SD} were measured by the Limnology Research Team of the University of Valencia. [Chl-*a*] data were obtained from the water samples by the spectrophotometric method. Samples were filtered through 0.4-0.6 µm GF/F glass

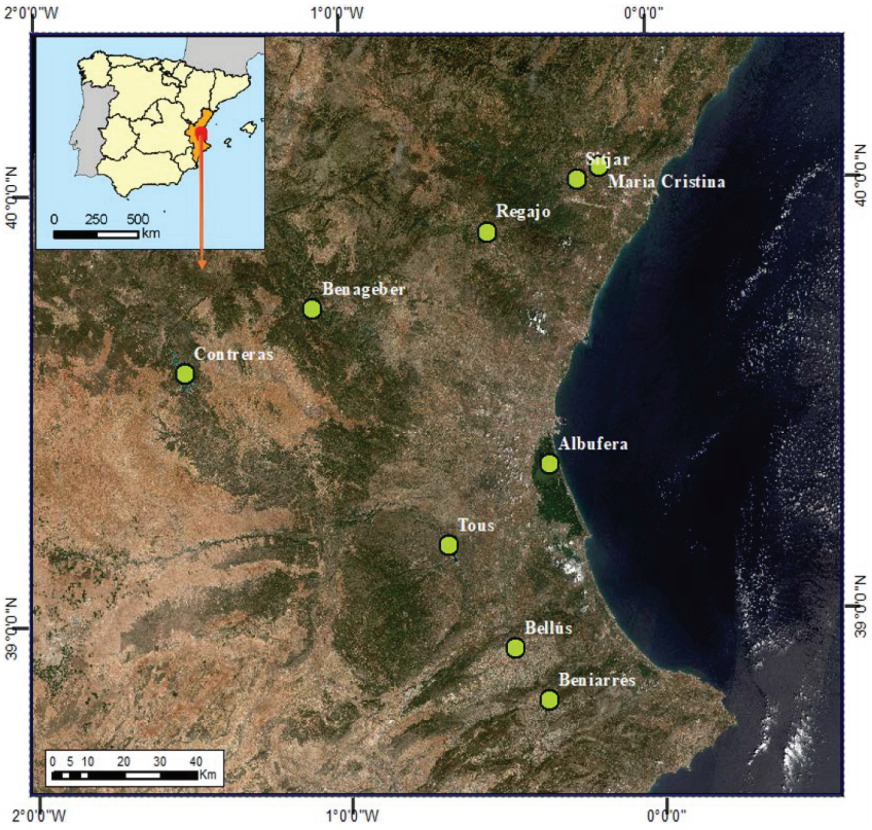


Figure 2. Location map of reservoirs and lakes in study in Valencia region. *Mapa de localización de los embalses y lagos en estudio en la región de Valencia.*

fiber filters, extracted according to standard methods (Shoaf & Lium, 1976) and the calculation methods by Jeffrey & Humphrey (1975) (Soria *et al.*, 2019). The range of *in situ* [Chl-*a*] measurements in the study area reservoirs was between 0.54 to 169 mg/m^3.

Z_{SD} data was measured with a metal disk called a Secchi disk. It is divided into four quarters, each one painted black and white. The measurement procedure for obtaining the Z_{SD} is to slowly lower it into the water until it disappears from sight. When that happens, it is possible to obtain the Z_{SD}. The *in situ* Z_{SD} measurement range in the reservoirs under study was between 0.25 to 10.5 m.

In situ reflectances

The *in situ* reflectances were calculated from above-water *in situ* measurements of: 1) The water-leaving radiance (L_w), 2) the sky radiance (L_{sky}) and 3) the downwelling solar irradiance (E_S). These were measured using a HandHeld2 radiometer, which has a wavelength range of 325 to 1075 nm and a spectral resolution of 1 nm; and an Ocean Optics (HR 4000) radiometer ranging from 200 to 1100 nm at a spectral resolution of ~0.2 nm. According to previous studies (Mobley, 1999) and recommendations done by international remote sensing ocean groups (Fargion & Mueller, 2000), the measurements have been taken with a view zenith angle of ~45 º and an azimuthal angle of ~135 º, with respect to the sun, to minimize the direct and diffuse sunlight reflected by the water surface (Mobley, 1999). Once the measurements were obtained, the *in situ* reflectance spectrum was convolved to the S2-MSI spectral bands (Table 1). This procedure was done to fit the *in situ* reflectance spectral resolution (~0.2 nm or 1 nm) to S2-MSI bandwidth using the Spectral Response Functions database (SRF v2.0) (ESA, 2018a). The SRF determine the position and width of each S2-MSI spectral band (D'Odorico *et al.*, 2013).

S2-MSI imagery

The Cloudless Level 1C S2-MSI images were downloaded from the Copernicus Open Access Hub of the European Spatial Agency (ESA, 2018b) to be concurrent with the *in situ* measurements. SNAP software v.5.0 (ESA, 2018c) was used for the image processing. For the specific case of the Albufera of Valencia, images were atmospherically corrected to Bottom of Atmosphere (BOA) reflectances using the Sen2Cor processor (ESA, 2018d). While this atmospheric correction method was designed for land products, it gives back optimum results in eutrophic waters like the Albufera of Valencia according to Ruescas *et al.* (2016) and Soria *et al.* (2017a). In this respect, we would highlight the importance of atmospheric correction over inland waters. In the ESAQS project, we are evaluating different atmospheric correction methods to obtain an optimum retrieval of the algorithms calibrated and validated to actual S2-MSI images for routine monitoring purposes.

Calibration and validation of established [Chl-*a*] algorithms

To calibrate the [Chl-*a*] algorithms and encompass the variability of the optically active components present in inland waters, we generated a calibration database of simulated data (N = 392) by using the radiative transfer model HydroLight (Mobley, 1994). We carried out simulations with a three-component Case-2 model with ranges of variables characteristics for oligotrophic to eutrophic waters: [Chl-*a*] (1-500 mg/m^3), Colored Dissolved Organic Matter (CDOM) (0.01-4 m^{-1}) and Non-Algal Particles (NAP) (0.1-100 g/m^3). To estimate [Chl-*a*] with S2-MSI spectral bands, we tested the Ocean Color 2_443 nm (OC2_443) and Ocean Color 2_490 nm (OC2_490) (replacing the original 483 band with the 490 band in S2-MSI), the Ocean Color 3 (OC3) (O'Reilly *et al.*, 2000) and the Dall'Olmo three-band algorithm (Dall'Olmo *et al.*, 2003), hereafter called the TBDO algorithm. Following the methodology of Ruiz-Verdú *et al.* (2016), the OC2 and OC3 algorithms, based on the ratio of green and blue bands, have been recalibrated for the S2-MSI spectral bands. The TBDO algorithm, which is based on the evaluation of the reflectance at red and near infrared bands (665, 705 and 740 nm), is used for the estimation of [Chl-*a*] in

inland and coastal water (Gilerson *et al.*, 2010). This algorithm has also been calibrated for the S2-MSI spectral bands and validated in a previous work in a hypertrophic coastal lagoon: the Albufera of Valencia (Guibaja *et al.*, 2016).

The [Chl-*a*] algorithms used in this work are described below. The general equation for the OC2 and OC3 models are:

$$\log_{10}[\text{Chl-}a] = a + bX + cX^2 + dX^3 \quad (1)$$

where X is calculated according to:

OC2_443 model (2)

$$X=\log_{10}[Rrs\_443/Rrs\_560]$$

OC2_490 model (3)

$$X=\log_{10}[Rrs\_490/Rrs\_560]$$

OC3 model (4)

$$X=\log_{10}[\max(Rrs\_443; Rrs\_490)/Rrs\_560]$$

The Rrs_443, Rrs_490 and Rrs_560 are reflectances of S2-MSI spectral bands.

The algorithm called OC2_490, fit to the S2-MSI bands, is set to the spectral band 490.

The three-band model of Dall'Olmo *et al.* (2003) (TBDO), takes the form:

$$[\text{Chl-}a] = aX^2 + bX + c \quad (5)$$

where

$$X=[Rrs\_740*((Rrs\_665)^{-1}-(Rrs\_705)^{-1})] \quad (6)$$

The Rrs_665, Rrs_705 and Rrs_740 are reflectances of S2-MSI spectral bands.

Once the different algorithms were calibrated from the HydroLight database, we validated it with the [Chl-*a*] field data, selecting the best method according to different statistical parameters. If a linear trend is maintained, but away from the 1:1 line, the algorithms were recalibrated.

Calibration and validation of Z_{SD} algorithms

The attenuation of light into water bodies and the vertical visibility (water transparency) is described by the Secchi disk depth (Z_{SD}). It is linked to two optical parameters: the vertical diffuse attenuation coefficient K_d (m^{-1}) and the attenuation coefficient c (m^{-1}) (Antoine, 2010). The vertical visibility is analogous to Z_{SD} and it is the inverse of the sum K_d + c (Doron *et al.*, 2007). Austin and Petzold, (1981) estimated K_d (490 nm) from blue-green water leaving radiances. This algorithm was modified by Kratzer *et al.*, (2008) to estimate Z_{SD} for remote sensing applications (Alikas & Kratzer, 2017). The ratio 490/560 is used to map transparency over clear waters, typically in open ocean waters (Mueller, 2000; Giardino *et al.*, 2001; Antoine, 2010; Soria *et al.*, 2017b). This ratio is a good estimator of Z_{SD} in clear waters in which the phytoplankton is the main contributor to the light attenuation in the water column.

To estimate the transparency of the water in this work, we have performed a process of re-calibration of the 490/560, and 490/705 ratios, plus the hereafter called "Koponen" method (Koponen *et al.*, 2001) after testing different algorithms. We have modified the original Alikas & Kratzer (2017) algorithm. In this paper, the band ratio are not raised to the exponent b. Instead, the b coefficient is added or subtracted depending on the direction of the slope, obtained in the relation between *in situ* Secchi disk depth (Z_{SD}) measurements and band ratio. The algorithm is detailed in equation 7:

$$Z_{SD} = e^{\wedge}(a*\ln[Rrs\_490/Rrs\_560] + b) \quad (7)$$

where

Rrs_490 and Rrs_560 are reflectances of S2-MSI spectral bands at 490 and 560 nm.

a and b are the coefficients derived by linear regression between the Napierian logarithm of *in situ* Z_{SD} and the Napierian logarithm of 490/560 ratio.

Alikas & Kratzer (2017) applied another model based on ratio 490/705. This model had a better result (R^2 = 0.73) in the Himmerfjärden lake with a 1.9 to 7.5 m Z_{SD} and a 1.2 to 11.6 mg/m^3 [Chl-*a*] range (Baltic Sea). The algorithm is detailed in equation 8:

$$Z_{SD} = e^{\wedge} (a*\ln [Rrs\_490/Rrs\_705] + b) \quad (8)$$

where

Rrs_490 and Rrs_705 are reflectances of S2-MSI spectral bands at 490 and 705 nm.

a and b are the coefficients derived by linear regression between Napierian logarithm of *in situ* Z_{SD} and the Napierian logarithm of 490/705.

In more turbid waters, where other optically active constituents such as NAP and CDOM are also present in significant concentrations, other spectral regions are more sensitive for estimating the Z_{SD} (Doron *et al.*, 2007). Since this was the present situation in some reservoirs studied in ESAQS, other algorithms were also tested, such as the Koponen *et al.*, (2001) algorithm. This algorithm was validated in Finnish Lakes where the water bodies generally present Z_{SD} values of less than 3 m. We have modified the original algorithm since we have not applied the 783 band. The modified Koponen algorithm is detailed in equation 9:

$$Z_{SD} = e^{\wedge} (a*\ln [Rrs\_560/Rrs\_705] + b) \quad (9)$$

where

a and b are the coefficients derived by linear regression between Napierian logarithm of *in situ* Z_{SD} and the Napierian logarithm of 560/705.

The Rrs_560 and Rrs_705 are reflectances of S2-MSI spectral bands.

RESULTS

The *C* Z_{SD} measurement range in the water bodies under study was between 0.25 to 10.5 m. The range of Limnetica, 38(1): 467-479 (2019). DOI: 10.23818/limn.38.26 [Chl-*a*] measurements in the study area water bodies was between 0.54 to 169 mg/m^3.

Calibration and validation of [Chl-*a*] algorithms

In the set of simulated data used for calibration, the input [Chl-*a*] presented a very high correlation, according the coefficient of determination (R^2), with the retrieved [Chl-*a*] in all algorithms: OC2_443 with a R^2 = 0.98 (Fig. 3a); OC2_490 with R^2 = 0.97; (Fig. 3b); OC_3 with R^2 = 0.97 (Fig. 3c); and TBDO with R^2 = 0.99 (Fig. 3d). The specific calibration coefficients of each algorithm were obtained by a polynomial fit between the simulated [Chl-*a*] database and reflectance band ratios. Table 2 summarizes the coefficients of each polynomial fit.

To validate these algorithms, the data measured in the field was used. Table 3 shown this dataset. According to Guibaja *et al.* (2015), the TBDO algorithm had an optimum performance on eutrophic and hypertrophic waters. In our study area, this type of water corresponds to reservoirs with [Chl-*a*] > 10 mg/m^3 like the Albufera of Valencia, and the Bellús and Beniarrés reservoirs (Table 3). For [Chl-*a*] < 10 mg/m^3, we have applied the OC2_443, OC2_490 and OC3 algorithm on ultraoligotrophic, oligotrophic and mesotrophic waters. These water types correspond to the Benageber, Tous, María Cristina, Contreras, Sitjar and Regajo reservoirs (Table 3). The error analysis was performed between the calibrated algorithms and *in situ* [Chl-*a*] measurements. For this analysis, we have applied the mean absolute error (MAE) and a systematic direction of the error (bias). Figure 4 shows the [Chl-*a*] measured *in situ* versus the calculated with the 4 algorithms. In the OC2_443, OC2_490 and OC3 algorithms, it can be seen that the slope presented values very far from line 1:1 (0.25 and 0.30) (Fig. 4a, b and c) with a very high bias. This means that in these cases, ocean color models are overestimating the actual chlorophyll concentration, with high error values. For this reason, it was necessary to perform a recalibration process to readjust each of the models. Table 4 shows the final equation, which was obtained from the recalibration process for the [Chl-*a*] estimation, replacing the x of each adjustment (Fig. 4 a, b and c) the formulas obtained in the calibration process. For this, the value of the line slope (coefficient a) was integrated logarithmically in equation 1.

According to Table 4 data, all three algorithms provide similar statistics, but the OC3 was the algorithm which gave back the lowest error equal

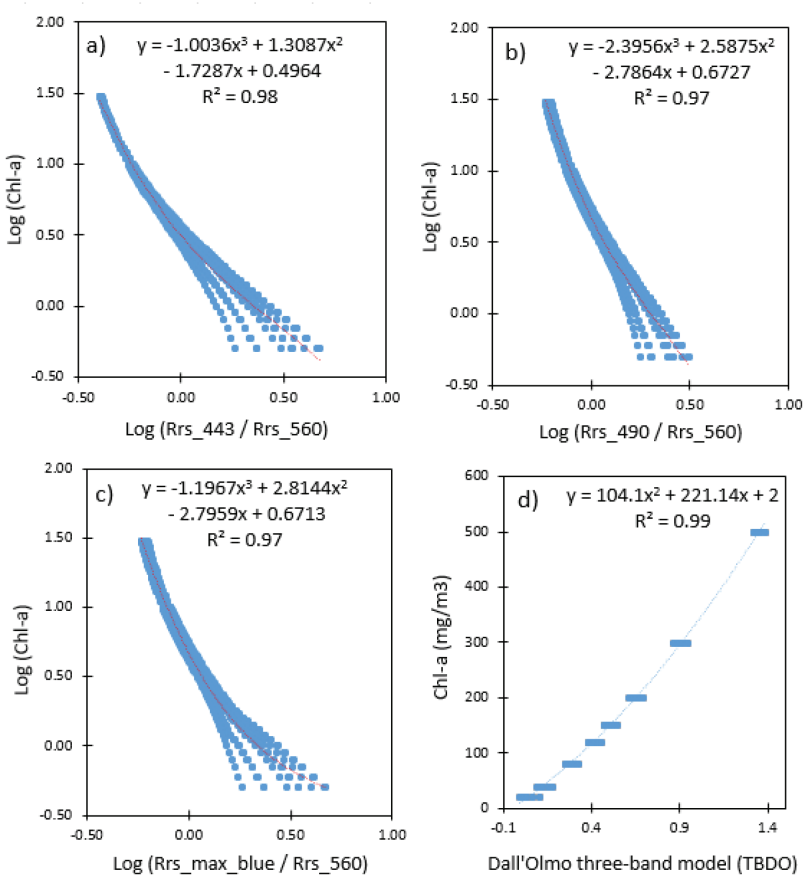


Figure 3. a) OC2_443 algorithm and simulated [Chl-*a*] polynomial fit. b) OC2_490 algorithm and simulated [Chl-*a*] polynomial fit. c) OC3 algorithm and simulated [Chl-*a*] polynomial fit, and d) TBDO algorithm and simulated [Chl-*a*] polynomial fit. *a) Ajuste polinómico entre el algoritmo OC2_443 y la [Chl-a] simulada. b) Ajuste polinómico entre el algoritmo OC2_490 y la [Chl-a] simulada. c) Ajuste polinómico entre el algoritmo OC3 y la [Chl-a] simulada, y d) Ajuste polinómico entre el algoritmo TBDO y la [Chl-a] simulada.*

Table 2. Coefficients obtained by polynomial fit for OC2_443, OC2_490, OC3 and TBDO algorithms for [Chl-*a*] estimation (in mg/m^3) from simulated HydroLight dataset. *Coeficientes obtenidos por ajuste polinomial para los algoritmos OC2_443, OC2_490, OC3 y TBDO para la estimacion de la [Chl-a] (en mg/m^3) a partir de una base de datos simulada con HydroLight.*

Algorithm	a	b	c	d
OC2_443	0.4964	-1.7287	1.3087	-1.0036
OC2_490	0.6727	-2.7864	2.5875	-2.3956
OC3	0.6713	-2.7959	2.8144	-1.1967
TBDO	104.1	221.1	2.0	

Table 3. Database of *in situ* [Chl-*a*] and Secchi depth disk measurements in Valencia region reservoirs and lakes according by field campaign date. *Base de datos de medidas* in situ *de [Chl-a] y profundidad del disco de Secchi en los embalses y lagos de la región de Valencia según fecha de campaña de campo.*

Reservoir	[Chl-*a*] (mg/m^3)	Z_{SD} (m)	Date
Albufera	51.3, 53, 54.2, 31.8, 39.1, 43.1		05/08/2015
Albufera	54.8, 56.3, 56.1,52.9,55.2, 58.3		27/08/2015
Albufera	148.9, 157.9, 156, 169.1, 159.4, 154, 93.3		30/11/2015
Albufera	135.6, 138.2, 121.5, 25, 128.7	0.31, 0.31, 0.33, 0.43, 0.32	12/03/2016
Albufera	141.8, 130.4, 140.1, 114.6, 78.1, 84.4, 83.2	0.25, 0.26, 0.3, 0.35, 0.26, 0.33	21/04/2016
Albufera	70.4, 68, 22.4, 10.7, 36.4, 43.5	0.3, 0.32, 0.33, 0.34,0.34, 0.37, 0.5	02/05/2016
Tous	1.27, 1.78, 3.1	5.9, 6, 5.8	27/12/2016
Bellús	31.8	1	16/01/2017
Contreras	2, 1, 0.79, 0.83	1, 1.3, 1.07	08/02/2017
Albufera	39.7, 64.5, 45.1, 47.3	0.27, 0.32, 0.34, 0.36	07/03/2017
Beniarrés	45.45	0.95	27/03/2017
Benagéber	2.48, 2.74	4, 5.2, 7.4	30/03/2017
Ma. Cristina	1.4, 1.33	5.2, 5.6	06/04/2017
Sitjar	0.54, 0.65	9.4, 10.5	06/04/2017
Bellús	61.39, 66.76, 68.01	0.55	15/06/2017
Regajo	8.6, 8.9, 10.2	2, 1.7	05/07/2017
Sitjar	0.68, 0.61	2.7, 3.15	23/10/2017
Benageber	5.4, 5.76, 4.57	3.4, 3.6, 4.1	26/10/2017
Beniarrés	11.13, 17.17	1.15, 1.4	07/11/2017
Tous	0.64, 0.72, 0.69	7.1, 8, 9.1	17/11/2017
Contreras	2.08, 2.42, 2.02, 0.86, 0.98	4.1, 5	30/11/2017
Tous	0.63, 0.58, 0.69	7, 7.75, 8.1	16/01/2018
Ma. Cristina	2.92, 2.72	0.75, 0.75	31/01/2018
Sitjar	0.59, 0.63	2.4, 2.2	31/01/2018
Benageber		4.3, 4.85, 5.5	23/02/2018
Albufera	84.5, 82.1, 81.6	0.31, 0.33, 0.30	07/03/2018
Bellús	41.54, 51.59, 49.09	0.45, 0.45, 0.5	22/03/2018
Regajo	5.57, 4.58, 4.63, 5.12	3, 3.75, 4, 4.25	11/05/2018
Benageber	4.91, 4.91, 4.85, 4.53	3.35, 3.35, 3.7, 3.75	16/05/2018

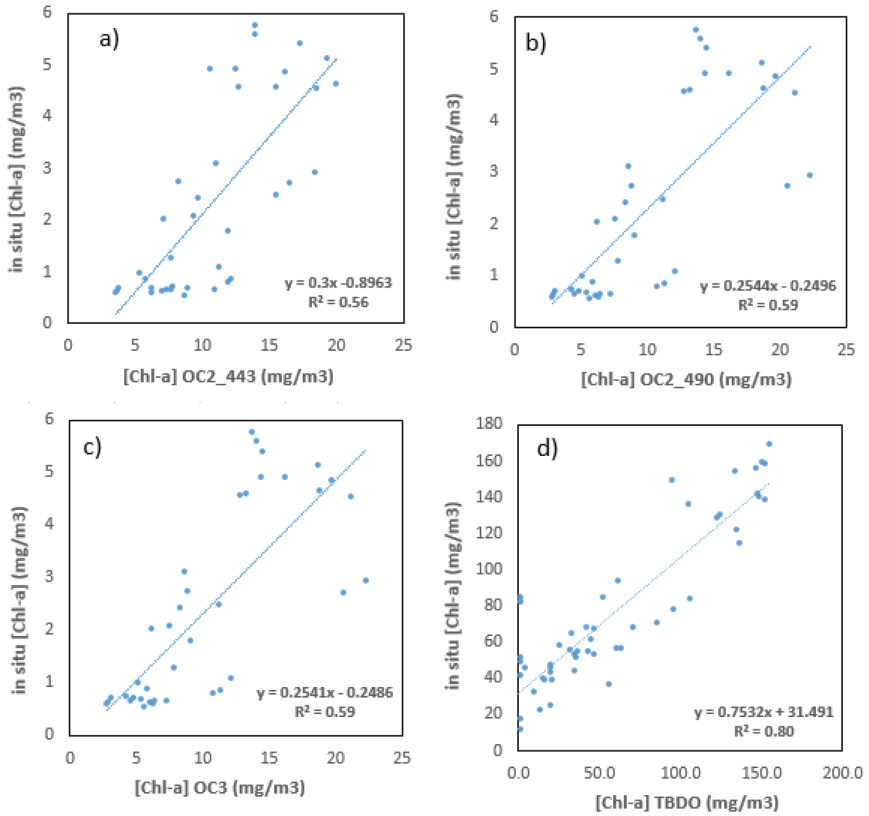


Figure 4. a) OC2_443 algorithm and *in situ* [Chl-*a*] validation. b) OC2_490 algorithm and *in situ* [Chl-*a*] validation. c) OC3 algorithm and *in situ* [Chl-*a*] validation, and d) TBDO algorithm and *in situ* [Chl-*a*] validation. *a) Validación del algoritmo OC2_443 en relación a la [Chl-*a*]* in situ*. b) Validación del algoritmo OC2_490 en relación a la [Chl-*a*]* in situ*. c) Validación del algoritmo OC3 en relación a la [Chl-*a*]* in situ*, and d) Validación del algoritmo TBDO en relación a la [Chl-*a*]* in situ*.*

to 0.89 mg/m^3. However, the OC2_490 model also has very good results, with the same correlation as OC3 and the advantage that it is easier to calculate, saving processing time and using only one blue band, versus OC3 that he need to use both.

For eutrophic waters with [Chl-*a*] > 10 mg/m^3, the TBDO algorithm (Fig. 4d) gave an $R^2 = 0.80$, a bias = 17.49 and an MAE = 23 mg/m^3, which is a relative low error for high [Chl-*a*] waters, taking into consideration the range of values for this group (10 to 169 mg/m^3). We consider it acceptable for the evaluation of the ecological status in these eutrophic or hypertrophic reservoirs (Guibaja *et al.*, 2016). The validated TBDO algorithm was applied to an S2-MSI image from the Albufera of Valencia acquired on 07/03/2018. This S2-MSI image was atmospherically corrected with Sen2-Cor. Fig. 5 shows the patterns of [Chl-*a*] concentration with medium to high values in nearly the entire lake with a [Chl-*a*] TBDO, between 75 and 90 mg/m^3. The same day, the *in situ* [Chl-*a*] measurements recorded in the Albufera of Valencia were between 81.6 to 84.5 mg/m^3 (Table 3). The green areas correspond to the water contributions with less chlorophyll-*a* coming from the irrigation channels renovating the water, while the central area corresponds to stagnant water, which leads to higher values of [Chl-*a*].

