# Remote sensing of phytoplankton biomass in oligotrophic and mesotrophic lakes: addressing estimation uncertainty through machine learning

**Mortimer Werther**


**Submitted to:**

**Biological and Environmental Sciences**
**University of Stirling, Scotland, United Kingdom**

**May 2022**

**For the degree of Doctor of Philosophy**

**Supervisors:**
**Prof. Dr. Evangelos Spyrakos**
**Prof. Dr. Andrew Tyler**
**Dr. Peter D. Hunter**
**Prof. Dr. Stefan G.H. Simis**
**Dr. Daniel Odermatt**

*Früher begriff ich nicht, warum ich auf meine Frage keine Antwort bekam, heute begreife ich nicht wie ich glauben konnte, fragen zu können. Aber ich glaubte ja gar nicht, ich fragte nur.*

Franz Kafka, Betrachtungen über Sünde, Leid, Hoffnung und den wahren Weg (1917-18)

**Statement of Originality**

I hereby confirm that this dissertation is an original piece of work conducted independently by the undersigned, and all work contained herein has not been submitted for any other degree. All research material has been duly acknowledged and cited.

Signature of Candidate:

Mortimer Werther

Date: 23rd of May 2022

# Abstract

Phytoplankton constitute the bottom of the aquatic food web, produce half of Earth's oxygen and are part of the global carbon cycle. A measure of aquatic phytoplankton biomass therefore functions as a biological indicator of water status and quality. The abundance of phytoplankton in most lakes on Earth is low because they are weakly nourished (i.e., oligotrophic). It is practically infeasible to measure the millions of oligotrophic lakes on Earth through field sampling. Fortunately, phytoplankton universally contain the optically active pigment chlorophyll-*a*, which can be detected by optical sensors. Earth-orbiting satellite missions carry optical sensors that provide unparalleled high spatial coverage and temporal revisit frequency of lakes. However, when compared to waters with high nutrient loading (i.e., eutrophic), the remote sensing estimation of phytoplankton biomass in oligotrophic lakes is prone to high estimation uncertainties. Accurate retrieval of phytoplankton biomass is severely constrained by imperfect atmospheric correction, complicated inherent optical property (IOP) compositions, and limited model applicability. In order to address and reduce the current estimation uncertainties in phytoplankton remote sensing of low - moderate biomass lakes, machine learning is used in this thesis.

In the first chapter the chlorophyll-*a* concentration (chla) estimation uncertainty from 13 chla algorithms is characterised. The uncertainty characterisation follows a two-step procedure: 1. estimation of chla from a representative dataset of field measurements and quantification of estimation uncertainty, 2. characterisation of chla estimation uncertainty. The results of this study show that estimation uncertainty across the dataset used in this chapter is high, whereby chla is both systematically under- and overestimated by the tested algorithms. Further, the characterisation reveals algorithm-specific causes of estimation uncertainty. The uncertainty sources for each of the tested algorithms are discussed and recommendations provided to improve the estimation capabilities.

In the second chapter a novel machine learning algorithm for chla estimation is developed by combining Bayesian theory with Neural Networks (NNs). The resulting Bayesian Neural Networks (BNNs) are designed for the Ocean and Land Cover Instrument (OLCI) and MultiSpectral Imager (MSI) sensors aboard the Sentinel-3 and Sentinel-2 satellites, respectively. Unlike established chla algorithms, the BNNs provide a per-pixel uncertainty associated with estimated chla. Compared to reference chla algorithms, gains in chla estimation accuracy > 15% are achieved. Moreover, the quality of the provided BNN chla uncertainty is analysed. For most observations (> 75%) the BNN uncertainty estimate covers the reference *in situ* chla value, but the uncertainty calibration is not constantly accurate across several assessment strategies. The BNNs are applied to OLCI and MSI products to generate chla and uncertainty estimates in lakes from Africa, Canada, Europe and New Zealand. The BNN uncertainty estimate is furthermore used to deal with uncertainty introduced by

prior atmospheric correction algorithms, adjacency affects and complex optical property compositions.

The third chapter focuses on the estimation of lake biomass in terms of trophic status (TS). TS is conventionally estimated through chla. However, the remote sensing of chla, as shown in the two previous chapters, can be prone to high uncertainty. Therefore, in this chapter an algorithm for the direct classification of TS is designed. Instead of using a single algorithm for TS estimation, multiple individual algorithms are ensembled through stacking, whose estimates are evaluated by a higher-level meta-learner. The results of this ensemble scheme are compared to conventional switching of reference chla algorithms through optical water types (OWTs). The results show that estimation of TS is increased through direct classification rather than indirect estimation through chla. The designed meta-learning algorithm outperforms OWT switching of chla algorithms by 5-12%. Highest TS estimation accuracy is achieved for high biomass waters, whereas for low biomass waters extremely turbid waters produced high TS estimation uncertainty. Combining an ensemble of algorithms through a meta-learner represents a solution for the problem of algorithm selection across the large variation of global lake constituent concentrations and optical properties.

# Acknowledgements

I was fortunate to experience great intellectual, technical, financial and familial support during my PhD. I owe deep gratitude to my internal doctoral supervisors from the University of Stirling, Prof. Dr. Evangelos Spyrakos, Prof. Dr. Andrew Tyler and Dr. Peter D. Hunter and my external supervisors Dr. Daniel Odermatt and Prof. Dr. Stefan G.H. Simis, for the offer to undertake this PhD research, the continuous trust in my abilities to perform it and the granted intellectual freedom. You have listened to ideas and questions with great patience over the years and were always available when I was stuck on a particular problem. Having five supervisors was a challenging aspect of my PhD that taught me in many ways how to prioritise and undertake my own research whilst still listening to and including your expertise. I could not have asked for a better supervisory team. Vagelis, thank you for the weekly meetings to discuss the research. I always felt supported during the different stages of my PhD. It truly made all the difference over the years.

Funding from this PhD was provided by the University of Stirling, the Swiss Federal Institute of Aquatic Science and Technology (EAWAG) and the European Union's Horizon 2020 research and innovation programme under grant agreement No. 776480 (MONOCLE) led by Stefan G.H. Simis (Plymouth Marine Laboratory). The EUMETSAT Copernicus Collaborative Exchange Award and University of Waikato are greatly acknowledged for making the field campaign in New Zealand 2020 possible. A big thank you to Claudia Giardino and Mariano Bresciani for enabling and supporting the field campaign 2019 in Italy.

I acknowledge all the wonderful colleagues in this field who went through this journey with me: Olivier Burggraaf, Moritz Lehmann, Daniela Gurlin, Adam Varley, Dalin Jiang, Matthew Blake, Cristian Silva-Perez, Tom Ovenden, Danni Thompson, Douglas Moore and Craig McDougall.

I learnt more than I could imagine at the Ocean Optics class 2019 and became part of a cohort all due to the great organisers, teachers and scientists Emmanuel Boss, Collin Roesler, Ken Voss, Ivona Cetinić, Curt Mobley and Sasha Kramer. Thank you for this unforgettable month in Maine.

Thank you, Oberon, for the music sessions and machine learning discussions over the years, and, Kevin, Simon, Ole and Felix for the stuff only oldschool friends can do. Bihari and the crew for the windsurf adventures that were both fun and essential for my health.

I cannot thank my parents Yvonne and Rainer enough for enabling and supporting my path in environmental science from the beginning, my siblings Harriet, Lesley and Jeremy and my lovely grandparents Heidi and Peter. Lastly and above all, I thank you Natalie, who I am fortunate to call my wife. Thanks for finding me in Stirling and the incredible years during our PhD journeys.

This PhD is devoted to you Natalie and our little one, Brennan.

# Contents

# List of Figures

# List of Tables



# List of Acronyms

| Acronym | Full text |
|---|---|
| AE | Adjacency effect |
| AC | Atmospheric correction |
| AOPs | Apparent optical properties |
| BNN | Bayesian neural network |
| C2RCC | Case 2 Region CoastColour |
| CDOM | Coloured dissolved organic matter |
| chla | Chlorophyll-*a* concentration |

| | |
|---|---|
| $CH_4$ | Methane |
| CI | Confidence interval |
| $CO_2$ | Carbon dioxide |
| CNNs | Convolutional neural networks |
| CZCS | Coastal Zone Color Scanner |
| DNR | Department of Natural Resources |
| DT | Decision tree |
| EnMAP | Environmental Mapping Analysis Program |
| EPOS | Earth and Planetary Observation Sciences |
| FLH | Fluorescence line height |
| GBM | Gradient boosting machines |
| GBDT | Gradient boosting decision trees |
| GSM | Garver-Siegel-Maritorena |
| HABs | Harmful algal blooms |
| HICO | Hyperspectral Imager for the Coastal Ocean |
| HPO | Hyper-parameter optimisation |
| IDEPIX | Identification of pixel properties |
| IOCCG | International Ocean-Colour Coordinating Group |
| IOPs | Inherent optical properties |
| ISM | Inorganic suspended matter |
| LASSO | Least Absolute Shrinkage and Selection Operator |
| LightGBM | Light gradient boosting machine |
| LIMNADES | Lake Bio-optical Measurements and Matchup Data for Remote Sensing |
| LOO | Leave-one-out |
| MAD | Mean absolute difference |
| MAE | Mean absolute error |
| MACD | Mean absolute calibration difference |
| MAPD | Median absolute percentage difference |
| MC | Monte carlo |
| MCI | Maximum chlorophyll index |
| MdAD | Median absolute difference |
| MdAPE | Median absolute percentage Error |
| MdSA | Median symmetric signed difference |
| MBR | Maximum band ratio |
| MERIS | Medium Resolution Imaging Spectrometer |
| MDN | Mixture density network |
| ML | Machine learning |
| mlogloss | Multi-class logarithmic loss |
| MLP | Multi-layer perceptron |
| MODIS | Moderate Resolution Imaging Spectrometers |

| | |
|---|---|
| MONOCLE | Multiscale Observation Networks for Optical monitoring of Coastal waters, Lakes and Estuaries |
| MOS | Modular Optoelectronic Scanner |
| MPH | Maximum peak height |
| MSI | MultiSpectral Imager |
| $N_2O$ | Nitrous oxide |
| NAP | Non-algal particles |
| NB | Naïve Bayes |
| NIR | Near-infrared |
| NN | Neural network |
| NLL | Negative-log likelihood |
| nm | Nanometer |
| NZ | New Zealand |
| OACs | Optically active components (or constituents) |
| OC | Ocean Colour |
| OLCI | Ocean and Land Colour Instrument |
| OWTs | Optical water types |
| OSM | Organic suspended matter |
| PACE | Plankton, Aerosol, Cloud, ocean Ecosystem |
| PICP | Percentage interval coverage probability |
| PML | Plymouth Marine Laboratory |
| PRISMA | Hyperspectral Precursor and Application Mission |
| QAAv6 | Quasi-analytical algorithm version 6 |
| ReLU | Rectified linear unit |
| RF | Random Forest |
| RMSD | Root mean square difference |
| RMSE | Root mean square error |
| $R_{rs}$ | Remote-sensing reflectance |
| RSR | Relative spectral response |
| RSS | Residual sum of squares |
| RT | Radiative Transfer |
| RVD | Residual value distribution |
| SAA | Semi-analytical algorithms |
| SCIF | Sun-induced chlorophyll-a fluorescence |
| SeaWiFs | Sea-viewing Wide Field-Of-View Sensor |
| SHAP | Shapley additive explanations |
| SRF | Spectral response function |
| SVM | Support vector machine |
| TOA | Top-of-atmosphere |
| TS | Trophic state (or status) |
| TSI | Trophic state index |

| | |
|---|---|
| TSM | Total suspended matter |
| UoStirling | University of Stirling |
| UV | Ultraviolet |
| var | Standard deviation |
| VIF | Variance inflation factor |
| VSF | Volume scattering function |
| WSLH | Wisconsin State Laboratory of Hygiene |
| XGBoost | eXtreme Gradient Boosting |
| $Z_{SD}$ | Water transparency |
| zSD | Secchi disk depth |
| 3SAA | 3-Step semi-analytical algorithm |

# Chapter 1: Introduction

Water is vital for all living organisms on Earth. Most freshwater on non-glaciated land is contained in over 300 million lakes larger than 0.1 hectare (Downing et al., 2006; Verpoorter et al., 2014). Lakes provide essential ecosystem services such as drinking water, sediment retention and processing, and are hotspots of biodiversity and endemism (Grizzetti et al., 2016; Sterner et al., 2020). Moreover, lakes have high cultural and commercial value and function as recreational areas. Lakes were described as 'sentinels, integrators and regulators' of climate and environmental change to reflect the inevitable impacts that a changing climate has on an aquatic ecosystem (Williamson et al., 2009). Among the most pervasive impacts of climate change on lakes are the direct and indirect effects of global warming. Lake temperature rise reduces ice cover, alters stratification and changes nutrient loadings (Maberly et al., 2020), with consequences on lake greenhouse gas fluxes (Demars et al., 2016), species distribution (Comte and Olden, 2017), food web interactions (Blois et al., 2013), phenology (Thackeray et al., 2016) and metabolic rate and balance (Yvon-Durocher et al., 2010). The functionality of lake ecosystem services and values is directly connected to water quality, available quantity and the severity of global warming (Vasistha and Ganguly, 2020).

Over 3 billion people are still at risk of using unsafe water for consumption and sanitary purposes because the quality and health status of their water resource is unknown (UN, 2021). Parameters used to measure water quality are not routinely collected in most countries (Figure 1.1). A reason for limited data availability is that effective lake monitoring through conventional field measurement technologies is costly, labour-intensive and usually not automated (Palmer et al., 2015; Tyler et al., 2016). Limited access to resources and inconsistencies in sampling and measurement strategies across organisations and states further constrain systematic understanding of how surface waters are changing at local, regional and global scales.

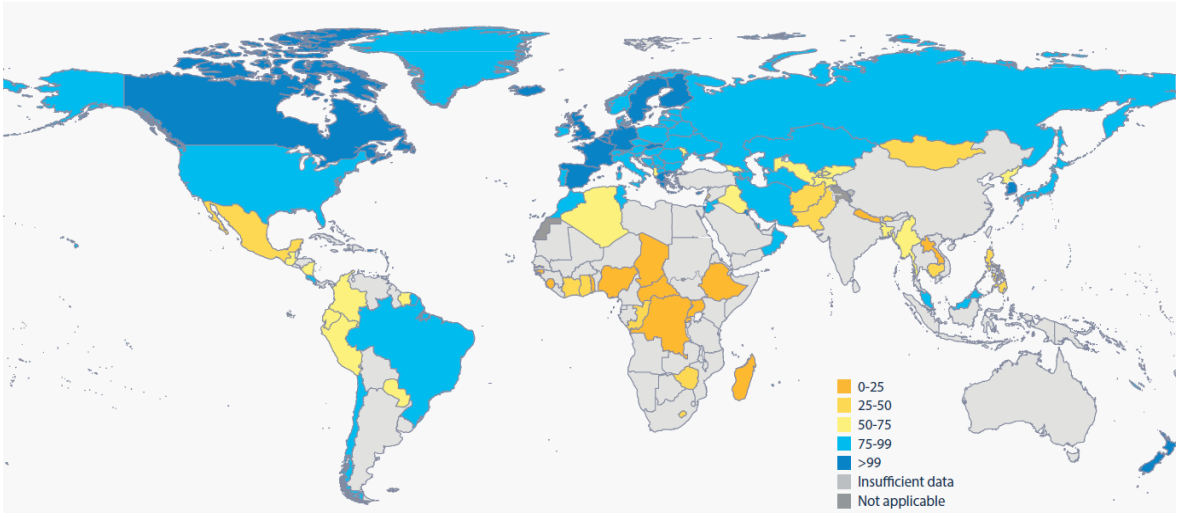


***Figure 1.1:*** *Proportion of national population using safely-managed drinking water in 2020 (in %). 138 countries had sufficient data on the accessibility, availability and quality of drinking water to produce a national estimate for this indicator in 2020. From UN (2021) under Creative Commons License.*

## 1.1 Earth observation through satellites

Satellite Earth observation provides an unparalleled number of measurements of aquatic ecosystems through high spatial coverage and temporal revisit frequency. It is therefore increasingly used to complement the monitoring and assessment of water quality (Chen et al., 2004; Greb et al., 2018; Kiefer et al., 2015; Mouw et al., 2015).

Satellites are devices launched into Earth's orbit in outer space. An orbit is a curved trajectory of an object around a planet, such as the trajectory of a satellite around Earth. In the 1600s, Johannes Kepler and Isaac Newton established the fundamental principles of orbital motion. Kepler presented what is known as Kepler's laws of motion between 1609 and 1619 (Russell, 1964):

1. Each planet's orbit is an ellipse with the Sun as the focus.
2. The line between the planet to the Sun sweeps out equal areas of space in equal amounts of time intervals.
3. The square of a planet's period is proportional to the size of its orbit.

Kepler's equations explain the kinematics of planetary motions, but Isaac Newton discovered the dynamic principles defining this motion later, which he published in his *Principia Mathematica* in 1687 (Barbour, 1989):

1. A body remains in a condition of rest or uniform motion along a straight line unless forces operating on it cause it to change that state.
2. The motion change is proportional to the active motive force. The shift in motion occurs in the direction of the straight line along which the force is operating.
3. The mutual interactions of two bodies are always equal and directed to opposing parts.

In *Principia Mathematica*, Newton also provides his basic law of gravity, which he expresses as: "any two point masses attract one another with a force proportional to the product of their masses and inversely proportional to the square of the distance between them." This law may be mathematically expressed for an Earth-orbiting satellite as:

$$\underset{F}{\rightarrow} = \frac{GMm}{r^2}, \tag{1.1}$$

where $\underset{F}{\rightarrow}$ is the force due to Newton's law of universal gravitational attraction that is exchanged by two bodies with masses $M$ (the Earth) and $m$ (the satellite) placed at a distance $r$ (Graziani et al., 2021). Here, $G$ = 6.6743x10$^{-11}$ m$^3$ kg$^{-1}$ s$^{-2}$ is the universal gravitational constant. Although Albert Einstein's general theory of relativity eventually addressed inconsistencies between observations and Newtonian dynamics, satellite Earth observation is fundamentally based on both Kepler's and Newton's laws.