Table 4. Equations for the [Chl-*a*] estimation (in mg/m^3) in the calibration process. *Ecuaciones para la estimación de la [Chl-a] (en mg/m^3) en el proceso de calibración.*

Algorithm	**Equation** ([Chl-*a*] in mg/m^3)	**MAE**	**R^2**
OC2_443	[Chl-*a*] = 10^(-0.02648-1.7287*X+1.3087*X^2-1.0036*X^3)-0.8963	0.93	0.56
OC2_490	[Chl-*a*] = 10^(0.078217-2.7864*X+2.5875*X^2-2.3956*X^3)-0.2496	0.90	0.59
OC3	[Chl-*a*] = 10^(0.076305-2.7959*X+2.8144*X^2-1.1967X^3)-0.2486	0.89	0.59

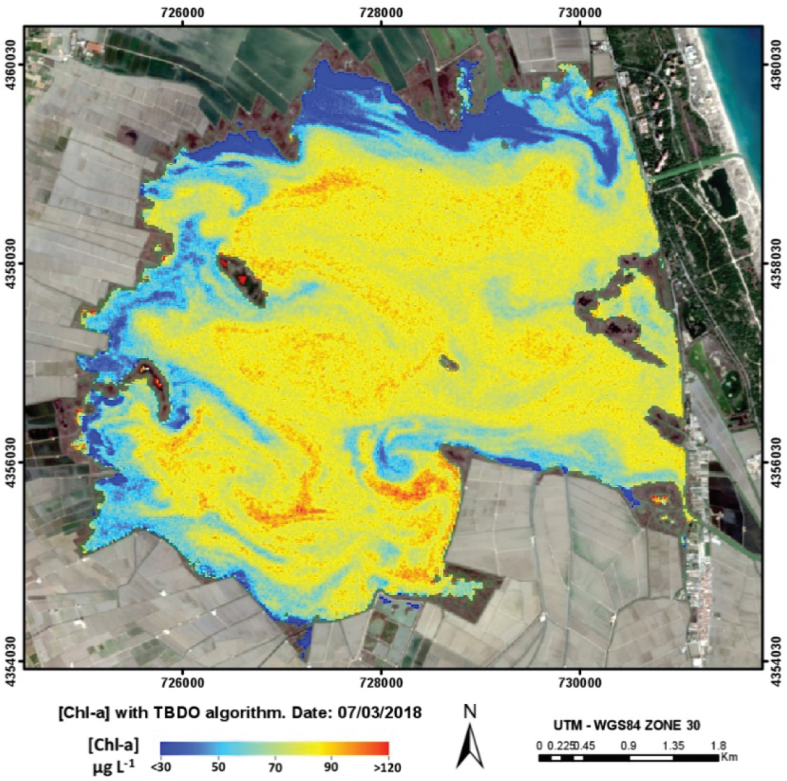


Figure 5. [Chl-a] map by TBDO algorithm in the Albufera of Valencia from S2-MSI image of 07/03/2018. *Mapa de [Chl-a] aplicando el algoritmo TBDO en la Albufera de Valencia obtenido de la imagen S2-MSI del 07/03/2018.*

Calibration and validation of Z_{SD} algorithms

The calibration-validation process of algorithms for the Z_{SD} estimation, has been based on the relationship between the 490/560, 490/705 and 560/705 ratios and *in situ* Z_{SD} measurements. Inspired by the algorithms described earlier, we have tested several regression models. The dataset collected in field campaigns over the study area consist of 79 measurements. It has a wide Z_{SD}

range from 0.25 to 10 m. For the calibration and validation process, the dataset was divided into 2 groups: 75 % of the data (N = 60) was used to calibrate and 25 % (N = 19) to validate the ratios. In a first attempt, we tested the three algorithms applying the trophic classification followed in the [Chl-*a*] algorithm calibration, differentiating ultraoligotrophic-mesotrophic waters and eutrophic-hypertophic waters, but the results were not optimum due to the limited dataset used.

Figure 6 shows the results of the calibration process. Here, the logarithm of measured Z_{SD} is represented, depending on the logarithm of each of the three indices. In Fig. 6 it is observed that the best fit is the linear one, and that for the calibration process, the 490/705 and 560/705 ratios showed the best performance in both cases.

To validate the Z_{SD} algorithms, we applied the coefficients obtained in the calibration process to the different band ratios, as explained earlier. The validation data set consists of 19 measurements (25 % of total). Figure 7 show the results. To select the best method, Table 5 shows the three formulas obtained in the calibration process and the statistics of the validation process. Considering all the statistics, the algorithm with the lowest error and the best correlation is 490/705.

According to these results, the 490/705 ratio is the most appropriate for estimating Z_{SD} with an error MAE of 0.88 m. We have applied this algorithm on the same S2-MSI scene that was used in the application of the [Chl-*a*] algorithm. We obtained a map that shows the different values of the water transparency of this hypertrophic lake (Fig. 8). For the Albufera of Valencia (Fig. 5 and Fig. 8), it can be observed that according to the 490/705 ratio, a vast majority of the lagoon had values between 0.30 to 0.36 meters. Once again, the correct match-up with the *in situ* Z_{SD} was confirmed. According to the data measured this day, the Z_{SD} was between 0.30 and 0.33 m (Table 3).

As can be seen in the [Chl-*a*] and the Z_{SD} maps, the expected relationship between the minimum value of chlorophyll concentration and maximum water transparency is given in areas near shoreline. In the northern region, the concentration of chlorophyll is minimum, taking into account the values normally measured in Albufera of Valencia (See Table 3). Here the values are between 50 to 70 mg/m^3, coinciding with maximum transparency, Z_{SD} up to 0.38 meters, again matching-up with the values registered in this lagoon (Table 3). This high transparency is probably due to the early opening of water gates for rice crops that usually starts in May.

DISCUSSION

To estimate [Ch-*a*] and according to the error analysis (MAE = 0.89 mg/m^3), the OC2_490

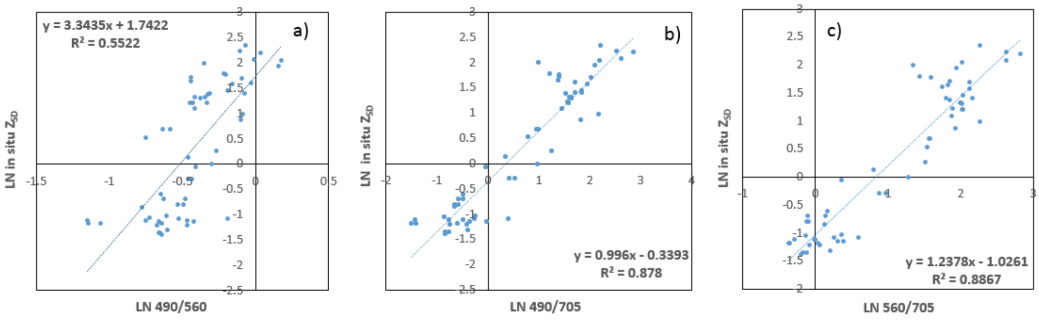


Figure 6. Calibration of the Z_{SD} algorithms. a) 490/560 algorithm and *in situ* ln(Z_{SD}) linear fit. b) 490/705 algorithm and *in situ* ln(Z_{SD}) linear fit. and c) 560/705 algorithm and *in situ* ln(Z_{SD}) linear fit. *Calibración de los algorimos para Z_{SD}. a) Ajuste lineal entre el algoritmo 490/560 y ln(Z_{SD})* in situ. *b) Ajuste lineal entre el algoritmo 490/705 y ln(Z_{SD})* in situ, *y c) Ajuste lineal entre el algoritmo 560/705 y ln(Z_{SD})* in situ.

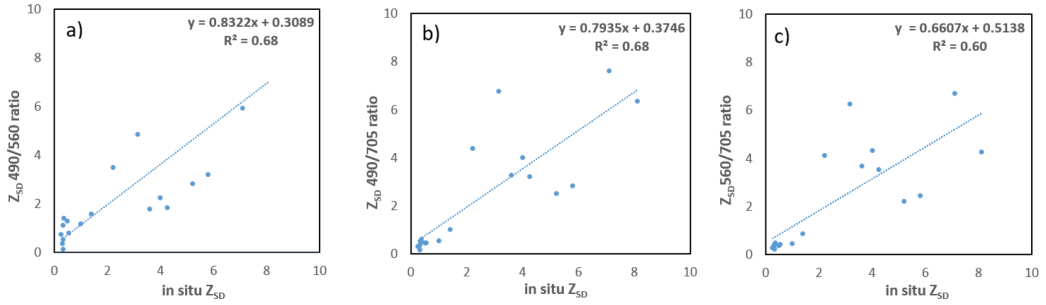


Figure 7. Validation of the Z_{SD} algorithms. a) Z_{SD} calculated with the 490/560 ratio in relation with *in situ* Z_{SD}. b) Z_{SD} calculated with the 490/705 ratio in relation with *in situ* Z_{SD} validation. c) Z_{SD} calculated with the 560/705 ratio in relation with *in situ* Z_{SD}. *Validación de los algorimos para* Z_{SD}. *a)* Z_{SD} *calculada con el ratio 490/560 en relación a* Z_{SD} *medida* in situ. *b)* Z_{SD} *calculada con el ratio 490/705 en relación a* Z_{SD} *medida* in situ. *c)* Z_{SD} *calculada con el ratio 560/705 en relación a* Z_{SD} in situ.

Table 5. Equations for the Z_{SD} estimation and statistics of the validation process. *Ecuaciones para la estimación de* Z_{SD} *y estadísticos del proceso de validación.*

Algorithm	**Equation** (Z_{SD} in m)	**R^2**	**MAE**	**Bias**
490/560	Z_{SD} = e^(3.3435*ln (Rrs_490/Rrs_560) + 1.7422)	0.68	1.16	0.12
490/705	Z_{SD} = e^(0.996*ln (Rrs_490/Rrs_705) - 0.3393)	0.68	0.88	0.15
560/705	Z_{SD} = e^(1.2378*ln (Rrs_560/Rrs_705) -1.0261)	0.60	0.96	0.35

algorithm should be preferred for low values ([Chl-*a*] = < 10 mg/m^3), as is the case of the reservoirs of Benagéber, Contreras, Sitjar, Tous, Benirarrés, María Cristina and Regajo. While Ocean Color algorithms have a longer tradition in ocean studies, the optimum results achieved in this study suggest they could have broad applicability on some types of inland waters. The validation of the TBDO algorithm gave a good result in eutrophic to hypertrophic water mass with [Chl-*a*] > 10 mg/m^3, such as the Albufera of Valencia and the Bellús reservoir. The results obtained in this study ratify the ones obtained by Guibaja *et al.* (2016) on the applicability of this algorithm on eutrophic to hypertrophic water masses. The match-up obtained between the TBDO application over the S2-MSI image and the values measured *in situ* the same day confirms the validity of the TBDO algorithm in water masses with high value of chlorophyll-*a*.

To estimate the Z_{SD}, we have obtained an optimum result with the 490/705 band ratio. While it is true that the environment and natural conditions between the Baltic Sea and the reservoirs of the Valencia region are different, the Z_{SD} range (0.25 to 10.5 m) measured in our study area is similar to that measured by Alikas & Kratzer (2017) in Himmerfjärden Lake (1.9 to 7.5 m). As happened with the [Chl-*a*], the match-up obtained between the 490/705 ratio application over the S2-MSI image and the values measured *in situ* the same day confirms the validity of this algorithm in water masses with lower water transparency value.

Regarding the applicability of the three algorithms by trophic states as we did for the [Chl-*a*] algorithms, we concluded that is not appropriate due to a) the limited dataset for the calibration-validation process and b) the composition of suspended solids of water, due to similar Z_{SD}

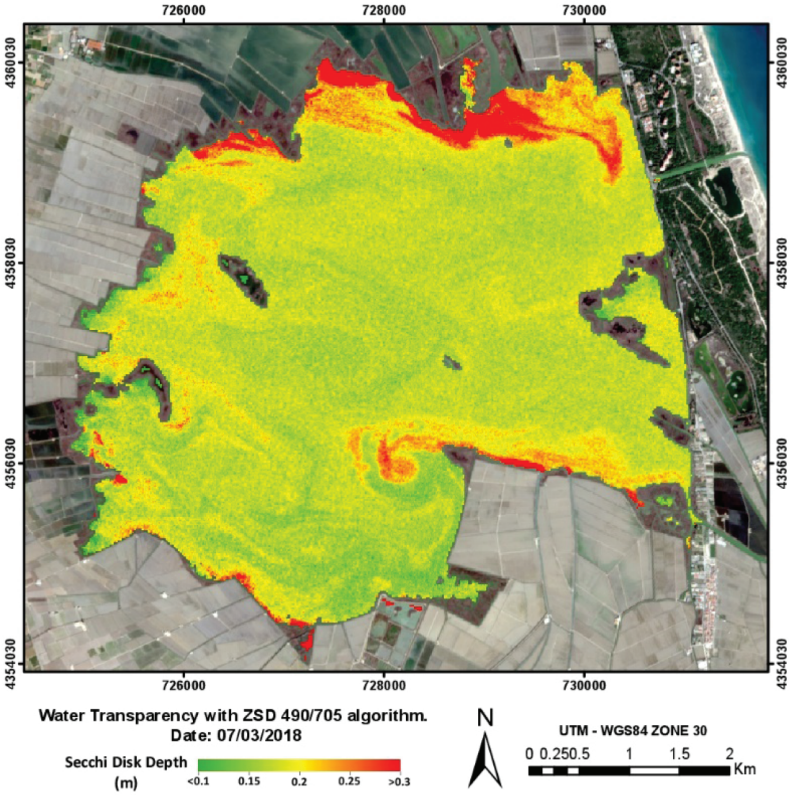


Figure 8. Z_{SD} map by 490/705 ratio in the Albufera of Valencia. S2-MSI image of 07/03/2018. *Mapa de Z_{SD} aplicando el ratio 490/705 en la Albufera de Valencia. Imagen S2-MSI del 07/03/2018.*

values have different [Chl-*a*]; this is produced by the presence of NAP giving turbidity to water but without pigments. In this way, Yang *et al.* (2013) also proposed two quasi-analytical algorithms to retrieve total absorption and backscattering coefficients based on a semi-analytical estimation model for clear and turbid inland waters, and these models could be proved in future works. There are many works were Z_{SD} is not well estimated, as Devi Prasad & Siddaraju (2012) in some Indian hypertrophic lakes; Borkman & Smayda (1998) in coastal areas of Rhode Island and Lahtrop (1992) were algorithms derived from samples in Lake Michigan cannot be extrapolated to Yellowstone and Jackson Lakes.

CONCLUSIONS

The accuracy of the algorithms estimated in the reservoirs studied so far is sufficient for most of the applications related to the ecological status assessment. The S2-MSI images showed that they are optimum for performing ecological monitoring on inland waters from space.

It is important to emphasize the relevance of the atmospheric correction method used: Sen2-Cor, which in our case was accurate enough, allows the algorithms to be applied to actual S2-MSI data for routine monitoring purposes, which is the ultimate goal of the ESAQS project. In this context, the next step of the research is to

evaluate and validate several atmospheric correction methods for different types of inland water in order to determine the one best suited for achieving the ESAQS objectives.

ACKNOWLEDGEMENTS

We would like to thank to the Generalitat Valenciana for the concession of the Prometeo program 2016/032 for the start-up and development of the ESAQS (Ecological Status of Aquatic Systems with Sentinel satellites) project. We also thank the Confederación Hidrográfica del Júcar for their support in the field campaigns carried out in this project, and to the Spanish National Institute for Aerospace Technology (INTA) for allowing us to generate the HydroLight database.

REFERENCES

ALIKAS, K. & S. KRATZER. 2017. Improved retrieval of Secchi depth for optically-complex waters using remote sensing data. *Ecological indicators*, 77: 218-227.

ANTOINE, D. 2010. Sentinel-3 Optical products and algorithm definition. OLCI Level 2 Algorithm Theoretical Basis Document Transparency Products in case 1 waters, 13.

AUSTIN, R. W. & T. J. PETZOLD. 1981. The determination of the diffuse attenuation coefficient of sea water using the Coastal Zone Color Scanner. *Oceanography from Space*, 239-256.

BORKMAN, D. G. & T. J. SMAYDA. 1998. Long-term trends in water clarity revealed by Secchi-disk measurements in lower Narragansett Bay, *ICES Journal of Marine Science*, 55(4), 668–679. DOI: 10.1006/jmsc.1998.0380

BRICAUD, A., A. MOREL, M. BABIN, H. CLAUSTRE. 1995. Variability in the chlorophyll-specific absorption coefficients of natural phytoplankton: Analysis and parametrization. *Journal of Geophysical Research*. DOI: 10.1029/95JC00463

CASPERS, H. 1984. OECD: Eutrophication of Waters. Monitoring, Assessment and Control. -154 pp. Paris: Organisation for Economic Co-Operation and Development 1982 (Publié en français sous le titre» Eutrophication des Eaux. Méthodes de Surveillance, d'Evaluation et de Lutte «). *International Review of Hydrobiology*, 69(2): 200-200.

DALL'OLMO, G., A. A. GITELSON & D. C. RUNDQUIST. 2003. Towards a unified approach for remote estimation of chlorophyll-a in both terrestrial vegetation and turbid productive waters. *Geophysical Research Letters*, 30:18

D'ODORICO, P., A. GONSAMO, A. DAMM & E. SCHAEPMAN. 2013. Experimental evaluation of Sentinel-2 Spectral Response Function for NDVI time-series continuity. *IEE Transactions on Geoscience and Remote Sensing*, 51(3): 1336-1348.

DEVI PRASAD, A.G. & SIDDARAJU, K., 2012. Application of CCME WQI to lakes of Mandya, Karnataka State, India. On. Internat. Interdispl. Res. J., 2(1): 108-114.

DORON, M., M. BABIN, A. MANGIN & O. HEMBISE. 2007. Estimation of light penetration, and horizontal and vertical visibility in oceanic and coastal waters from surface reflectance. *Journal of Geophysical Research: Oceans*, 112(C6). DOI: 10.1029/2006JC004007

ESA 2018a. Spectral Response Function. (Available on: https://earth.esa.int/web/sentinel/user-guides/sentinel-2-msi/document-library/-/asset_publisher/Wk0TKajiISaR/content/sentinel-2a-spectral-responses). Last time consulted: 01/02/2018.

ESA 2018b. Sentinel images. (Available on: https://scihub.copernicus.eu/dhus/). Last time consulted: 01/02/2018.

ESA 2018c. SNAP software. Version 5.0. (Available on: http://step.esa.int/main/download/). Last time consulted: 01/02/2018.

ESA 2018d. Sen2Cor plugin. Version 2.4.0. (Available on: http://step.esa.int/main/third-party-plugins-2/sen2cor/). Last time consulted: 01/02/2018.

FARGION, G. S., and J. L. MUELLER (eds.) 2000. Ocean Optics Protocols for Satellite Ocean Color Sensor Validation, Vol. Revision 2. *National Aeronautical and Space Administration*. Goddard Space Flight Center. Greenbelt, Maryland, 184 p.

GIARDINO, C., M. PEPE, P. A. BRIVIO, P.

GHEZZI & E. ZILIOLI. 2001. Detecting chlorophyll, Secchi disk depth and surface temperature in a sub-alpine lake using Landsat imagery. *Science of the Total Environment*, 268(1-3): 19-29.

GILERSON, A. A., A. A. GITELSON, J. ZHOU, D. GURLIN, M. WESLEY, I. IANNOU & S. AHMED. 2010. Algorithms for remote estimation of chlorophyll-a in coastal and inland waters using red and near infrared bands. *Optic Express*, 18(23): 24109-24125.

GUIBAJA, G., A. RUIZ-VERDÚ, S. ROMO, J. M. SORIA, C. TENJO, M. PEREIRA-SANDOVAL, J. DELEGIDO, R. PEÑA & J. MORENO. 2016. Mapping water quality in the Albufera of Valencia lake with the new Sentinel-2 Earth Observation satellite. XVIII Iberian Congress of Limnology. IRTA. July 4-8, 2016. Tortosa, Spain.

JEFFREY, S. T. & G. F. HUMPHREY. 1975. New spectrophotometric equations for determining chlorophylls a, b, c1 and c2 in higher plants, algae and natural phytoplankton. *Biochemie und Physiologie der Pflanzen*. 167: 191-194.

KOPONEN, S., J. PULLIAINEN, K. KALLIO & M. HALLIKAINEN. 2001. Lake water quality classification with airborne hyperspectral spectrometer and simulated MERIS data. *Remote Sensing of Environment*, 79(1): 51-59.

KRATZER, S., C. BROCKMANN & G. MOORE. 2008. Using MERIS full resolution data to monitor coastal waters – A case study from Himmerfjärden, a fjord-like bay in the northwestern Baltic Sea. *Remote Sensing of Environment*, 112(5): 2284-2300.

LATHROP, R. G. 1992. Landsat Thematic Mapper monitoring of turbid inland water quality. *Photogrammetric Engineering and Remote Sensing*, 58: 465-470.

MOBLEY, C. 1994. Light and Water: Radiative Transfer in Natural Waters. *Academic Press*.

MOBLEY, C. 1999. Estimation of the remote-sensing reflectance from above-surface measurements. *Applied optics*, 38(36): 7442-7455.

MUELLER, J. L. 2000. SeaWIFS algorithm for the diffuse attenuation coefficient, K (490), using water-leaving radiances at 490 and 555 nm. *SeaWiFS postlaunch calibration and validation analyses*, 3: 24-27.

O'REILLY, J. E. et al. 2000. Ocean color chlorophyll-a algorithms for SeaWiFS, OC2, and OC4: Version 4. *SeaWi-FS postlaunch calibration and validation analyses*, Part, 3: 9–23.

PREISENDORFER, R. 1986. Secchi disk science: Visual optics of natural waters1. *Limnology and Oceanography*. Volume 31, Issue 5. DOI: 10.4319/lo.1986.31.5.0909

RUIZ-VERDÚ, A., J. C. JIMÉNEZ, X. LAZZARO, C. TENJO, J. DELEGIDO, M. PEREIRA, J. SOBRINO & J. MORENO. 2016. Comparison of MODIS and LANDSAT-8 retrievals of chlorophyll-*a* and water temperature over Lake Titicaca. XVIII Iberian Congress of Limnology. IRTA. July 4-8, 2016. Tortosa, Italy.

RUESCAS, A. B., M. PEREIRA-SANDOVAL, C. TENJO, A. RUIZ- VERDÚ, F. STEINMETZ & L. DE KEUKELAERE. 2016. Sentinel-2 atmospheric correction inter-comparison over two lakes in Spain and Peru-Bolivia. CLEO (Colour and Light in the Ocean from Earth Observation). ESA-ESRIN. September 6-8, 2016. Frascati, Rome, Italy.

SORIA, X., J. DELEGIDO, E. P. URREGO, M. PEREIRA-SANDOVAL, E. VICENTE, A. RUIZ-VERDÚ, J. M. SORIA, R. PEÑA, C. TENJO & J. MORENO. 2017a. Validación de algoritmos para la estimación de la Clorofila-a con Sentinel-2 en la Albufera de Valencia. Teledetección: Nuevas plataformas y sensores aplicados a la gestión del agua, la agricultura y el medio ambiente. XVII Congreso de la Asociación Española de Teledetección. October 3-7, 2017. Murcia, Spain. 293-296.

SORIA, X., E. VICENTE, C. DURÁN, J. M. SORIA, R. PEÑA. 2017b. Uso de imágenes Landsat-8 para la estimación de la profundidad del disco de Secchi en aguas continentales. Teledetección: Nuevas plataformas y sensores aplicados a la gestión del agua, la agricultura y el medio ambiente. XVII Congreso de la Asociación Española de Teledetección. October 3-7, 2017. Murcia, Spain. 293-296.

SORIA, X., E. P. URREGO, M. PEREIRA-SANDOVAL, A. RUIZ-VERDÚ, R. PEÑA, J. M. SORIA, J. DELEGIDO, E. VICENTE & J.

MORENO. 2019. Monitoring the ecological state of a hypertrophic lake (Albufera of València, Spain) using multitemporal Sentinel-2 images. *Limnetica*, 38(1): 457-469. DOI: 10.23818/limn.38.26

SHOAF, W. T. & B. W. LIUM. 1976. Improved extraction of chlorophyll *a* and *b* from algae using dimethyl sulphoxide. *Limnology and. Oceanography*, 21: 926-928.

STRAMSKI, D., A. BRICAUD, A. MOREL. 2001. Modeling the Inherent Optical Properties of the Ocean Based on the Detailed Composition of the Planktonic Community. *Applied Optics*. 40 (18): 2929-2945. DOI: 10.1364/AO.40.002929

YANG, W., B. MATSUSHITA, J. CHEN, K. YOSHIMURA & T. FUKUSHIMA. 2013. Retrieval of inherent optical properties for turbid inland waters from remote-sensing reflectance. *IEEE Transactions on Geoscience and Remote Sensing*, 51: 3761–3773. DOI: 10.1109/TGRS.2012.2220147

ZHENG, G. and P.M. DIGIACOMO. 2017. Detecting phytoplankton diatom fraction based on the spectral shape of satellite-derived algal light absorption coefficient. Limnology and Oceanography. *Association for the Sciences of Limnology and Oceanography*. DOI: 10.1002/lno.10725

Con el apoyo de:

GOBIERNO DE ESPAÑA MINISTERIO PARA LA TRANSICIÓN ECOLÓGICA Fundación Biodiversidad

Artículo 2

"Evaluation of Atmospheric Correction Algorithms over Spanish Inland Waters for Sentinel-2 Multispectral Imagery Data"

Este artículo fue publicado en *Remote Sensing* en 2019. Esta revista tuvo en 2019 un factor de impacto de 4.509 y la posición de 9/50 (Q1) en la categoría Remote Sensing.

remote sensing

MDPI

Article

Evaluation of Atmospheric Correction Algorithms over Spanish Inland Waters for Sentinel-2 Multi Spectral Imagery Data

Marcela Pereira-Sandoval [1,*], Ana Ruescas [1], Patricia Urrego [1], Antonio Ruiz-Verdú [1], Jesús Delegido [1], Carolina Tenjo [1], Xavier Soria-Perpinyà [2], Eduardo Vicente [2] and Juan Soria [2] and José Moreno [1]

[1] Image Processing Laboratory, Laboratory of Earth Observation. Universitat de València, C/Catedrático José Beltrán, 2, 46980 València, Spain; ana.b.ruescas@uv.es (A.R.); patricia.urrego@uv.es (P.U.); antonio.ruiz@uv.es (A.R.-V.); jesus.delegido@uv.es (J.D.); carolina.tenjo@uv.es (C.T.); jose.moreno@uv.es (J.M.)

[2] Institut Cavanilles de Biodiversitat i Biología Evolutiva, Universitat de València, C/Catedrático José Beltrán, 2, 46980 València, Spain; javier.soria-perpina@uv.es (X.S.-P.); eduardo.vicente@uv.es (E.V.); juan.soria@uv.es (J.S.)

* Correspondence: marcela.pereira@uv.es; Tel.: +34-963-543-679

Received: 23 April 2019; Accepted: 18 June 2019; Published: 21 June 2019

check for updates

Abstract: The atmospheric contribution constitutes about 90 percent of the signal measured by satellite sensors over oceanic and inland waters. Over open ocean waters, the atmospheric contribution is relatively easy to correct as it can be assumed that water-leaving radiance in the near-infrared (NIR) is equal to zero and it can be performed by applying a relatively simple dark-pixel-correction-based type of algorithm. Over inland and coastal waters, this assumption cannot be made since the water-leaving radiance in the NIR is greater than zero due to the presence of water components like sediments and dissolved organic particles. The aim of this study is to determine the most appropriate atmospheric correction processor to be applied on Sentinel-2 MultiSpectral Imagery over several types of inland waters. Retrievals obtained from different atmospheric correction processors (i.e., Atmospheric correction for OLI 'lite' (ACOLITE), Case 2 Regional Coast Colour (here called C2RCC), Case 2 Regional Coast Colour for Complex waters (here called C2RCCCX), Image correction for atmospheric effects (iCOR), Polynomial-based algorithm applied to MERIS (Polymer) and Sen2Cor or Sentinel 2 Correction) are compared against in situ reflectance measured in lakes and reservoirs in the Valencia region (Spain). Polymer and C2RCC are the processors that give back the best statistics, with coefficients of determination higher than 0.83 and mean average errors less than 0.01. An evaluation of the performance based on water types and single bands–classification based on ranges of in situ chlorophyll-a concentration and Secchi disk depth values- showed that performance of these set of processors is better for relatively complex waters. ACOLITE, iCOR and Sen2Cor had a better performance when applied to meso- and hyper-eutrophic waters, compare with oligotrophic. However, other considerations should also be taken into account, like the elevation of the lakes above sea level, their distance from the sea and their morphology.

Keywords: atmospheric correction; complex inland water; Sentinel- 2 MSI; water type classification

1. Introduction

The radiance over water measured by satellite sensors has two important contributors: (1) the water itself, due to the interaction of sunlight with the optically active constituents (OACs), namely pure water, phytoplankton, Colored Dissolved Organic Matter and (CDOM) and Non-algal Particulates

(NAP) or suspended sediments; and (2) the atmosphere, made up of atmospheric gases and aerosols. The latter represents about 90% of the signal measured by the satellite sensor [1]. This high proportion indicates that the correction of the impact of the atmosphere on the signal captured by the satellite sensor is a fundamental step in the processing chain for obtaining accurate water quality variables. Therefore, low uncertainty surface reflectance products are essential for obtaining an adequate retrieval from remote sensing [2,3]. These retrievals have been produced regularly in recent decades with more or less accuracy in different regions of the planet and for several water types. Prieur and Sathyendranath [4] determined that it was possible to separate the water masses into two cases according to the correlation between the different OACs, the Case 1 waters (or oceanic waters) and Case 2 waters (or coastal and inland waters). Wang [5] has determined that for typical Case 1 waters, the water-leaving reflectance contributions at Top Of Atmosphere (TOA) to the signal measured by sensors varies for different bands of the visible spectrum, with the highest atmospheric contribution in the blue-green bands close to 12%. For Case 2 waters, several situations have been found and Wang [5] remarks that the percentage of water-leaving reflectance contributions at TOA differs depending on whether the waters are dominated by suspended sediments or CDOM. These OACs have a very distinct effect, with Total Suspended Matter (TSM) producing a reflectance increase in the green and red bands, while CDOM increases absorption (reducing reflectance) in the blue bands.

Obtaining an accurate estimation of the water-leaving reflectance and therefore good input for retrieving water quality concentrations, is then an important step. This atmospheric correction (AC) procedure subtracts the atmospheric contribution (aerosols scattering effects), sunglint and whitecaps from the TOA signal [6]. There is increasing interest in evaluating the performance of different atmospheric correction (AC) processors on inland and coastal waters for obvious reasons: they constitute our main sources of drinkable water, are widely used for recreation and contain high biodiversity (40% of the marine and freshwater biomass [7]). For the present analysis, we have selected six AC processors to be applied on Sentinel-2 data: Atmospheric correction for OLI 'lite' (ACOLITE), Case 2 Regional Coast Colour (here called C2RCC), Case 2 Regional Coast Colour for Complex waters (here called C2RCCCX), Image correction for atmospheric effects (iCOR), Polynomial-based algorithm applied to MERIS (Polymer) and Sentinel 2 Correction (Sen2Cor). All these processors are free for the users and are quite simple to implement and apply. Retrievals obtained by each processor are validated with in situ measurements.

The work developed here is part of the research done within the Ecological Status of AQuatic Systems with Sentinel satellites (ESAQS) project, funded by the Prometeo Programme (Generalitat Valenciana, Spain). The aim of ESAQS is to develop and validate algorithms for estimating ecological quality indicators of water bodies in the region of Valencia (south-eastern Spain); for example chlorophyll-a (Chl-a), Secchi disk depth (Z_{sd}), Colored Dissolved Organic Matter (CDOM) and suspended solids [8,9]. In this context, several small-sized lakes and some reservoirs located in the SE of Spain have been selected and biophysical parameters and water-leaving reflectances have been analyzed using satellite imagery and in situ data from field campaigns. The lakes studied here can be classified into several water types based on their chlorophyll-a content and water transparency. In some cases, the water type in a lake changes visibly throughout the year due to the influence of the surrounding farming and industrial activities on the water content. The study area is the Jucar hydrographic basin. Geographically, it is located in the central-eastern part of the Iberian Peninsula. It is a transitional area located between a mountainous interior, the Iberian and Baetic systems and the coastal alluvial plain. In the alluvial platform lies the natural reserve of the Albufera of Valencia, one of most important freshwater lagoons with an area of approximately 21,120 hectares. The predominant climate in the area is typical semi-arid Mediterranean, with dry, warm summers and mild winters. The elevations of these lakes and reservoirs ranges from 0 (Albufera lagoon) to about 700 mamsl (Contreras reservoir) and their distance to the Mediterranean Sea ranges from 3 to more than 17 km (see details in Table 1). The geographical environment gives these inland waters particular shapes, so the determination of such morphometric parameters as the shoreline development index can

help us describe the shape and physical characteristics of these lakes, lagoons and reservoirs [10]. Twenty-one cloudless Sentinel 2-MultiSpectral Imagery Level 1 scenes were free-downloaded from the Copernicus Open Access Hub of the European Space Agency (ESA) [11] to match-up with the ESAQS project field campaign days.

Similar types of exercise have been previously done before using S2-MSI data [12]; for example, in the Atmospheric Correction Inter-comparison eXercice (ACIX) [3]. ACIX is an international initiative by the Committee on Earth Observation Satellites (CEOS), the National and Aeronautic and Space Administration (NASA) and ESA, the aim of which is "to explore the different aspects of every AC processor and the quality of the surface reflectance products". The performance of ACOLITE and SeaDAS at coastal sites has been evaluated by ACIX on the coast of Romania, Italy and Belgium. The Global Lakes Sentinel Services (GLASS), funded by the European Commission (EC), researched, applied and compared several commercial and open source AC algorithms over European inland waters (Finland, Netherlands, Estonia, Italy and Sweden) [13]. Other initiatives [14] have analyzed several AC processors like C2RCC, ACOLITE, Polymer and the "standard" Sentinel 2 Correction (Sen2Cor) in the Baltic Sea. In Latin America, Souza Martins et al. [15] applied and assessed the performance of the Second Simulation of a Satellite Signal in the Solar Spectrum (6SV), ACOLITE and Sen2Cor AC processors to S2-MSI over Amazon floodplain lakes.

2. Study Area, Data and Approaches

2.1. Study Area

Since the beginning of the ESAQS Project in December 2016, many field campaigns have been carried out in the Valencia region. These field campaigns were done on cloud-free days coinciding with Sentinel-2 satellite overpasses within a time window of $\pm$3 h. Eight reservoirs and one coastal lagoon have been monitored: Benagéber, Bellús, Beniarrés, Contreras, M^a Cristina, Regajo, Sitjar, Tous and the lagoon of the Albufera of Valencia. Figure 1 shows their location on the Spanish Mediterranean coast. Figure 2 shows the irregular and diverse shapes of these small water masses and Table 1 provides details about the surface area—calculated with maximum water level height- and meters above mean sea level (mamsl), as well as the distance to the sea and the shoreline ratio [10]. This information can be useful to explain some of the behaviours of the AC algorithms on the different water masses. Most of the field campaigns were carried out in autumn, winter and spring and only four in summer due to more favourable weather conditions and availability of staff and equipment. Among others, two biophysical parameters measured in the field campaigns are the Chl-a and Z_{sd}. These were measured by the Limnology Research Team of the University of Valencia [8]. Chl-a data were obtained from water samples by the spectrophotometric method. Samples were filtered through 0.4–0.6 μm GF/F glass fiber filters, extracted according to standard methods by Shoaf and Lium [16] and the calculation methods of Jeffrey and Humphrey [17]. Z_{sd} data was measured with a Secchi disk. It is divided into four quarters, each one painted black and white. The measurement procedure for obtaining the Z_{sd} is to lower it slowly in the water until it disappears from sight. When that happens, it is possible to obtain the Z_{sd} [9]. Table 2 describes the in situ minimum and maximum values of the Chl-a and Z_{sd} parameters, together with other statistics.

Table 1. Main characteristics of the lagoons, lakes and reservoirs.

Reservoir lagoon or lake	Surface Area (km^2)	Distance to Sea (km)	Meters above Mean Sea Level	Shoreline Development Ratio (Index)
Albufera	22	1.3	0	1.4
Bellús	8	31	159	1.8
Benagéber	12.06	78	530	4.1
Beniarrés	2.6	30	320	3.2
Contreras	27.1	103	670	6
M^a Cristina	3.25	17	138	2.9
Regajo	0.83	41	406	3.2
Sitjar	3.17	22	168	3.1
Tous	9.8	39	163	4.5

Table 2. In situ chlorophyll-a and Secchi disk depth value ranges measured in Ecological Status of AQuatic Systems with Sentinel satellites project field campaigns.

In Situ Measured Parameter	Number of Samples	Minimum Value	Maximum Value	Standard Deviation
Chl-a (mg/m^3)	99	0.54	169	48.76
Z_{sd} (m)	74	0.25	10	2.79

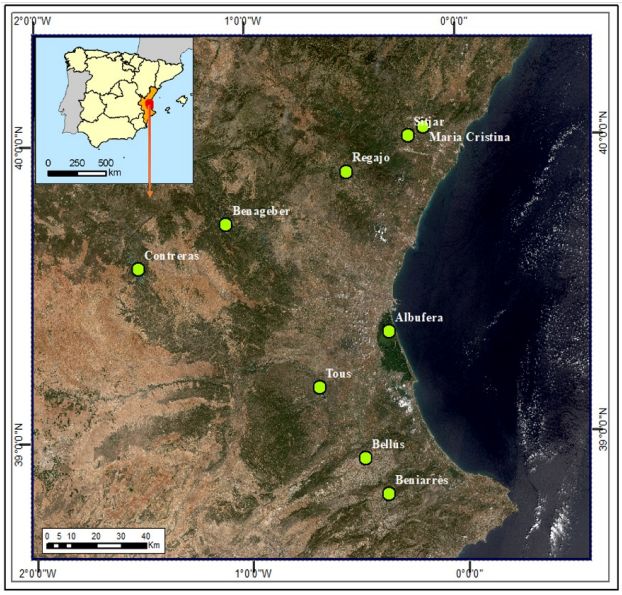


Figure 1. Reservoirs selected for carrying out field campaigns in the Valencia region, Spain.

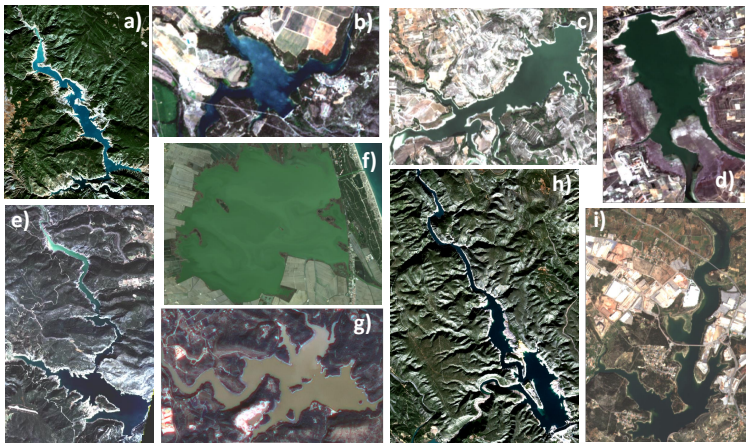


Figure 2. Reservoirs in study area: (**a**) Contreras, (**b**) Regajo, (**c**) Beniarrés, (**d**) Bellús, (**e**) Benagéber, (**f**) Albufera of Valencia, (**g**) Sitjar, (**h**) Tous and (**i**) M^{a}. Cristina.

The study area covers a wide gradient of trophic states, from ultraoligotrophic to hypertrophic. Following the trophic classification scheme for lake waters proposed by the Organisation for Economic Cooperation and Development [18], we applied this scheme to our reservoirs, lakes and lagoon based on value ranges of Chl-a and Z_{sd} measured in situ. This classification is shown in Table 3. The total number of properly measured points is 53. For each reservoir one to five measuring points were taken at a suitable distance from shoreline to avoid mixed pixels (land-water mixed and bottom reflectance contamination). In the same set of lagoons, lakes and reservoirs, Pereira-Sandoval et al. [9] demonstrated that due to the variability presented in the Chl-a values, a better adjustment in the result of the algorithms was obtained if the algorithms were calibrated and applied according to the trophic level presented. For this reason and in order to simplify the analysis presented here, we have grouped the water types into three main classes (Table 3):

Table 3. Classification scheme by water type applied over lakes and lagoons in study area.

Water Type	Description	Chl-a (mg/m^3)	Secchi (m)
Type 1	Ultraoligotrophic-to-oligotrophic	Chl-a < 2.5	$Z_{sd} > 3$
Type 2	Mesotrophic-to-eutrophic	2.5 < Chl-a < 25	$0.7 < Z_{sd} < 3$
Type 3	Hypertrophic	Chl-a > 25	$Z_{sd} < 0.7$

The aggregation into three water types is later used in several statistical analyses and plotting tests in order to understand how the atmospheric correction processors are affected by the different water components. Thus, of the 29 field campaigns carried out in lakes and reservoirs in the Valencia region, eight have data that belong to water Type 1, seven to water Type 2 and fourteen to water Type 3, see Table 4, last column.

Table 4. In situ data measurements in lakes and reservoirs.

Reservoir lagoon or lake	[Chl-a] (mg/m^3)	Z_{sd} (m)	Date (dd-mm-yyyy)	Water Type
Albufera	31.8–54.2	0.3	05-08-2015	3
Albufera	52.9–58.3	0.3	27-08-2015	3
Albufera	93.3–169.1	0.3	30-11-2015	3
Albufera	25–138.2	0.3–0.4	12-03-2016	3
Albufera	78.1–141.8	0.2–0.3	21-04-2016	3
Albufera	10.7–70.4	0.3–0.5	02-05-2016	3
Tous	1.2–3.1	5.8–6	27-12-2016	2
Bellús	31.8	1	16-01-2017	3
Contreras	0.7–2	1–1.3	08-02-2017	1
Albufera	39.7–64.5	0.2–0.3	07-03-2017	3
Beniarrés	45.4	0.9	27-03-2017	3
Benagéber	2.4–2.7	4–7.4	30-03-2017	2
M^a Cristina	1.3–1.4	5.2–5.6	06-04-2017	1
Sitjar	0.5–0.6	9.4–10.5	06-04-2017	1
Bellús	61.3–68	0.5	15-06-2017	3
Regajo	8.6–10.2	1.7–2	05-07-2017	3
Sitjar	0.6	2.7–3.1	23-10-2017	1
Benagéber	4.5–5.7	3.4–4.1	26-10-2017	2
Beniarrés	11.1–17.1	1.1–1.4	07-11-2017	3
Tous	0.6–0.7	7.1–9.1	17-11-2017	1
Contreras	0.8–2.4	4.1–5	30-11-2017	1
Tous	0.5–0.6	7–8.1	16-01-2018	1
M^a Cristina	2.7–2.9	0.7	31-01-2018	2
Sitjar	0.5–0.6	2.2–2.4	31-01-2018	1
Benagéber	2–2.4	4.3–5.5	23-02-2018	2
Albufera	81.6–84.5	0.3	07-03-2018	3
Bellús	41.5–51.5	0.4–0.5	22-03-2018	3
Regajo	4.5–5.5	3–4.2	11-05-2018	2
Benagéber	4.5–4.9	3.3–3.7	16-05-2018	2

2.2. Satellite Data

The Sentinel-2 satellites are part of the Copernicus Programme (European Commission and European Space Agency), which have on-board the MultiSpectral Instrument (from now on S2-MSI). Though S2-MSI was designed for land studies, it is possible to use it for water studies thanks to its optimized spatial resolution (10–20 m), good radiometric resolution, adequate band configuration and short revisit time (5 days using the Sentinel-2A and Sentinel-2B satellites at the equator and 2–3 days at mid latitudes), making it an optimal instrument for remotely monitoring lakes and reservoirs with a reduced surface area, as well as coastal waters. MultiSpectral Instrument (MSI) imagery Level1 (L1) from Sentinel-2A and Sentinel-2B are processed and evaluated. In this work, 21 cloudless S2-MSI L1 scenes were free-downloaded from the Copernicus Open Access Hub of the European Space Agency [11] in match-up with the ESAQS project field campaign days (see details in Table 4). The images were resampled to 10 m as part of the preprocessing. The difference between the campaign days (29) and the final number of S2-MSI L1 processed is due either to problems derived from the imagery (e.g., containing haze or cirrus clouds not observed from the ground) or to equipment issues (spectroradiometer malfunctioning).

2.3. Above-Water Radiometry Measured In Situ

All measured stations selected in the different lagoons and lakes were made at a distance of at least 3 pixels (30 m) from the shoreline to avoid or at least reduce the effect of mixed pixels, bottom reflectance and adjacency contamination from the surrounding land and breaking surf near shorelines [19]. The water-leaving radiance (L_w) was obtained from measurements taken from the bow

of a boat by sequentially measuring the total observed radiance (L_t), which includes the contributions of diffuse sunlight reflected by the water surface (L_g) and the sky radiance (L_{sky}) used for the calculation of L_g. For the field measurements, we used a ASD FieldSpec® HandHeld 2 spectroradiometer, which has a wavelength range of 325 to 1075 nm and a spectral resolution of 1 nm and an Ocean Optics (HR 4000) spectrometer ranging from 200 to 1100 nm at a spectral resolution of 0.2 nm. In accordance with recommendations by experts and the existing protocols in the bibliography [20–22], the measurements were carried out with a zenith angle of 40° and an azimuth angle of 135° to minimize sunglint perturbations. The radiometer's field of view (FOV) was delimited using a FOV of 8°. Once these general issues were defined, between three to five points were measured for each reservoir and lake. The number differs depending on the shape of the water mass and its surface area (see Figure 2 and Table 4). For each point, the measurement procedure followed consisted of taking five measurements of the water-leaving radiance (L_w) and total downward irradiance E_d using a reflectance plaque made of Spectralon® (L_{ref}). This plaque is used to normalize the uncalibrated radiance measurements from E_d. We used a gray reflectance plaque (25% nominal reflectance) that is required for the minimization of changes of illumination conditions during measurements. With these data, the in situ remote sensing reflectance (R_{rs}) is calculated from the following equations:

$$R_{rs} = L_w / E_d \tag{1}$$

$$L_w = L_t - L_g \tag{2}$$

$$L_g = rho * L_{sky} \tag{3}$$

$$R_{rs} = (L_t - -L_g) / E_d \tag{4}$$

to obtain the L_w, it is necessary to remove the radiance measured from the sky L_g from the observed radiance L_t. The result of that difference is divided by the total downward irradiance E_d.

Rho is the surface reflectance used to correct the radiance of the sky. In accordance with wind speed data measured in the ESAQS field campaigns (values equal to or less than 5 m per second) and consistent with the reflectance factor provided by Mobley 1999 and Mobley 2015 [20,23], we have decided to use a constant *rho* factor value equal to 0.028.

$$E_d = (L_{ref} / R_{ref}) / pi \tag{5}$$

the E_d term is obtained through the division between L_{ref} (the reference radiance measured using the reflectance plaque) and R_{ref} of the plaque, provided by the vendor. Up to this point, we have worked with radiance units; therefore, it is necessary to cancel out the units dividing the terms L_w and E_d by the value of *pi* to obtain the units of steradian. Finally to obtain the spectral remote sensing reflectance, we have applied the following equation:

$$R = (R_{rs} * pi) \tag{6}$$

where to obtain R_{rs}, it is necessary to multiply the terms previously obtained by the value of *pi*. Once the in situ remote sensing reflectance spectra is obtained, it is convoluted to the S2-MSI spectral bands using the Sentinel-2 Spectral Response Functions S2-SRF (SRF v2.0) [24].

2.4. Atmospheric Correction Approaches

The requirements for the selection of the atmospheric correction algorithms are based on their availability and cost zero, the ease for understanding their implementation and the possibility of correcting inherent effects like sunglint and adjacency of land pixels. All these processors are able to correct for the effect of the aerosol contributions and remove them from the water-leaving radiance with different levels of accuracy.

ACOLITE is an atmospheric correction processor for coastal and inland waters developed by the Management Unit of the Mathematical Model of the North Sea (MUMM) in Belgium [25]. It performs atmospheric correction using the dark spectrum fitting approach by default but it can be configured to use the exponential extrapolation [26–29] approach. In this work, we have applied the ACOLITE v.beta. 20180925 processor using the default approach in order to check its performance for the different water types. According to Reference [30], the algorithm would work over clear and mixed clear/turbid waters for most sensors but it would require the presence of SWIR bands to work over scenes with only turbid waters.

Case 2 Regional Coast Colour (C2RCC) is a development of the original Case 2 Regional processor [31,32]. It relies on a database of radiative transfer simulations of water-leaving reflectance (water signal) and related Top-Of-Atmosphere (TOA) radiances (satellite signal). The inversion of water signal and satellite signal is performed by neural networks. A characterization of optically complex waters through its inherent optical properties (IOPs) is used along with the coastal atmospheres to parameterize radiative transfer models for the atmosphere over the water body. The C2RCC has been improved to cover extreme ranges of scattering and absorption, now using a 5-component bio-optical model. The in-water modelling uses a Hydrolight model and the atmospheric radiative transfer is based on the SOS model [33] with aerosol properties derived from AERONET measurements. C2RCC has two versions: a version called the *normal net* (here C2RCC), with typical ranges of IOPS; and an *extreme net* version (here C2RCCCX), for extreme ranges of absorption and scattering. It is available on the SeNtinel Application Platform (SNAP v.6.0) [34]. The images were processed here according to the default processing parameters.

iCOR, previously known as OPERA [35,36], is a generic scene and sensor atmospheric correction algorithm for land and water targets. The following steps are performed: (i) identification of land and water pixels; (ii) land pixels are used to derive Aerosol Optical Thickness (AOT) based on an adapted version of the method developed in Reference [37] in the SCAPE-M algorithm; (iii) an adjacency correction is performed using SIMEC [36] over water and fixed background ranges over land targets; and (iv) the radiative transfer equation is solved. iCOR uses MODTRAN 5 [38] Look Up Tables (LUT) to perform the atmospheric correction and needs information about the solar and viewing angles (Sun Zenith Angle (SZA), View Zenith Angle (VZA) and Relative Azimuth Angle (RAA)) and a digital elevation model (DEM). The images were processed here according to the default processing parameters, applying the SIMEC adjacency correction. The present version of iCOR (v.1.0.0) does not correct for sunglint effects.

Polymer (Polynomial based algorithm applied to MERIS [39]) is an atmospheric correction algorithm for processing oceanic waters with and without the presence of sunglint. Polymer is a physical model based on a spectral optimization method called *spectral matching*. It applies all the spectral bands to make the atmospheric and sunglint correction; therefore, AC is not exclusively based on the NIR signal. The images were processed here according to default processing parameters in Polymer v.4.6. Polymer is used by the Ocean Colour- Climate Change Initiative [40]. This processor contains quality flags but they differ from C2RCC (Table 5).