## 1.2 Aquatic remote sensing

Sensors mounted on Earth observation satellites are designed to monitor and survey the planet's surfaces. This is known as remote sensing. Passive, optical remote sensors are commonly employed in aquatic remote sensing. These optical sensors include imaging spectrometers with several cameras. Typically, the sensors are "push-broom scanners" with the ability to actively tilt the sensor's cameras to reduce Sun glint contamination. Passive optical sensors can detect emerging radiation from water bodies in a variety of visible and near-infrared wavelengths of the electromagnetic spectrum. The electromagnetic spectrum is the range of all types of radiation.

Aquatic remote sensing started with the experimental Coastal Zone Color Scanner (CZCS) over the coastal ocean in the 70s (Hovis et al., 1980), and has since significantly advanced the understanding of biological and physical oceanography. The first remote sensing experiments over inland waters were however only carried out in the 90s (Dekker and Peters, 1993; Gitelson et al., 1993; Mittenzwey et al., 1992; Ritchie et al., 1990; Zilioli et al., 1994). Due to the success of the CZCS programme, follow-up missions in the 90s designed for ocean and coastal waters were launched, such as the Modular Optoelectronic Scanner (MOS) (Zimmermann et al., 1993), the Sea-viewing Wide Field-Of-View Sensor (SeaWiFs) (Hooker and Esaias, 1993), the Medium Resolution Imaging Spectrometer (MERIS) (Rast et al., 1999) and Moderate Resolution Imaging Spectrometers (MODIS) (Pagano and Durham, 1993). Despite the fact that most of these sensors had limited spatial resolution (>= 300 m per satellite pixel) and thus could not adequately resolve many inland waters, they provided the critical data streams that allowed experimental techniques to progress to a near-operational stage in the last two decades (Greb et al., 2018; Odermatt et al., 2018). The Landsat missions, which were originally created for remote sensing over land, have been readily explored for use over inland waters due to their excellent spatial resolution of 30 m in blue to near-infrared bands. Regardless of the fact that the available bands for aquatic remote sensing are suboptimal, the Thematic Mapper (TM) on Landsat 4-5, the Enhanced Thematic Mapper (ETM+) on Landsat 7, and the Operational Land Imager (OLI) on Landsat 8 were all employed effectively in remote sensing of lakes (Giardino et al., 2001; Manzo et al., 2015; Ritchie et al., 1990; Sharaf et al., 2019; Smith et al., 2021). In recent years, space agencies globally have started to launch fleets of Earth observation satellites. Most prominent is the European Copernicus programme with 6 unique Sentinel missions carrying over 15 different sensors (Thépaut et al., 2018). For aquatic remote sensing specifically, the Ocean and Land Colour Instrument (OLCI) aboard the Sentinel-3 A and B satellites provides observations at 300 meter spatial resolution measured in 21 bands of the optical domain between 400 and 1020 nm. Also sensors designed for land applications with higher spatial resolution, such as the Sentinel-2 MultiSpectral Imager (MSI), were used to measure smaller inland water systems.

Aquatic remote sensing is fundamentally based on radiometry, the science of measuring electromagnetic radiation. Sun light carries energy that arrives on Earth as radiation. The path radiation travels from the Sun to a water body is altered by the atmosphere, its composition and the

manifold interactions of objects, gases, and aerosols. Radiative transfer (RT) and related equations provide a physical description of these processes and pathways between the atmosphere and a water body (Gordon et al., 1975).

When the human eye detects radiation, its sensitivity changes dynamically. For scientific purposes, a more objective measure that also extends the restricted human capability of sensing visible light is required. Through detectors, the flow of radiation as a function of direction at any location within a water body can be measured. By adding a filter, polarizer, and diffuser in front of a detector, the wavelength dependence and polarisation of radiation can be determined (Figure 1.2). Radiation measurements were used to develop descriptions of light in water (Kirk, 1994; Mobley, 1994).

Spectral radiance is the fundamental quantity that describes light in radiometric terms as it defines the spatial ($x$), temporal ($t$), directional ($\xi$) and wavelength ($\lambda$) structure of the measured light field:

$$L(x, t, \hat{\xi}, \lambda) = \frac{\partial^4 Q}{\partial t, \partial A, \partial \Omega, \partial \lambda}. \quad (1.2)$$

The measurement of radiance $L$ is a function of the amount of radiant energy $Q$ that enters a suitable radiometer in an area $\partial A$ in a solid angle $\partial \Omega$ of the radiometer's field of view during an interval time $\partial t$ in wavelength $\partial \lambda$. All other radiometric quantities can be derived from the radiance $L$ (Mobley, 2022).

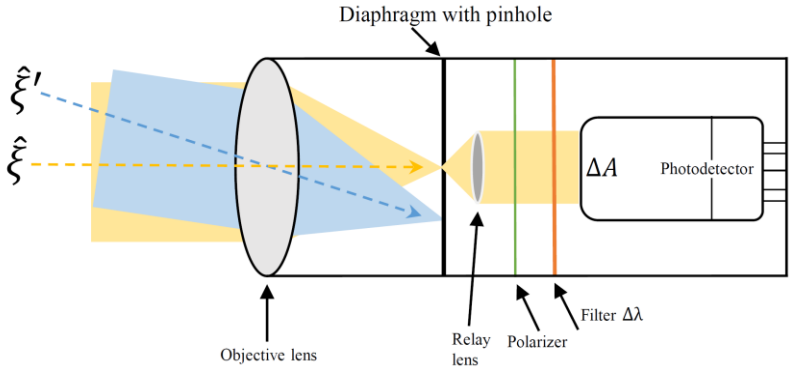


***Figure 1.2:*** *Instrument elements of a collimated radiometer. Re-drawn from Mishchenko (2014) and Mobley (2022) under Creative Commons License.*

Optical radiometric detectors installed on spaceborne carriers measure radiance emerging from the water surface, column, and the atmosphere. A small fraction of measured radiance contains information about water column optical properties (i.e., absorption, scattering and attenuation) and components that can be related to water quality indicators. The derivation of water quality indicators from measured radiance is complex, because inland and coastal waters usually contain a multitude of different dissolved and particulate substances. Over 40 years ago, a broad classification into Case-1 (open ocean) and Case-2 waters (inland and coastal waters) was established to separate the two cases (Morel and Prieur, 1977). Case-1 waters are optically dominated by phytoplankton and its associated organic products. Besides living phytoplankton organisms, in Case-2 waters additional optically active components (OACs) are present:

1. Organic and inorganic suspended particulates, such as solids, sediments, living and dead matter aggregated under the term total suspended matter (TSM) or suspended particulate matter (SPM).
2. Dissolved organic matter from anthropogenic, terrigenous, and littoral sources, including decayed phytoplankton cells and zooplankton grazing products that dissolved over time, referred to as coloured dissolved organic matter (CDOM) (Kirk, 1976).
3. Particles such as minerals from biogenic substances (e.g. shells), heterotrophic organisms (bacteria, viruses) and detrital organics such as cell debris and faecal known as detritus or non-algal particles (NAP) (Sosik, 2008).

Compared to ocean waters, high optical complexity, extreme variation in constituent variability and concentration, complex atmospheric compositions and the close proximity to land as well as bottom effects on the underwater radiance distribution pose great challenges for satellite remote sensing over inland waters (Bukata et al., 1995; Greb et al., 2018; IOCCG, 2019; Pahlevan et al., 2021; Palmer et al., 2015; Tyler et al., 2016).

### 1.2.1 Inherent optical properties

Through radiometry, light fields can be described quantitatively. For aquatic remote sensing it is of interest to infer the optical properties of OACs in the considered water body through which light propagates. When a photon interacts with matter, it has varying outcomes. The photon can be absorbed in a process called absorption (a), whereby its energy is converted to another form such as heat. It can also change direction, through a process called scattering ($\beta$) that occurs with or without a change of wavelength. Scattering is described by the volume scattering function (VSF) (Mobley, 1994). The VSF describes the angular distribution of scattered light resulting from the water and the OAC scattering processes. The absorption and scattering properties of water are described through the *inherent* optical properties (IOPs) that depend solely on the ambient light field (Gordon and Clark, 1980).

IOPs can be illustrated geometrically (see Figure 1.3). Consider a small volume of water $\Delta V$, that has thickness $\Delta r$ and is illuminated by a beam of light $\Phi_i$ at some wavelength $\lambda$ (in W nm$^{-1}$). Then,

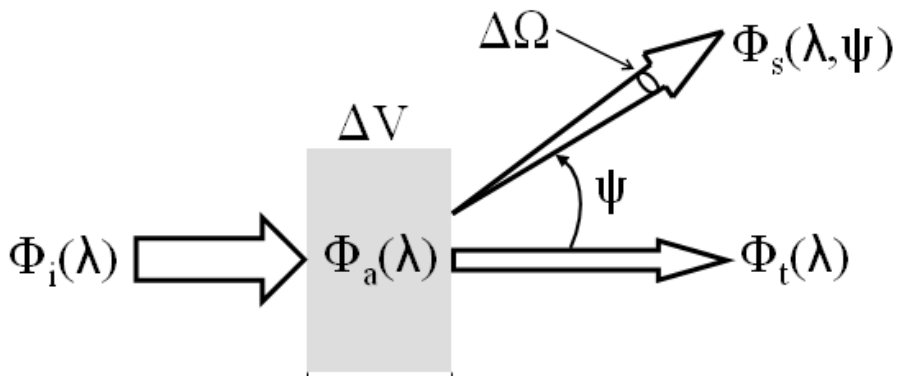


***Figure 1.3:*** *Geometry of inherent optical properties. From Mobley (2022) under Creative Commons License.*

some part $\Phi_a(\lambda)$ of the incident light is absorbed within the volume of water, whereas some part $\Phi_s(\lambda, \psi)$ is scattered out of the beam at an angle $\psi$ into an annular solid angle $\Delta\Omega$. The remaining beam $\Phi_t(\lambda)$ is transmitted through the volume with no change in direction. In aquatic optics, the absorption and scattering coefficients are used, which are respectively the absorptance and scatterance of light per unit distance in water (Mobley, 2022).

Lakes contain many forms of dissolved and particulate substances due to the growth and decay of biological material and because they receive constant input from their catchments. OACs vary both in their composition and concentration. Consequently, the optical properties of lakes show large temporal, spatial and vertical heterogeneity (Minaudo et al., 2021; Nouchi et al., 2018).

All the different OACs determine light propagation in water define the total absorption and VSF. The total IOPs are the sum of the individual component IOPs. Hence, if a water body contains *N* different and independent constituents, then the total absorption can be written as:

$$a_{tot} = a_w + \sum_{i=1}^{N} a_i\,, \tag{1.3}$$

where subset $i$ indicates a constituent and $a_w$ is the absorption of water. To model IOPs, four absorption components are commonly considered: $a_w$, whose absorption coefficient is invariant of location and water body; phytoplankton ($\phi$), often parametrised by its chlorophyll-*a* pigment concentration (see section 1.2); CDOM and NAP. Thus Eq. (1.3) can be further specified as:

$$a_{tot} = a_w + a_\phi + a_{CDOM} + a_{NAP}, \tag{1.4}$$

all of which vary with wavelength, and except for water, with location. The same partitioning of total IOPs into the sums of constituent contributions applies to the VSF in the same manner:

$$\beta_{tot} = \sum_{i=1}^{N} \beta_i\,. \tag{1.5}$$

Details about the VSF can be found in Gordon (2019) and Mobley (2022).

### 1.2.2 Apparent optical properties

The physical properties of a component, such as its particle size, affect IOPs. *Apparent* optical properties (AOPs) refer to the up- and downwelling radiances that depend both on IOPs and on the geometric structure of the radiance distribution. Therefore, AOPs can be used to deduce information about the concentration of an OAC, such as phytoplankton, in aquatic ecosystems. Spectral remote-sensing reflectance ($R_{rs}$) has become the main AOP (Mobley, 1999; O'Reilly et al., 1998) and is defined as:

$$R_{rs}(\theta, \phi, \lambda) = \frac{L_w(\theta, \phi, \lambda, 0^+)}{E_d(\theta, \phi, \lambda, 0^+)}\,(\mathrm{sr}^{-1}). \tag{1.6}$$

The argument $0^+$ indicates a measurement above the water surface, i.e. in air. All quantities depend on zenith ($\theta$) and azimuth direction ($\phi$), as illustrated in Figure 1.4, and vary with wavelength $\lambda$.

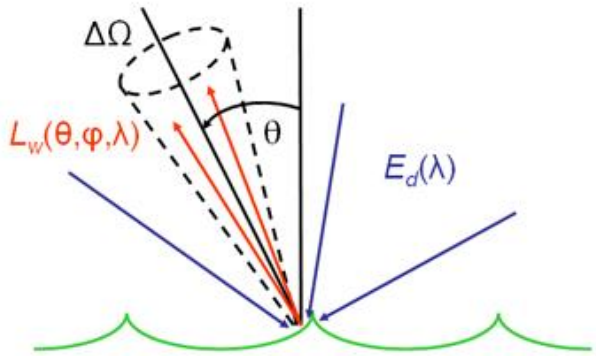


***Figure 1.4:*** *Downwelling and water-leaving radiances that constitute the remote-sensing reflectance. From Mobley (1994) under Creative Common License.*

$E_d$ is the downwelling planar irradiance, which can be quantified by a single detector that measures light incident on a surface. $L_w$ is the water-leaving radiance entering the water medium from above the water surface, which is then absorbed and scattered back through the water surface into air. $L_w$ cannot be measured directly, and thus has to be estimated from the upwelling radiance $L_u$ minus the water surface reflected radiance (including sun and sky glint) $L_{surf}$:

$$L_w(\theta, \phi, \lambda, 0^+) = L_u(\theta, \phi, \lambda, 0^+) - L_{surf}(\theta, \phi, \lambda). \quad (1.7)$$

A measurement of $L_u$ may also be made at some distance below the water surface from which it then has to be extrapolated through the surface. A wide variety of instrument setups and methodologies exist to derive $R_{rs}$ (Lee et al., 2013; Mobley, 1999; Simis and Olsson, 2013). $R_{rs}$ can be measured *in situ*, modelled through radiative transfer models and derived from satellite sensors after correction for atmospheric and air-water effects on the radiance distributions.

The established Case-1 and Case-2 classification system to separate the ocean from optically complex waters is limited for lake and coastal water remote sensing, therefore further classifications were developed (Jerlov, 1968; Mélin and Vantrepotte, 2015; Moore et al., 2014, 2001; Morel, 1988; Spyrakos et al., 2018). Mobley (2004) were already in favour of abandoning the artificial distinction between Case-1 and Case-2 waters for optical remote sensing purposes. Nowadays, most common for inland and coastal waters is the clustering into distinct optical water types (OWTs) that are determined by clustering $R_{rs}$ into optical clusters (Spyrakos et al., 2018). The resulting optical clusters share similar AOPs, IOPs and OAC concentrations and can guide algorithm development and validation.

### 1.2.3 Atmospheric correction

For aquatic satellite remote sensing, the aim is to obtain $R_{rs}$ free from atmospheric disturbance. The atmospheric correction (AC) process is thus an essential element of aquatic satellite remote sensing. A satellite sensor at the top-of-atmosphere (TOA) viewing a water body measures upwelling radiances with contributions by three main sources: the atmosphere, the water surface and the water column contents. Solar radiance in the atmosphere is absorbed and scattered by gases and aerosols

(denoted here as $L_{atm}$) including multiple scattering between aerosols. In addition, downwelling irradiance is reflected by the water surface $L_{surf}$ back into the direction of the sensor and lastly some radiance originates from within the water column, namely $L_w$. The total radiance $L_{tot}$ measured by the sensor at TOA is the sum of the individual radiance contributions:

$$L_{tot} = L_{atm} + L_{surf} + L_w. \tag{1.8}$$

Removing the contributions by $L_{atm}$ and $L_{surf}$ from $L_{tot}$ to retain $L_w$ formally defines the AC process. All effects on $L_{tot}$ depend on viewing direction and wavelength but are omitted here for simplicity. $L_{surf}$ and $L_w$ refer to their measurement at TOA, not just above the water surface.

The three contributors in Eq. (1.7) are grouped terms for more complex radiance processes. The radiances absorbed and scattered by atmospheric gases or aerosols can be split by their origin. The scattering regime is defined through the particle size. The molecular, elastic scattering by atmospheric gases is typically called Rayleigh radiance in aquatic remote sensing, $L_R$ (Gordon and Wang, 1992). Likewise, aerosol scattering $L_a$ is commonly modelled as a distinct process. If there were no aerosols, then $L_{atm} = L_R$. Since gases and aerosols in the atmosphere are not isolated, they lead to multiple scattering of the radiances and denoted as $L_A$. Like $L_{atm}$, also $L_{surf}$ can be further separated into contributions by sun glint $L_{glint}$ and sky radiance $L_{sky}$ reflected by the water surface. In addition, given the natural influence of wind and the dynamic movements of the water's surface, the sun and sky radiances reflected by whitecaps and foam $L_{WF}$ must be incorporated. Thus Eq. (1.8) further expands into:

$$L_{tot} = L_R + [L_A] + L_{glint} + L_{sky} + L_{WF} + L_w. \tag{1.9}$$

Several studies reformulated Eq. (1.9) to add the direct and diffuse transmittance between the water surface and TOA along the viewing direction (Gordon and Wang, 1994; Wang, 1999) or the viewing path of a satellite sensor (Franz et al., 2007).

After decades of research AC is considered accurate over clear ocean waters. There a number of commonly used approaches which primarily differ in their aerosol removal methodology (Frouin et al., 2019; Gao et al., 2000; Wang et al., 2021). For ocean waters, $L_w$ is generally, with exceptions, small in the red and near-infrared (NIR) and maritime aerosols dominate. For inland waters however, $L_w$ in red-NIR is often significant due to higher water constituent concentrations as well as terrestrial aerosols (Moses et al., 2017; Tyler et al., 2016). Moreover, the radiance distribution over lakes may also be affected by the neighbouring terrestrial environment, known as the adjacency effect (AE). This effect occurs when light from land objects surrounding a lake is reflected into the sensor viewing path through atmospheric scattering, which thus modifies at-sensor radiance recorded over a lake. AE decreases the spectral contrast between the water body and the area surrounding it. Whether the AE occurs depends on the areal topography, viewing and illumination geometries of the sensor and aerosol type and density (Reinersman and Carder, 1995; Santer and Schmechtig, 2000). AE over inland and coastal waters can affect satellite-measured radiances as far as 20 km offshore and varies with wavelength (Bulgarelli and Zibordi, 2018; Odermatt et al., 2008; Sterckx et al., 2011).