Sentinel 2 Correction (Sen2Cor) is designed exclusively for Sentinel-2 Level 2A land products [41]. It is based on the dark dense vegetation approach (DDV) [42]. This method assumes that the vegetation is sufficiently dark and the ratio between the bottom of the atmosphere reflectance at different wavelengths is constant. This algorithm requires some pixels in the image to correspond to dense dark vegetation. Once the presence of such pixels is established, the algorithm automatically chooses these pixels, derives the AOT and corrects the image [43]. One main difference with the preceding processors is that it considers a lambertian surface, while the air water interface has a specular reflection [44]. The images were processed using the default mode.

Table 5. Flags used from C2RCC, C2RCCCX and Polymer.

AC Processor	Flag	Meaning
C2RCC, C2RCCCX	Rtosa_OOS	The input spectrum to the atmospheric correction neural net was out of the scope of the training range and the inversion is likely to be wrong
	Rtosa_OOR	The input spectrum to the atmospheric correction neural net out of training range
	Rhow_OOS	The Rhow input spectrum to the IOP neural net is probably not within the training range of the neural net and the inversion is likely to be wrong.
	Rhow_OOR	One of the inputs to the IOP retrieval neural net is out of training range
	Cloud_risk	High downwelling transmission indicates cloudy conditions
Polymer	!bitmask & 1023 == 0	invalid pixels

3. Preparation of Match-Ups

Once the S2-MSI data atmospherically corrected by processors were obtained, we extracted the pixel value reflectance on a 3-by-3 window centered on the geographical coordinates of the in situ stations. To assess the performance between the imagery reflectance and the in situ reflectance, we did two tests: (i) using all match-ups without applying quality control flags; (ii) applying quality control flags. The selection criteria for match-ups included a time window of $\pm$3 h between the satellite overpass and the in situ measurement times. The image pixels are extracted over a 3-by-3 S2-MSI macro-pixel (30-by-30 m) centered on in situ measurements for ACOLITE, C2RCC, C2RCCCX, iCOR, Sen2Cor and Polymer processed scenes. We calculated the average reflectance for each macro-pixel, applying a regular outlier calculation to remove suspicious pixels from the average. We recalculated the mean and standard deviation with the valid pixels remaining and those were used to determine the coefficient of variation, which has to be below 15% to assure homogeneity. Finally, the number of remaining pixels within the macro-pixel are counted and if the number is higher than half the original macro-pixel (i.e., 5 pixels over 9), the macro-pixel is valid and taken into account [45,46]. The total number of macro-pixels (N) left by type of processing is shown in Table 6, middle column.

The description of the flags used to mask out invalid pixels when applying quality control flags included in C2RCC, C2RCCCX and Polymer- is specified in Table 5. With the unmasked pixels left, the outlier calculation is carried out as explained in the previous paragraph. The total number of macro-pixels left by processor is shown in the right-hand column of Table 6. The average number of match-ups is around 50 points, C2RCCCX has the minimum number of match-ups (37), while Sen2Cor has the maximum with 62 points (See Figure 3).

Table 6. Number of macro-pixels by AC processor to be used in the match-up analysis.

AC Processor	N Total	N Flagged
ACOLITE	56	
C2RCC	53	43
C2RCCX	37	27
iCOR	60	
Sen2Cor	62	
Polymer	52	40

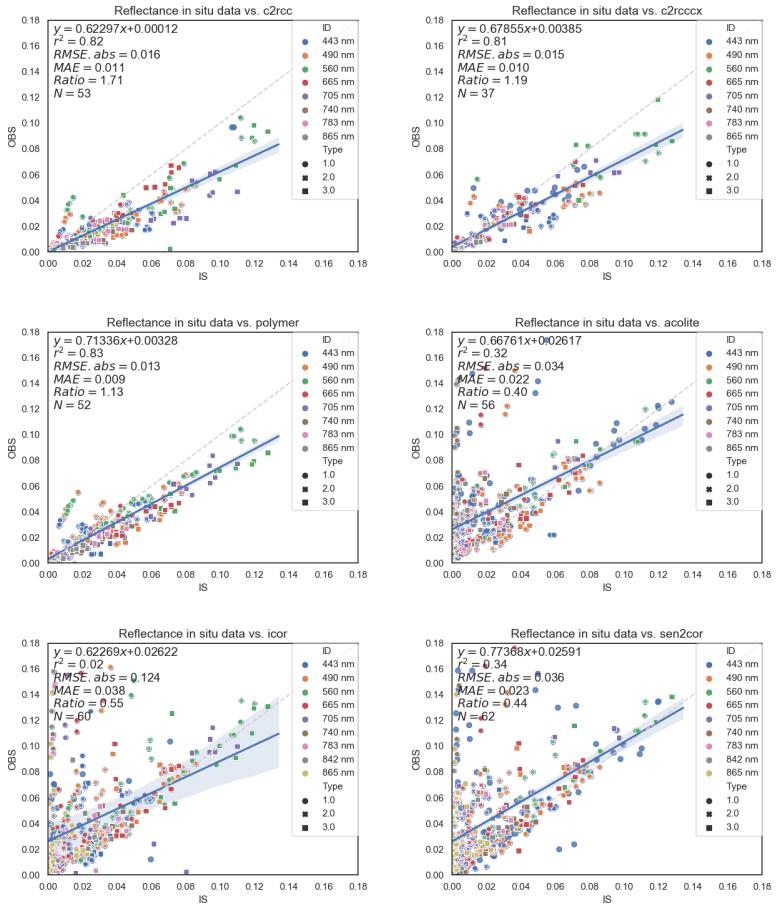


Figure 3. Match-ups of the total spectrum by AC processor, including all data. IS corresponds to reflectance measured in situ and OBS corresponds to the reflectance derived by the S2-MultiSpectral Imagery (MSI) sensor. Each S2-MSI band is depicted by a specific color. The water type is symbolized by a circle for Type 1 (ultraoligotrophic-to-oligotrophic), a cross for Type 2 (mesotrophic-to-eutrophic) and a square for Type 3 (hypertrophic).

We have evaluated the performance of the S2-MSI reflectances when compared to the in situ reflectances and plotted the results and statistics per AC processor. We used ordinary least square metrics: the coefficient of determination (R^2) and the root mean square error (RMSE). We also derived other metrics to evaluate non-Gaussian distributions in order to understand the systematic error and accuracy through the bias and the mean absolute error (MAE) [47].

4. Results

It does not seem feasible to develop a universal algorithm for deriving bio-optical parameters in variable and complex waters. For that reason, a previous classification of the water types is a good compromise for improving the inversion of bio-optical parameters [48]. According to the results

obtained by Pereira- Sandoval et al. [9], we decided to apply a water type classification, grouping the reservoir data into three types based on the Chl-a and Z_{sd} values: ultraoligotrophic-to-oligotrophic (Type 1), mesotrophic-to-eutrophic (Type 2) and hypertrophic (Type 3) defined in Section 2.1. This pre-classification would help us understand the results of the performance of each AC processor when taking into account the water type.

4.1. Results with All the Match-Ups

Figure 3 shows the scatter plots with the spectral performance by each AC processor. The relationship between the reflectance measured in situ (IS) is plotted against the reflectance derived by the MSI sensor (OBS). Each S2-MSI band is depicted by a specific color. The water type is symbolized by a circle for Type 1, a cross for Type 2 and a square for Type 3.

In general, it is possible to distinguish two major groups: in the first group, C2RCC, C2RCCCX and Polymer show quite a good performance, with small dispersion and measurements aligned to the 1:1 lines, though with clear biases; and in a second group, ACOLITE, iCOR and Sen2Cor depict less favourable results. C2RCC, C2RCCCX and Polymer show low mean average errors (MAE) from 0.0089 (Polymer) to 0.01 (C2RCC) and coefficients of determination (R^2) ranging from 0.81 (C2RCCCX) and up to 0.83 (Polymer), when plotting all bands together. The plots also show a satisfactory adjustment to the 1:1 line, especially for C2RCCCX and Polymer in Types 2 and 3 waters (mesotrophic and hypertrophic waters). However, C2RCCCX seems more restrictive in the procedure with a lower number of match ups (37) compared to C2RCC (53) and Polymer (52). Two common patterns are shown in these AC processors: C2RCC, C2RCCCX and Polymer show an overestimation of some data in the visible bands (443, 490 and 560 nm), especially for Type 1 waters (ultraoligotrophic-to-oligotrophic) while the bulk of the data is below the 1:1 line. In the plots with the ACOLITE, iCOR and Sen2Cor processing, results are not so good, with a lower accuracy range (MAE) between 0.022 for ACOLITE to 0.038 for iCOR and R^2 values from 0.02 for iCOR, to 0.34 for Sen2cor. One remarkable issue is that the accuracy seems higher or lower depending on the water type. For these three processors, it is possible to identify two behaviours in the dataset: a cloud of points with high dispersion above 1:1 line for in situ reflectance below 0.04 values; and a second cloud mainly pertaining to water of Type 3 (squared symbol), apparently with a better fit to the 1:1 line for practically all bands.

In order to assess the spectral dependency errors for all processors, Figure 4 shows the statistical analysis applied to all spectrum bands by AC processor: R^2, MAE, bias and the absolute RMSE per band are shown. Some patterns can be identified: C2RCC, C2RCCCX and Polymer–the green, red and yellow lines respectively- present the highest R^2 values across all bands. Visible bands show the best result for Polymer and C2RCC. Both processors have similar values but Polymer has a lower MAE range between 0.009 to 0.0014 (Figure 4, top right) with R^2 from 0.29 to 0.91 (Figure 4, top left). C2RCC presents a MAE range between 0.007 to 0.020 and R^2 values from 0.48 to 0.94. In red-NIR bands, Polymer presents a MAE range from 0.002 to 0.009 with a R^2 over 0.84. C2RCC shows a MAE between 0.02 to 0.013 and R^2 over 0.76, confirming the good performance of both processors.

The bias values of C2RCC, C2RCCCX and Polymer show negative values for the entire spectrum range, indicating a slight overestimation of the reflectance obtained by these three AC processors (Figure 4, bottom left). Sen2Cor and ACOLITE show a low positive bias, while iCOR increases the bias in the red and NIR bands quite impressively. Finally, the RMSE shows range values from 0 to above 0.1, reaching 0.22 (Figure 4, bottom right). For ACOLITE, iCOR and Sen2Cor–dark blue, magenta and light blue lines respectively- the statistical results are notably less satisfactory compared to the previous group of processors. The R^2 values in the blue-green bands are very low, between 0.18 to 0.45. The NIR bands show the worst performance, with values close to 0 for wavelengths longer than 740 nm. The MAE range is slightly higher than for the previous group of AC, between 0.02 to 0.025, indicating less accuracy. ACOLITE and Sen2Cor have very regular MAE values (0 to 0.02) for all bands. A similar trend is shown in the RMSE line, with ACOLITE and Sen2Cor presenting a similar trend close to 0.03 for all bands. iCOR is the AC processor that shows the worst performance in this analysis.

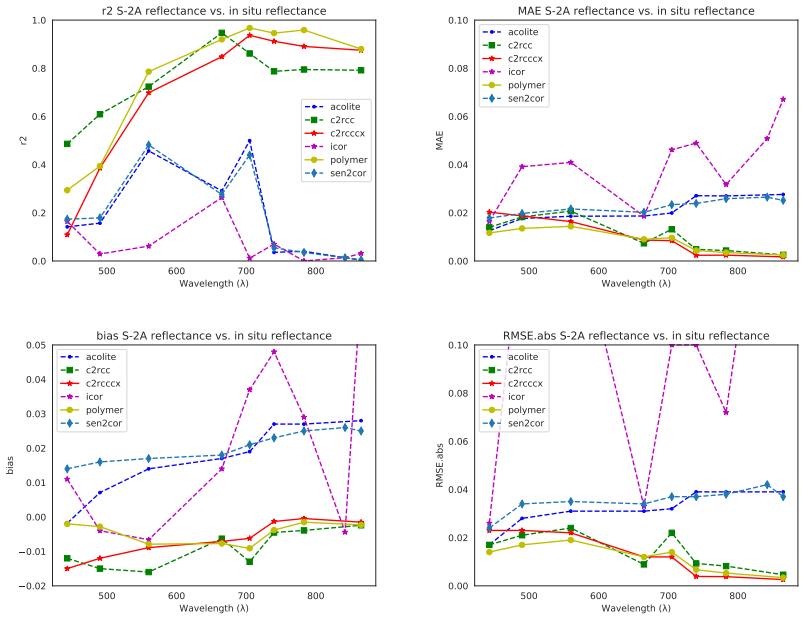


Figure 4. Summary of statistics per wavelength, all data without filtering using flags; from left to right, up to bottom: R^2 (coefficient of determination), MAE (Mean Absolute Error), bias and RMSE (Root Mean Square Error).

4.2. Results with All the Match-Ups by Water Type

Figure 5 shows the results of the performance of the AC processors by water type. The first three lines of plots show the C2RCC, C2RCCCX and Polymer performance. In accordance with the general results shown in Figure 3, Polymer and C2RCC validate their good results. For water Type 1 (ultra-to-oligotrophic), Polymer maintains good results, with a MAE of 0.007 and R^2 value equal to 0.81, with the best performance. For water Type 2 (mesotrophic-to eutrophic), C2RCC and Polymer show similar performance, though Polymer has less absolute error (MAE 0.004) than C2RCC (0.007). The coefficient of determination is quite high for both processors, achieving values of 0.93. For water Type 3 (hypertrophic), Polymer again shows a quite good adjustment, but MAE increases to 0.015, closer to the one obtained in the general results of 0.008 (Figure 3). C2RCCCX seems to perform better for water types 2 and 3. The second group of processors does not perform as well, as already indicated. ACOLITE, Sen2Cor and iCOR in general show lower accuracy, but again, at least ACOLITE and Sen2Cor show some improvement in water types 2 and 3 (R^2 values between 0.40 and 0.55 and MAE ranges from 0.012 to 0.022). For oligotrophic waters, the results are quite poor for all these processors. iCOR still shows the worst performance in water Type 2 with a big dispersion of data, practically no correlation and higher errors.

4.3. Results with All the Match-Ups by Water Type and Spectral Band

In order to better understand the results, Table 7 gives further details by separating each water type per AC processor and S2-MSI spectral band. For water Type 1, from 443 to 665 nm bands, Polymer has a lower MAE (0.004 to 0.015) and a higher R^2 range (0.55 to 0.89) than C2RCC, with MAE range of 0.006 to 0.020 and R^2 range 0.39 to 0.83. The highest MAE values correspond to the 490 nm band with values of 0.015 and 0.020, respectively for Polymer and C2RCC. In NIR bands, both processors give

similar results, with slight better performing of C2RCC. Nevertheless, the errors are smaller than in visible bands. The R^2 range for Polymer are between 0.64 to 0.82 and for C2RCC are between 0.72 to 0.86, except for band 856 nm. This band has the lowest coefficient of determination, 0.006 and 0.12, correspondingly. In water Type 2, the visible spectral region shows optimum results in Polymer. Here, MAE is between 0.005 to 0.008 and the R^2 values range from 0.61 to 0.99. It is followed by C2RCC but with larger errors than Polymer (MAE between 0.011 to 0.018) and R^2 range between 0.72 to 0.98. In NIR spectrum, C2RCC and Polymer have similar results: MAE values between 0.001 to 0.006 and R^2 values between 0.92 to 0.99. In this water type the 865 nm band, in both AC processors, has a better performance than in water Type 1. Finally, for water Type 3, Polymer shows better results than C2RCC in the visible spectrum with a MAE range between 0.007 to 0.018 and R^2 range from 0.88 to 0.90. In NIR bands, Polymer has the better performance of the two processors, with MAE between 0.07 to 0.019. Band 865 nm performance is quite better here compared with C2RCC and C2RCCCX.

The ACOLITE and Sen2Cor improvements in water Type 2 (mesotrophic-to eutrophic) and Type 3 (hypertrophic) as mentioned in Section 4.2, are concentrated in the visible bands. ACOLITE has lower errors than Sen2Cor does, with MAE between 0.009 to 0.016 and a R^2 from 0.21 to 0.87. Sen2Cor errors range between 0.015 to 0.0024 and the R^2 between 0.10 to 0.80. iCOR numbers are very low for R^2 in all the bands and the three water types, however some slight improvements can be seen in the most complex waters (Type 3), with MAE between 0.011 to 0.019 and R^2 from 0.24 to 0.46 in the visible bands.

4.4. Results Applying Quality Flags

As mentioned, only C2RCC (both nets) and Polymer processors have quality flags that can be used to identify possible invalid reflectance (see Table 5 for the flags tested here). If these flags are applied, the number of match-ups is considerably reduced for C2RCCCX (27), but Polymer and the normal net of C2RCC keep a similar number of points (around 40). For all cases, the retrieval slightly underestimates the observations, but in general a good performance is observed, with R^2 values from 0.78 (C2RCCCX) up to 0.83 (Polymer), accuracy range (MAE) between 0.0074 to 0.01 and quite good adjustment to the 1:1 line (scatter plots not shown here). Polymer shows a better performance than C2RCC in the visible spectrum. For C2RCCCX, the restricted control of the quality flags provides fewer available match-ups, but results look similar to the normal net results, improving the statistics from the green to the NIR, lowering accuracy in the blue bands. Because the application of the quality flags changes the number of match-ups but did not noticeably affect the statistics, we finally decided to work with the whole dataset, which allowed us to compare results with the other three processors that do not raise flags.

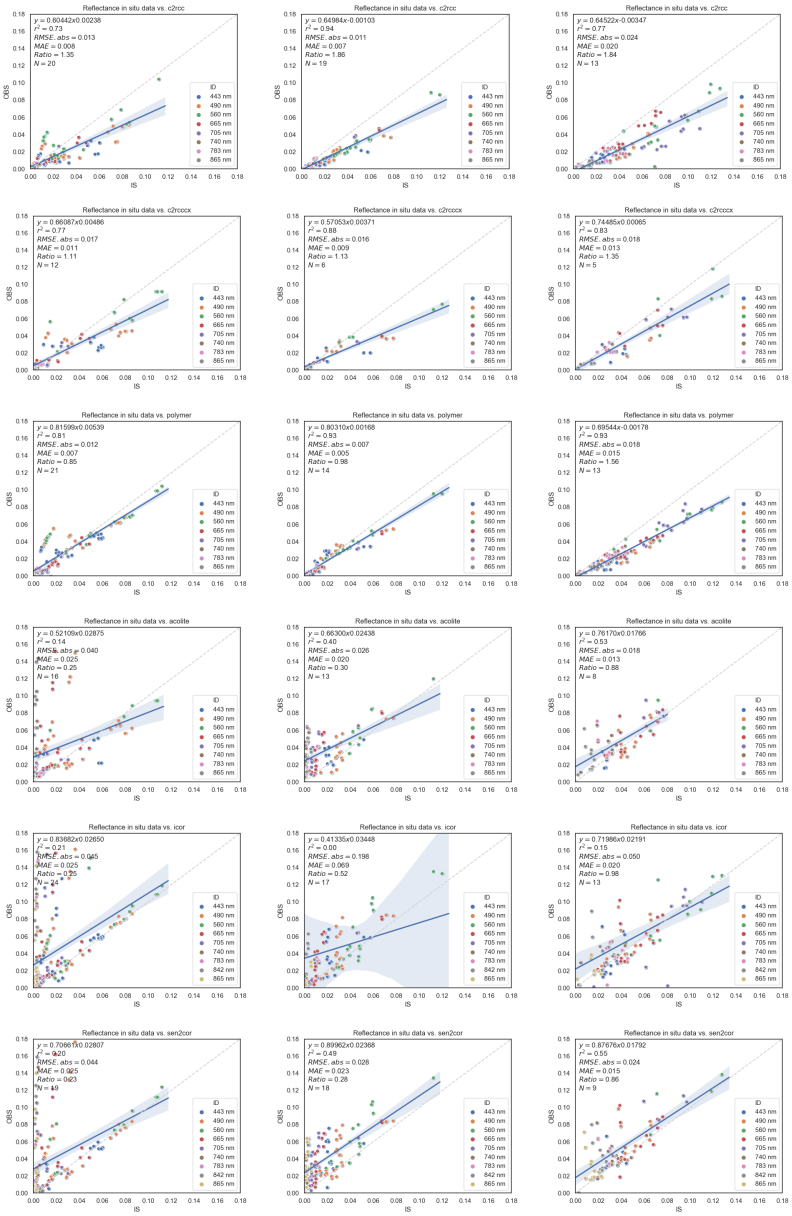


Figure 5. Summary of statistics per AC and water type. (**Left** column): Type 1; (**middle** column): Type 2 and (**right** column): Type 3.

Table 7. Scatter plots for the six AC processors by water type.

Type Water	Band	C2RCC		C2RCCCX		Polymer		ACOLITE		iCOR		Sen2Cor	
		R^2	MAE	R^2	MAE	R^2	MAE	R^2	MAE	R^2	MAE	R^2	MAE
1	443	0.39	0.015	0.09	0.022	0.61	0.011	0.12	0.018	0.14	0.019	0.65	0.010
	490	0.58	0.020	0.41	0.021	0.55	0.015	0.07	0.026	0.13	0.025	0.09	0.026
	560	0.75	0.015	0.71	0.015	0.78	0.012	0.29	0.029	0.36	0.026	0.34	0.028
	665	0.83	0.006	0.83	0.005	0.89	0.004	0.10	0.024	0.14	0.021	0.10	0.024
	705	0.86	0.002	0.83	0.003	0.82	0.003	0.06	0.023	0.06	0.024	0.02	0.029
	740	0.70	0.001	0.70	0.001	0.64	0.001	0.04	0.031	0.05	0.023	0.02	0.029
	783	0.72	0.000	0.69	0.001	0.65	0.001	0.07	0.030	0.08	0.027	0.51	0.035
	842									0.07	0.030	0.06	0.033
	865	0.12	0.001	0.47	0.000	0.06	0.002	0.27	0.031	0.08	0.032	0.10	0.030
2	443	0.72	0.011	0.11	0.017	0.61	0.008	0.21	0.010	0.20	0.015	0.10	0.024
	490	0.79	0.013	0.70	0.013	0.87	0.005	0.64	0.009	0.04	0.075	0.50	0.015
	560	0.97	0.018	0.91	0.012	0.99	0.007	0.87	0.009	0.07	0.077	0.80	0.016
	665	0.98	0.007	0.99	0.010	0.99	0.006	0.70	0.016	0.60	0.015	0.55	0.018
	705	0.97	0.004	0.99	0.005	0.98	0.006	0.58	0.021	0.10	0.150	0.36	0.019
	740	0.94	0.001	0.97	0.001	0.96	0.001	0.14	0.027	0.09	0.087	0.05	0.024
	783	0.92	0.001	0.96	0.000	0.98	0.001	0.18	0.028	0.03	0.068	0.04	0.025
	842									0.10	0.110	0.10	0.026
	865	0.80	0.000	0.86	0.001	0.72	0.001	0.20	0.029	0.11	0.183	0.11	0.025
3	443	0.51	0.015	0.03	0.018	0.47	0.015	0.34	0.008	0.24	0.011	0.06	0.014
	490	0.68	0.022	0.26	0.019	0.56	0.019	0.59	0.009	0.27	0.013	0.26	0.012
	560	0.68	0.032	0.61	0.021	0.90	0.026	0.82	0.010	0.46	0.019	0.51	0.016
	665	0.93	0.009	0.73	0.010	0.88	0.018	0.51	0.009	0.30	0.015	0.32	0.014
	705	0.53	0.040	0.77	0.014	0.89	0.019	0.75	0.009	0.63	0.023	0.41	0.019
	740	0.33	0.016	0.63	0.004	0.88	0.009	0.12	0.015	0.00	0.076	0.04	0.015
	783	0.36	0.014	0.54	0.005	0.88	0.007	0.10	0.016	0.02	0.011	0.05	0.014
	842									0.04	0.014	0.02	0.015
	865	0.35	0.008	0.54	0.003	0.82	0.004	0.00	0.016	0.02	0.010	0.30	0.015

5. Discussion

The radiance reaching the satellite sensor experiences multiple interactions within the atmosphere and the aquatic systems. Around seventy to ninety percent of a sensor-measured signal over the ocean and water bodies comes from the atmosphere. This high amount of "noise" has to be removed to obtain the water-leaving reflectance, the real value coming from the surface. To the inherent complexity of Case 2 waters, other external factors such as sunglint, wind speed, geometry of the observations and angles of the sun, together with possible land adjacency effects, are added to the sum of the sunlight detected by the sensor. This fact has led to the emergence of several AC processors that focus on oceanic, coastal and inland waters [5]. This is the reason reflectance derived from different AC processors must be analyzed and validated. These validation exercises are usually done over coastal waters and only in a few cases it is done over inland waters in the recent literature [3,12–15,19].

Atmospherically corrected S2-MSI data with six different AC processors are matched-up here to in situ measured reflectance in order to test, analyze and validate their performance. ACOLITE, C2RCC, C2RCCCX, iCOR, Polymer and Sen2Cor were tested over a set of reservoirs and lakes in the Eastern Iberian Peninsula. We deemed it necessary to apply a previous water type classification based on the trophic status of the lakes, according to the Chl-a and Z_{sd} values measured in situ: Type 1 ultra-to-oligotrophic, Type 2 mesotrophic-to-eutrophic and Type 3 hypertrophic [18]. This water type classification would help to understand the performance of the different AC processor, because this factor can affect the results in several ways: for instance, the AC algorithm may have not been trained for a certain range or another IOPs, or it might influence on the magnitude of the adjacency effect (AE) in the visible bands [49].

General statistics have shown solid results for Polymer and C2RCC. As Warren et al. [12] pointed out, this good performance from the two algorithms, which use quite different approaches (see Section 2.4), means that Polymer and C2RCC reproduced the spectral shape of the in situ data better than the others. The accuracy of these results improved markedly after the application of the classification per water type, particularly for water Types 2 and 3 (increasingly complex waters). A more detailed analysis by bands (see Table 7) allowed us to identify the strengths and weaknesses of each AC processor according to the spectral region. A detailed discussion of the results per AC method follows.

5.1. Polymer, C2RCC and C2RCCCX

In the work of Steinmetz et al. [50], the Polymer algorithm (version 4.1) retrievals compared with MERIS simulated reflectance, showed results in Case 1 water of R^2 = 0.98 at 443 nm and R^2 = 0.93 at 560 nm. Steinmetz et al. validated Polymer using *in situ* data from the Satellite Intercomparison for Marine Biology and Aerosol Determination (SIMBADA) and with the MEris MAtchup In-situ Database (MERMAID), obtaining R^2 values equal to 0.78 for 443 nm, 0.62 for 490 nm and 0.89 for 560 nm. Those results are in the same range of values as the ones observed here most notably in Type 2 (see Table 7). Figure 3 shows a particular cloud of points in bands 443 nm, 490 nm and 560 nm, which is remarkably far from the rest. This cloud belongs to the points measured in the Contreras reservoir (Type 1) on 21 September 2018. This reservoir is located at an altitude of 700 mamsl and it is possible that this elevation could influence the atmospheric correction calculations due to the pressure change and the different distance to the sensor. Regrettably, in neither Reference [50] nor in the ATDB v1. Polymer Atmospheric Correction Algorithm [51] is there any mention about altitude induced issues. In Contreras, negative values were accounted for in the retrievals of band 865 nm. This led us to do a more exhaustive analysis in this reservoir and allowed us to identify the same situation on 30 November 2017 and on 13 June 2018 in bands 740 nm and 783 nm. Another important factor is related to wind speed. In Steinmetz et al. [50] and Reference [51], the authors point out an observation about the initial correction: "the wind speed at each pixel is not known accurately, therefore, a mis-estimation of the wind speed will lead to a mis-estimation of reflectances and can possibly lead to negative values of reflectances". An over-correction of the atmospheric model, subtracting too much from the total

signal, can produce these negative values. Ruescas et al. [52] applied Polymer to the Lake Mayor of the Lake Titicaca (Peru and Bolivia) (Chl-a near to 6 mg/m^3) at an altitude of almost 4 km and negative values were also obtained from band 665 nm. But in general Polymer shows a good performance. Qin et al. [14] evaluated Polymer in the Baltic Sea using MERIS. The authors obtained R^2 values between 0.6 to 0.81 in the visible bands. For NIR, the present results are better in water Types 2 and 3 of our lakes (see Table 7), with higher coefficients of determination and lower errors than those obtained in the Baltic Sea.

C2RCC generally shows a good performance, ranking second behind Polymer, with slight differences per water type and bands. In water Type 1, the cloud of points previously described in Polymer is visible here too (Figure 3). The coefficient of determination and the errors are worse than Polymer for bands 443 to 560 nm (Table 7). This behaviour of high MAE reappears in the other two water types, which indicates poor performance of the algorithm in the blue bands. This might be due to insufficient correction of a strong Rayleigh scattering by atmospheric gases below 500 nm. These results in Type 1 are in line with the validation made by the Case 2 eXtreme project [53]. The accuracy seems to improve in the green and red bands, but there is another decrease in the 865 nm band, coincident with the results in the rest of the AC algorithms applied. For water Types 2 and 3 the errors are a bit higher than for Polymer, as we summarized in Section 4.3. The C2RCC algorithm is under constant evolution and we are aware that there is a new version being testing that improves the accuracy for the blue bands and corrects for the sunglint effect (Roland Doerffer's personal communication). We are looking forward to testing this new version in our waters.

5.2. ACOLITE, Sen2Cor and iCOR

We would like to emphasize the Sen2Cor and the ACOLITE results obtained in water Type 2 and Type 3. In the visible bands, their performance is relatively good after excluding band 443 nm. The ACOLITE range of errors are in accordance with those obtained by Souza et al. [15] using S2-MSI data in what they called "bright and dark lakes". The differences in the error values could be due to the version of ACOLITE that they use. The ACOLITE processor using the dark spectrum fitting has been validated by Vanhellemont and Ruddick over Pléiades images in turbid coastal waters of Zeebrugge (Brussels) with good results [30]. Furthermore, the authors validated its good performance in the Thames Estuary [29]. Pereira-Sandoval et al. [9] applied the dark spectrum fitting approach on a specific water Type 3 (Albufera of Valencia) and had a suitable correlation coefficient and a low MAE. In the correction procedure, for each band, the darkest spectrum is fitted to different aerosol models in accordance with Moses [54]. It is known that variations in the altitude of inland waters from the mean sea level might introduce uncertainties in the estimation of aerosol content within the atmospheric column. In our study area, the lakes and reservoirs are at different altitudes, from 0 mamsl at the Albufera lagoon to more than 700 mamsl at the reservoir of Contreras. Therefore, one possible explanation of the moderately bad results obtained with ACOLITE and other processors in some cases, could be related to the altitude variable, since it might be affecting the estimation of the aerosol optical depth. Vanhellemont [29] has also warned about how processing a whole scene with the same aerosol model to estimate the path reflectance is likely insufficient. This idea is also supported by Pahlevan et al. and others [55,56] who said that if the aerosol properties are not well represented in the aerosol models over inland waters, high uncertainties are to be expected. The new version of ACOLITE updates the Look-up table (LUT) to be applied in dark spectral fitting, with default continental and maritime aerosol models, but with the possibility of supporting the addition of more aerosol models if required [29]. This means that the altitude of the different lakes within one scene, together with their location and distance to the coast, should be taken into account. Vanhellemont [30] stressed that to use the sample of a dark pixel over land would significantly benefit the atmospheric correction. However, ACOLITE does not apply an adjacency correction approach, in which, considering the necessity of having dark pixels from shadowy land, land pixels could contribute to the contamination of the reflectance of neighbouring water pixels. This is in fact a problem that can be extrapolated to other processors.

At wavelengths above 700 nm, the impact of adjacency effects is particularly strong since neighbouring land pixels show distinctly higher reflectance [36]. Sterckx et al. illustrated that adjacency effects occur even for pixels several hundred metres away from the shoreline. Santer and Schmechting [57] indicated that for similar solar elevation, adjacency effects become negligible (less than 0.1%) only for a distance greater than 5 km from the shoreline. Thus, a correction of adjacency effects appropriate for the water/land environment is essential for reliable spectra when using approaches based in some degree on the darkest pixel correction. Both ACOLITE and Sen2Cor had in 740 and 783 nm higher values than the in situ measurements. Once established this fact, the adjacency correction may not be so necessary in some cases, that is, the perturbations caused by the adjacency effect show a great dependence on the position of the sun with respect to land, with the Fresnel value increasing if the sun is over the land portion of the image, which is not the case in many of the scenes of our study area [49].

Sen2Cor has non-negligible results in the visible spectra for water Type 2, excluding band 443 nm. Errors are generally higher, between 0.015 to 0.026, than results obtained by other authors using this AC processor [15]. This poor performance could be attributed to the algorithm approach, with the use of the DDV method, which is clearly better adapted to the atmospheric correction of land scenes. The best results obtained by Souza could be explained by the high presence of Amazon forest (darker vegetation) around the floodplain lakes. Other issues should be taken into consideration. As Bulgarelli and Zibordi [49] pointed out, the different geometry of the sensor observations, sun angles and water and land use type influence the adjacency effect contributions. Ruescas et al. [52] evaluated Sen2Cor over the Albufera lagoon, an extremely eutrophic lake with reflectance values that could approximate those of land and they had good response from the processor. According to the algorithm approach, the conclusion made was that the good response might actually be due to the high chlorophyll-a concentration (>50 mg/m^3) all year round in that water mass , which can reduce the adjacency effect, especially when the crops surrounding it are green during the spring and summer seasons [49].

In comparison with the previous processors, iCOR results were very poor. We had only moderate values in water Type 3 between bands 449 nm and 705 nm. iCOR is the only processor that uses an adjacency correction prior to the atmospheric correction. Bad results over water Type 1 and Type 2 could be explained by this factor. In the work of De Keukelaere et al. [58], observations are made over different lakes of Europe. They reported promising results for S2-MSI, except for band 443 nm, with and without adjacency correction, especially for Lake Marken. In the NIR, results are less promising, but with the SIMilarity Environment Correction (SIMEC) on, the adjacency correction seems to have a positive effect. De Keukelaere et al. [58] pointed out several important issues with iCOR. First, the surface reflectance should be representable by a linear combination of two pure green and a bare soil endmembers and the ocean and inland water do not meet this requirement unless there is some land within the scene. They recommended the user set an atmospheric optical thickness (AOT) value appropriate for each area when this requirement is not met. A second matter is that iCOR sets a fixed rural aerosol model as default. In an upcoming phase of iCOR, a valid water based AOT retrieval in combination with the current land based implementation will be done. So it is expected that this will reduce errors caused by extrapolation of AOT over large water bodies. Another issue related to the adjacency effect is that SIMEC should be used with caution "in high turbid waters, in waters with macrophyte growth or specific algae blooms or in areas where bottom effects are significant in the NIR (optically shallow waters)". Several of those cases could be present on our study area, so a new water type classification taking these variables into account, could be a path to follow in future research activities.

5.3. Other Considerations

Sunglint is the reflection of sunlight off the water surface at the same angle a sensor is viewing this surface. The smoother the water surface, the higher the visibility of this silvery effect. But often water surfaces are in motion due to waves and currents, so the sunlight gets scattered in many directions, blurring the areas and causing the component of the radiance received by the sensor to be higher than

the water-leaving radiance from sub-surface features. Since Sentinel-2 operates at near-nadir viewing angles, the probability of sunglint polluted radiance increases [59]. Of the six atmospheric correction processors tested here, Polymer includes detection and correction of sunglint. Future versions of C2RCC will also include this feature, but it is still under development. In any case, since our study areas are located in southeastern Spain and satellite overpasses in this area are usually around 10:50 local time (relative orbits 51 and 94 in descending mode), the sunglint contamination, if any, is visible in the right part of the scenes, not in the left where the lakes lie in most of the scenes. Only the Benageber and Contreras reservoirs, when using scenes on the 94 relative orbit, could be affected by sunglint. And these two cases are probably the most easily affected by other issues like adjacency effects and altitude related problems due to their high elevation and shapes (see Figure 2).

A mention of signal-to-noise ratio (SNR) could be useful as well. Higher radiometric resolutions and SNRs are required to describe the low range of reflectance value over water [60]. Pahlevan et al. [55] analyzed and compared Sentinel-2 MSI and Landsat 8 OLI SNRs and demonstrated that "MSI requires spatial aggregations to 20 or 30 m to yield reasonably smooth products comparable to those of OLI for typical TOA radiances". Resampling the S2-MSI to 20 or 30 m before processing should then be considered. Another factor related to S2-MSI is the stripes that appear in many images, which fortunately is less pronounced over turbid inland waters.

Based on the radiometric performance and the good spatial and temporal resolution of S2-MSI, its applicability for inland water studies is guaranteed. The high coefficient correlation and lower error values obtained with some atmospheric correction processors like Polymer and C2RCC support the applicability of S2-MSI for inland water studies.

6. Conclusions

The aim of this exercise is to assess the water reflectance accuracy obtained with S2-MSI through the reflectance analysis obtained by six atmospheric correction processors. The statistical linear analysis shows that Polymer and C2RCC are the processors with the highest correlation coefficients and lowest errors when comparing in situ measurements and satellite reflectance. The statistical analysis is also performed over the different water types classified according to chlorophyll-a concentration and Secchi disk depth values of in situ measurements. The water type pre-classification using basic biophysical parameters can help us select an appropriate atmospheric correction processor. Furthermore, thanks to the water type classification, it is possible to highlight the failures in the performance of AC approaches like ACOLITE and Sen2Cor in the clearest waters. The analysis made by spectral bands made it possible to distinguish certain strengths and weaknesses of the processors in the visible and near infrarred spectrum. Many other questions arose in the discussion section. For instance, if beyond the water type, it would be advisable to analyse the effect of other parameters like the shape of the lake, the altitude and the distance to the coast. This latter point is related to the aerosol model used within the ACs, because with greater distance to the sea, the aerosol type would change from maritime to continental. Since there are two AERONET (Aerosol Robotic Network) stations within the area (Burjassot, Aras de los Olmos), it might be possible to know the predominant aerosol type. We are aware that the algorithms used are under constant development and improvements are on the way, which means that future assessments will be necessary. The high variability of our dataset could help to understand better the performance of the different ACs, taking into account several new variables. Finally, due to the good results obtained with Polymer and C2RCC in water reflectance, it is possible to support the applicability of S2-MSI for inland water quality estimation, which will be applied and validated in future work. The combination of the two Sentinel-2 satellites plus the Landsat mission will increase the revisit time of measurements over lakes, helping monitor many processes -like algal blooms, which take place over a short time and are very dynamic- and which up to now have been difficult to detect and monitor with satellites.

Author Contributions: M.P.-S. processed the S2-MSI imagery with the set of atmospheric correction procesors, participated in the fieldwork and processed the field data. M.P.-S. and A.R. were responsible for the statistical analysis and wrote the bulk of the manuscript. P.U. participated in the fieldwork. A.R.-V. and J.D. supported the writing of the manuscript. C.T. automated the processing of field data. X.S.-P., E.V. and J.S. measured and processed the in situ data and carried out the laboratory analysis. J.M. is the leader of the ESAQS project.

Funding: The funding of the ESAQS project was done through the Prometeo program 2016/032 from Generalitat Valenciana.

Acknowledgments: We would like to thank the Generalitat Valenciana for the funding of the ESAQS project through the Prometeo program 2016/032. We also thank the Confederación Hidrográfica del Júcar for their support in the field campaigns and to the Spanish National Institute for Aerospace Technology (INTA) for helping us to generate the HydroLight database.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. IOCCG. *Atmospheric Correction for Remotely-Sensed Ocean Colour Products*; Wang, M., Ed.; Technical Report; IOCCG: Dartmouth, NS, Canada, 2010.
2. Vermote, F.; Kotchenova, S. Atmospheric correction for the monitoring of land surfaces. *J. Geophys. Res.-Atmos.* **2008**, *113*. [CrossRef]
3. Doxani, G.; Vermote, E.; Roger, J.C.; Gascon, F.; Adriensen, S.; Frantz, D.; Hagolle, O.; Hollstein, A.; Kirches, G.; Li, F.; et al. Atmospheric Correction Inter-Comparison Exercise. *Remote Sens.* **2018**, *10*, 352. [CrossRef]
4. Prieur, L.; Sathyendranath, S. An optical classification of coastal and oceanic waters based on the specific spectral absorption curves of phytoplankton pigments, dissolved organic matter, and other particulate materials. *Limnol. Oeanogr.* **1981**, *26*, 671–689. [CrossRef]
5. Wang, M. Atmospheric correction of Ocean Color RS observations. In Proceedings of the IOCCG Summer Lecture Series, Villefranche-sur-Mer, France, 21 July–2 August 2014; pp. 1–58.
6. Gordon, H.R.; Wang, M. Retrieval of water-leaving radiance and aerosol optical thickness over the oceans with SeaWiFS: A preliminary algorithm. *Appl. Opt.* **1994**, *33*, 443–452. [CrossRef] [PubMed]
7. Mograne, M.A.; Jamet, C.; Loisel, H.; Vantrepotte, V.; Mériaux, X.; Cauvin, A. Evaluation of Five Atmospheric Correction Algorithms over French Optically-Complex Waters for the Sentinel-3A OLCI Ocean Color Sensor. *Remote Sens.* **2019**, *11*, 668. [CrossRef]
8. Soria-Perpinya, X.; Urrego, P.; Pereira-Sandoval, M.; Ruíz-Verdú, A.; Peña, R.; Soria, J.; Delegido, J.; Vicente, E.; Moreno, J. Monitoring the ecological state of a hypertrophic lake (Albufera of València, Spain) using multitemporal Sentinel-2 images. *Limnetica* **2019**, *38*, 457–469.
9. Pereira-Sandoval, M.; Urrego, P.; Ruíz-Verdú, A.; Delegido, J.; Soria, J.; Perpinyà, J.; Vicente, E.; Moreno, J. Calibration and validation of algorithms for the estimation of the chlorophyll-a concentration and Secchi depth in inland waters with Sentinel-2. *Limnetica* **2019**, *38*, 471–487.
10. Aronow, S. Shoreline development ratio. In *Beaches and Coastal Geology*; Springer: Boston, MA, USA, 1982; pp. 754–755.
11. Copernicus-ESA. Copernicus Open Access Hub. Available online: https://scihub.copernicus.eu/dhus/#/home (accessed on 1 December 2018).
12. Warren, M.; Simis, S.; Martinez-Vicente, V.; Poser, K.; Bresciani, M.; Alikas, K.; Spyrakos, E.; Giardino, C.; Ansper, A. Assessment of atmospheric correction algorithms for the Sentinel-2A MultiSpectral Imager over coastal and inland waters. *Remote Sens. Environ.* **2019**, *225*, 267–289. [CrossRef]
13. Deliverable 3.2: Atmospheric Correction Harmonisation—GLaSS Project. Available online: https://www.glass-project.eu/assets/Deliverables/GLaSS-D3.2.pdf (accessed on 1 December 2018).
14. Qin, P.; Simis, S.; Gavin, T. Radiometric validation of atmospheric correction for MERIS in the Baltic Sea based on continuous observations from ships and AERONET-OC. *Remote Sens. Environ.* **2017**, *200*, 263–280. [CrossRef]
15. Souza Martins, V.; Faria Barbosa, C.; Sander de Carvalho, L.; Schaffer Ferreira Jorge, D.; Lucia de Lobo, F.; de Moraes Novo, E.M.L. Assesment of atmospheric correction methods for Sentinel-2 MSI images applied to Amazon floodplain lakes. *Remote Sens.* **2017**, *9*, 322.

16. Shoaf, W.; Lium, B. Improved extraction of chlorophyll a and b from algae using dimethyl sulfoxide. *Limnol. Oeanogr.* **1976**, *21*, 926–928. [CrossRef]
17. Jeffrey, S.T.; Humphrey, G.F. New spectrophotometric equations for deter-mining chlorophylls a, b, c1 and c2 in higher plants, algae and natural phytoplankton. *Biochem. Physiol. Pflanz.* **1975**, *167*, 191–194. [CrossRef]
18. Casper, H. Eutrophication of Waters, Monitoring, Assessment and Control. *Int. Rev. Hydrobiol.* **1984**, *69, 200*.
19. Kuhn, C.; de Matos Valerio, A.; Ward, N.; Loken, L.; Sawakuchi, H.O.; Kampel, M.; Richey, J.; Stadler, P.; Crawford, J.; Striegl, R.; et al. Performance of Landsat-8 and Sentinel-2 surface reflectance products for river remote sensing retrievals of chlorophyll-a and turbidity. *Remote Sens. Environ.* **2019**, *224*, 104–118. [CrossRef]
20. Mobley, C.D. Estimation of the remote-sensing reflectance from above-surface measurements. *Appl. Opt.* **1999**, *38*, 7442–7455. [CrossRef] [PubMed]
21. Zibordi, G.; Melin, F.; Berthon, J. A regional assesment of OLCI data products. *IEEE Geosci. Remote Sens.* **2018**, *15*, 1490–1494. [CrossRef]
22. Mueller, J.L.; Morel, A.; Frouin, R.; Davis, C.; Arnone, R.; Carder, K.; Lee, Z.; Steward, R.; Hooker, S.; Mobley, C.; et al. *Ocean Optics Protocols For Satellite Ocean Color Sensor Validation, Revision 4, Volume III: Radiometric Measurements and Data Analysis Protocols*; Technical Report; NASA: Washington, DC, USA, 2003.
23. Mobley, C.D. Polarized reflectance and transmittance properties of wind-blown sea surfaces. *Appl. Opt.* **2015**, *54*, 4828–4849. [CrossRef] [PubMed]
24. ESA. Spectral Response Function v2.0. Available online: https://earth.esa.int/web/sentinel/user-guides/sentinel-2-msi/document-library/-/asset_publisher/Wk0TKajiISaR/content/sentinel-2a-spectral-responses (accessed on 1 February 2018).
25. RBINS. Acolite Atmospheric Correction Processor. Available online: https://odnature.naturalsciences.be/remsem/software-and-data/acolite (accessed on 1 December 2018).
26. Vanhellemont, K.; Ruddick, K. Turbid wakes associated with offshore wind turbines observed with Landsat 8. *Remote Sens. Environ.* **2014**, *145*, 105–115. [CrossRef]
27. Vanhellemont, K.; Ruddick, K. Advantages of high quality SWIR bands for ocean colour processing: Examples from Landsat-8. *Remote Sens. Environ.* **2015**, *161*, 89–106. [CrossRef]
28. Vanhellemont, K.; Ruddick, K. Acolite for Sentinel-2: Aquatic applications of MSI imagery. In Proceedings of the ESA Living Planet Symposium, Prague, Czech Republic, 9–13 May 2016; pp. 1–8.
29. Vanhellemont, K. Adaptation of the dark spectrum fitting atmospheric correction for aquatic applications of the Landsat and Sentinel-2 archives. *Remote Sens. Environ.* **2019**, *225*, 175–192. [CrossRef]
30. Vanhellemont, K.; Ruddick, K. Atmospheric correction of metre-scale optical satellite data for inland and coastal water applications. *Remote Sens. Environ.* **2018**, *216*, 586–597. [CrossRef]
31. Doerffer, R.; Schiller, H. The MERIS Case 2 water algorithm. *Int. J. Remote Sens.* **2007**, *28*, 517–535. [CrossRef]
32. Brockmann, C.; Doerffer, R.; Peters, M.; Stelzer, K.; Embacher, S.; Ruescas, A. Evolution of the C2RCC neural network for Sentinel 2 and 3 for the retrieval of ocean colour products in normal and extreme optically complex waters. In Proceedings of the Living Planet Symposium 2016, Prague, Czech Republic, 9–13 May 2016.
33. Chami, M.; Dilligeard, E. Radiative transfer model for the computation of radiance and polarization in an ocean-atmosphere system: Polarization properties of suspended matter for remote sensing. *Appl. Opt.* **2001**, *40*, 2398–2416. [CrossRef] [PubMed]
34. ESA. SNAP. Available online: http://step.esa.int/main/download/ (accessed on 1 December 2018).
35. Vito. iCOR. Available online: https://blog.vito.be/remotesensing/icor_available (accessed on 1 December 2018).
36. Sterckx, S.; Knaeps, S.; Kratzer, S.; Ruddick, K. SImilarity Environment correction (SIMEC) applied to MERIS data over inland and coastal waters. *Remote Sens. Environ.* **2015**, *157*, 96–110. [CrossRef]
37. Guanter, L.; González-Sanpedro, M.; Moreno, J. A method for the atmospheric correction of ENVISAT/MERIS data over land targets. *Int. J. Remote Sens.* **2007**, *28*, 709–728. [CrossRef]
38. Berk, A.; Anderson, G.P.; Acharya, P.K.; Bernstein, L.S.; Muratov, L.; Lee, J.; Fox, M.; Adler-Golden, S.M.; Chetwynd, J.H., Jr.; Hoke, M.L.; et al. MODTRAN5: 2006 update. *Proc. SPIE* **2006**, *6233*, 62331F.
39. HYGEOS. Polymer. Available online: https://www.hygeos.com/polymer (accessed on 1 December 2018).