Because of the added complexity due to land proximity, varying aerosol types, non-negligible reflectance in red/NIR and complex water constituent compositions, ACs designed for the ocean had to be adapted for inland and coastal waters. Similarly, land-based approaches were adapted (Vermote et al., 1997). The approaches can be broadly separated into two categories. First, two-step methods in which absorption and scattering due to gases are removed, followed by an approximation of the aerosol type and its contribution to $L_{atm}$. Example ACs are ACOLITE (Vanhellemont and Ruddick, 2018), iCOR (De Keukelaere et al., 2018; Wolters et al., 2021), POLYMER (Steinmetz et al., 2011), SeaDAS (Franz et al., 2015), GRS (Harmel et al., 2018) and MEETC2 (Saulquin et al., 2016). Second, machine learning neural networks (Brockmann et al., 2016; Fan et al., 2021, 2017) that were trained with datasets generated through RT simulations based on models such as Hydrolight (Mobley et al., 2009) and AccuRT (Stamnes et al., 2018). Many of these approaches were tested and compared in a recent atmospheric correction intercomparison exercise (Pahlevan et al., 2021b).

## 1.3 Phytoplankton biomass

Satellite remote sensing of water quality is primarily based on the assumption that $R_{rs}(\lambda)$ can be precisely obtained through accurate AC for the inversion into IOPs or OAC concentrations. The most relevant OAC biologically is phytoplankton. Phytoplankton are single-cell, floating microorganisms that constitute the bottom of the aquatic food web. Through their optically active pigment chlorophyll they harvest light and (together with water and carbon dioxide) through photosynthesis, release oxygen while also generating organic material high in energy content. Phytoplankton are thus the primary producers of food for other aquatic organisms such as zooplankton and prey fish (Reynolds, 1984). Moreover, as part of the ocean and inland waters, they contribute significantly to the global carbon budget and produce half of the Earth's annual oxygen (Calbet and Landry, 2004).

Chlorophyll has many types, such as chlorophyll-*a*, -*b*, -*c*. Chlorophyll-*a* is found in all phytoplankton. The most common approach is to measure phytoplankton biomass through a measure of chlorophyll-*a* concentration (chla), in unit milligram of chlorophyll per cubic meter. Compared to other optical measurements of phytoplankton, chla is the best indicator of available radiance for photosynthesis (Huot et al., 2007). Lakes are inherently dynamic systems and phytoplankton biomass varies with depth (Longhi and Beisner, 2009), season (Minaudo et al., 2021; Talling, 1986) and per specific system. A surface chla measurement, as obtained through optical aquatic remote sensing, is therefore a proxy of phytoplankton biomass.

### 1.3.1 Trophic status

Phytoplankton growth in lakes is not only related to available sunlight, water and carbon dioxide for photosynthesis, but also nutrient availability. Phosphorus (P) and nitrogen (N) are essential nutrients for animal and plant growth. Whereas the ocean is considered nutrient-barren, lakes are in constant exchange with their environment and therefore prone to higher nutrient contents. Human activities

cause a discharge of P and N from point sources such as wastewater treatment plants into freshwater systems (Tong et al., 2020). P and N may also be induced through non-point sources. For example, storms flush nutrients from surrounding land into the catchments of lakes (Stockwell et al., 2020). Moreover, N and P are commonly added to soils for crop production (as fertilisers or animal manures) and form a by-product of biosolids and composts (Gold and Sims, 2004). Combustion of fossil fuel generates biologically available N, indirectly introduced into the watershed through atmospheric-water gas exchanges (Jonson et al., 2017).

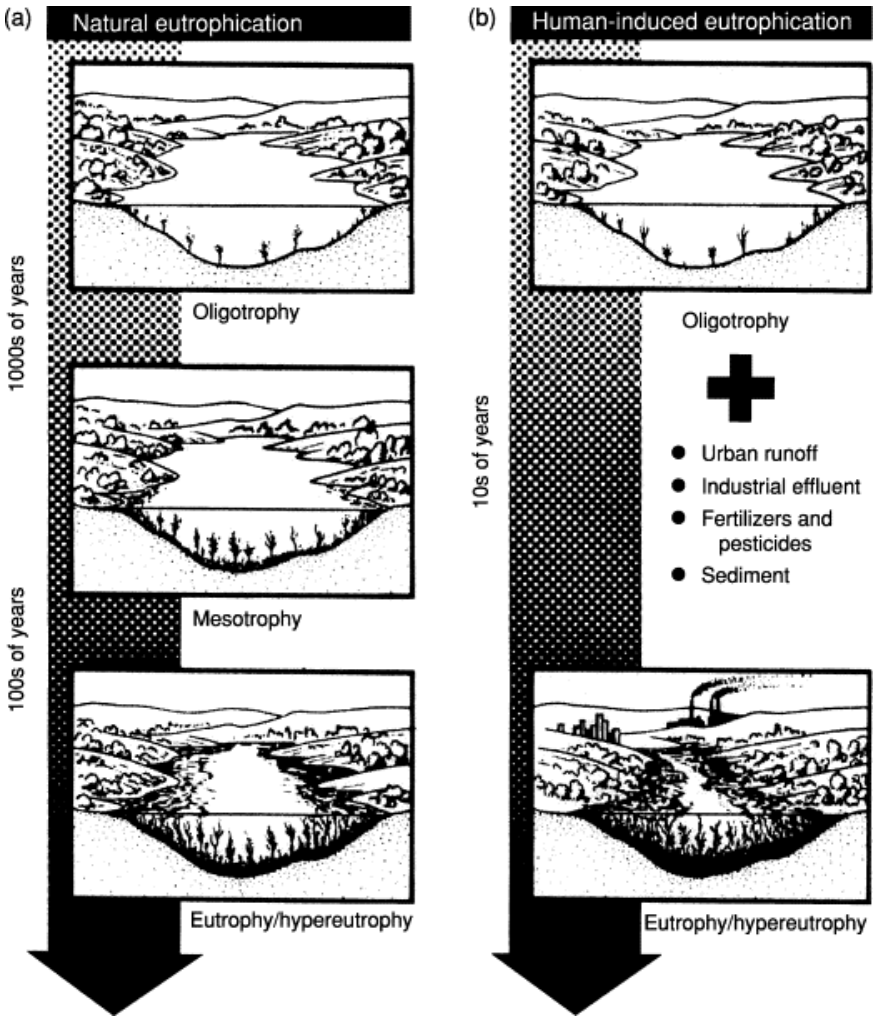


***Figure 1.5:*** *Eutrophication progression. (a) natural eutrophication, (b) human-induced eutrophication. From Gold and Sims (2004) under Creative Commons License.*

An increase in nutrient loading is known as eutrophication and associated with cascading effects in biogeochemical processes. In lakes slow eutrophication is often natural and may balance over time, but through artificially increased levels of nutrients, phytoplankton can grow excessively (Figure 1.5). The consequences of eutrophication are manifold: a decrease in levels of dissolved oxygen (Foley et al., 2012), increased turbidity, harmful algal bloom (HABs) formations (Heisler et al., 2008) and in extreme cases death of fish. High levels of eutrophication can fundamentally change aquatic biodiversity and composition (Anderson et al., 2002).

Many aspects of freshwater ecosystems are related to nutrient supply, therefore nutrient status is the basis for widely used trophic (i.e., nutrient level) classification systems applied to water bodies (Carlson, 1977). Low-nutrient waters are classified as 'oligotrophic' (i.e., poorly nourished); moderate-levels as 'mesotrophic' (moderately nourished) and excessive-nutrient concentrations as 'eutrophic' (highly nourished). Extreme nutrient-enrichment is concurrently classified as 'hypereutrophic'. Because phytoplankton biomass is directly related to nutrient levels, chla is a globally widespread indicator of trophic status (Michelutti et al., 2010).

### 1.3.2 Low, moderate and high biomass lakes

Phytoplankton pigments compete with other absorbers such as aquatic vegetation, CDOM and NAP for available light. In blue – green wavelengths, CDOM and NAP absorb strongly at 443 nm close to the $a_\phi$ absorption maximum. While CDOM and NAP absorption decreases exponentially with wavelength, the magnitude and shape of $R_{rs}(\lambda)$ in blue – green can be mostly due to these two IOPs. Figure 1.6 features a comparison of the absorption coefficients from 400 – 750 nm from water, phytoplankton, CDOM and NAP with varying chla across four sites in Fremont Lake (USA). Below 10 mg m$^{-3}$ at 443 nm CDOM and NAP absorb stronger than phytoplankton. The phytoplankton

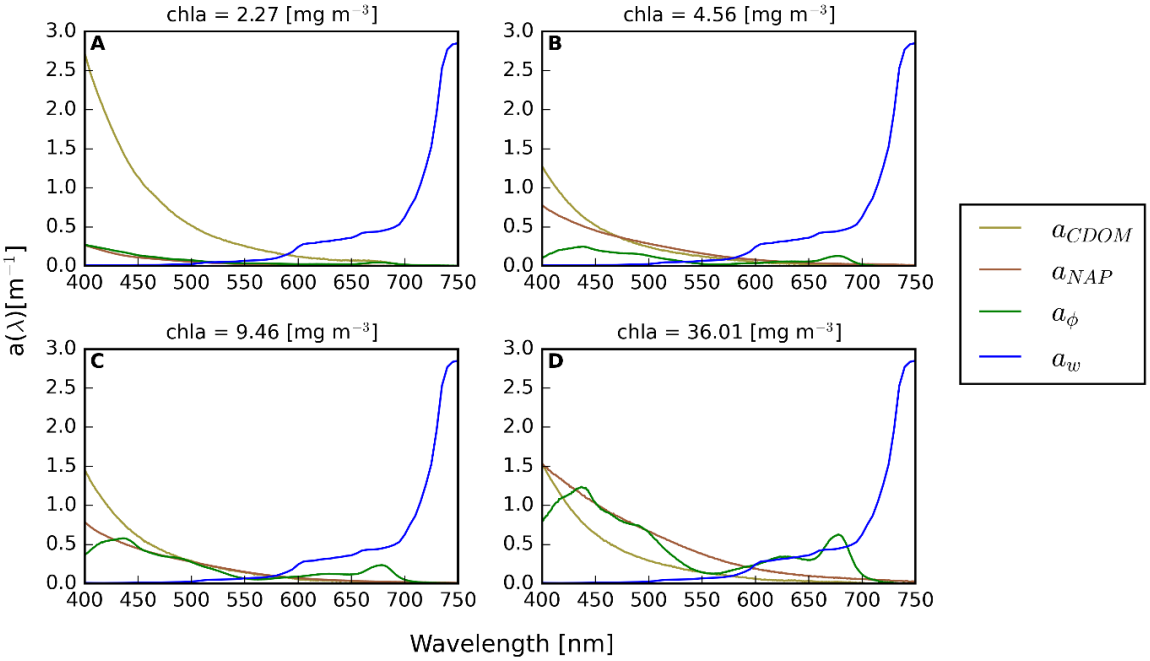


***Figure 1.6:*** *Comparison of the absorption coefficients of phytoplankton ($a_\phi$), NAP ($a_{NAP}$), CDOM ($a_{CDOM}$) and water ($a_w$) from four sampling sites in Fremont Lake with chla of 2.27 mg m$^{-3}$ (A), 4.56 mg m$^{-3}$(B), 9.46 mg m$^{-3}$(C) and 36.01 mg m$^{-3}$ (D). TSM for sites A, B, C <= 5 g m$^{-3}$. Re-drawn from Gurlin et al. (2011) with permission.*

absorption peak near 675 nm is affected by CDOM and NAP absorption at 2.27 mg m$^{-3}$ but only marginally with increasing chla. At approximately 10 mg m$^{-3}$ of chla, the contribution from phytoplankton backscattering to the water-leaving radiance starts to significantly increase.

Gitelson et al. (1999) found that for different phytoplankton groups with increasing scattering due to algal biomass reflectance in NIR > 700 nm also increases. In higher biomass waters peaks in $R_{rs}(\lambda)$ near 560 and 700 nm are evident. While chla-related peaks in $R_{rs}(\lambda)$ in high biomass waters are clearly visible they are less prominent with lower chla or may be absent entirely. Reflectance spectra used in this dissertation for lakes with chla <= 10 mg $m^{-3}$ and chla >= 10 & <= 100 mg $m^{-3}$ are displayed in Figure 1.7. It is obvious from Figure 1.6 and 1.7 that significant differences in both IOPs and resulting $R_{rs}(\lambda)$ between low and high biomass waters exist, which consequently affect the estimation of OACs. For these reasons, remote sensing algorithm development and validation studies have frequently distinguished between low (chla < 10 mg $m^{-3}$), moderate (chla 10 - 25 mg $m^{-3}$) and high (chla > 25 mg $m^{-3}$) biomass waters (e.g., Blondeau-Patissier et al., 2014; Gilerson et al., 2010; Odermatt et al., 2012; Ruddick et al., 2001).

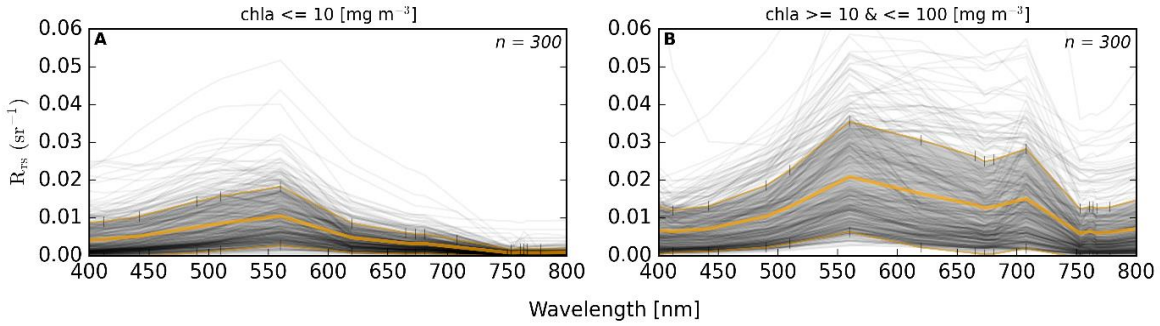


***Figure 1.7:*** *300 randomly selected $R_{rs}(\lambda)$ in OLCI resolution from lakes contained in the datasets of chapter 2 (A) and chapter 4 (B) used in this dissertation. Average chla is 2.5 $mg\ m^{-3}$ (A) and 34.07 $mg\ m^{-3}$ (B). Solid orange lines depict mean +- standard deviation and black vertical lines the band positions of the OLCI bands.*

## 1.4 Remote sensing of phytoplankton chlorophyll-*a*

Remote sensing chla algorithms are usually separated into (semi-) empirical and (semi-) analytical algorithms. Purely empirical algorithms use statistical models to estimate chla from $R_{rs}(\lambda)$. Datasets are necessary to build a statistical model. Purely empirical models make no assumptions about the relationships between IOPs and AOP or their governing properties. Semi-empirical algorithms employ statistical models but focus on specific spectral absorption or scattering phenomena, discovered through prior experiments that can be related to chla. Conversely, (semi- or quasi-) analytical algorithms rely on the inversion of RT equations to derive IOPs from AOPs and subsequently relate a derived IOP to chla. These algorithms are in practice not entirely analytical (thus the terms ‘semi-‘ and ‘quasi-‘) because they involve some form of statistical optimisation routine or parameters that were derived from empirical datasets.

### 1.4.1 Empirical algorithms

Phytoplankton are efficient light absorbers in blue and red wavelengths owing to the presence of intracellular pigments. Compared to minerals, they do not effectively scatter light. Phytoplankton water content is lower and their particle size is smaller than that of minerals, which show a larger backscattering-to-scattering ratio (Aas, 1996; Stramski et al., 2004; Stramski and Kiefer, 1991). Therefore, the most obvious approach to remotely sense phytoplankton is to base the estimation on the amount of absorbed light.

The amount of phytoplankton pigments, their concentration and composition, influence the magnitude and shape of $a_\phi(\lambda)$, which varies with wavelength. As explained, $a_\phi(\lambda)$ has two maxima at approximately 443 and 675 nm. $a_\phi(\lambda)$ at these two peaks is linearly related to chla through a power law function in ocean waters (Bricaud et al., 2004, 1998). The relationship between $a_\phi(\lambda)$ at these two wavelengths and chla is typically weaker for inland waters. Effects such as pigment packaging, whereby the absorbing phytoplankton pigments are "packaged" into discrete particle clusters rather than being uniformly distributed throughout the cell can impact the relationship (Kirk, 1994). The scaling of $a_\phi(\lambda)$ to chla further depends on the phytoplankton type, size and accessory pigments (Johnsen et al., 1994; Morel and Bricaud, 1981; Simis et al., 2007).

A change in phytoplankton absorption in blue through increasing phytoplankton concentration was found to be accompanied by a colour shift to green (Morel and Prieur, 1977). Using this empirical observation, O'Reilly et al. (1998) conceived a blue/green maximum band ratio (MBR), known as the Ocean Colour (OC) algorithm. The OC MBR relates reflectance band ratios to chla using a single fourth-order polynomial:

$$Chla\_OC = 10^{(a+bX+cX^2+dX^3+eX^4)}, \tag{1.10}$$

where the polynomial fit coefficients $a, \dots, e$ are usually estimated from an *in situ* dataset of observations and $X$ is:

$$X = log_{10}\left(\frac{R_{rs}(\lambda_b)}{R_{rs}(\lambda_g)}\right). \tag{1.11}$$

In Eq. 1.11, $\lambda_b$ *and* $\lambda_g$ refer to a blue and green wavelength, respectively. The blue wavelength is determined as the maximum $R_{rs}(\lambda)$ value from a set of bands (depending on the OC version) and the green wavelength value is the band that is within the 545 and 570 nm range. Multiple OC versions exist with various band combinations. Taking the OC4 version as an example, Eq. 1.10 becomes:

$$X = log_{10}\left(\frac{max[R_{rs}(443), R_{rs}(490), R_{rs}(510)]}{R_{rs}(560)}\right). \tag{1.12}$$

The latest MBR update (OC6) divides the maximum $R_{rs}(\lambda)$ blue band value by the mean of $R_{rs}(560)$ and $R_{rs}(665)$ (O'Reilly and Werdell, 2019). The OC algorithms are semi-empirical algorithms frequently used in ocean, inland and coastal water studies until today (Matsushita et al., 2015; Neil et al., 2019; Sayers et al., 2015).