40. Sathyendranath, S.; Grant, M.; Brewin, R.; Brockmann, C.; Brotas, V.; Chuprin, A.; Doerffer, R.; Dowell, M.; Farman, A.; Groom, S.; et al. *ESA Ocean Colour Climate Change Initiative (Ocean Colour cci): Global Dataset of Inherent Optical Properties (IOP) Gridded on a Sinusoidal Projection, Version 3.1*; ESA: Paris, France, 2018.
41. ESA. SNAP-Sen2Cor. Available online: http://step.esa.int/main/third-party-plugins-2/sen2cor/ (accessed on 1 December 2018).
42. Kaufman, Y.; Sendra, C. Algorithm for automatic atmospheric corrections to visible and near-IR satellite imagery. *Int. J. Remote Sens.* **1988**, *9*, 1357–1381. [CrossRef]
43. Ouaidrari, H.; Vermote, E. Operational atmospheric correction of Landsat TM data. *Remote Sens. Environ.* **1997**, *70*, 4–15. [CrossRef]
44. Gao, B.C.; Montes, M.J.; Davis, C.O.; Goetz, A.F. Atmospheric correction algorithms for hyper-spectral remote sensing data of land and ocean. *Remote Sens. Environ.* **2009**, *113*, 17–24. [CrossRef]
45. Bailey, S.W.; Werdell, P.J. A multi-sensor approach for the on-orbit validation of ocean color satellite data products. *Remote Sens. Environ.* **2006**, *102*, 12–23. [CrossRef]
46. Müller, D.; Krasemann, H.; Brewin, R.J.; Brockmann, C.; Deschamps, P.Y.; Doerffer, R.; Fomferra, N.; Franz, B.A.; Grant, M.G.; Groom, S.B.; et al. The Ocean Colour Climate Change Initiative: I. A methodology for assessing atmospheric correction processors based on in-situ measurements. *Remote Sens. Environ.* **2015**, *162*, 242–256. [CrossRef]
47. Seegers, B.N.; Stumpf, R.P.; Schaeffer, B.A.; Loftin, K.A.; Werdell, P.J. Performance metrics for the assessment of satellite data products: An ocean color case study. *Opt. Express* **2018**, *26*, 7404–7422. [CrossRef] [PubMed]
48. Xue, K.; Ma, R.; Wang, D.; Shen, M. Optical Classification of the Remote Sens. Reflectance and Its Application in Deriving the Specific Phytoplankton Absorption in Optically Complex Lakes. *Remote Sens.* **2019**, *11*, 184. [CrossRef]
49. Bulgarelli, B.; Zibordi, G. On the detectability of adjacency effects in ocean color Remote Sens.of mid-latitude coastal environments by SeaWiFS, MODIS-A, MERIS, OLCI, OLI and MSI. *Remote Sens. Environ.* **2018**, *209*, 423–438. [CrossRef] [PubMed]
50. Steinmetz, F.; Deschamps, P.Y.; Ramon, D. Atmospheric correction in presence of sunglint: Application to MERIS. *Opt. Express* **2011**, *19*, 9783–9800. [CrossRef] [PubMed]
51. Steinmetz, F.; Ramon, D.; Deschamps, P. *Ocean Colour Climate Change (OC CCI)-Phase One*; Technical Report; ESA-ESRIN: Frascati, Italy, 2016.
52. Ruescas, A.B.; Pereira-Sandoval, M.; Tenjo, C.; Ruiz-Verdú, A.; Steinmetz, F.; Keukelaere, L.D. Sentinel-2 Atmospheric Correction inter-comparison over two lakes in Spain and Peru-Bolivia. In Proceedings of the Colour and Light in the Ocean from Earth Observation (CLEO), ESA-ESRIN, Frascati, Italy, 6–8 September 2016.
53. Nechad, B.; van der Zande, D.; Hieronymi, M.; Kraseman, H.; Mueller, D.; Stenimetz, F.; Tilstone, G.; Simis, S.; Brockmann, C.; Ruescas, A.; et al. *C2X Product Validation Report*; Technical Report; ESA: Paris, France, 2017.
54. Moses, W.J.; Sterckx, S.; Montes, M.J.; Keukelaere, L.D.; Knaeps, E. Chapter 3 Atmospheric Correction for Inland Waters. In *Bio-Optical Modeling and Remote Sensing of Inland Waters*; Elsevier: Amsterdam, The Netherlands: 2017; pp. 69–100.
55. Pahlevan, N.; Sarkar, S.; Franz, B.; Balasubramanian, S.; He, J. Sentinel-2 MultiSpectral Instrument (MSI) data processing for aquatic science applications: Demonstrations and validations. *Remote Sens. Environ.* **2017**, *201*, 47–56. [CrossRef]
56. Ahmad, Z.; Franz, B.A.; McClain, C.R.; Kwiatkowska, E.J.; Werdell, J.; Shettle, E.P.; Holben, B.N. New aerosol models for the retrieval of aerosol optical thickness and normalized water-leaving radiances from the SeaWiFS and MODIS sensors over coastal regions and open oceans. *Appl. Opt.* **2010**, *49*, 5545–5560. [CrossRef] [PubMed]
57. Santer, R.; Schmechtig, C. Adjacency effects on water surfaces: Primary scattering approximation and sensitivity study. *Appl. Opt.* **2000**, *39*, 361–375. [CrossRef]
58. De Keukelaere, L.; Sterckx, S.; Adriaensen, S.; Knaeps, E.; Reusen, I.; Giardino, C. Atmospheric correction of Landsat-8/OLI and Sentinel-2/MSI data using iCOR algorithm: Validation for coastal and inland waters. *Eur. J. Remote Sens.* **2018**, *51*, 525–542. [CrossRef]

59. Harmel, T.; Chami, M.; Tormos, T.; Reynaud, N.; Danis, P. Sunglint correction of the Multi-Spectral Instrument (MSI)-SENTINEL-2 imagery over inland and sea waters from SWIR bands. *Remote Sens. Environ.* **2018**, *204*, 308–321. [CrossRef]
60. Matthews, M.W. A current review of empirical procedures of remote sensing in inland and near-coastal transitional waters. *Int. J. Remote Sens.* **2011**, *32*, 6855–6899. [CrossRef]

Artículo 3

"Supervised Classifications of Optical Water Types in Spanish Inland Waters"

Este artículo fue publicado en *Remote Sensing* en 2022. Esta revista tuvo en 2021 (último dato disponible) un factor de impacto de 5.349 y la posición de 14/57 (Q1) en la categoría Remote Sensing.

remote sensing

MDPI

Article

Supervised Classifications of Optical Water Types in Spanish Inland Waters

Marcela Pereira-Sandoval [1,*,†], Ana B. Ruescas [1,2,†], Jorge García-Jimenez [2,†], Katalin Blix [3,†], Jesús Delegido [1] and José Moreno [1]

[1] Image Processing Laboratory, Universitat de València, 46980 Paterna, Spain
[2] Department of Geography, Universitat de València, 46010 Valencia, Spain
[3] Department of Physics and Technology, University of Tromsø—The Arctic University of Norway, 9019 Tromsø, Norway
* Correspondence: marcela.pereira@uv.es
† These authors contributed equally to this work.

Abstract: Remote sensing of lake water quality assumes there is no universal method or algorithm that can be applied in a general way on all inland waters, which usually have different in-water components affecting their optical properties. Depending on the place and time of year, the lake dynamics, and the particular components of the water, non-tailor-designed algorithms can lead to large errors or lags in the quantification of the water quality parameters, such as the suspended mineral sediments, dissolved organic matter, and chlorophyll-a concentration. Selecting the most suitable algorithm for each type of water is not a simple matter. One way to make selecting the most suitable water quality algorithm easier on each occasion is by knowing ahead of time the type of water being handled. This approach is used, for instance, in the Lake Water Quality production chain of the Copernicus Global Land Service. The objective of this work is to determine which supervised classification approach might give the most accurate results. We use a dataset of manually labeled pixels on lakes and reservoirs in Eastern Spain. High-resolution images from the Multispectral Instrument sensor on board the ESA Sentinel-2 satellite, atmospherically corrected with the Case 2 Regional Coast Colour algorithm, are used as the basis for extracting the pixels for the dataset. Three families of different supervised classifiers have been implemented and compared: the K-nearest neighbor, decision trees, and support vector machine. Based on the results, the most appropriate for our study area is the random forest classifier, which was selected and applied on a series of images to derive the temporal series of the optical water types per lake. An evaluation of the results is presented, and an analysis is made using expert knowledge.

Keywords: Sentinel-2; optical water types; supervised classification; ocean color; inland waters

Citation: Pereira-Sandoval, M.; Ruescas, A.B.; García-Jimemez, J.; Blix, K.; Delegido, J.; Moreno, J. Supervised Classifications of Optical Water Types in Spanish Inland Waters. *Remote Sens.* **2022**, *14*, 5568. https://doi.org/10.3390/rs14215568

Academic Editors: Karsten Rinke, Caren Binding, Peter Hunter and Shushanik Asmaryan

Received: 29 August 2022
Accepted: 29 October 2022
Published: 4 November 2022

1. Introduction

Remote sensing techniques are getting used increasingly often by official institutions for monitoring lakes and freshwater reservoirs, especially in view of the demand from the European Union through the Water Framework Directive [1]. These inland waters are optically more complex than ocean waters. Due to their complexity, many of the ocean color techniques that work in open waters are not valid for these waters, which are usually less dynamic but more heterogeneous in composition. To obtain accurate water quality products in inland waters, the first step should be a water type classification, as the optical properties of lakes vary in time and space depending on weather conditions, biological composition, and physical attributes [2]. For instance, an optical water type approach is used in the Lake Water Quality production chain of the Copernicus Global Land Service [3]. A single method for obtaining quality variables in all lakes, even in the same lake, may not work well [4]. For this reason, a prior optical classification can help determine the type of water existing at a certain time, and thus assign one algorithm or another to extract more

precisely the concentration of chlorophyll (phytoplankton) or the amount of suspended sediment or dissolved organic matter [5].

Inland water masses can be categorized based on their nutrient richness, which, along with their availability of light, affects the growth of vegetation and phytoplankton. When the availability of nutrients (phosphates, nitrates, and sulfates) decreases, water is shown as more transparent and clearer and is considered oligotrophic. Mesotrophic waters have a low turbidity and a medium level of nutrients. Eutrophic waters are rich in nutrients, leading to high concentrations of phytoplankton that can cause algal blooms. There are also hypertrophic cases, usually affected by human activity from fertilizer use in the surrounding fields [6–9]. Some lakes and reservoirs can be broadly classified as one type or another, but in many cases, there can be, just to mention a few possible cases, a seasonal or one-off response to a massive influx or a lack of nutrients; episodes of intense precipitation; or the opening or closing of dams in artificial reservoirs. We will therefore assume that the optical properties of lakes change with time and from natural or man-made events. The lakes and reservoirs of the Iberian Peninsula show a variety of these cases [10]. For example, algae blooms can occur in the spring, when the water temperature rises. Hot summers can lead to intense blooms that can last for weeks and even sometimes cause damage, as happens with the cyanobacteria blooms in the Albufera lagoon of Valencia [11,12]. Continuous monitoring helps to quickly detect these cases. This can only be performed efficiently using remote sensing as a complementary method to on-site observations.

The vast majority of the classifications of optical water types (OWT) in the literature are based on the characteristics of the spectral response of the surface remote sensing reflectance ($R_{rs}(\lambda)$). The optical characteristics of the water body point to the ecological patterns and its diversity, which means the type of water being monitored can be inferred by analyzing several factors related to the ($R_{rs}(\lambda)$) signal, such as the wavelength where the highest reflectance or absorption peak is found, the slope of the response curve along the wavelengths, and the magnitude of the signal. The development and tests of optical water type classifications have been carried out in several projects, such as the Global Lakes Sentinel Services1 (GLaSS, EU), Diversity-II (ESA), and CyanoAlert (EU) [13–15]. Occasionally, the results have been integrated into software, such as the SeNtinel Application Platform (SNAP) [16]. However, some of these optical water type classifications are intended for wavelengths adapted for the Sentinel-3 OLCI and ENVISAT MERIS [17,18]. With the launch of the ESA (European Space Agency) satellite Sentinel-2 Multi Spectral Instrument (S2-MSI), new possibilities opened up for using satellite images with a high spectral, temporal, and spatial resolution, whose benefits can be used to monitor small bodies of water.

The objective of this work is to establish a robust classification approach for determining optical water types based on the remote sensing reflectance from S2-MSI data. The measured reflectances are taken over optically different inland waters of the eastern Iberian Peninsula, with a variety of water dynamics. The end goal is to obtain information about the water quality of these lakes, which can be done using the information provided by the optical water types, either using the data by itself or before applying the atmospheric correction. OWTs can also provide useful information for adjusting the in-water algorithms used to estimate water quality parameters, such as chlorophyll-a (Chl-a) or cyanobacteria [19,20]. The innovation of our approach resides in the use of the so-called supervised learning classifiers. In this type of learning, the algorithm receives previously classified inputs, in this case the optical water types, from which it learns and applies said classification to new data. The work presented here is organized as follows: Section 2 is dedicated to the materials and classification methods used in this investigation, including the presentation of the study area. Section 3 shows the results with several subsections focused on the statistical results of the classifiers, examples of the application of the best performing classifier on a single lake, and time series of the OWTs on another four lakes in the study area. Section 4 is dedicated to the discussion of the results. Finally, Section 5 sums up the conclusions and points to future work.

2. Materials and Methods

2.1. Study Area

The study area is located in the eastern Iberian Peninsula, where several lakes and reservoirs are used as case studies: the Albufera lagoon of Valencia; the Benagéber, Bellús, Beniarrés, Contreras, Maria Cristina, Regajo, Sitjar, and Tous reservoirs in the Comunitat Valenciana; the Mediano and Sotonera reservoirs in the province of Huesca; and the Toba reservoir in the province of Cuenca (Figure 1). Most of these reservoirs are located within the Jucar river basin and have been monitored by the project Ecological Status of AQuatic systems with Sentinel Satellites (ESAQS) funded by the Prometeo Programme 2016/032 (Generalitat Valenciana, Spain). Following the indicators measured on several field campaigns [21], these reservoirs cover a wide gradient of trophic states and are classified from ultraoligotrophic to hypertrophic (Chl-a between 0.5 and 169 mg m^{-3}, Z_{sd} between 0.25 and 10.5 m, phycocyanin between 4 and 320 mg m^{-3}, suspended solids between 0.3 and 91 mg m^{-3}). More information about the characteristics and dynamics of the lakes can be found in [21,22].

We have selected lakes and reservoirs that show a very distinctive spectral signature in different time periods, or even different spectral response within the lake, in order to demonstrate how the classification applies on a per-pixel basis. For instance, we looked into water masses where there had been periods of intense rainfall and subsequent runoff that generated plumes in the water of different nature and where their progression across the length of the water body could be seen. Except in the case of the Albufera lagoon, where cyanobacteria dominate most of the year, the selected water masses are reservoirs at the head of river basins and are surrounded by mountainous areas with major slopes, so they collect a large amount of sediments from runoff after precipitation.

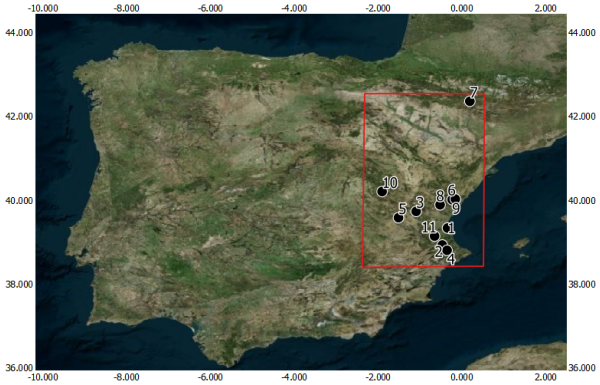


Figure 1. Map of the location of the water bodies used to obtain samples. (**1**) Albufera lagoon. (**2**) Bellús reservoir. (**3**) Benagéber reservoir. (**4**) Beniarrés reservoir. (**5**) Contreras reservoir. (**6**) Maria Cristina reservoir. (**7**) Mediano reservoir. (**8**) Regajo reservoir. (**9**) Sitjar reservoir. (**10**) Toba reservoir. (**11**) Tous reservoir. Source: IGN and Sentinel-2 cloudless, 2019.

2.2. Processing Sentinel-2 MSI Imagery

The Sentinel-2 mission is part of the Copernicus program, funded by the European Commission and ESA, where the European Space Agency contributes through launching and maintaining the satellites, in addition to collecting and pre-processing the data. The main characteristics of the mission and the MSI instrument can be found on the dedicated ESA Sentinel online web page [23]. In this work, S2-MSI Level-1C (L1C) images are downloaded and processed to Level 2 (L2). The radiometric values per pixel of the L1C are TOA reflectance (top of atmosphere). It is necessary to convert these values to surface reflectance and remove the effect of the atmosphere on the dataset. Although ESA provides

products atmospherically corrected with the Sen2Cor processor (Level-2A), the application of a non-specific atmospheric correction (AC) for inland water translates into a less accurate bottom reflectance. For this reason, we use a water-specific AC approach, the Case 2 Regional Coast Colour (C2RCC) algorithm v1.1 [24]. This neural net-based algorithm can be found in SNAP and is easily applied to the L1C images. The C2RCC processor is based on an extensive database of simulated water reflectances and their corresponding TOA radiance, with more than 5 million data. Neural networks are trained to invert the spectrum for the atmospheric correction, deriving from this process the signal from the water (water-leaving radiance) and simultaneously deriving the in-water concentrations from the inherent optical properties (IOPs). C2RCC allows for the conversion of the water-leaving radiances into R_{rs}. A review of the bibliography shows that C2RCC is the atmospheric correction that gives the best results when obtaining information on continental water masses with characteristics similar to those in our study area [2,25,26]. Specifically, in the work of Pereira-Sandoval et al. [22], it was determined that C2RCC obtained the highest coefficients of determination and the lowest deviations in relation to in situ measured reflectance on several of the lakes investigated here (average of R^2 of 0.82, RMSE of 0.016, and MAE of 0.011). In another work carried out by Urrego et al. [27], a validation process to assess the performance of biophysical parameters, such as Chl-a and total suspended matter (TSM) concentrations derived from C2RCC applied to S2-MSI imagery, confirmed the good performance of the neural net approach for waters with Chl-a below 10 mg m^{-3} and TSM lower than 10 mg L^{-1}. The C2RCC allows for the retrieval of 8 of the 12 bands of S2-MSI (R_{rs}), plus the inherent optical properties (IOPs), concentrations of constituent (Chl-a and TSM), and attenuation coefficients (K_{ds}).

2.3. Definition of Optical Water Types

An adaptation of the classes proposed by Uudeberg et al. [25,28] and Soomets et al. [2] is used as a basis. The first step involves analyzing the spectral response shape from the set of S2-MSI images corrected by C2RCC. After analyzing the spectral response and using expert knowledge of the biophysical parameters in the lakes of interest, four categories of water types are adapted: clear, moderate, turbid, and very turbid. The category "brown" is discarded because the work of Uudeberg et al. [25] is based on the boreal lakes of Estonia and Latvia, and we did not observe similar ($R_{rs}(\lambda)$) responses in this particular Mediterranean area for the cases analyzed. The identified and defined types are (Figure 2):

1. Clear: The ($R_{rs}(\lambda)$) signal is weak, with a small peak between 490 and 560 nm. These are waters with low optically active constituent concentrations and high transparency. The higher absorption is in the red and infrared (600–700 nm), and the water appears with very dark or black colors. The content of substances, such as phytoplankton, is possible but in very low quantities. For this class, the Chl-a C2RCC range found is between 0.01 and 14.0 mg m^{-3}, with an average of 2.15 mg m^{-3}, and the TSM values are below 22.2 mg L^{-1}, with an average of 4.3 mg L^{-1}.
2. Moderate: The maximum reflectance is located between 560 and 700 nm. Unlike the "Clear" category, the peak is wider due to a higher presence of substances with optical properties, although there is no one particular matter that dominates. If the signal peak is present at wavelengths 600 nm or longer, it can be linked to the presence of Chl-a, though suspended sediments and other Non-Algal Particles (NAP) in small quantities might also be present. For this class, the Chl-a range calculated with C2RCC is between 1.5 and 38 mg m^{-3}, with an average of 19 mg m^{-3}; and the TSM values are between 1.3 and 123 mg L^{-1}, with an average of 65 mg L^{-1}. In our study area, this class has the greatest number of measurements.
3. Turbid: The ($R_{rs}(\lambda)$) peak can be found closer to green and red wavelengths (500–665 nm), with a secondary peak around 700 nm. In general, reflectances are higher in magnitude than they are in the other classes. The main in-water components of this type of water are related to phytoplankton, such as chlorophyll-a pigments. For this class, the C2RCC Chl-a range is between 1.0 and 280 mg m^{-3}, with an average of 17.11 mg m^{-3};

the TSM values can reach 150 mg L^{-1}, with an average of 43 mg L^{-1}. In this case, only 8 measurements have Chl-a values > 200 mg m^{-3}, and they are in correspondence with low TSM values.

4. Very Turbid: The highest ($R_{rs}(\lambda)$) values can be found in longer wavelengths, reaching 740 nm and with a clear signal beyond 800 nm. This large spectral amplitude is due to the high concentrations of suspended sediments, mainly (but not only) to non-organic mineral particles. Waters with large amounts of phytoplankton, such as those caused by massive blooms of cyanobacteria, also cause great turbidity in the water, so this type of case could include two types of water with concentrations from different sources. For this class, the C2RCC Chl-a range can reach 364 mg m^{-3}, with an average of 27 mg m^{-3}, and the TSM values 150 mg L^{-1}, with an average of 114 mg L^{-1}. Only 7 points have Chl-a values > 200 mg m^{-3}, and they are in correspondence with same high TSM values.

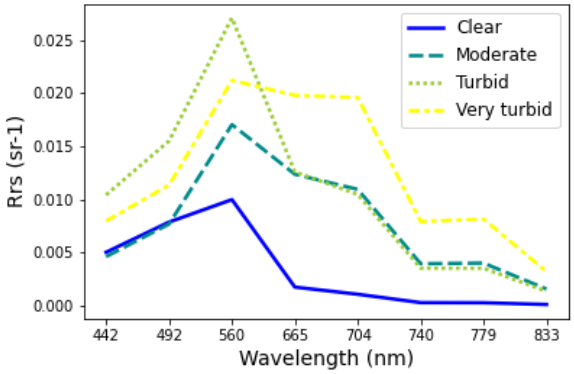


Figure 2. Spectral response of each type of OWT obtained from the average value of all training samples collected.

2.4. The Sampling Procedure

All the bands obtained from the application of the C2RCC algorithm to the S2-MSI imagery are used as inputs. Thus, a total of 21 bands are used to generate the features: R_{rs} from B1 to B8 (8); inherent optical properties (IOPs) related to the scattering of light, the backscattering of white (*iop_bwit*) and other particles (*iop_bpart*), and the total backscattering (*iop_btot*) (3); the absorption of the detritus (*iop_adet*) (1); the total absorption (*iop_atot*) (1); the absorption of the chlorophyll pigments (*iop_adg*,*iop_apig*) (2); the absorption of the yellow substance (*iop_agelb*) (1); the attenuation coefficients and the depth of the water column from which 90% of the water-leaving irradiance comes (*Kd*489, *kdmin*, and *kd_z*90*max*) (3); and the chlorophyll-a and total suspended matter concentrations (*con_chl* and *conc_tsm*) (2).

Seventeen corrected C2RCC S2-MSI images from 2017 to 2020, in different seasons of the year, are used to extract the pixel values centered on 8470 points based on 3 × 3 macropixel extraction, from which screening of valid pixels is performed following [29]. Many of them are randomly selected, others are matched-up to in situ measured reflectance and concentrations (Chl-a, TSM, Secchi disk (Z_{sd}), phycocyanin). The spectral signal of the extracted pixels is analyzed together with visual interpretation and the previously acquired knowledge of the dynamics of the lakes, and the pixels are labeled manually.

Figure 2 shows the median values per class calculated on the labeled data. The collection of the samples is carried out using the SNAP software. Python Jupyter Notebooks are developed for the screening of the original macropixels [29] and the calculation of the median, both on the macropixels and the resulting clusters after labeling with the four

classes. The database compiled is not balanced: the "moderate" class is the most sampled (nearly 4000 points), in contrast with the "very turbid" with roughly 300 points. The "moderate" and "turbid" classes are more common in our lakes compared with the "clear" and "very turbid" classes. A large part of the data comes from the Albufera lagoon. This water mass is rather round and large, which thus reduces the error introduced by the adjacency effect occurring, especially in pixels closer to the shores, and makes it quite easy to obtain good quality data. However, this lagoon is hypereutrophic and, for most of the year, there is some type of matter in suspension. That is the reason most of the labeled samples were considered "moderate" or "turbid". On other clearer and narrower water bodies, the effect of adjacency is plausible, which makes it more difficult to obtain reliable samples without a good screening of non-valid pixels. Clear waters have very weak reflectance values, especially at wavelengths greater than 665 nm, and very turbid waters occur only during very specific events, which means they do not have a great recurrence nor is it so easy to obtain good quality data from these lakes.

2.5. Classification Algorithms

In continental waters, the application of classification algorithms for OWT has generally been performed using unsupervised methods [18,30–33]. The supervised classification approaches used here are based on a training dataset with areas (pixels) known to belong to a certain type of class and which are later used to execute the classification and predict the belonging to a certain class on new data. The a priori knowledge of the dynamics of the lakes and reservoirs of our study area, acquired during field work, observation, and analysis of several types of data, provided a valuable source of information. The classification algorithms used here are framed as the so-called "supervised learning classifiers". The three most well-known families of algorithms selected are the K-nearest neighbor, decision trees, and support vector machines. According to similar experiences and optimal results obtained by different authors [34–37], and after evaluating their ease of implementation and their accuracy (see Table 1), six machine learning approaches are tested: K-nearest neighbor (KNN), decision trees (DTC), random forest (RFC), linear support vector machine (linSVC), radial basis function SVC (rbfSVC), and polynomial SVC (polySVC).

1. The K-nearest neighbor (KNN) is very easy to implement and does not require training prior to making the predictions, thus reducing computing times. It calculates the distance of a new data point to all other data points using selected approaches (Euclidean, Manhattan, etc.). K is the number of points with the least distance to the new point. These new data are assigned to the class to which the majority of the K-nearest points belong. The main disadvantage of this classifier is that it can fail if there are too many dimensions (large dataset). K is selected after testing and evaluation, with K = 5 being quite common in the literature, though some testing can be performed to determine which value is recommended for a specific dataset.
2. Among decision trees, we selected the basic expression (DTC here) and the random forest classifier (RFC) [38]: RFCs are classical algorithms based on the ensemble of hundreds or thousands of classification trees. The main idea of RFC is that the results obtained by averaging simple classifiers (trees in this case) can obtain better results than the ones obtained by single, more sophisticated, and powerful classifiers. Important and relevant features of DTC and RFC are that they are fast to train and in making predictions; that they are easily parallelizable, making them especially suitable for new hardware composed of multi-core systems; and that they provide a feature ranking (or variable importance) which indicates the most influential features. They are known for obtaining quite good results for remote sensing classification problems.
3. The support vector machine for classification, SVC [39], is a well-established, state-of-the-art algorithm for nonlinear binary and multiclass classification. It works by mapping the input (training) samples into a high-dimensional Hilbert space using a (in principle unknown) nonlinear function where a linear classification is performed by defining a hyperplane that separates the different classes with maximum margin.