When phytoplankton pigments absorb sun light a very small fraction of this absorbed light, referred to as the quantum yield $\phi_f$, is fluoresced. The fluorescence of substances in most cases refers to the emission of radiation at a longer wavelength than that at which it was absorbed. In this case, it is known as sun-induced chlorophyll-*a* fluorescence (SICF). Radiance from the fluorescence emission at the measurement wavelength can be extracted using the semi-empirical fluorescence line height (FLH) algorithm (Gower et al., 1999; Neville and Gower, 1977). The FLH algorithm estimates the magnitude of SICF at $R_{rs}(681)$ above a baseline between $R_{rs}(665)$ and $R_{rs}(708)$ (Gower et al., 1999):

$$FLH = R_{rs}(681) - \left[R_{rs}(708) + \left(R_{rs}(665) - R_{rs}(708)\right) \times \left(\frac{\lambda_{708} - \lambda_{681}}{\lambda_{708} - \lambda_{665}}\right)\right]. \quad (1.13)$$

The output is a difference in $R_{rs}(\lambda)$ which is then scaled to chla:

$$Chla\_FLH = a + b \times FLH, \quad (1.14)$$

where a and b are estimated from an *in situ* dataset. The FLH algorithm assumes the presence of the SICF peak at $\lambda_{681}$, which changes with frequency and length of sunlight exposure. A review on fluorescence inland water approaches was recently published by Gupana et al. (2021).

In addition to phytoplankton absorption and fluorescence the backscatter signal has a relationship to chla. As discussed, backscatter of phytoplankton becomes significant only for chla > 10 mg $m^{-3}$. Several algorithms target the 560 and 700 nm peaks to establish a relationship between the red-edge scattering signal near 700 nm and chla (Gitelson, 1992; Gons, 1999). When biomass increases, chla is empirically correlated to the height and position of the peak around 700 nm. The peak is primarily caused by increased particulate (rather than pigment) scattering which is offset by increased water absorption at wavelengths larger than 700 nm and the adjacent chla absorption band at 665 nm. Many two, three or four band ratio algorithms form ratios from the 700 nm and 665 nm bands (e.g. Gurlin et al., 2011; Mishra and Mishra, 2014). Other algorithms include these red bands in peak height approaches.

One example is the maximum peak height (MPH) algorithm that also uses SICF (Matthews et al., 2012). First, the algorithm selects the maximum signal of three bands ($R_{rs}(681)$, $R_{rs}(708)$ or $R_{rs}(753)$) to generate two maximum reflectance magnitude variables ($R_{max,0}, R_{max,1}$) and corresponding peak positions ($\lambda_{Rmax0}, \lambda_{Rmax1}$). Additional variables, such as the SICF peak, are calculated from red through NIR wavelengths (for more details on all variables see Matthews et al. (2012)). The two reflectance magnitudes ($R_{max,0}, R_{max,1}$) are used to derive corresponding $MPH_{0,1}$ variables for the calculation of chla, for example $MPH_0$:

$$MPH_0 = R_{max,0} - R_{rs}(665) - \left[\left(R_{rs}(885) - R_{rs}(665)\right) \times \left(\frac{\lambda_{max} - \lambda_{665}}{\lambda_{885} - \lambda_{665}}\right)\right]. \quad (1.15)$$

Through various conditional expressions each observation is tested for cyanobacteria occurrence. If the algorithm determines an observation is cyanobacteria-free, a polynomial is used to produce estimates of chla through usage of the $MPH_0$ variable:

$$Chla\_MPH_0 = 5.24 \times 10^9 MPH_0^4 - 1.95 \times 10^8 MPH_0^3 + 2.46 \times 10^6 MPH_0^2 + 4.02 \times 10^3 MPH_0 + 1.97. \quad (1.16)$$

In waters flagged for cyanobacteria occurrence the $MPH_1$ variable is employed:

$$Chla\_MPH_1 = 22.44 \times exp(35.79\ MPH_1). \quad (1.17)$$

The MPH equations can be calibrated for regional conditions as exercised in Pitarch et al. (2017). Chapter 2 and recent work by Schaeffer et al. (2022) shows that the MPH starts to excel at chla > 10 mg $m^{-3}$, but is ineffective for chla < 10 mg $m^{-3}$.

### 1.4.2 Analytical algorithms

As exemplified in section 1.2.2, in lakes non-phytoplankton materials compete strongly with phytoplankton in contribution to the total light absorption coefficient $a_{tot}(\lambda)$. Because of the influence of CDOM and NAP absorption, a simple empirical relationship between blue and green wavelengths and chla becomes severely degraded in waters with high CDOM and NAP. The difficulty originates from the fact that the relationship between the $R_{rs}(\lambda)$ and the IOPs of each water component is not unique. IOPs are cumulative, meaning that several water component combinations can lead to the same reflectance spectrum making it indistinguishable. Because $R_{rs}(\lambda)$ solutions are not unique in practice the inverse problem of aquatic remote sensing is said to be ill-posed or ambiguous (Babin, 2003; Defoin-Platel and Chami, 2007; Sydor et al., 2004).

Analytical algorithms have been designed to iteratively solve for IOP components by developing an appropriate bio-optical forward model, relating IOPs to $R_{rs}(\lambda)$. Werdell et al. (2018) outline two overall pathways for deriving component phytoplankton absorption: (1) simultaneous solving for $a_\phi(\lambda)$, $a_{CDOM}(\lambda), a_{NAP}(\lambda)$ and $b_{bp}(\lambda)$, the particulate backscattering coefficient; and (2) a determination of $a_{tot}(\lambda)$, and $b_{bp}(\lambda)$ to then decompose $a_{tot}(\lambda)$ into $a_\phi(\lambda)$. Both approaches are usually referred to as semi-analytical approaches because they use some form of empiricism (such as datasets, parameters or empirical relationships). SAA based on simultaneous partitioning of IOPs assign constant spectral shape values for $a_w(\lambda)$, and $b_{bw}(\lambda)$, then parametrise the spectral dependency of the IOPs of non-water constituents, i.e. the mass-specific constituent absorption coefficients, such as $a^*_\phi(\lambda)$, and retrieve the magnitudes of these constituents. Differences between the approaches include the definition of the IOP spectral shape, the relationship defined between $R_{rs}(\lambda)$ and IOPs and the statistical optimisation to calculate the magnitude of the component IOPs (Gege, 2014; Giardino et al., 2012). Several solution approaches exist such as non-linear optimisation (Kuchinke et al., 2009), linear inversion (Hoge and Lyon, 1996; Hoogenboom et al., 1998), spectral deconvolution (Lee et al., 2002) and bulk inversion (Loisel and Stramski, 2000).

SAA that employ a two-step solution for partitioning $a_{tot}(\lambda)$ into subcomponents begin with decomposition of $a_{nw}(\lambda)$ into $a_p(\lambda)$ (the particulate absorption coefficient) and $a_{CDOM}(\lambda)$ and then subsequently separate $a_p(\lambda)$ further into $a_\phi(\lambda)$ and $a_{NAP}(\lambda)$ (Jorge et al., 2021; Roesler et al., 1989; Zheng and Stramski, 2013). Limitations to both approaches include the assumptions about model outputs, the use of ancillary input data, generalised parameterisations, and spectral shape definitions.

## 1.5 Machine learning

Machine learning (ML) algorithms belong to the category of empirical approaches. The core concepts of ML are introduced separately because this dissertation frequently uses ML. In this work a ML "model" refers to the underlying statistical model, whereas "algorithm" includes the model and all elements included in the procedure to estimate phytoplankton (e.g., transformations of variables and values). ML is based on statistical models that learn a complex estimation function of a target variable, such as chla, through a measurement dataset called the training dataset. More philosophically, learning defined by Mitchell, (1997) is "a computer program said to learn from experience E, with respect to some class of tasks T and a performance measure P, if its performance at tasks in T, as measured by P, improves with experience E". Possible tasks, performances, and experiences are unlimited. Following the explanations given in Goodfellow et al. (2016), the task, performance and experience can be used to illustrate machine learning. In this work they are brought into the context of aquatic remote sensing.

### 1.5.1 Definition of learning tasks, T

A learning task can be described in terms of how a ML algorithm should process an observation. In aquatic remote sensing the input is typically a $R_{rs}(\lambda)$ spectrum, whose individual wavelengths are treated as features or variables. Statistically the $R_{rs}(\lambda)$ measurement is a vector $x \in \mathbb{R}^n$, where each entry $x_i$ of the vector is a wavelength.

This dissertation employs the two most prominent tasks in ML:

- **Classification.** The ML model is required to specify which of $n$ classes an input $R_{rs}(\lambda)$ belongs to. To solve this task the learning procedure must produce a function $f^*(x) -> \{1, \ldots, n\}$. If successful, then the model correctly assigned a $R_{rs}(\lambda)$ described by vector $x$ to class $y$. Chapter 4 contains the classification of $R_{rs}(\lambda)$ into trophic status classes of inland and nearshore waters.
- **Regression.** The ML algorithm estimates a numerical value for an input $R_{rs}(\lambda)$. Whereas in classification the target is categorical, in regression the target is continuous. In Chapter 3 the ML algorithm estimates chla from $R_{rs}(\lambda)$.

### 1.5.2 Measurement of performance, P

To gauge algorithm performance a quantitative measure must be chosen. The performance measure often varies with the task T and the learning algorithm. In classification, the classification accuracy measures how many observations were correctly classified, or vice versa, how many observations were misclassified, i.e. the misclassification rate or error. Moreover, it is important that the performance of the developed algorithm is evaluated on independent systems that were not part of model training. This may seem obvious, but in practice has a set of associated problems. Strategies to assure independence are outlined in Chapters 3 and 4 of this dissertation.

### 1.5.3 Algorithm experience, E

ML algorithms are either supervised or unsupervised, depending on their experience during the learning process. In unsupervised learning, the algorithm experiences a dataset containing features that enables to learn useful properties about the dataset. Unsupervised algorithms were used to cluster a dataset of $R_{rs}(\lambda)$ into OWTs (Moore et al., 2014; Spyrakos et al., 2018). Most common are supervised learning algorithms which not only experience a dataset of features for exploration, but in which each observation is associated with a target label. A label is either a continuous value or a defined class. Supervised learning algorithms are primarily used for the estimation of IOPs, OACs or classes through $R_{rs}(\lambda)$.

### 1.5.4 Linear models and least squares

The simplest ML algorithms are linear models. Given a $R_{rs}(\lambda)$ with features $\{x_1, x_2, \ldots, x_p\}$, the output value $\hat{y}$ can be estimated via the model:

$$\hat{y} = f(x) = \hat{\beta}_0 + \sum_{i=1}^{p} x_i \hat{\beta}_i. \tag{1.18}$$

$\hat{\beta}_0$ is the intercept (also known as bias in ML) and $\hat{\beta}_i's$ are the unknown coefficients. Least squares is a common approach to fit a model to training observations. In least squares fitting, the coefficients $\beta$ of the model minimise the residual sum of squares (RSS):

$$RSS(\beta) = \sum_{i=1}^{n} (y_i - x_i\beta)^2$$

$$= \sum_{i=1}^{n} \left( y_i - \beta_0 - \sum_{j=1}^{p} x_{ij} \hat{\beta}_j \right)^2. \tag{1.19}$$

Note that Eq. (1.18) makes no assumptions about the model validity of (1.17). It simply measures the average lack of fit and is thus applicable to a range of models. Standard $R_{rs}(\lambda)$ values can form the input, but they may also be transformed through the logarithm or the square-root of the inputs. Moreover, the input may be subject to basis expansions such as $x_2 = x_1^2, x_3 = x_1^3$, leading to a polynomial representation enabling arbitrarily complex linear models. Linear models are linear in their parameters, even though the relationship between features and the target variable can be non-linear.

### 1.5.5 Non-linear models

The linear equation has one basic form, namely the addition of each included (also transformed) term. Non-linear models can take many forms. In aquatic remote sensing especially decision trees (DTs), neural networks (NNs) and support vector machines (SVMs) are common non-linear models used for the estimation of chla (Cao et al., 2020; González Vilas et al., 2011; Matarrese et al., 2008;

Pahlevan et al., 2020). DTs and NNs are used in the following chapters and therefore the basics are explained in more detail.

### 1.5.5.1 Decision Trees

DTs divide the feature space $X$ through a set of possible feature values $x_1, x_2, \ldots, x_p$ into $M$ distinct and non-overlapping regions $R_1, R_2, \ldots, R_M$. For each region, DTs fit a simple model (referred to as a constant $c$). As for linear models the RSS can be used as a fit measure:

$$\sum_{i=1}^{M} \sum_{i \in R_M} \left(y_i - \hat{y}_{R_M}\right)^2 . \tag{1.20}$$

A standard regression tree model can thus be defined as:

$$\hat{y} = f(x) = \sum_{m=1}^{M} c_m \cdot 1_{(x \in R_m)} . \tag{1.21}$$

It is computationally infeasible to consider each partition of the feature space into $M$ possible regions. For this reason a top-down approach to split the feature space is a common approach, also known as recursive binary splitting (Strobl et al., 2009). Figure 1.8 shows an example with two input features $(x_1, x_2)$, for which a regression DT is constructed.

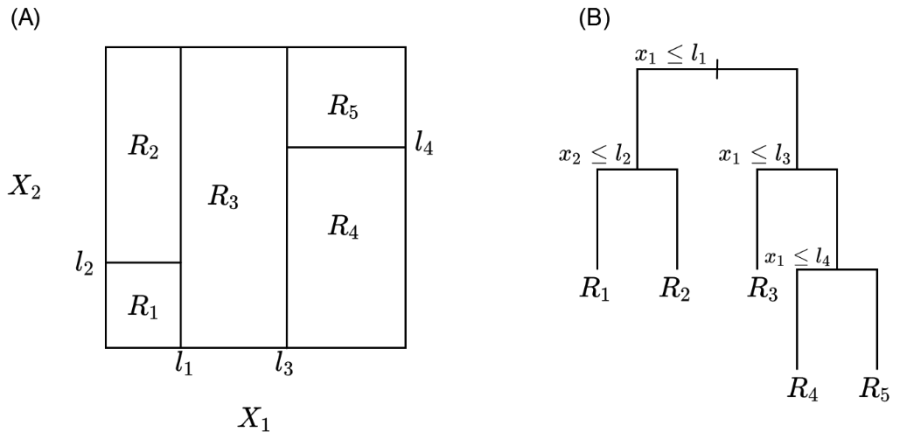


***Figure 1.8:*** *A regression DT obtained through recursive binary splitting. $R_1, R_2, \ldots, R_5$ represent the regions, $l_1, l_2, \ldots, l_4$ a set of rectangles used for the splitting. (A) Output of recursive binary splitting in an example with two features. (B) The resulting DT of the partitions displayed in (A). Re-drawn from Hastie et al., 2001.*

Formally, the DT displayed in Figure 1.8 estimates the target variable $\hat{y}$ with a constant $c_m$ in each region $R_1, R_2, \ldots, R_5$:

$$\hat{y} = \hat{f}(x) = \sum_{m=1}^{5} c_m \cdot 1\{(x_1, x_2) \in R_m\} . \tag{1.22}$$

Estimation capabilities of standard DTs can be enhanced through bagging (Breiman, 1996). Bagging starts with the generation of $N$ different bootstrapped, smaller training sets from the provided entire

set of training observations. DTs are then trained on the *n*-th bootstrapped set to obtain a model for each set: $\hat{f}^n(x)$. The estimates from each DT model can subsequently be combined, typically through averaging:

$$\hat{y} = \hat{f}_{bagging}(x) = \frac{1}{n}\sum_{n=1}^{N} \hat{f}^n(x)\,. \tag{1.23}$$

A popular decision tree variation is the RandomForest (RF) (Breiman, 2001). A RF is based on bagging in combination with random sampling of features. In each region split the model is only allowed one of *k* input features. A new set of *k* features is used for each split. In RFs, the DTs are not allowed to consider a majority of the available features (e.g., the wavelengths of the $R_{rs}(\lambda)$ vector). The rationale is to decorrelate the trees from each other and to reduce the influence of single strong features. RFs are used in Chapter 2.

The most common approach applied to DTs is boosting. Boosting works similar to bagging, except that a DT is grown sequentially: a DT grows with information obtained through a previous tree (Freund and Schapire, 1996). Whereas in bagging a new DT is fit to a new bootstrapped sample of the full training set, in boosting a tree is fit to the full training set. Boosting starts with an initial model that is built to estimate the target variable. Then, a new DT is fit to the residuals of the initial model, i.e., inaccurate regions of the overall DT. The updated DT is then included in the fitted function to update the residuals and the estimation is repeated. Improvements in $\hat{f}(x)$ are achieved through targeting of areas where the overall DT performs weakly. Boosting thus constructs a strong single model from several weak models (Breiman, 2004).

DTs enhanced through bagging or boosting have so-called hyper-parameters. Hyper-parameters guide the learning behaviour of the model. For boosted regression trees these are commonly:

- The maximum depth of the tree. Unlike bagging, boosting can produce very large and complex trees which tend to overfit on the provided set of observations (high variance)
- A learning rate to control model construction through boosting
- The number of splits in each tree, which controls the overall complexity of the boosted DT

Most prominent hyper-parameter approaches include a search through a grid (Liashchynskyi and Liashchynskyi, 2019), a random combination (Bergstra et al., 2012) or Bayesian optimisation (Wu et al., 2019). To find optimal hyper-parameters a part of the overall training set is usually set aside and called the validation set (not to be confused with the independent test set).

Boosting in combination with DTs is used in XGBoost (Chen and Guestrin, 2016) and LightGBM (Ke et al., 2017). XGBoost and LightGBM are employed in Chapter 3 and include several additional specifics which are omitted here for simplicity. DTs used in RFs or boosting schemes for inland waters were, for example, developed for the Landsat-8 Operational Land Imager (OLI) over Chinese waters (Cao et al., 2020) or with hyperspectral data collected from the European Elbe river (Keller et al., 2018).

### 1.5.5.2 Neural Networks

A neural network (NN), also known as multilayer perceptron (MLP), refers to a ML model loosely inspired by the human brain, wherein information is passed through connected neurons. As for DTs the aim is to approximate some function $f^*$. Unlike a DT however, a NN defines a mapping of $y = f^*(x; \theta)$ and learns the value of the parameter $\theta$ that results in the most accurate function approximation. The network term in NN derives from the combination of many different functions. For example, if four different functions $f^{(1)}, f^{(2)}, f^{(3)}, f^{(4)}$ are connected, they form a combined, chained function: $f(x) = f^{(4)}(f^{(3)}(f^{(2)}(f^{(1)}(x))))$ (Goodfellow et al., 2016). In this chain, $x$ is the input layer, $f^{(1)}$, $f^{(2)}$, $f^{(3)}$ the intermediate layers and $f^{(4)}$ the output layer. The intermediate layers are called hidden layers because they are not observable from the input features and the target variable(s). Each input feature is typically connected to a neuron of the first layer (Figure 1.9). The number of input features defines the width and the number of layers the depth of the model. To represent non-linear functions, NNs learn transformations $\phi$ of the input $(x)$. Such a model takes the form $y = f(x; \theta; w) = \phi(x; \theta)\, w$. Parameters $\theta$ are learned through a broad class of functions $\phi$, and weights $w$ map $\phi(x)$ to the output $y$. One may again use RSS$(\theta)$ as a measure of fit for a regression NN. The generic approach to minimising the fit function for NNs is through gradient descent (Ruder, 2017). The reason is that because of the chain of NN layers, a gradient can be derived using chain rules for differentiation. The gradient descent can be computed in a forward and backward pass through the network by keeping track of local changes in each neuron (Leung and Haykin, 1991).

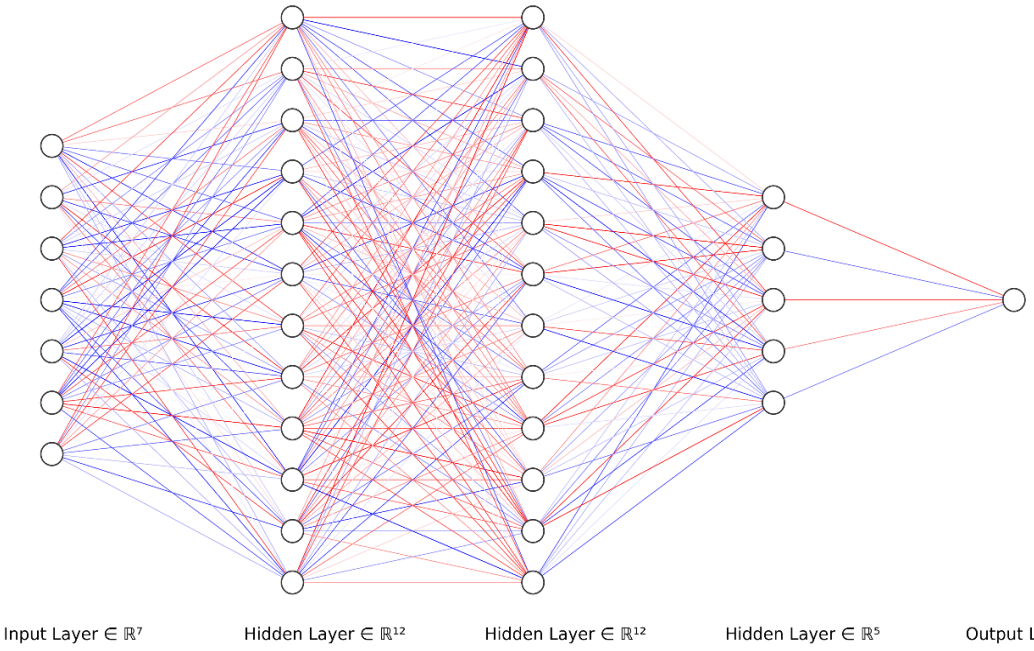


***Figure 1.9:*** *A neural network of the form $f(x) = f^{(4)}(f^{(3)}(f^{(2)}(f^{(1)}(x))))$, where x is the input layer (7 input features) connected to the first layer of neurons. $f^{(3)}, f^{(2)}, f^{(1)}$ represent the three hidden layers with 12-12-5 neurons and $f^{(4)}$ the output layer with a single neuron. Colors indicate the weights between the neurons (randomly set); red represents negative and blue positive weights.*

The adoption of NNs for oceanic waters started in the late 1990s (Buckton et al., 1999; Keiner and Yan, 1998; Krasnopolsky and Schiller, 2003; Schiller and Doerffer, 1999). NNs developed for coastal waters followed (Doerffer and Schiller, 2007; González Vilas et al., 2011; Hieronymi et al., 2017; Schiller, 2007; Schiller and Doerffer, 1999; Schroeder et al., 2007) and were applied to inland waters, but with mixed success (Binding et al., 2011; Neil et al., 2019; Philipson et al., 2016; Schaeffer et al., 2022). In recent years NNs were specifically developed for inland water OAC estimation with more complex extensions to the standard MLP. Examples are convolutional neural networks (CNNs) (Maier et al., 2021), mixture density networks (MDNs) (Pahlevan et al., 2020; Smith et al., 2021) or Bayesian neural networks (BNNs, as used in this dissertation Chapter 3). The amount of NNs designed specifically for inland waters remains small when compared to band ratios and other empirical approaches.

## 1.6 Uncertainties in aquatic remote sensing

The quality of reflectance data and derived downstream products such as chla varies per satellite scene. The distribution of gases and aerosols in the atmosphere changes rapidly, while the water surface and column are constantly changing through wind, waves and underwater mixing processes. Scene-specific conditions may deviate from the conditions considered during the design of an estimation algorithm, leading to uncertainty in the retrieval result. In the estimation of chla from satellite-derived $R_{rs}(\lambda)$ three main uncertainty sources exist:

1. Observational error
2. Atmospheric correction
3. The retrieval algorithm

Observational error refers to the systematic error associated with the physical process of taking a measurement (IOCCG, 2019). An example is the measurement procedure of an IOP or OAC in the laboratory, which has a systematic error (or flaw) that is unknown during the measurement. The systematic error has to be identified and a proposed correction exercised. Though, even in a (theoretical) scenario where systematic error can be excluded, simple measurement variations such as an inconsistent amount of water used for filtration will ultimately propagate into the measurement. Laboratory measurements of IOPs or OACs are often used for developing new algorithms and to validate their application. Thus, observational error is an important source of uncertainty that should be considered, and ideally quantified already during the measurement process. Uncertainty is part of laboratory chla (Claustre et al., 2004; Hooker et al., 2005; McKee et al., 2014), field-based radiometric measurements (Mélin and Franz, 2014; Zibordi and Voss, 2014) and varies per measurement methodology (IOCCG, 2019).

AC is the largest source of uncertainty as the retrieval of chla depends directly on the accuracy of satellite-derived $R_{rs}(\lambda)$ (Guanter et al., 2010; Pereira-Sandoval et al., 2019). Over inland waters the estimation of the aerosol contribution is the most inaccurate part of the AC process, as aerosols

strongly vary spatially and may be of terrestrial rather than aquatic origin. Small errors in the aerosol type and density property estimation may lead to large uncertainties in $L_w$, especially in the blue because here aerosol optical properties vary most significantly (Gordon and Wang, 1994; Pahlevan et al., 2021). Moreover, the water-leaving reflectance $L_w$ in NIR may not be negligible, and the AE can alter the radiance along the path to the sensor (Bulgarelli et al., 2014).

The retrieval algorithm refers to uncertainty that is introduced through the assumptions made when developing and applying an algorithmic procedure (Brewin et al., 2015; Salama and Stein, 2009). Any algorithm ultimately has an underlying model, whether physics-based or statistical. A model makes assumptions about valid conditions, and the assumptions may be invalid when applied to new, unknown observations. In the best case, the model can accurately deal with unknown conditions, but increased retrieval uncertainty is a common outcome because the encountered conditions deviate from those considered in the design phase of the algorithm.

## 1.7 Objectives

### 1.7.1 Scope of research

Previous remote sensing research over low biomass lakes is characterised by mixed success in the estimation of accurate chla. A sensitivity analysis conducted by Ruddick et al. (2001) on red/NIR band ratio approaches revealed that the relative uncertainty in chla retrievals was significant at low chla (< 10 mg $m^{-3}$) and when particulate backscatter was low. Neil et al. (2019) evaluated 47 chla algorithm configurations using an *in situ* dataset of 185 water bodies with chla ranging between 0.1 and 13.000 mg $m^{-3}$. The overall estimation accuracy for the dataset observations was 0.87 ($R^2$) when combining the best-performing chla algorithms through OWTs. The study found that chla retrieval for waters > 10 mg $m^{-3}$ was mostly accurate. Conversely, the highest chla retrieval uncertainties were associated with low biomass waters (chla < 10 mg $m^{-3}$). The chla retrieval inaccuracies for low biomass conditions reported by Ruddick et al. (2001) and Neil et al. (2019) are in line with other studies over the U.S. Great lakes (Gons et al., 2008, 2002; Mouw et al., 2015), Canadian (Binding et al., 2011) and Scandinavian lakes (Kallio et al., 2015; Kutser et al., 2016). In contrast, for the Swiss and Italian perialpine lakes more accurate chla estimation was reported (Bresciani et al., 2011, 2018; Giardino et al., 2001; Odermatt et al., 2008, 2010; Pepe et al., 2001).

Although the optical properties of low biomass lakes have been studied for years, the mixed estimation results and causes for algorithmic uncertainties are not fully understood. Moreover, an algorithm that can be applied to low biomass lakes across multiple regions does not exist. Although chla estimation uncertainty can be high, none of the developed chla algorithms provides an indication of confidence.

Over the last two decades field measurements of AOPs, IOPs and OACs increased in number and were made openly available through significant community-wide data sharing efforts. Large datasets

are the primary requirement for the development of ML algorithms, which are explored in this dissertation to address phytoplankton estimation uncertainties in low biomass lakes across multiple nations.

### 1.7.2 Research questions

The central research question addressed in the second chapter is:

- What are the reasons for the chla retrieval uncertainty of established and recently developed algorithms in low - moderate biomass lakes?

The second chapter "Characterising retrieval uncertainty of chlorophyll-*a* algorithms in oligotrophic and mesotrophic lakes and reservoirs" addresses the retrieval uncertainties of established and recently-developed chla algorithms over low biomass lakes and reservoirs (chla < 10 mg $m^{-3}$). Previously, significant retrieval uncertainty was found throughout many studies, but none of the studies investigated the underlying causes on a more representative dataset to detect fundamental issues that persist between water bodies. Therefore, a dataset of 346 observations from European, American and South African low biomass lakes and reservoirs was compiled. Recent and well-established empirical and SAA algorithms were applied to the dataset to obtain estimates of chla. To address the retrieval uncertainties, a RF model was fit to the residual values of each algorithm using a set of explanatory variables, such as IOPs and OACs other than chla. The RF residual model of each chla algorithm was subsequently investigated through Shapley Additive Explanations (SHAP) (Lundberg et al., 2020). Through SHAP it was possible to measure the variable impact on a model to characterise chla algorithm retrieval uncertainty. For the characterisation, relationships between the explanatory variables and residuals of each chla algorithm were analysed.

The following research questions are addressed in the third chapter:

- Do Bayesian neural networks (BNNs) improve the estimation of chla in oligo- and mesotrophic inland waters over reference algorithms?
- How can the provided uncertainty be used to improve chla products?

In the third chapter "A Bayesian approach for remote sensing of chlorophyll-a and associated retrieval uncertainty in oligotrophic and mesotrophic lakes" BNNs for the Sentinel-3 Ocean and Land Colour Instrument (OLCI) and Sentinel-2 MultiSpectral Imager (MSI) were developed. For the development, a semi-global dataset of lakes and reservoirs was compiled (1755 observations from 182 water bodies; chla dataset median: 5.11 mg $m^{-3}$, standard deviation: 10.76 mg $m^{-3}$). The BNNs are based on Bayesian probability theory (Ghahramani, 2015) in combination with Monte Carlo Dropout (Srivastava et al., 2014) to estimate chla and associated uncertainty. The BNNs were evaluated using randomised dataset splits and regional assessments. Moreover, a 3 ½ year-time series (including match-ups) using both OLCI and MSI over the LéXPLORE platform moored in Lake

Geneva were produced to investigate BNN performance over time. Several OLCI and MSI products were then generated to investigate how the BNN deals with uncertainty introduced through AC, the adjacency effect, and other sources that affect the chla estimation.

The fourth chapter addresses the following research questions:

- Is it possible to estimate trophic status of a water body directly from $R_{rs}(\lambda)$?
- Can an ensemble of classification algorithms improve trophic status estimation over conventional switching of multiple chla algorithms per OWT?

The fourth chapter, entitled "Meta-classification of remote sensing reflectance to estimate trophic status of inland and nearshore waters", uses the entire LIMNADES database (https://limnades.stir.ac.uk) to build a meta-learning algorithm to ultimately classify $R_{rs}(\lambda)$ into trophic state classes directly, thereby circumventing chla retrieval altogether. The meta-learner in this study is based on four individual base-learners, namely MLP NN, XGBoost, LightGBM and Naïve Bayes classifiers. The meta-learner is applied to a dataset of independent observations (n = 567) that were not used during the training and compared versus estimated trophic status obtained through switching of chla algorithms per OWT.

## 1.8 Structure

Chapter 1 briefly summarises the physical concepts of aquatic remote sensing, provides a background on chla estimation and uncertainties through empirical (including those based on ML), and SAA algorithms and states the research objectives and questions that are addressed in the dissertation chapters. Chapters 2 - 4 are self-contained data chapters. These chapters were all peer-reviewed and are either already published or accepted for publication. Conclusions from the chapters are drawn in the Synopsis Chapter 5.

# Chapter 2: Characterising retrieval uncertainty of chlorophyll-*a* algorithms in oligotrophic and mesotrophic lakes and reservoirs

*This chapter was published as: Werther, M., Odermatt, D., Simis S.G.H, Gurlin, D., Jorge, D.S.F., Loisel, H., Hunter, P.D., Tyler, A.N., Spyrakos, E., 2022. Characterising retrieval uncertainty of chlorophyll-a algorithms in oligotrophic and mesotrophic lakes and reservoirs. ISPRS Journal of Photogrammetry and Remote Sensing 190, 279 – 300. This chapter's version differs from the journal paper.*

## Abstract

Remote sensing product uncertainties for phytoplankton chlorophyll-a (chla) concentration in oligotrophic and mesotrophic lakes and reservoirs were characterised across 13 existing algorithms using an *in situ* dataset of water constituent concentrations, inherent optical properties (IOPs) and remote-sensing reflectance spectra ($R_{rs}(\lambda)$) collected from 53 lakes and reservoirs (346 observations; chla concentration < 10 mg $m^{-3}$, dataset median 2.5 mg $m^{-3}$). Substantial shortcomings in retrieval accuracy were evident with median absolute percentage differences (MAPD) > 37% and mean absolute differences (MAD) > 1.82 mg $m^{-3}$. Using the Hyperspectral Imager for the Coastal Ocean (HICO) band configuration improved the accuracies by 10 - 20% compared to the Ocean and Land Colour Instrument (OLCI) configuration. Retrieval uncertainties were attributed to optical and biogeochemical properties using machine learning models through SHapley Additive exPlanations (SHAP). The chla retrieval uncertainty of most semi-analytical algorithms was primarily determined by phytoplankton absorption and composition. Machine learning chla algorithms showed relatively high sensitivity to light absorption by coloured dissolved organic matter (CDOM) and non-algal pigment particulates (NAP). In contrast, the uncertainties of red/near-infrared algorithms, which aim for lower uncertainty in the presence of CDOM and NAP, were primarily explained through the total absorption by phytoplankton at 673 nm ($a_{\phi}(673)$) and variables related to backscatter. Based on these uncertainty characterisations we discuss the suitability of the evaluated algorithm formulations, and we make recommendations for chla estimation improvements in oligo- and mesotrophic lakes and reservoirs.