The most relevant characteristics of this type of classifiers are that (i) they obtain a sparse solution where only the most relevant training samples are used, which are called support vectors; (ii) they are able to deal with noisy input samples; (iii) they generalize well, i.e., they are able to make good predictions on new, unseen samples; and (iv) they are relatively fast to train, and once a model is obtained, very fast for making predictions on new input samples.

The accuracy of the algorithms is measured by means of the overall accuracy (OA) and the Cohen's Kappa values (Kappa). The overall accuracy is expressed as the sum of the number of correctly classified values divided by the total number of values. The correctly classified values are located along the upper-left to lower-right diagonal of the confusion matrix. The total number of values is the number of values in either the truth or predicted-value arrays. If the entire set of predicted labels for a sample strictly match with the true set of labels, then the subset accuracy is 1.0; otherwise, it is 0.0 (scikit-learn.org (accessed on 1 January 2020)):

$$accuracy(y, \hat{y}) = \frac{1}{n_{samples}} \sum_{i=0}^{n_{samples}-1} 1(\hat{y}_i = y_i) \quad (1)$$

where $(y, \hat{y})$ are the measured and the predicted values, respectively.

The Kappa statistic (or value) is a metric that compares observed accuracy with expected accuracy (random chance). The Kappa statistic is used not only to evaluate a single classifier but also to evaluate classifiers amongst themselves. In addition, it takes into account random chance (agreement with a random classifier), which generally means it is less misleading than simply using accuracy as a metric. Computation of observed and expected accuracy is integral to comprehending the Kappa statistic and is most easily illustrated through the use of a confusion matrix.

$$Kappa = (observed\_accuracy - expected\_accuracy)/(1 - expected\_accuracy) \quad (2)$$

Landis and Koch [40] considered a Kappa of range 0–0.20 as slight, 0.21–0.40 as fair, 0.41–0.60 as moderate, 0.61–0.80 as substantial, and 0.81–1 as almost perfect agreement. The role of the Kappa value in this work is to serve as an auxiliary indicator for accuracy. Limitations of this value have been reported by [41].

3. Results

As has been mentioned, one of the main objectives of this work is to test different classifiers and see how they distribute each category during different events and on various lakes and reservoirs of the study area. The classifications generated can help better understand the seasonal and spatial variations of the water masses being studied and can serve as a basic support in the monitoring programs of lakes and reservoirs. In addition to being used as a final product, classifications can also be considered an intermediate product that helps in the subsequent selection of the water quality data extraction algorithm generated and adapted to specific types of water [20].

3.1. Comparison of Classifiers

Several tests are performed using the Python scikit-learn classifiers package. The original dataset is split into the training and test sets: 80%, or 6776 data points and 21 bands, are used for training; and 20%, or 1694 data points, are used for testing. To evaluate the performance of the classifiers, the overall accuracy (OA), Kappa, and confusion matrices are calculated. The input features are standardized when specific classifiers needed the data in similar ranges (KNN and SVC classifiers). The variety of the bands in the set is quite large because we used not only the information provided by the reflectances but also the IOPs and concentrations calculated by the C2RCC algorithm. Table 1 shows the accuracy of

the various classifiers implemented. We also test whether exclusively using the R_{rs} would also lead to good results.

Table 1. Statistics of the tested models. KNN refers to K-nearest neighbor; DTC, decision tree classifiers; RFC, random forest classifiers; linSVC is the linear support vector classifier; rfbSVC is the non-linear radial basis function (RBF) kernel for the SVC; and polySVC, the polynomial kernel for the SVC.

Classifier Tests					
	Specifications	**All Bands**		**Only Rrs**	
		OA	**Kappa**	**OA**	**Kappa**
KNN	2 neighbors	0.98	0.96	0.96	0.94
DTC		0.96	0.93	0.95	0.91
RFC	max_depth:70; max_leaf:34; min_split:2; n_estimators:430; min_samples_leaf:4	0.96	0.92	0.94	0.90
linSVC	lin (gamma:0.01; C:1000)	0.94	0.89	0.92	0.85
rbfSVC	rbf (gamma:0.1; C:10000)	0.98	0.96	0.96	0.93
polySVC	poly (degree:9)	0.93	0.86	0.91	0.81

The statistics show that using all 21 features gives better metrics in general, even if the differences when using only the R_{rs} are not striking. The training and processing times do not change dramatically, which is the reason we consider training the models with all the available information, assuming some redundancy. The configuration by the model is performed by tackling some of the hyperparameters (see "specifications" in Table 1). See Annex A to gain deeper knowledge on how this was conducted. The confusion matrices for all the models are also derived. Table 2 shows the confusion matrix of the RFC model. The confusion matrix was generated with the test dataset. As mentioned before, the input dataset is separated into a training set (80%), for the learning process of the models, and the test set (20%), an independent group of data used exclusively to check the algorithm performance. The test data are then classified by the model, and the comparison with the labeled data is performed by class in a confusion matrix. The results of the predictions ("Predicted") are compared with the control ("Real") or test. The overall accuracy is quite high (0.96), but we see that the commission and omission errors of the class "very turbid" are higher than those of the other classes (0.29 and 0.15). This is also visible for the other classifiers, for instance, the DTC confusion matrix shows a producer's accuracy (PA) of 0.86 (not shown here) for the "very turbid" class (see Table A1 in Appendix A for the KNN).

Table 2. Confusion matrix of RFC with all bands. EC is error of commission; EO is error of omission; PA is the producer's accuracy; UA is the user's accuracy.

	Predicted						
Real	**Clear**	**Moderate**	**Turbid**	**Very Turbid**	**Total**	**EC**	**PA**
clear	206	2	3	0	211	0.02	0.98
moderate	21	1044	17	0	1082	0.04	0.96
turbid	10	3	333	14	360	0.07	0.93
very turbid	3	0	3	35	41	0.15	0.85
Total	240	1049	353	49			
EO	0.14	0.0	0.06	0.29			
UA	0.86	1.0	0.94	0.71			
Overall accuracy						1694	0.96

3.2. Feature Importance

The RFC and the SVC linear and radial basis function (RBF) kernel allow us to check for the relevance of each feature used as input in the classifications [42]. This analysis is performed by calculating the increase in the model's prediction error after permuting that specific feature. The permutation importance of a feature uses a baseline metric, defined by

scoring, that is evaluated on a dataset defined by the X. Next, a feature column from the validation set is permuted and the metric is evaluated again. The permutation importance is defined to be the difference between the baseline metric and the metric from permutating the feature column. If this change in value makes the error increase, the feature is relevant because the model relied on it more for the prediction. Figure 3 shows the importance of each of the 21 features for the RFC model: the reflectances in the range from blue to green (B1–B3, 443–560 nm), together with the near infrared (B8A, 864 nm) appear as relevant as the scattering IOPs (*iop_bwit* and *iop_bpart*) and the absorption IOPs (*iop_adet* and *iop_atot*). The matter concentrations (*conc_chl* and *conc_tsm*) and the bands in the red part of the spectrum (B4–B7, 664–780 nm) are less significant.

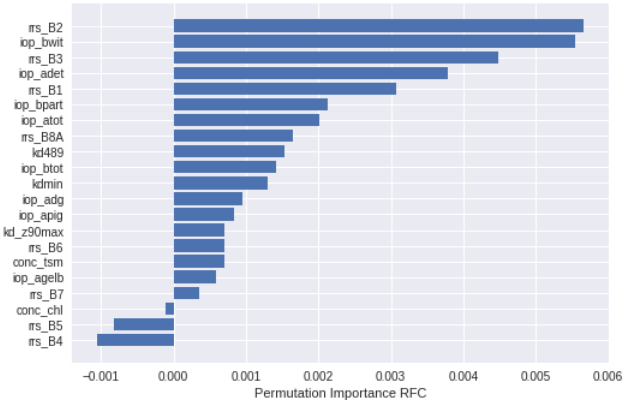


Figure 3. Feature importance derived from RFC.

The SVC (linear and RBF) feature importance is also derived, and the plots are available in Appendix B (Figure A3). There are some differences between the kernels used. The linear SVC (linSVC) gives more weight to the first four reflectance bands, followed by the scattering IOPs, while in the RBF (rfbSVC), the absorption IOPs have a more prominent role, with the IOPs related to the scattering of light (*iop_bwit*, *iop_bpart*, and *iop_btot*), together with the absorption of the detritus (*iop_adet*), the total absorption (*iop_atot*), and the absorption of the chlorophyll pigments (*iop_adg* and *iop_apig*) being the most relevant features. The reflectances seem to have a strong weight too, highlighting bands B2 (490 nm) and B3 (560 nm), followed by B1 (443 nm). The in-water concentrations of the total suspended matter and chlorophyll-a (*conc_tsm* and *conc_chl*) seem to be less significant. The depth of the penetration of light (*Kd_z90max*) and the absorption of the yellow substance (*iop_agelb*) seem to be less suited to this four-class classification. This can be partially explained by the lack of training data related to the CDOM absorption-dominant lakes, which are not represented in our training dataset.

3.3. Application of Classifiers: Performance in Contreras Reservoir

We apply the classification approaches to the lakes in the study area. As an example, we show in Figure 4 the results of the classifications using the KNN, DTC, and RFC and the three tested SVC kernel (linear, RBF, and polynomial) classifiers on the Contreras reservoir on 11 May 2019. In Figure 5, the Chl-a and TSM derived from the C2RCC algorithm are also shown on the same day for reference. The first thing we should remark is that an area in the southeastern part of the lake considered to be water is in reality the shadow of a cloud, and it shows an error on the screening of non-valid pixels when the C2RCC is applied. The four algorithms show three classes in the lake, but there are obvious differences. The DTC and RFC results are similar, with the DTC being a bit more noisy. The DTC classifies

the pixels of the SE as Class 3, while the RFC assigns those pixels to Class 1 or Class 2. The SVC results differ from them, with a dominance of Class 1 or "Clear" in the linear kernel, except in the northern part, while the rfbSVC and the polynomial kernel detect the turbidity in the SE part in a way more similar to the DTC and RFC. It is surprising to see that the shadowy area in the SE has been classified as "Turbid" or Class 3 by the rfbSVC. The KNN gives similar results to the polySVC in the central and SE areas of the lake but classifies the northern part as in the DTC and RFC (the "Turbid" and "Moderate" classes are found). With a visual analysis alone, it is difficult to establish which classification gives results closer to reality, and we cannot really determine which one is better for all areas of the lake and for all the images analyzed.

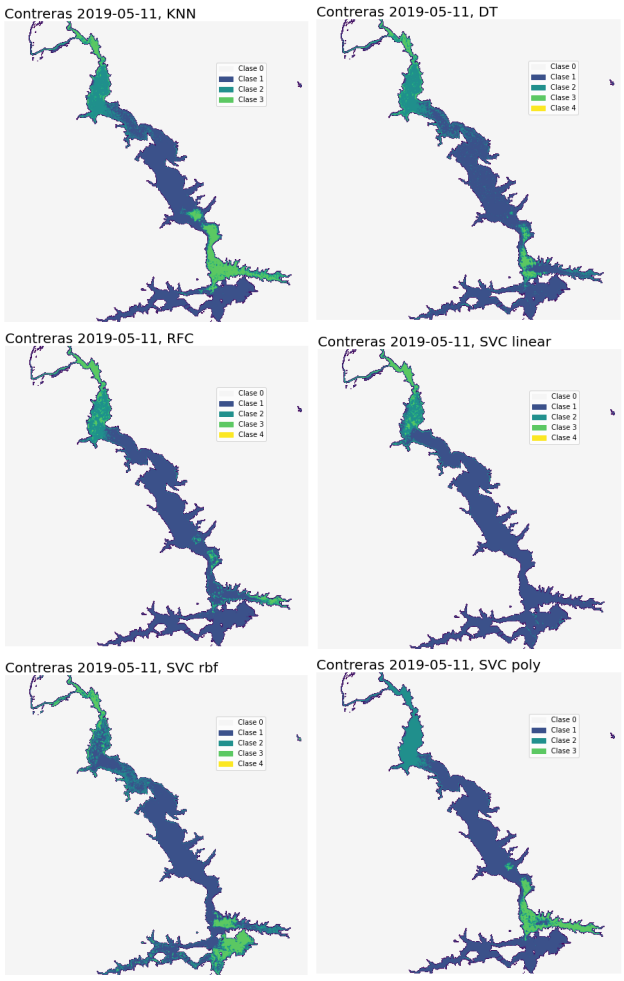


Figure 4. Comparison of the models in one image, 11 May 2019, Contreras reservoir: left top, KNN; right top, DTC; left center, RFC; right center, linSVC; left bottom, rbfSVC; right bottom, polySVC. Class 1: Clear, Class 2: Moderate, Class 3: Turbid, Class 4: Very Turbid. Class 0 is not really a class. It is assigned to the background.

Fortunately, the Contreras reservoir has been monitored by the University of Valencia on several field campaigns since 2017 (February 2017, May 2017, November 2017, June 2018, and September 20018). Among the biophysical parameters measured in situ, Chl-a, the content of the sediments, and the Secchi disk depth (Z_{sd}) measurements have been carried out. During these campaigns, the reservoir has presented low values of Chl-a (less than 2.5 mg m^{-3}), a Secchi disk between 0.95 and 4.5 m; and a range of sediment concentrations between 1.45 and 28 mg L^{-1}. According to these, the Contreras waters have been classified as ultraoligotrophic to oligotrophic throughout the year, with occasional cases of mesotrophic to eutrophic. In order to verify the separation in the area/classes made by the classifiers, we have matched those different areas with the Chl-a and TSM products derived from the C2RCC algorithm for our complete time series (Figure 6). The lake is separated into four areas: North, Center, South–Center, and South–East. In each area, a series of points are marked and the average values of the Chl-a and TSM are extracted for the available days. Focusing on the day analyzed in this example, in the N area, the values of Chl-a are 6.8 mg m^{-3} on average, and the TSM is 14.38 mg L^{-1} on average. In the central area (C), Chl-a reaches values of 2.4 mg m^{-3} and TSM 4.1 mg L^{-1}. In the SC area, the Chl-a has a 3.5 mg m^{-3} average and TSM 7.3 mg L^{-1}. Finally, the SE area shows Chl-a average values of 6.1 mg m^{-3} and TSM of 14.5 mg L^{-1}. From these results, we can infer that the classes derived by the RFC classifier are closer to the real case than the classifications made by the other approaches. We have obtained good statistics results with the RFC, and we are able to obtain information about the relevance of the bands in the model, thus making the approach a bit more explainable. Other studies have also found the RFC to be the best classifier for various types of datasets, followed closely by SVCs [43,44]. In addition, the RFC is scale invariant and does not require pre-processing, whereas the SVC needs scaling and normalization of the datasets. The RFC is computationally complex, but less so than the SVCs, in addition to being highly tunable. Bearing all these reasons in mind, we used the RFC for the time-series analysis of the status of three other lakes apart from the Contreras: the Sotonera and Mediano reservoirs and the Albufera lagoon.

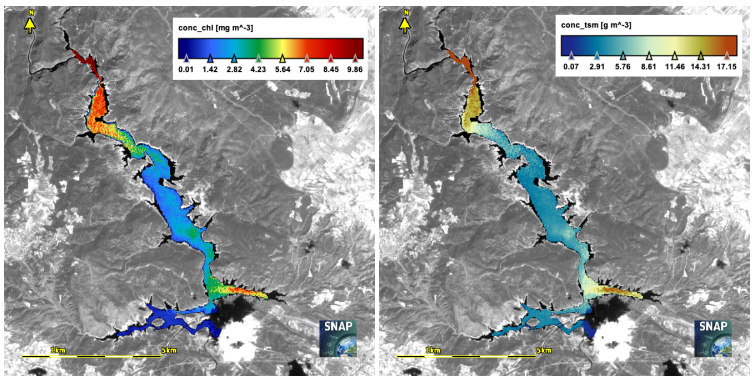


Figure 5. Ch-a and TSM concentration derived on the 11 May 2019 in the Contreras reservoir.

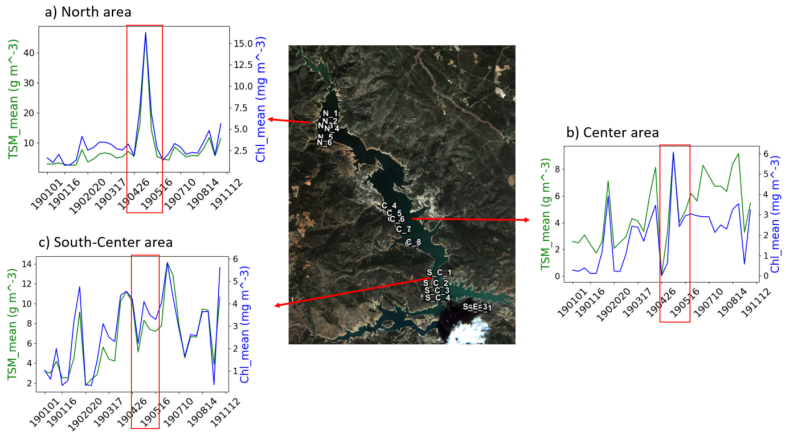


Figure 6. Time-series and areas analysis of the Contreras reservoir: (**a**) North, (**b**) Center, and (**c**) South–Center. Each figure shows the annual evolution of the average value of TSM (green line) and Chl-a (blue line). The red rectangles indicate the date analyzed, 11 May 2019.

3.4. Time-Series Analysis

The use of a time series of OWTs can help visualize possible changes to later apply the proper in-water algorithms in the retrieval of the concentration of the different components (e.g., Chl-a, TSM, and CDOM) based on specific water types. Four water bodies are analyzed here using the RFC results: the Sotonera, Contreras, and Mediano reservoirs and the Albufera lagoon. We have selected several days over 2019 as a representation of the four seasons of the Mediterranean climate. Figure 7 shows the results obtained by applying the RFC in the Sotonera reservoir. This water body is nourished by the waters of the Gállego Canal and the Sotón River, of very different origins, so the composition of its waters is also highly variable. The reservoir is intensively used for irrigation and human consumption, making its level very unstable, dropping down to 10 m in the summer [45], which, in addition to some strong slopes in its margins, generates a high risk of erosion in the catchment area of the lake. The result of the erosion and transport of materials ends up producing large TSM volumes in the water. The effect of these sediments on the water quality can be easily seen: there is not one image in which we cannot see turbidity in the water. Although officially it is considered mesotrophic [45], which indicates the possible development of biological communities, the high turbidity of the water by suspended inorganic particulates prevents greater trophic development, as it should be according to its levels of nutrients. The image on 3 February shows a turbid mass (Class 3) throughout the reservoir; on 26 May, there is a predominance of Class 1 at the south (clear water) while the area on the north appears turbid (Class 3). On 15 July, the turbidity class has extended to the south, and according to the auxiliary information, there is a high content of Chl-a. Finally, on 26 October, we see the lowest water level of the lake after a rather dry summer. Turbid Class 3 covers the water mass.

The time-series plot on the Contreras reservoir (Figure 8) tells a different story. It is one of the most optically stable water masses studied here, with predominantly clear or moderate waters due to its oligotrophic state, in which the regularly low Chl-a content (lower than 2.5 mg m^{-3}) does not affect it enough to change the water into the turbid types in most cases. Occasionally, there is an input of large amounts of sediment during events of precipitation and runoff. For instance, from 19 to 21 December 2019, a rainfall episode of great magnitude dragged a huge quantity of materials from the surrounding slopes, which ended up in the Cabriel riverbed and brought a significant amount of particulates into the waters of the Contreras reservoir. This type of event leads to an

important modification in the dynamics and ecological state of the water. In the current context of climate change, in which the increase in the frequency, intensity, and seasonal variability of extreme rainfall events is proven [46,47], a high recurrence of events of this type can lead to the modification of ecological thresholds. These thresholds are conditioned not only by the availability of nutrients but also by the usual optical properties and the development of biological communities in the photic zone. With increased sediment and reflectance of the waters, the photic zone, of vital importance for primary production, decreases in depth, modifying the bio-optical conditions and therefore possibly destroying the adapted biological communities and inducing a change in the trophic state of the waters. This type of analysis was also performed on the Mediano reservoir (Figure 8, right), where the behavior is similar to Contreras, because both are at the head of river courses. Here, the contributions of the sediments and materials surrounding the lakes are important, but seasonal. However, sediment enters the Mediano more frequently and with more evident color changes. During the summer (4 August 2019) and autumn (12 November 2019), Class 3 seems to be dominant. The values of Chl-a are not high enough, which indicates that the main concentration of the TSM reaching this reservoir consists of particles of minerals of inorganic origin (SPIM, Suspended Particulate Inorganic Matter). The fact that the reservoir is located at the head of the river prevents the arrival of nutrients of anthropogenic origin, so in this type of reservoir, the turbidity is closely related to sediment clogging rather than to problems of bacteria or eutrophication. A common pattern that we can establish between the Contreras and Mediano reservoirs is that, in both, the turbid-type water values increase throughout the summer, with its maximum presence in the month of August due to the typical summer storms in this area of the Mediterranean.

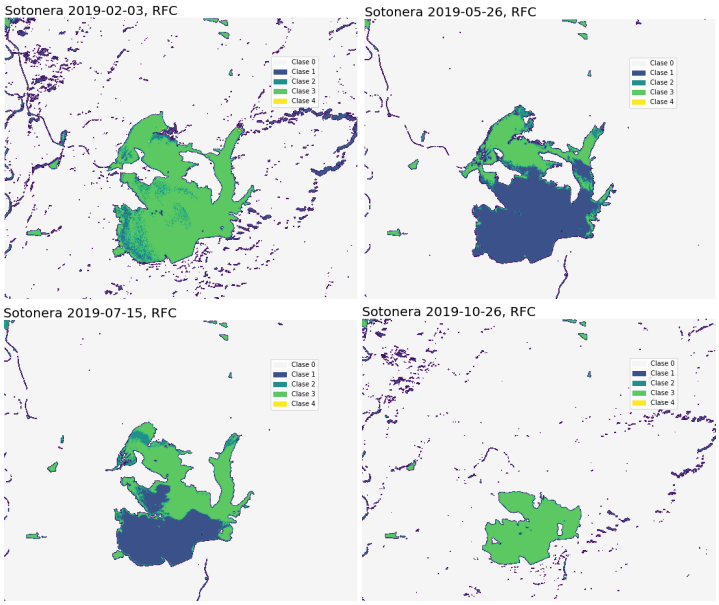


Figure 7. La Sotonera reservoir: temporal analysis of water changes using RF algorithm.

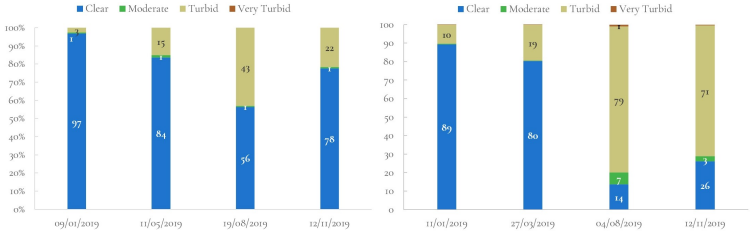


Figure 8. Temporal series of optical water types in Contreras (**left**) and Mediano (**right**) lakes.

The Albufera lagoon is a special case, where cyanobacteria is the dominant taxonomic class of phytoplankton. The Albufera is located south of the city of Valencia, between the mouths of the Turia and Jucar rivers, with an area of 2500 ha and an average depth of 1–3 m [48]. It is one of the most important wetlands in the western Mediterranean. Three ecosystems can be differentiated: the lake, which gives the park its name; the marsh and its rice fields (14500 ha); and the Devesa sandbank that separates the lake and the sea. However, its location, surrounded by the metropolitan area of Valencia and a large area of rice fields, traced by a network of ditches and canals, results in an excessive input of nutrients, fertilizers, and urban and industrial discharges that, together with the decrease in the flow of water entering the lagoon from its hydrographic basin due to climatic change conditions coupled with the increase in water concessions in the upper areas of the Jucar River, poses a growing threat to this emblematic space. A hypertrophic state is manifested in an explosion of phytoplankton that greens and darkens the water, in recurrent cycles of anoxia and in a radical fall in fish production, particularly of the species of greatest commercial value. Its dynamics follow patterns that depend on surface inputs and the management of the water system according to the needs of the rice crops that surround it. The average annual Chl-a since the late 1980s is 150–160 mg m^{-3}. In the images shown in Figure 9, we can observe the water covering the rice fields surrounding the lake at the beginning of the year. The rice fields are flooded in November and stay covered by water until the end of February when the gorges, until now closed, are opened and the fields dry. The water classes on 11 January 2019 are classified from moderate to very turbid. As mentioned earlier, the water level is about 20 cm above the annual average and the fields below the lagoon level are flooded to promote the decomposition of organic remains. In the summer, the turbid class (Class 3) seems to be dominant. Sòria-Perpinyà et al. [48] carried out a study about the presence of phycocyanin during 2016 and 2017 and determined that phycocyanin presents a strong decline in the summer period, which indicates a decrease in the primary production motivated by the depletion of nutrients after the production peak in the late spring.

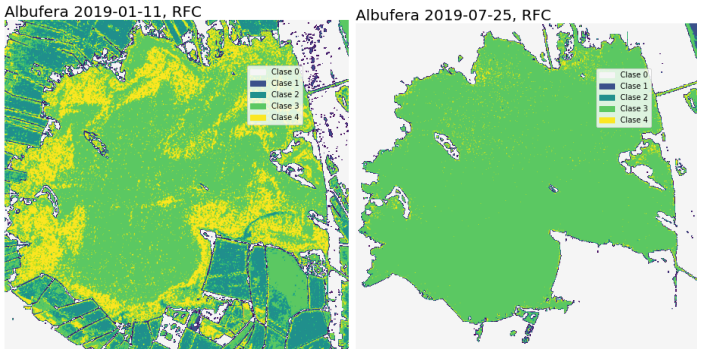


Figure 9. Albufera of Valencia lagoon: temporal analysis of water changes using RF algorithm.

4. Discussion

Examining the results obtained using the different classifiers, we see that the OWT supervised classifications show good statistics in general, with pixels assigned into the four defined classes with more or less accuracy. However, we are aware that knowing the actual constituents of the water at any one determined moment is not so straightforward because similar spectral curves may indicate different compositions or constituents in the water. This makes it complex to quickly infer the process or processes that can affect the water masses.

One of the main weaknesses of the database generated here is that it is only based on very good, easily classifiable spectral curves. However, nature usually does not respond to such defined patterns. The collection of the training samples and the manual labeling process were performed mainly in one year (2019) on different lakes but with a clear dominance of the samples taken for the "moderate" class (Class 2). This was partially corrected when the models were trained because the SVCs and RFC allow the classes to be balanced internally.

It is also sometimes difficult to distinguish between moderate and turbid waters because the only thing distinguishing these two classes is a higher magnitude of the signal peak at 560 nm. The various tests performed on the types and number of inputs of the training models were conducted solely to examine how a classifier with a higher quantity of input data but with less specific characterization of each sample would behave in comparison with a classifier with less data but a better water characterization (i.e., adding IOPs, concentrations, and geometries). What became clear is that thanks to the biophysical parameters derived by the C2RCC algorithm, the more highly detailed water characterization greatly facilitated the distinction of the water types, especially when very similar spectral curves were obtained, corresponding to different constituents in the water. For instance, although two bodies of water can be classified as very turbid, knowing the chlorophyll content derived by the C2RCC made it possible to determine if the turbid event is produced by decaying algal blooms or if it is only being produced by suspended inorganic particulate matter. We can venture that the addition of auxiliary information, such as the contents in suspended particulate organic matter (SPOM) or phycocyanin, could be helpful for making further distinctions of more water types.

We cannot really draw a conclusion as to the most appropriate supervised classifier based on the statistics or the analysis made on the images. In some cases, one model works better than the others, but the results change within different water masses. We decided to use the RFC for several reasons: the strong literature support, how easily understood it is compared to other complex models, and finally, because it is possible to tune the hyperparameters to make it more adjustable to the input database, thus avoiding overfitting. The RFC can handle binary features, categorical features, and numerical features. Very little pre-processing needs to be performed because there is no need to scale or transform the features. The feature importance permutation analysis is also a relevant reason for selecting this model as it makes the results a bit more understandable. Other reasons are: it works very well with high dimensionality; it is parallelizable (faster computation time); training times are fast; it handles unbalanced data, thus minimizing the overall error rate; and each decision tree in the forest has a high variance but low bias. Concerning this last point, all the trees in the random forest are averaged and, consequentially, the variance is averaged as well, so we obtain a low bias and a moderate variance model.

The OWT classification can be used as a final product for the analysis of water changes [18], or as an intermediate step for selecting the in-water algorithm, or even the most appropriate atmospheric correction for each water type [49]. Ideally, the OWT classification should be performed on non-atmospherically corrected data, which would save time and loops in the processing of large datasets. Work is being done in this direction, testing the bottom of the Rayleigh reflectance as input instead of the fully corrected L2 data.

5. Conclusions

A classification of optical water types, as a monitoring or diagnostic tool of the initial state of water bodies, has been implemented here. We have trained several supervised classification models (the KNN, DTc, and RFC and variations of SVCs) with a manually labeled dataset with four different water types. The features for the identification of the four classes were based on the spectral curves and the information derived from IOPs and concentrations calculated by the C2RCC neural net over Sentinel-2 images. The ability to apply supervised classification algorithms to obtain the OWT has been demonstrated, achieving optimal results in lakes and reservoirs with different levels of complexity in the SE of the Iberian Peninsula. Although the statistical results of the different classifiers are very similar, the RFC model has been the one selected for the processing of the time series. The analysis of the OWT time series of several lakes has been used to validate the classification results, in comparison with the expected dynamics of the water masses. The information provided by the OWT can be used as a final product, for instance, for detecting different types of algae blooms or water dynamics. It can also be used as an intermediate product for gaining further knowledge on the status of the water and for generating more accurate reflectance or water quality concentration parameters. Future work will be focused on enhancing the labeled dataset. To achieve this, more work must be done for the manual labeling, continuing with the methodology used here. Trying online machine learning models is another possibility, in which data become available in a sequential order and are used to update the best predictor/classifier for future data at each step (incremental learning [50]). Another topic of interest is the knowledge transfer on lakes with similar characteristics in other geographical regions, e.g., Lake Balaton in Hungary, or testing the models on more extreme water masses, such as the Mar Menor in Murcia (Spain), for different times of the year and conditions.

Author Contributions: Conceptualization and writing, M.P.-S. and A.B.R.; methodology, A.B.R., J.G.-J. and K.B.; formal analysis, M.P.-S. and A.B.R.; resources, M.P.-S. and J.G.-J.; data curation, A.B.R.; review and editing, J.D.; review, editing, and funding acquisition, J.M. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Acknowledgments: We would like to thank Antonio Ruiz-Verdú for improving this paper with helpful comments and constructive suggestions.

Conflicts of Interest: The authors declare no conflict of interest.

Appendix A

The KNN configuration consisted of changing the K value to see how this variation would affect the metrics. As can been seen in Figure A1, either a value K of 2 or 3 would lead to low-value errors. The accuracy value in Table 1 shows that the difference between neighbor 1 and 3 is only 0.004, but due to the fast processing of the algorithm, it is worth using more neighbors in order to improve the accuracy.

The confusion matrix of the KNN classifier implementation with all the bands is shown in Table A1. Both errors of omission (misclassified) and commission (false positives) occur in all cases, except for class "very turbid", less than 0.1. The producer's accuracy of this class is the lowest (0.63) due to the higher commission error (7 pixels assigned to class "turbid") and the user's accuracy is 0.87. It seems clear that the lower number of samples, especially for the "very turbid" class, is introducing some errors. Still, the overall accuracy is 0.97, which is considered very good.

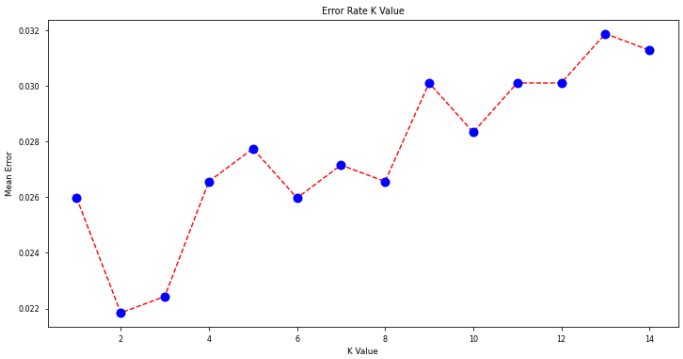


Figure A1. K-value-associated error of the training dataset.

The goal of the support vector classifier (SVC) is to create the best line or decision boundary that can segregate n-dimensional space into classes in a way that a new data point can be easily put in the correct category in the future. This best-decision boundary is called a hyperplane. The SVC chooses the extreme points/vectors that help in creating the hyperplane. These extreme cases are called support vectors, which is the reason the algorithm is named the support vector machine (SVM). Within the scikit.learn package in Python, there are two types: SVC linear and SVC non-linear (the radial basis function (RBF) kernel and the polynomial kernel). Besides the kernels, the gamma parameter defines *"how far the influence of a single training example reaches, with low values meaning 'far' and high values meaning 'close'. The gamma parameters can be seen as the inverse of the radius of influence of samples selected by the model as support vectors"*. The C parameter trades off the correct classification of the training examples against the maximization of the decision function's margin, that is, C behaves as a regularization parameter. Figure A2 is a heatmap of the classifier's cross-validation accuracy as a function of C and gamma. The gamma parameter has a big influence on the model. If gamma is too large (>1.0) or too small ($<1 \times 10^{-3}$), the radius of the area of influence of the support vectors only includes the support vector itself and no amount of regularization with C will be able to prevent overfitting or smoothing effects. Smooth models (lower gamma values) can be made more complex by increasing the C values up to one limit. However, it is usually recommended to use the smallest C values possible. There is a trade-off in between very high C values, which increase the fitting time, and lower C values, which may increase the prediction time (more support vectors needed). For each type of kernel (linear, RBF, or polynomial), the gamma and C parameters were analyzed and selected, as shown in Table 1.

Table A1. Confusion matrix KNN with all bands. EC is error of commission; EO is error of omission; PrAc is the producer's accuracy; UsAc is the user's accuracy.

	Predicted						
Real	**Clear**	**Moderate**	**Turbid**	**Very Turbid**	**Total**	**EC**	**PrAc**
clear	196	8	7	0	211	0.07	0.93
moderate	11	1069	2	0	1082	0.01	0.99
turbid	4	2	350	4	360	0.03	0.97
very turbid	5	3	7	26	41	0.37	0.63
Total	223	1080	354	36			
EO	0.08	0.01	0.04	0.13			
UsAc	0.91	0.99	0.96	0.87			
Total accuracy						1694	0.97

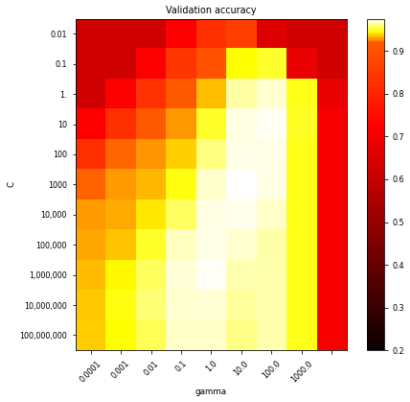


Figure A2. SVC radial basis function parametrization.

Tree-based models split the data multiple times according to certain cutoff values in the features, creating new subsets of data that can be split again until the so-called leaf nodes are achieved. The prediction of these final nodes is the average of the training data of each specific node. There are various algorithms that can grow a tree, with the differences being: the structure of the tree, the criteria for the splits, when to stop splitting, and how to estimate the simple models within the leaf nodes. Decision trees are a non-parametric supervised learning method. The subset method by default in the scikit.learn module is the Gini index, which marks how "impure" a node is. A pure node will be classified as only one class. Here, we used the best splitter, and the rest of the parameters of the algorithm were left by default. More tests were conducted using a specific type of decision tree, the random forest classifier (TRFC). These algorithms had worked with success in a test on Sentinel-2 images for OWT classification, and we conducted some tests with the current dataset. The main issue with the decision trees, and the RFC in particular, is that they can be interpretable while keeping them short. In order to find the best estimated parameters, we applied a randomized search of hyperparameters to select the best parameters and tuned the model accordingly. Table 1 shows the final maximum depth (70), maximum numbers of leafs (35), minimum splits (35), and number of estimators (430), with bootstrap and using the maximum features. The interpretability of the tree is certainly difficult in the present case because the number of terminal nodes increases quickly with depth. The more terminal nodes and the deeper the tree, the more difficult it becomes to understand the decision rules of the tree.

Appendix B

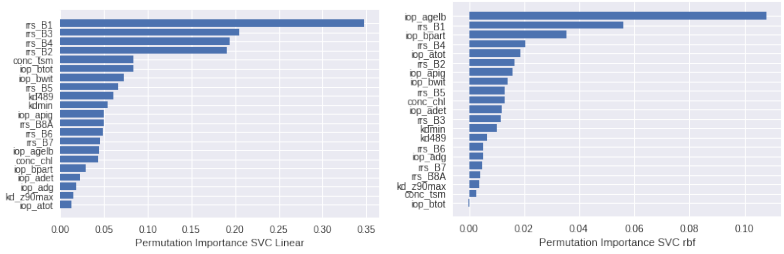


Figure A3. Feature importance derived from SVC.

References

1. European Parliament 2000. Available online: https://ec.europa.eu/environment/water/water-framework/index_en.html (accessed on 8 May 2022).
2. Soomets, T.; Uudeberg, K.; Jakovels, D.; Zagars, M.; Reinart, A.; Brauns, A.; Kutser, T. Comparison of Lake Optical Water Types Derived from Sentinel-2 and Sentinel-3. *Remote Sens.* **2019**, *11*, 2883. [CrossRef]
3. Copernicus Global Land Service. Available online: https://land.copernicus.eu/global/products/lwq (accessed on 8 May 2022).
4. Odermatt, D.; Danne, O.; Philippson, P.; Brockmann, C. Diversity II water quality parameters for 300 lakes worldwide from ENVISAT (2002–2012): A new global information source for lakes. *Earth Syst. Sci. Data* **2018**, *10*, 1527–1549. Available online: https://essd.copernicus.org/articles/10/1527/2018/ (accessed on 8 May 2022). [CrossRef]
5. Reinart, A.; Paavel, B.; Tuvikene, L. Effect of coloured dissolved organic matter on the attenuation of photosynthetically active radiation in Lake Peipsi. *Proc. Est. Acad. Sci. Biol./Ecol.* **2004**, *53*, 88–105.
6. Mañosa, S.; Mateo, R.; Guitart, R. A review of the effects of agricultural and industrial contamination on the Ebro Delta biota and wildfire. *Environ. Monit. Assess.* **2001**, *71*, 187–205. [CrossRef]
7. Soria, J.M. Past, present and future of la Albufera of Valencia Natural Park. *Limnetica* **2006**, *25*, 135–142. [CrossRef]
8. Serrano, L.; Reina, M.; Martín, I.; Reyes, A.; Arechederra, D.L.; Toja. J. The aquatic systems of Doñana (SW Spain): Watersheds and frontiers. *Limnetica* **2006**, *25*, 11–32. [CrossRef]
9. Soria, J.M.; Caniego, G.; Hernández-Sáez, N.; Dominguez-Gomez, J.A.; Erena, M. Phytoplankton Distribution in Mar Menor Coastal Lagoon (SE Spain) during 2017. *J. Mar. Sci. Eng.* **2020**, *8*, 600. [CrossRef]
10. Sòria-Perpinyà, X.; Vicente, E.; Urrego, P.E.; Pereira-Sandoval, M.; Tenjo, C.; Ruiz-Verdú, A.; Delegido, J.; Soria, J.M.; Peña, R.; Moreno, J. Validation of Water Quality Monitoring Algorithms for Sentinel-2 and Sentinel-3 in Mediterranean Inland Waters with In Situ Reflectance Data. *Water* **2021**, *13*, 686. [CrossRef]
11. Romo, S.; García-Murcia, A.; Villena, M.J.; Sánchez, V.; Ballester, A. Tendencias del fitoplancton en el lago de la Albufera de Valencia e implicaciones para su ecología, gestión y recuperación. *Limnetica* **2008**, *27*, 11–28. [CrossRef]
12. Romo, S.; Soria, J.; Fernández, F.; Ouahid, Y.; Barón-Solá, A. Water residence time and the dynamics of toxic cyanobacteria. *Freshw. Biol.* **2013**, *58*, 513–522. [CrossRef]
13. Global Lakes Sentinel Services. Available online: https://un-spider.org/es/links-and-resources/gis-rs-software/glass-global-lakes-sentinel-services (accessed on 8 May 2022).
14. Diversity II. Available online: http://www.diversity2.info/ (accessed on 8 May 2022).
15. CyanoAlert. Available online: https://www.cyanoalert.com/ (accessed on 8 May 2022).
16. Sentinel Application Platform. Available online: https://step.esa.int/main/download/snap-download/ (accessed on 8 May 2022).
17. Moore, T.S.; Campbell, J.W.; Feng, H. A fuzzy logic classification scheme for selecting and blending ocean color algorithms. *IEEE Trans. Geosci. Remote Sens.* **2001**, *39*, 1764–1776. [CrossRef]
18. Moore, T.S.; Dowell, M.; Bradt, S.; Ruiz-Verdú, A. An optical water type framework for selecting and blending retrievals from bio-optical algorithms in lakes and coastal waters. *Remote Sens. Environ.* **2014**, *143*, 97–111. [CrossRef] [PubMed]
19. Le, C.; Li, Y.; Zha, Y.; Sun, D.; Huang, C.; Zhang, H. Remote estimation of chlorophyll a in optically complex waters based on optical classification. *Remote Sens. Environ.* **2011**, *115*, 725–737. [CrossRef]
20. Elveld, M.A.; Ruescas, A.B.; Hommersonm, A.; Moore, T.S.; Peters, S.W.M.; Brockmann, C. An optical classification tool for Global Lake Waters. *Remote Sens.* **2017**, *9*, 420. [CrossRef]
21. Pereira-Sandoval, M.; Urrego, P.E.; Ruiz-Verdú, A.; Tenjo, C.; Delegido, J.; Sorià-Perpinyà, X.; Vicente, E.; Soria, J.M.; Moreno, J. Calibration and validation of algorithms for the estimation of chlorophyll-a concentration and Secchi depth in inland water with Sentinel-2. *Limnetica* **2019**, *38*, 471–487. [CrossRef]
22. Pereira-Sandoval, M.; Ruescas, A.; Urrego, P.E.; Ruiz-Verdú, A.; Delegido, J.; Tenjo, C.; Sorià-Perpinyà, X.; Vicente, E.; Soria, J.M.; Moreno, J. Evaluation of atmospheric correction algorithms over Spanish inland waters for Sentinel-2 Multispectral Imagery Data. *Remote Sens.* **2019**, *11*, 1469. [CrossRef]
23. Sentinel-2. Sentinel Online. Available online: https://sentinel.esa.int/web/sentinel/missions/sentinel-2 (accessed on 8 May 2022).
24. Brockmann, C.; Doerffer, R.; Peters, M.; Stelzer, K.; Embacher, S.; Ruescas, A. Evolution of C2RCC neural network for Sentinel 2 and 3 for the retrieval of ocean colour products in normal and extreme optically complex waters. In Proceedings of the Living Planet Symposium 2016, Prague, Czech Republic, 9–13 May 2016. Available online: http://step.esa.int/docs/extra/EvolutionoftheC2RCC_LPS16.pdf (accessed on 8 May 2022).
25. Uudeberg, K.; Ansko, I.; Poru, G.; Ansper, A.; Reinart, A. Using optical water types to monitor changes in optically complex inland and coastal waters. *Remote Sens.* **2019**, *11*, 2297. [CrossRef]
26. Warren, M.A.; Simis, S.G.H.; Martinez-Vicente, V.; Poser, K.; Bresciani, M.; Alikas, K.; Spyrakos, E.; Giardino, C.; Ansper, A. Assessment of atmospheric correction algorithms for the Sentinel-2A MultiSpectral Imager over coastal and inland waters. *Remote Sens. Environ.* **2019**, *225*, 267–289. [CrossRef]
27. Urrego, E.P.; Delegido, J.; Tenjo, C.; Ruiz-Verdú, A.; Soriano-Gonzalez, J.; Pereira-Sandoval, M.; Sorià-Perpinyà, X.; Vicente, E.; Soria, J.M.; Moreno, J. Validation of chlorophyll-a and total suspended matter products generated by C2RCC processor using Sentinel-2 and Sentinel-3 satellites in inland waters. In Proceedings of the XX Congress of the Iberian Association of Limnology, Murcia, Spain, 26–29 October 2020.

28. Uudeberg, K.; Aavaste, A; Köks, K.L.; Ansper, A.; Uusöe, M.; Kangro, K.; Ansko, I.; Ligi, M.; Toming, K.; Reinart, A. Optical Water Type Guided Approach to Estimate Optical Water Quality Parameters. *Remote Sens.* **2020**, *12*, 931. [CrossRef]
29. Bailey S.W.; Werdell, P.J. A multi-sensor approach for the on-orbit validation of ocean color satellite data products. *Remote Sens. Environ.* **2006**, *102*, 12–23. [CrossRef]
30. Shi, K.; Li, Y.; Li, L.; Lu, H.; Song, K.; Liu, Z.; Xu, Y.; Li, Z. Remote chlorophyll-a estimates for inland waters based on a cluster-based classification. *Sci. Total. Environ.* **2013**, *444*, 1–15. [CrossRef] [PubMed]
31. Bi, S.; Li, Y.; Xu, J.; Liu, G.; Song, K.; Mu, M.; Liu, H.; Miao, S.; Xu, J. Optical classification of inland waters based on an improved Fuzzy C-Means method. *Opt. Express* **2019**, *27*, 34838–34856. [CrossRef] [PubMed]
32. Botha, E.J.; Anstee, J.M.; Sagar, S.; Lehmann, E.; Medeiros, T.A.G. Classification of Australian Water bodies across a Wide Range of Optical Water Types. *Remote Sens.* **2020**, *12*, 3018. [CrossRef]
33. Du, Y.; Song, K.; Liu, G. Monitoring Optical Variability in Complex Inland Waters Using Satellite Remote Sensing Data. *Remote Sens.* **2022**, *14*, 1910. [CrossRef]
34. Bourel, M.; Crisci, C.; Martinez, A. Consensus methods based on machine learning techniques for marine phytoplankton presence-absence prediction. *Ecol. Inform.* **2017**, *42*, 46–54. [CrossRef]
35. Chou, J.-S.; Ho, C.-C.; Hoang, H.-S. Determining quality of water in reservoir using machine learning. *Ecol. Inform.* **2018**, *44*, 57–75. [CrossRef]
36. Watanabe, F.S.Y.; Miyoshi, G.T.; Rodrigues, T.W.P.; Bernardo, N.M.R.; Rotta, L.H.S.; Alcântara, E.; Imai, N.N. Inland water's trophic status classification based on machine learning and remote sensing data. *Remote Sens. Appl. Soc. Environ.* **2020**, *19*, 100326. [CrossRef]
37. Grendaite, D.; Stonevicius, E. Machine Learning Algorithms for Biophysical Classification of Lithuanian Lakes Based on Remote Sensing Data. *Water* **2022**, *14*, 1732. [CrossRef]
38. Ho, T.K. Random decision forests. In Proceedings of the 3rd International Conference on Document Analysis and Recognition, Montreal, QC, Canada, 14–16 August 1995; Volume 1, pp. 278–282. [CrossRef]
39. Cortes, C.; Vapnik, V. Support-vector networks. *Mach. Learn.* **1995**, *20*, 273–297. [CrossRef]
40. Landis, J.R.; Koch, G.G. The measurement of observer agreement for categorical data. *Biometrics* **1977**, *33*, 159–174. [CrossRef]
41. Pontius, R.G., Jr.; Millones, M. Death to Kappa: Birth of quantity disagreement and allocation disagreement for accuracy assessment. *Int. J. Remote Sens.* **2011**, *32*, 4407–4429. [CrossRef]
42. Breiman, L. Random forests. *Mach. Learn.* **2001**, *45*, 5–32 . [CrossRef]
43. Fernandez-Delgado, M.; Cernadas, E.; Barro, S. Do we need hundreds of classifiers to solve real world classification problems? *J. Mach. Learn. Res.* **2014**, *15*, 3133–3181.
44. Park, J.W.; Korosov, A.A.; Babiker, M.; Won, J.S.; Hansen, M.W., Kim, H.C. Classification of sea ice types in Sentinel-1 synthetic aperture radar images. *Cryosphere* **2020**, *14*, 2629–2645. [CrossRef]
45. Confederación Hidrográfica del Ebro. *Diagnóstico y Gestión Ambiental de Embalses en el ámbito de la Cuenca Hidrográfica del Ebro*; Ministerio de Medio Ambiente: Embalse de la Sotonera, Spain, 1996. Available online: https://www.chebro.es/documents/20121/48992/Informe_Final_Embalse_de_la_Sotonera_1996.pdf/b4844842-a49b-209b-96a6-7bb799d125af (accessed on 8 May 2022).
46. Dourte, D.R.; Fraisse, C.W.; Bartels, W.L. Exploring changes in rainfall variability in the Southeastern U.S.: Stakeholder engagement, observations, and adaptation. *Clim. Risk Manag.* **2015**, *7*, 11–19. [CrossRef]
47. Yilmaz, A.; Hossain, I.; Perera, B. Effect of climate change and variability on extreme rainfall intensity-frequence-duration relationship: A case study of Melbourne. *Hydrol. Earth Syst. Sci.* **2014**, *18*, 4065–4076. [CrossRef]
48. Sorià-Perpinyà, X.; Vicente, E.; Urrego, P.; Pereira-Sandoval, M.; Ruiz-Verdú, A.; Delegido, J.; Soria, J. M.; Moreno, J. Remote sensing of cyanobacterial blooms in a hyoertrophic lagoon (Albufera of València, Eastern Iberian Peninsula) using multitemporal Sentinel-2 images. *Sci. Total. Environ.* **2020**, *698*, 134305. [CrossRef]
49. Stelzer, K.; Simis, S.; Selmes, N.; Muller, D. Copernicus Global Land Operations "Cryosphere and Water CGLOPS-2". Product User Manual. Framework Service Contract N° 199496 (JRC). 2020. Available online: https://land.copernicus.eu/global/sites/cgls.vito.be/files/products/CGLOPS2_PUM_LWQ100_S2_v1.2.0_I1.03.pdf (accessed on 8 May 2022).
50. Geng X.; Smith-Miles, K. Incremental Learning. In *Encyclopedia of Biometrics*; Springer: Boston, MA, USA, 2009. [CrossRef]


VNIVERSITAT
DE VALÈNCIA