## 2.1 Introduction

Optical remote sensors are used to observe the major optically active water constituent (OAC) concentrations in lakes and reservoirs. The phytoplankton pigment chlorophyll-a (chla) is used as a proxy of phytoplankton biomass (Gitelson et al., 1993; Mittenzwey et al., 1992) and net primary production (Carlson, 1977; Huot et al., 2007; Poikane et al., 2010) and thereby to assess the

ecological integrity of aquatic ecosystems. In order to map chla consistently in time and space, algorithms are needed to relate the observed signal at the sensor to the concentration of this pigment. Several studies have shown that multiple algorithms in switching and blending (Liu et al., 2021; Neil et al., 2019; Spyrakos et al., 2018) or ensemble schemes (Werther et al., 2021) outperform individual algorithms across the optical-biogeochemical diversity of natural waters. There is, therefore, a common need to express algorithmic uncertainty associated with each technique, to inform appropriate algorithm selection and to further their development. This requires the assessment of a range of approaches against a representative data set. Remote-sensing reflectance, $R_{rs}(\lambda)$, is determined by inherent optical properties (IOPs) of water and OACs (Gordon et al., 1988; Mobley, 1999). The estimation of chla from $R_{rs}(\lambda)$ can be analytically approached by exploiting this relationship (through forward and inverse modelling), or chla can be empirically associated with the variability of the $R_{rs}(\lambda)$ signal in one or more wavebands (Schalles, 2006). With increasing phytoplankton abundance in ocean waters, the main observed effect is a shift of the reflectance maximum from blue to green wavebands caused by a combination of increased absorption and particulate backscattering processes (Morel and Prieur, 1977). These empirical observations were translated into the widely used ocean colour (OC) algorithms (O'Reilly et al., 1998; O'Reilly and Werdell, 2019), of which several variations exist for specific sensor configurations and regional effects. Another empirical approach relates chla concentration to sun-induced fluorescence of chla associated with photosystem II in the region around 685 nm (Gower et al., 1999; Gupana et al., 2021; Neville and Gower, 1977). The height of the fluorescence peak above a baseline between 650 and 730 nm is then assumed to correlate linearly with chla concentration (Gower et al., 2004). Several other empirical and semi-empirical algorithms relate chla concentration to variability in the red/near-infrared (NIR) area of the reflectance spectrum, which is caused by expression of chla absorption in a narrow region around 675 nm. This expression can be offset against NIR reflectance near 700 nm, which has weak combined absorption by phytoplankton and water (Dall'Olmo et al., 2003; Gilerson et al., 2007; Gilerson et al., 2010; Gitelson, 1992; Moses et al., 2009). A further group of empirical algorithms were specifically designed for chla estimation in waters with high algal biomass and the identification of algae blooms, based on peak height methods like the fluorescence line height, exploiting variability in the spectral shape of the reflectance peak variation near 700 nm (Binding et al., 2013; Matthews et al., 2012). Besides band arithmetic approaches, machine learning (ML) algorithms are increasingly trained for use in inland waters. ML algorithms are based on non-linear regression models and fit within the empirical algorithm category, as they are developed with training datasets generated from field and/or simulated observations (Bricaud et al., 2004; Hieronymi et al., 2017; Pahlevan et al., 2020). Contrary to purely empirical algorithm formulations, analytical and semi-analytical chla algorithms start from physics-based inversion of IOPs from $R_{rs}(\lambda)$, such as the absorption by phytoplankton ($a_{\phi}(\lambda)$), which can be subsequently scaled to chla concentration (Lee et al., 2002). Most contemporary semi-analytical algorithms (SAA) invert sub-surface remote-

sensing reflectance ($r_{rs}(\lambda, sr^{-1})$) obtained through conversion of above-water $R_{rs}(\lambda)$. SAA show many variations in the inversion for $a_\phi(\lambda)$. They vary in their (empirical) definition of the $a_\phi(\lambda)$ spectral shape, the method to calculate the magnitude of $a_\phi(\lambda)$ and the defined relationship between $r_{rs}(\lambda)$ and $a_\phi(\lambda)$ (Werdell et al., 2018). If retrieval errors were constant across a wide chla concentration range, larger relative errors would occur in the low concentration part of the considered range. Several studies have shown that the issue with low chla concentration estimates is also more pressing in absolute terms, with widely applied empirical algorithms and SAA revealing large uncertainties in oligo- and mesotrophic systems. For example, (Mouw et al., 2013) evaluated seven SAAs in Lake Superior (North America), an oligotrophic system with OAC absorption dominated by CDOM. The evaluated SAA algorithms were unable to effectively retrieve chla concentration due to inaccurate $a_\phi(\lambda)$ estimation ($r^2$< 0.2). On a large in situ dataset (> 2900 observations), (Neil et al., 2019) evaluated 47 chla algorithms, which revealed that the largest retrieval uncertainties for the dataset were associated with oligo- and mesotrophic conditions. More accurate chla estimation was achieved for Swiss perialpine lakes. (Odermatt et al., 2010, 2008) demonstrated that for low chla concentrations (chla < 5 mg $m^{-3}$), lake-parameterised empirical and SAA algorithms can provide precise estimates (correlation coefficients between 0.58 - 0.94). Comparable outcomes for Italian perialpine lakes were achieved by (Bresciani et al., 2018; Giardino et al., 2001; Pepe et al., 2001). The results of previous studies show that chla retrieval accuracy in oligo- and mesotrophic waters is inconsistent across regions. To improve the chla retrieval across different regions, it is thus necessary to characterise the fundamental causes of chla retrieval uncertainty. Uncertainty of chla estimates (hereafter retrieval uncertainty) can originate from three sources (Merchant et al., 2017). First, observation uncertainty caused by random and systematic errors associated with the physical process of making measurements using optical sensors (IOCCG, 2019). Second, uncertainty introduced in the calculation of $R_{rs}(\lambda)$ from the top-of-atmosphere signal measured by a satellite sensor through an atmospheric correction (AC) algorithm (Moses et al., 2017; Pahlevan et al., 2021b). And third, algorithmic uncertainty stemming from the empirical and/or physical assumptions made and translated into a procedure to retrieve chla concentration (Brewin et al., 2015; Salama and Stein, 2009).

To aid the study of algorithmic uncertainties over oligo- and mesotrophic systems, we focus on the retrieval of chla concentration from *in situ* measured $R_{rs}(\lambda)$. It is known that significant observation uncertainties are intrinsic to *in situ* $R_{rs}(\lambda)$ (Mélin et al., 2016; Mélin and Franz, 2014; Zibordi and Voss, 2014) and chla quantification methods (Claustre et al., 2004; Hooker et al., 2005; McKee et al., 2014b; Sørensen et al., 2007). It is, however, common practice to validate chla retrieval algorithms against *in situ* reference measurements whose observation uncertainty is considered low in comparison to satellite-derived estimates (Sørensen et al., 2007). Yet, these uncertainties are rarely specified in large inland and coastal water *in situ* datasets, and clarification of their origin is not an established practice for algorithm development and validation (McKinna et al., 2021).

Observation and algorithmic uncertainty sources are commonly unknown, thus additional information that can be attributed to the retrieval uncertainty is necessary. The accuracy of chla algorithms can be assessed through calculation of the residual value, which is the difference between an algorithm estimate and a measured *in situ* reference value. The frequency of the obtained residual values forms a residual value distribution (RVD). To improve our understanding of the underlying sources of algorithmic uncertainty, an algorithm's RVD must first be attributed to measurement conditions or the presence of specific substances which may hinder the intended function of the algorithm. The measurement conditions and substances can be represented by a set of explanatory variables that were co-measured with $R_{rs}(\lambda)$ and chla concentration. Examples are IOPs, OACs other than chla or spectral band ratios.

To enable characterisations of retrieval uncertainty, we build ML regression models (herein uncertainty models) through a set of explanatory variables. The constructed uncertainty models quantify the retrieval uncertainty from a chla algorithm expressed through the RVD. Subsequently, we investigate the impact of the explanatory variables on an uncertainty model to characterise algorithmic uncertainty through specific observation conditions. For this purpose, we use the SHapley Additive exPlanations (SHAP) procedure (Lundberg and Lee, 2017; Lundberg et al., 2020, 2018). Through SHAP we aim to identify uncertainty drivers of 13 chla algorithms tested against a dataset of 53 waterbodies.

In summary, the research objectives of this study are to: 1) determine sensitivities between empirical and semi-analytical algorithms when applied to the varying spectral resolutions of the Hyperspectral Imager for the Coastal Ocean (HICO) and Ocean and Land Colour Instrument (OLCI), 2) characterise the retrieval uncertainties of the tested algorithms and ultimately, 3) reveal if any of the evaluated algorithms are sufficiently suitable for application to specific observation conditions under the presence of CDOM and NAP absorption across oligo- and mesotrophic lakes and reservoirs.

## 2.2 Data

### 2.2.1 *In situ* data description

A dataset comprised of oligo- and mesotrophic lakes and reservoirs was compiled from three data sources: (i) U.S. Wisconsin Department of Natural Resources (DNR), (ii) the global LIMNADES database (Lake Bio-optical Measurements and Matchup Data for Remote Sensing (https://limnades.stir.ac.uk/)) and (iii) and an internal dataset from the University of Stirling (UoStirling) (Table 2.1).

Only samples including information on $R_{rs}(\lambda)$, chla concentration, total suspended matter (TSM) dry weight, absorption by CDOM at 443 nm ($a_{CDOM}(443)$) and water transparency ($Z_{SD}$) were selected, with chla concentration ≤ 10 mg $m^{-3}$ and TSM dry weight ≤ 4 g $m^{-3}$. We limit this study

***Table 2.1:*** *Datasets and their water bodies included in this study.*

| Dataset name | Inland water system name (n = 53) | Number of observations (n = 346) | Prior data use |
|---|---|---|---|
| U.S. Wisconsin DNR | Big Saint Germain Lake, Big Sand Lake, Butternut Lake, Fence Lake, Geneva Lake, Green Lake, Lac Courte Oreilles, Lac Vieux Desert, Lake Chippewa, Lake Mendota, Lake Wissota, Metonga Lake, Pelican Lake, Rainbow Flowage, Rock Lake, Round Lake, Shawano Lake, Trout Lake | 46 | (Pahlevan et al., 2021b, 2021a, 2020) and see section "2.1.1. U.S. Wisconsin DNR" for more information and references. |
| LIMNADES | Branched Oak Lake, Crescent Lake, Cuerda del Pozo, Diamond Pond, East Twin Lake, Lake Erie, Ewill Lake, Fivemile Pond, Forest Lake, Fresmond, Lake Garda, Ginger Cove, Goose Pond, Granite Lake, Great Salt Lake, Groton Pond, Iznajar Reservoir, Loskop Reservoir, Lake Maggiore, Lake Mantova, Lake Okoboji, Lake Ontario, Lake Paeijaenne, Lake Peipsi, Lake Pyhäjärvi, Lake Trasimeno, Lake Vesijärvi, Lake Winnipeg | 276 | (Binding et al., 2008, 2013, 2011, 2010; Bresciani et al., 2011; Giardino et al., 2015, 2014, 2013, 2005; Gitelson et al., 2009; Guanter et al., 2010; Gurlin et al., 2011; Kallio et al., 2015; Kutser et al., 2013; Manzo et al., 2015; Matthews, 2014; Matthews and Bernard, 2013; Ruiz-Verdú et al., 2008, 2005; Schalles, 2006) |
| University of Stirling | Lake Biel, Coniston Water, Derwent Water, Lake Geneva, Loch Lomond, Loch Ness, Lake Vänern | 24 | (Aulló-Maestro, 2019) |

to oligo- and mesotrophic lakes and reservoirs whose chla concentration does not exceed 10 mg $m^{-3}$ because up to this threshold chla estimation uncertainties were found to vary and increase most substantially (Gilerson et al., 2010; Mouw et al., 2013; Neil et al., 2019; Shi et al., 2013; Smith et al., 2018). Across the U.S. state of Wisconsin in situ measurements were made at 18 lakes between spring 2014 and autumn 2016 (n = 46). The dataset was used in recently conducted algorithm development and inter-comparison studies (Pahlevan et al., 2021a, 2021b, 2020). Collection methods are detailed in Appendix 1. The datasets and methods used to measure the water constituent parameters and IOPs in LIMNADES are described in (Spyrakos et al., 2018) and publications associated with individually contributed datasets (see references in Table 2.1). IOP measurements were not available for all observations in this study and often only at 443 nm and 673 nm. $R_{rs}(\lambda)$ was measured just above the water surface at varying spectral resolutions, and provided at a common 1 nm resolution, interpolated from the contributed $R_{rs}(\lambda)$ rather than individual radiometric components. Further in situ samples were collected in seven lakes in the United Kingdom, Sweden

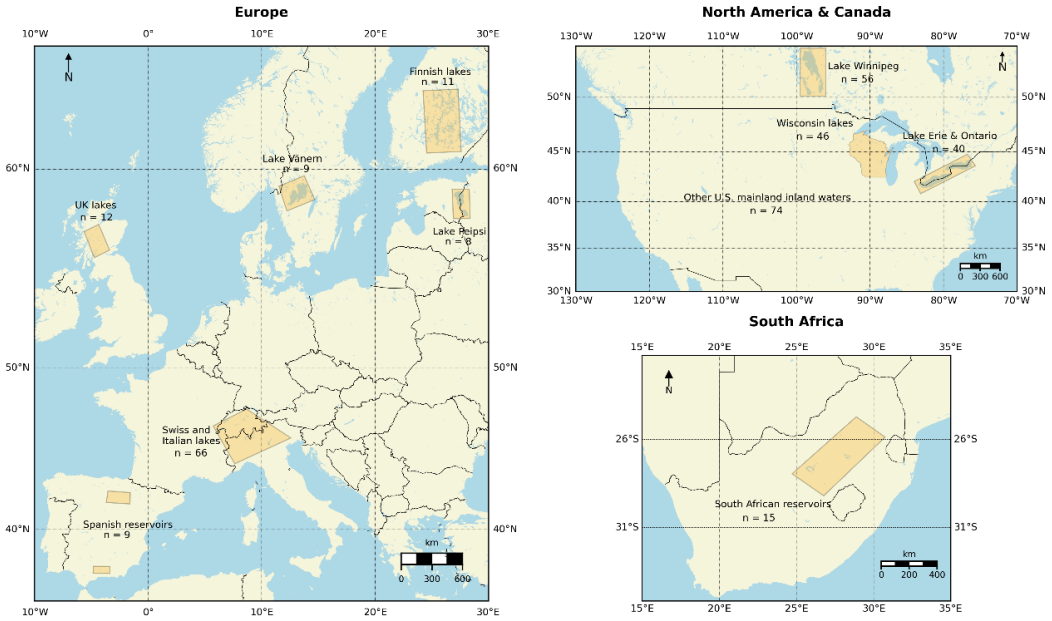


***Figure 2.1:*** *Locations of the measurement sites of this study. n is the number of observations taken in the indicated area. See Table 2.1 for details about the dataset.*

and Switzerland between 2013 and 2016 (Table 2.1). A detailed description of the measurements is provided in Appendix 2. $R_{rs}(\lambda)$ from the Wisconsin DNR and UoStirling datasets were interpolated to 1 nm spectral resolution to match the LIMNADES measurements. All three datasets were spectrally convolved to the band configurations of the hyperspectral HICO and multispectral OLCI sensors. Each chla algorithm tested in this study was used with the OLCI band configuration. HICO was included to test algorithms designed for hyperspectral configurations for the influence of the additionally available bands on both retrieval performance and associated uncertainty. The details of the spectral convolution are described in (Werther et al., 2021). The resulting dataset consisted of 346 observations from lakes and reservoirs located in the boreal and tundra zones of North America, Canada, and Europe. In addition, moderate-altitude and perialpine lakes (Italy and Switzerland) and clear water reservoirs of Spain and South Africa were included (Figure 2.1).

## 2.3 Methods

The analysis of algorithm retrieval uncertainty consists of three parts (Figure 2.2). First, explanatory variables are selected to construct an uncertainty model. Second, chla concentration is estimated for the collated dataset and the RVD is derived per algorithm. The uncertainty model is then fit to the RVD of the algorithms to quantify the retrieval uncertainty. Finally, the uncertainty model outcome is forwarded to SHAP to explain the variable impact on the model and to identify drivers of the quantified retrieval uncertainty.

### 2.3.1 Explanatory variables and uncertainty model definition

The initial set of explanatory variables consisted of the following measurements and indices: $a_{CDOM}(443)$, $a_{NAP}(443)$, $a_{\phi}(443)$ and $a_{\phi}$ $(673)$, $\frac{R_{rs}(400)}{R_{rs}(673)}$, TSM and ISM (the inorganic fraction of suspended matter), $\frac{chla}{TSM}$ and $Z_{SD}$. This set of explanatory variables was based on 89 out of 346 dataset records for which all the variables were available. These components of the dataset characterise the relative magnitude of IOPs in a sample, and partially its biogeochemical composition, as well as how biological components could be acclimated to ambient light conditions. Furthermore, they relate to water column clarity and phytoplankton variability.

ML in aquatic remote sensing is conventionally used to build a predictive model to estimate some quantity from $R_{rs}(\lambda)$ such as phytoplankton biomass (Pahlevan et al., 2020; Werther et al., 2021). Here we used a non-linear RandomForest (RF) ML approach to model expected complex RVDs. Given the explanatory variables as inputs we fit a unique RF to the RVD of each chla algorithm. We then investigated the RVD quantification by the RF and how the explanatory variables, used as the input, impacted the RF model. Therefore, RFs in this study were used as regression models that relate the explanatory variables to the retrieval uncertainty of each chla algorithm.

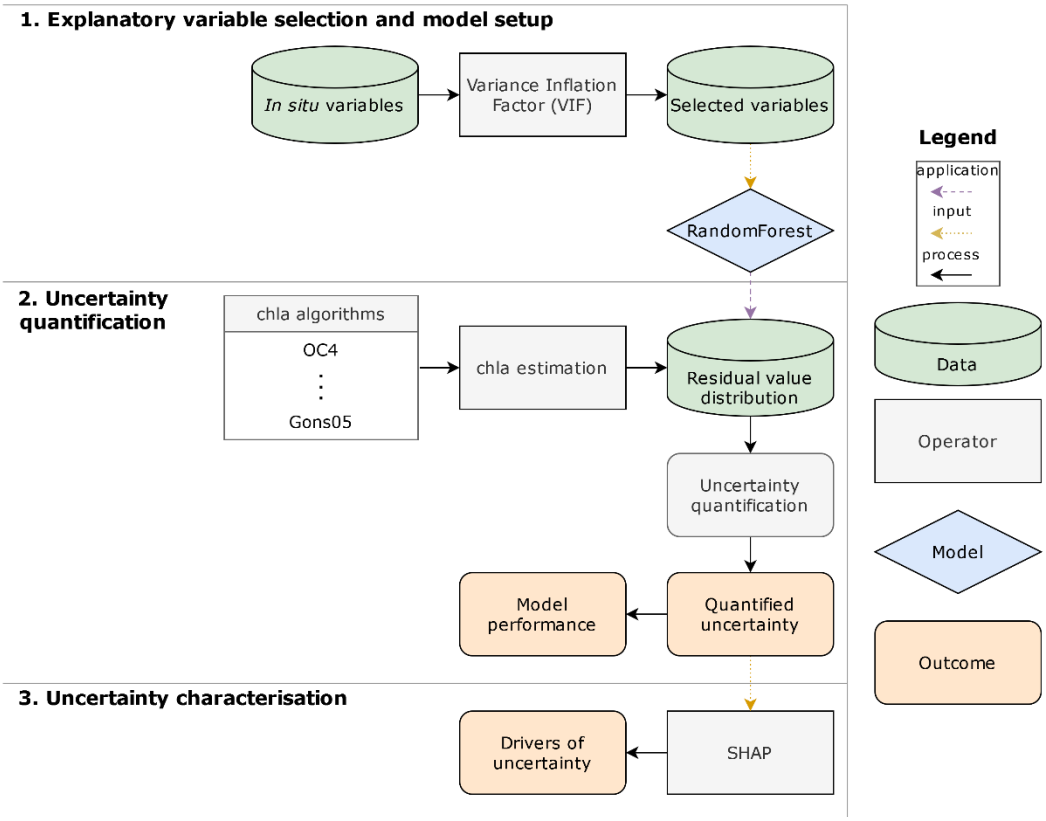


***Figure 2.2:*** *Uncertainty analysis scheme to characterise the retrieval uncertainties of the tested chla algorithms.*

A RF model is based on decision trees (DT), which split a variable space consisting of $c$ variables $c_1, c_2, \ldots, c_n$ into $j$ distinct, unique regions $R_1, R_2, \ldots, R_j$ (Breiman, 2001). DTs are constructed from a sample drawn with replacement (also known as bagging) (Breiman, 1996). A RF exposes

several hyper-parameters, most notably the number and maximum depth of the constructed trees, which were tuned as they guide the model construction procedure (Bergstra and Bengio, 2012). The aim was to fit a chla algorithm RVD as accurately as possible through a RF. Since the RFs are not used for prediction, the fitting was optimised through v-fold cross validation (where v was optimised for each RF model) and a search over a large pre-defined grid of the hyper-parameters was performed. Appendix 3 features a comparison of ML approaches for the uncertainty modelling task. To reduce multicollinearity between explanatory variables (e.g., $a_\phi(443)$ and $a_\phi(673)$), the initial set of explanatory variables was filtered using the variance inflation factor (VIF). This procedure regressed each variable *i* against all others, with the VIF per variable calculated as:

$$VIF_i = \frac{1}{1 - r^2}, \tag{2.1}$$

where $r^2$ is the coefficient of determination. A VIF value > 10 for an explanatory variable indicates high multicollinearity, in which case the variable that resulted in the strongest decrease of VIF was excluded (Miles, 2014). The number of variables excluded/kept depends on the considered dataset and the explanatory variables which ultimately determine the degree of multicollinearity. In practice other variables such as the spectral slopes of IOPs or other spectral ratios than $\frac{R_{rs}(400)}{R_{rs}(673)}$ can be used as explanatory variables. Variables should provide new knowledge about the observation conditions not already covered and must be sufficiently de-correlated to pass the VIF selection.

### 2.3.2 Chlorophyll-a retrieval and uncertainty quantification

A complete list of the chla algorithms evaluated in this study is provided in Table 2.2. We optimised the coefficients of the algorithms where possible (denoted as -opt) with the present dataset (denoted as k, l, m, n, o in Table 2.2). Several red/NIR approaches exist that use the 665 and 708 nm bands (Gilerson et al., 2010; Mishra and Mishra, 2012; Moses et al., 2009). For this study we selected the Gons05, G11 and MPH algorithms as they are commonly used over inland waters. Where previous studies already suggest merging of approaches (Schaeffer et al., 2022; Smith et al., 2021). We have only included their component algorithms to enable attribution of uncertainty to algorithm-specific causes.

The $a_\phi(443)$ and $a_\phi(673)$ estimates of the Quasi-Analytical Algorithm Version 6 (QAAv6), Garver-Siegel-Maritorena (GSM) and the 3-Step Semi-Analytical Algorithm (3SAA) algorithms were scaled to chla concentration using the relationship by (Bricaud et al., 1998):

$$chla = \left(\frac{a_\phi(\lambda)}{\boldsymbol{v}}\right)^{\frac{1}{w}}, \tag{2.2}$$

where $v$ and $w$ were estimated from the dataset of this study through non-linear least squares fitting. SAA performance highly depends on known parameter values. The slope of CDOM absorption *S*

***Table 2.2**: Chlorophyll-a algorithms used in this study.*

| **Algorithm number** | **Algorithm abbreviation** | **Equation** | **k** | **l** | **m** | **n** | **o** | **Reference(s)** |
|---|---|---|---|---|---|---|---|---|
| I | OC3 -opt | $Chla\_OC3 = 10^{(\boldsymbol{k} + \boldsymbol{l}X + \boldsymbol{m}X^2 + \boldsymbol{n}X^3 + \boldsymbol{o}X^4)}$ $X = log_{10}\left(\frac{max[R_{rs}(443), R_{rs}(490)]}{R_{rs}(560)}\right)$ | 0.088171 | -0.65440 | 8.56389 | 17.80220 | 5.89746 | (O'Reilly et al., 1998; O'Reilly and Werdell, 2019) |
| II | OC4 -opt | $Chla\_OC4 = 10^{(\boldsymbol{k} + \boldsymbol{l}X + \boldsymbol{m}X^2 + \boldsymbol{n}X^3 + \boldsymbol{o}X^4)}$ $X = log_{10}\left(\frac{max[R_{rs}(443), R_{rs}(490), R_{rs}(510)]}{R_{rs}(560)}\right)$ | 0.136257 | -1.084227 | 9.14503 | 4.79545 | -46.36124 | |
| III | OC6 -opt | $Chla\_OC6 = 10^{(\boldsymbol{k} + \boldsymbol{l}X + \boldsymbol{m}X^2 + \boldsymbol{n}X^3 + \boldsymbol{o}X^4)}$ $X = log_{10}\left(\frac{max[R_{rs}(412), R_{rs}(443), R_{rs}(490), R_{rs}(510)]}{mean[R_{rs}(560)/ R_{rs}(665)]}\right)$ | 0.48767 | -1.84059 | -0.27403 | 7.55868 | -3.59798 | |
| IV | G11 -opt | $Chla\_G11 = \boldsymbol{k}X^2 + \boldsymbol{l}X + \boldsymbol{m}$ $X = \frac{R_{rs}(708)}{R_{rs}(665)}$ | 12.0259 | -4.9116 | 1.2115 | | | (Gurlin et al., 2011) |

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| V | FLH | $FLH = R_{rs}(681) - (R_{rs}(708) + (R_{rs}(665) - R_{rs}(708)) \times (\frac{(708-681)}{(708-665)}))$ | | | | | | (Gower et al., 1999) |
| VI | MPH | $Chla\_MPH\_0 = 5.24 \times 10^9 MPH_0^4 - 1.95 \times 10^8 MPH_0^3 + 2.46 \times 10^6 MPH_0^2 + 4.02 \times 10^3 MPH_0 + 1.97$<br>$Chla\_MPH\_1 = 22.44 \times exp(35.79\ MPH_1)$ | | | | | | (Matthews et al., 2012; Matthews and Odermatt, 2015) |
| VII | MDN HICO | $chla^e\_mdn = \mu_i(R_{rs}): i = argmax\ \sigma(R_{rs})$ | | | | | | (Pahlevan et al., 2021b, 2020) |
| VIII | MDN OLCI | | | | | | | |
| IX | QAAv6 (443) | $a_\phi(\lambda) = a(\lambda) - a_{dg}(\lambda) - a_w(\lambda)$ | | | | | | (Lee et al., 2002) |
| | QAAv6 (673) | | | | | | | |
| X | GSM HICO | $a_\phi(\lambda) = chla \times \boldsymbol{k}$, | 0.029 (443 nm) and 0.018 | 0.017 | 1.0337 | | | (Garver and Siegel, 1997; Maritorena et al., 2002) |
| XI | GSM OLCI | $a_{CDOM}(\lambda) = a_{CDOM}(\lambda_0)\ exp\ (-\boldsymbol{l}(\lambda - \lambda_0))$ | | | | | | |

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| | | $b_{bp}(\lambda) = b_{bp}(\lambda_0)\left(\frac{\lambda}{\lambda_0}\right)^{\boldsymbol{m}}$ | (673 nm) | | | | | |
| XII | 3SAA | Chla_3SAA = $q_i/(q_2 + q_3 + q_4)$, where i = 2, 3 or 4 for pico-, nano- and microplankton absorption, respectively | | | | | | (Jamet et al., 2012; Jorge et al., 2021; Loisel et al., 2018; Loisel and Stramski, 2000) |
| XIII | Gons05 | $Chla\_Gons = \left[\left(\frac{R_{rs}(708)}{R_{rs}(665)}\right) \times (0.7 + b_b) - 0.4 - b_b^{\boldsymbol{k}}\right]/\boldsymbol{l}$<br>$b_b = \frac{0.6 \times a_w(778) \times R_{rs}(778)}{0.082 - 0.6 \times R_{rs}(778)}$ | 0.016 | 1.063 | | | | (Gons et al., 2008, 2005, 2002) |

(0.017 $m^{-1}$) was estimated from our dataset using the method described in (Stedmon et al., 2000). The mass-specific chla absorption coefficients $a_{\phi}^{*}(443)$ and $a_{\phi}^{*}(673)$ were estimated from observations where $a_{\phi}(\lambda)$ was available (n=89), as 0.029 $m^2g^{-1}$ and 0.018 $m^2g^{-1}$, respectively. Table 2.3 lists the scaling coefficients used in Eq. 2.2 for the SAA algorithms.

In contrast to the empirical algorithms, as well as to the Gons05 algorithm, the QAAv6, GSM and 3SAA algorithms were not primarily developed to assess chla concentration but IOPs such as $a_{\phi}(\lambda)$. These three algorithms have been thoroughly assessed over a large range of predominantly ocean and coastal water datasets. The latest IOP (not chla) retrieval exercise of the QAAv6, GSM and 3SAA algorithms was performed over 1020 observations and showed MAPD values of 28%, 40% and 31%, respectively (Jorge et al., 2021). Since chla concentration retrieval via SAA depends on $a_{\phi}(\lambda)$, for these three algorithms we also display the SAA $a_{\phi}(\lambda)$ estimates versus the respective *in situ* reference $a_{\phi}(\lambda)$ values. For the uncertainty analysis we use the GSM and 3SAA standard chla outputs as described in the original publications.

The Mixture Density Networks (MDNs) and Maximum Peak Height (MPH) algorithms were only applied to a subset of the dataset. Part of the dataset used in the training of the MDNs overlaps with the Wisconsin DNR and LIMNADES datasets of this study. We thus removed overlapping observations from this study as the evaluation would otherwise not be independent. After removal, 246/346 (71%) observations remained for application of the MDNs. MPH requires $R_{rs}(885)$, however for most observations in the dataset the wavelength range is restricted to 800 nm, hence limiting the evaluation of the MPH to 91 out of 346 observations.

***Table 2.3:*** *$a_{\phi}(\lambda)$ to chla concentration scaling coefficients used in Eq. 2.2 for the QAAv6, GSM and 3SAA algorithms derived from the dataset of this study.*

| **Algorithm number** | **Algorithm configuration** | **v** | **w** |
|---|---|---|---|
| IX | QAAv6 (443) | 0.0150 | 2.1117 |
| IX | QAAv6 (673) | 0.0033 | 2.8459 |
| X | GSM HICO (443) | 0.0015 | 3.4482 |
| XII | 3SAA (443) | 0.0027 | 3.0552 |

### 2.3.3 Uncertainty characterisation

To overcome the drawbacks of standard variable importance measures (Altmann et al., 2010; Grömping, 2009; Strobl et al., 2008) we use the SHAP procedure (Lundberg and Lee, 2017, 2016; Lundberg et al., 2020). Technical details about SHAP and standard variable importance measures are provided in Appendix 4.

We quantify the RVD of a chla algorithm with an uncertainty model $f(c^*)$, whose input is an explanatory variable vector $c = c^*$ (Figure 2.3). In Shapley game theory, S $\subseteq n = \{1,\dots,n\}$ is an ordered subset that consists of |S| variables (the coalition). Herein, a contribution function $v(S)$ maps subsets of variables to their real number, which is defined as the marginal contribution of an ordered subset of variables to the model's estimate. A Shapley value $\phi_c(v)$ is the contribution that an explanatory variable c provides to the model:

$$\phi_c(v) = \phi_c = \sum_{S \subseteq n \setminus \{c\}} \frac{|S|!\,(n - |S| - 1)!}{n!}\big(v(S \cup c) - v(S)\big), c = 1, \dots, n. \qquad (2.3)$$

In Eq. 2.3, the ordered subsets (coalitions) of variables are formed one at a time. The marginal contribution of each variable to the uncertainty model towards the estimation of the output value is calculated through $(v(S \cup c) - v(S))$. Then, for each variable, the average of this contribution over all orderings of all possible ordered subsets is computed. This yields the weighted mean over the uniquely ordered subsets. The sign of $\phi_c$ provides information about an explanatory variable's effect on the output. A positive value indicates that a variable c increases the model's estimated output value for an input observation. Conversely, a negative value indicates a decrease of the output value

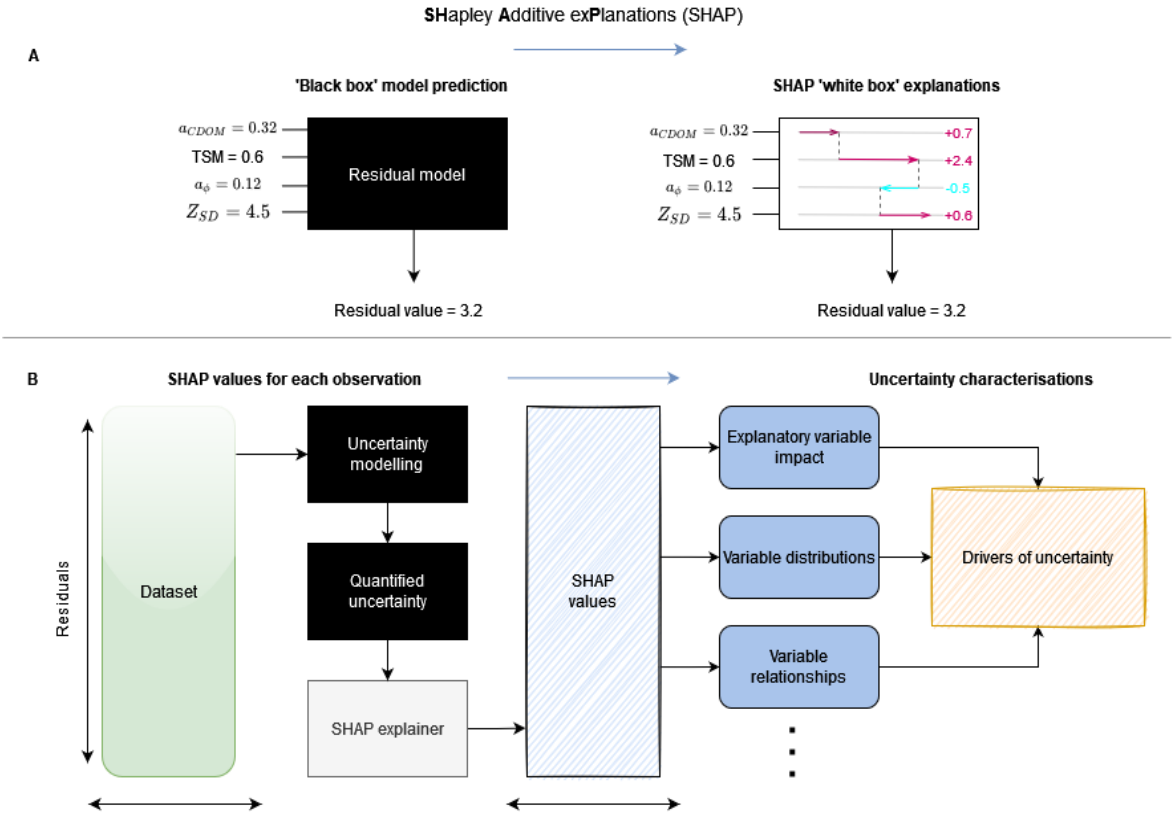


***Figure 2.3:*** *SHapley Additive exPlanations (SHAP) procedure. (A) Conceptual difference between standard black box ML model estimates and SHAP white box explanations. (B) Retrieval uncertainty is quantified through uncertainty models. The model is then provided to the SHAP explainer to calculate SHAP values. SHAP values provide the impact an explanatory variable had on an uncertainty model, enable to draw variable distributions and to investigate explanatory variable relationships. These drivers of uncertainty are used to characterise retrieval uncertainties. Adapted from Lundberg et al. (2020).*

(see white box in Figure 2.3A). The values of a variable may thus be related to both low and high residual values. The magnitude of $\phi_c$ indicates the impact a variable has on a model. Variable candidates can be identified by ranking the variables in order of impact on the model.

### 2.3.4 Performance metrics

Chla algorithm performance was expressed in terms of standard metrics such as bias, mean absolute difference (MAD), median absolute percentage difference (MAPD) and root mean square difference (RMSD) between the algorithm estimates (E) and the *in situ* chla observations (O) (Morley et al., 2018; Pahlevan et al., 2020; Seegers et al., 2018). We also report the slope of the regression to infer whether an algorithm systematically over- or underestimated chla or $a_\phi(\lambda)$ over the entire range of values. The MAD, MdAD (median absolute difference, no percentage) and RMSD metrics were used to measure the RVD fit accuracy of the RFs. All metrics were first calculated in logarithmic space and then transformed back to linear scale for display following recommended practice by Seegers et al. (2018). Linear scales enable to directly compare the chla performance metrics to the derived RVDs of the algorithms and fit accuracy of a RF uncertainty model.

## 2.4 Results

### 2.4.1 Optical and biogeochemical properties of the dataset

*In situ* data distributions show tailed distributions of chla (limited to 10 mg $m^{-3}$) with a median of 2.5 mg $m^{-3}$ and standard deviation (std. dev) of 2.57 mg $m^{-3}$ (Figure 2.4A). TSM concentrations (Figure 2.4B) covered a broad range (limited to 4 g $m^{-3}$) with a median of 1.87 g $m^{-3}$ (0.96 g $m^{-3}$ std. dev). $a_{CDOM}(443)$ was distributed around a median of 0.41 $m^{-1}$ (0.71 $m^{-1}$ std. dev) (Figure 4C), while $Z_{SD}$ had a median of 3.9 m (2.65 m std. dev) (Figure 2.4D). Absorption contributions of $a_\phi(\lambda)$, $a_{NAP}(\lambda)$ and $a_{CDOM}(\lambda)$ attribute the highest fraction of absorption to $a_{CDOM}(\lambda)$ and

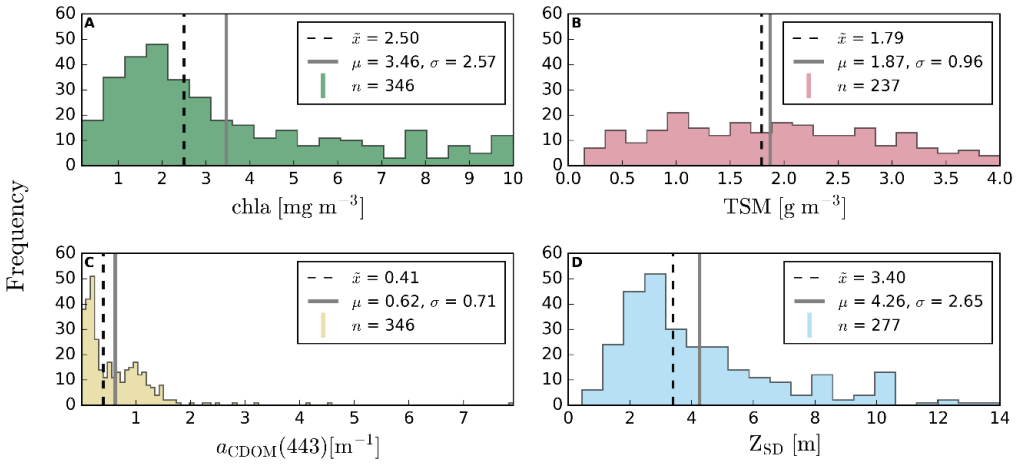


***Figure 2.4:*** *Parameter distributions of the dataset. (A) Chla, (B) TSM, (C) $a_{CDOM}(443)$ and (D) $Z_{SD}$. Denoted are the median ($\hat{x}$), mean ($\mu$) and standard deviation ($\sigma$) of the respective parameter.*

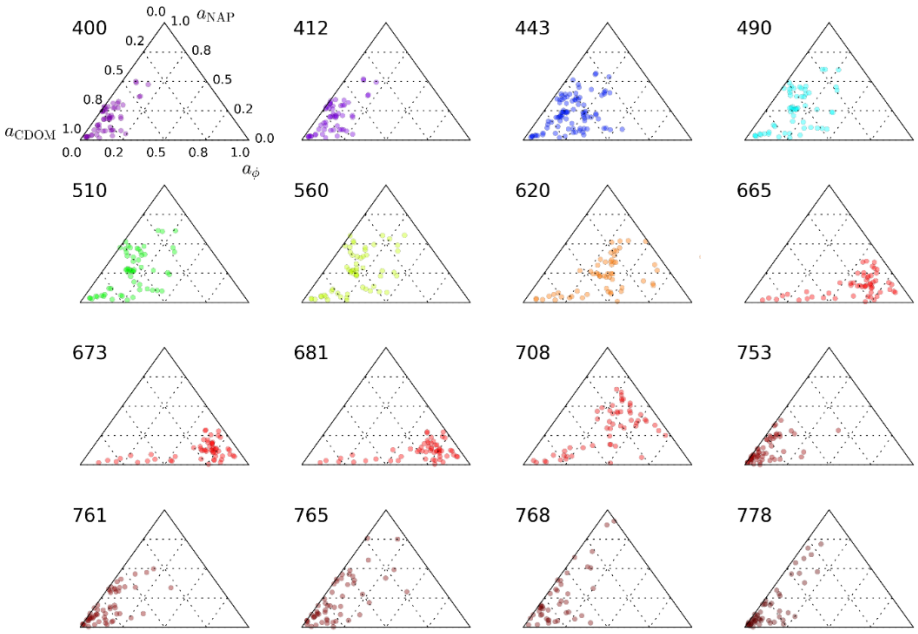


***Figure 2.5:*** *Ternary diagrams of absorption ($a_{\phi}(\lambda)$, $a_{NAP}(\lambda)$ and $a_{CDOM}(\lambda)$) for OLCI bands. Measurements where $a_p(\lambda)$ < 0.01 were excluded.*

$a_{NAP}(\lambda)$ from 400 through 560 nm (Figure 2.5). $a_{\phi}(\lambda)$ only contributed about 20% to the total non-water absorption budget but was still the major absorption fraction in red OLCI bands. $R_{rs}(\lambda)$ of the band configurations of HICO and OLCI are displayed in Figure 2.6, with a mean reflectance magnitude < 0.01 $sr^{-1}$ (Figure 2.6 A, B). Reflectance peaks in the red area of the spectrum were minor at 685 nm (fluorescence domain) and at longer wavebands (absorption and scattering domain). The relationships of $a_{\phi}(443)$ and $a_{\phi}(673)$ to chla concentration for this dataset were weak - moderate (Figure 2.7 A, B). Low TSM concentrations were weakly associated with low chla concentrations (Figure 2.7 C), whereas higher TSM concentrations did not exhibit a linear relationship to increasing chla concentration. Similarly, chla concentration was not correlated to $Z_{SD}$ below < 4 - 5 m.

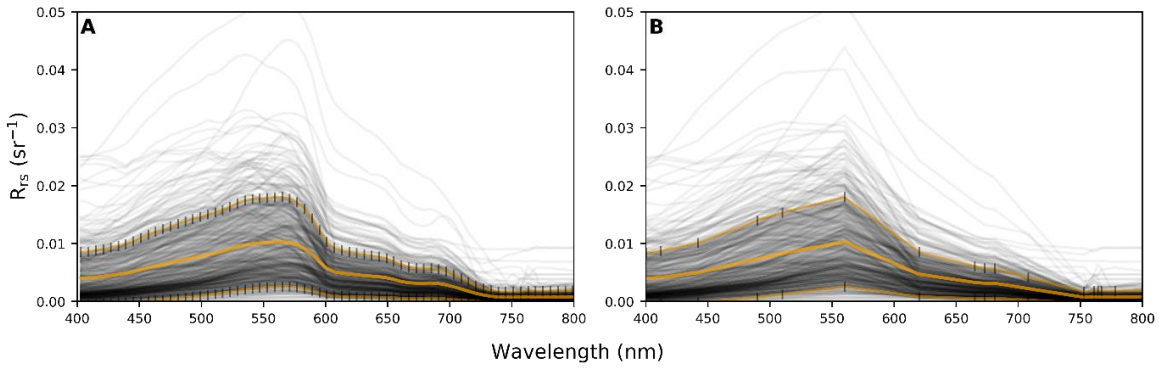


***Figure 2.6:*** *$R_{rs}(\lambda)$ of the dataset. (A) Hyperspectral HICO resolution. (B) Multispectral OLCI resolution. Solid orange lines depict mean +- standard deviation and black vertical lines the band positions of the sensors.*

A clearer trend emerged between records with low chla concentration and increased transparency at $Z_{SD} > 6$ m (Figure 2.7 D).

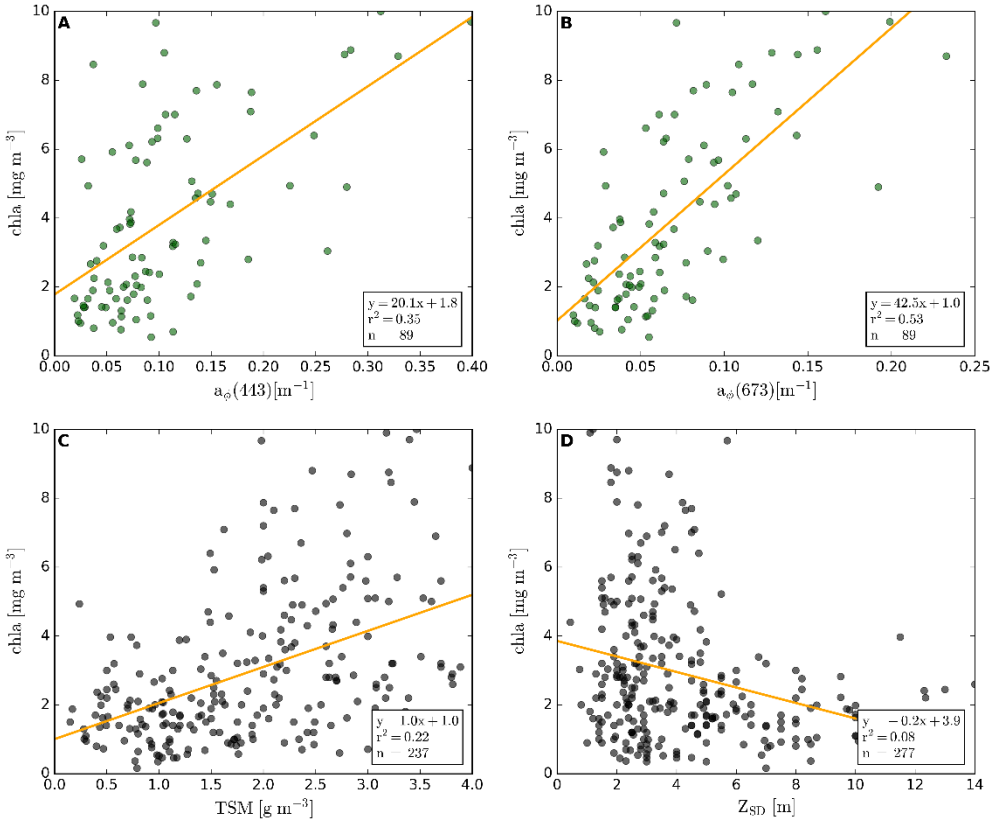


***Figure 2.7:*** *Relationships of chla concentration to other dataset parameters. Chla versus (A)* $a_\phi(443)$*, (B)* $a_\phi(673)$*, (C) TSM, (D)* $Z_{SD}$*.*

## 2.4.2 Chlorophyll-a algorithm performance

Chla concentration was estimated from all records in the dataset using the algorithms listed in Table 2.2 (Figure 2.8, 2.9). Performance metrics for each chla algorithm are detailed in Table 2.4. In addition, $a_\phi(443)$ and $a_\phi(673)$ values produced by the QAAv6, GSM and 3SAA algorithms are compared against $a_\phi(\lambda)$ *in situ* reference observations (Figure 2.10).

***Table 2.4:*** *Performance metrics of the chla algorithms applied to the full dataset (n = 346). A retrieval failure means that no chla concentration was returned for an observation. Grey rows indicate algorithm configurations used in the uncertainty analysis.*

| Algorithm number | Algorithm configuration | Retrieval failures | Slope | Bias | RMSD | MAD | MAPD |
|---|---|---|---|---|---|---|---|
| I | OC3 -opt | 0 | 0.37 | 1 | 1.89 | 1.97 | 39.45% |
| II | OC4 -opt | 0 | 0.36 | 1 | 1.90 | 1.96 | 38.52% |
| III | OC6 -opt | 0 | 0.36 | 1 | 1.90 | 1.97 | 39.56% |
| IV | G11 -opt | 0 | 0.31 | 1.25 | 1.97 | 1.95 | 36.88% |

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| V | FLH | 0 | 0.13 | 1.32 | 2.27 | 2.45 | 48.62% |
| VI | MPH | 0 | 0.06 | 0.76 | 2.19 | 2.56 | 52.45% |
| VII | MDN HICO | 0 | 0.71 | 1.61 | 2.23 | 2.27 | 53.29% |
| VIII | MDN OLCI | 0 | 0.73 | 1.87 | 2.50 | 2.73 | 63.41% |
| IX | QAAv6 ($a_\phi(443)$ to chla) | 41 | 0.33 | 1.28 | 1.97 | 1.98 | 40.28% |
| | QAAv6 ($a_\phi(673)$ to chla) | 16 | 0.20 | 1.33 | 2.11 | 2.20 | 48.42% |
| X | GSM HICO | 0 | 0.55 | 1.36 | 2.55 | 2.74 | 53.96% |
| XI | GSM OLCI | 0 | 0.59 | 1.77 | 2.75 | 3.09 | 62.16% |
| XII | 3SAA | 0 | 0.41 | 2.71 | 2.95 | 3.55 | 73.54% |
| XIII | Gons05 | 10 | 0.81 | 1.27 | 2.62 | 2.76 | 71.30% |

The GSM HICO and 3SAA algorithms produced a chla standard estimate via their non-linear optimisation routine. For these algorithms we also provide the results of scaling estimated $a_\phi(443)$ to chla concentration. This enables a direct comparison to the $a_\phi(443)$ estimates of the QAAv6 algorithm.

Chla retrieval through the OC3 -opt, OC4 -opt and OC6 -opt band ratios was only partially accurate across the considered range (MAD 1.97 mg $m^{-3}$, MAPD > 38%). The OC algorithms overestimated chla up until 5 mg $m^{-3}$ and consistently underestimated it thereafter. G11 -opt slightly outperformed the OC algorithms (MAD 1.95 mg $m^{-3}$ and MAPD of 36.88%) through more accurate retrieval > 5 mg $m^{-3}$, but like the OC algorithms also overestimated chla in the range below 5 mg $m^{-3}$. Similar to G11 -opt, Gons05 chla estimates started to become more accurate towards higher chla concentrations (> 5 mg $m^{-3}$). The FLH algorithm was unable to detect a fluorescence signal for most observations in the dataset which resulted in a poor relationship to chla. Like the FLH, the MPH algorithm lacked sensitivity over the chla range considered here and was unable to reproduce chla concentrations > 3 mg $m^{-3}$. The MDN HICO configuration outperformed its MDN OLCI counterpart by 10 - 20% (depending on the considered metric). Until 5 mg $m^{-3}$ of chla, both MDN HICO and OLCI over- or underestimated chla concentration and then consistently overestimated it for chla higher than 5 mg $m^{-3}$ by 1 – 3 mg $m^{-3}$.

The QAAv6 algorithm yielded $a_\phi(443)$ and $a_\phi(673)$ (hereafter QAAv6 (443) and QAAv6 (673)) which included 41 and 16 negative $a_\phi(443)$ and $a_\phi(673)$ values, respectively. The negative values could not be used in the power law function of Eq. 2.2 and were therefore excluded from the algorithm performance metrics but kept in the $a_\phi(\lambda)$ comparison (Figure 2.10). For the 305 remaining observations, QAAv6 (443) estimated chla similarly to the OC -opt and G11 -opt algorithms. Like the MDNs, the GSM HICO configuration outperformed the OLCI configuration by approximately 10%, but the algorithm was biased towards overestimation of chla concentration. The retrieval differences of the multispectral 3SAA algorithm were higher than the ones of the hyperspectral GSM HICO configuration (MAD 3.55 mg $m^{-3}$ and 2.74 mg $m^{-3}$, respectively).

For the QAAv6, GSM and 3SAA algorithms large retrieval differences (MAD >= 2.78 $m^{-1}$, MAPD > 53%) were found in the comparison of algorithm estimated $a_\phi(\lambda)$ versus *in situ* $a_\phi(\lambda)$ (Figure 2.10).

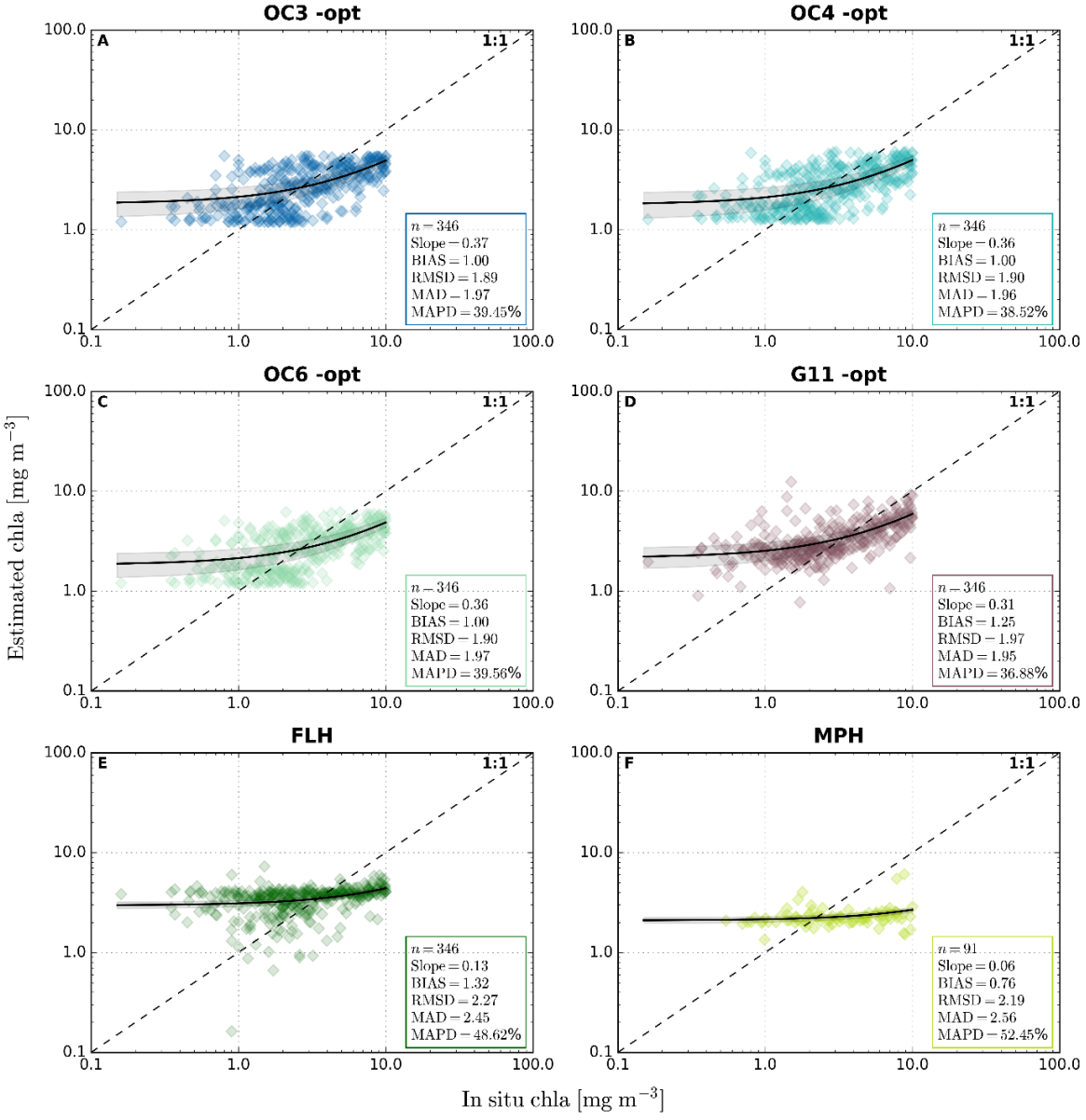


***Figure 2.8:*** *Chla concentration estimates of the OC -opt (I – III), G11 -opt (IV), FLH (V) and MPH (VI) chla algorithms. Black solid lines represent regression lines with 95% confidence intervals.*

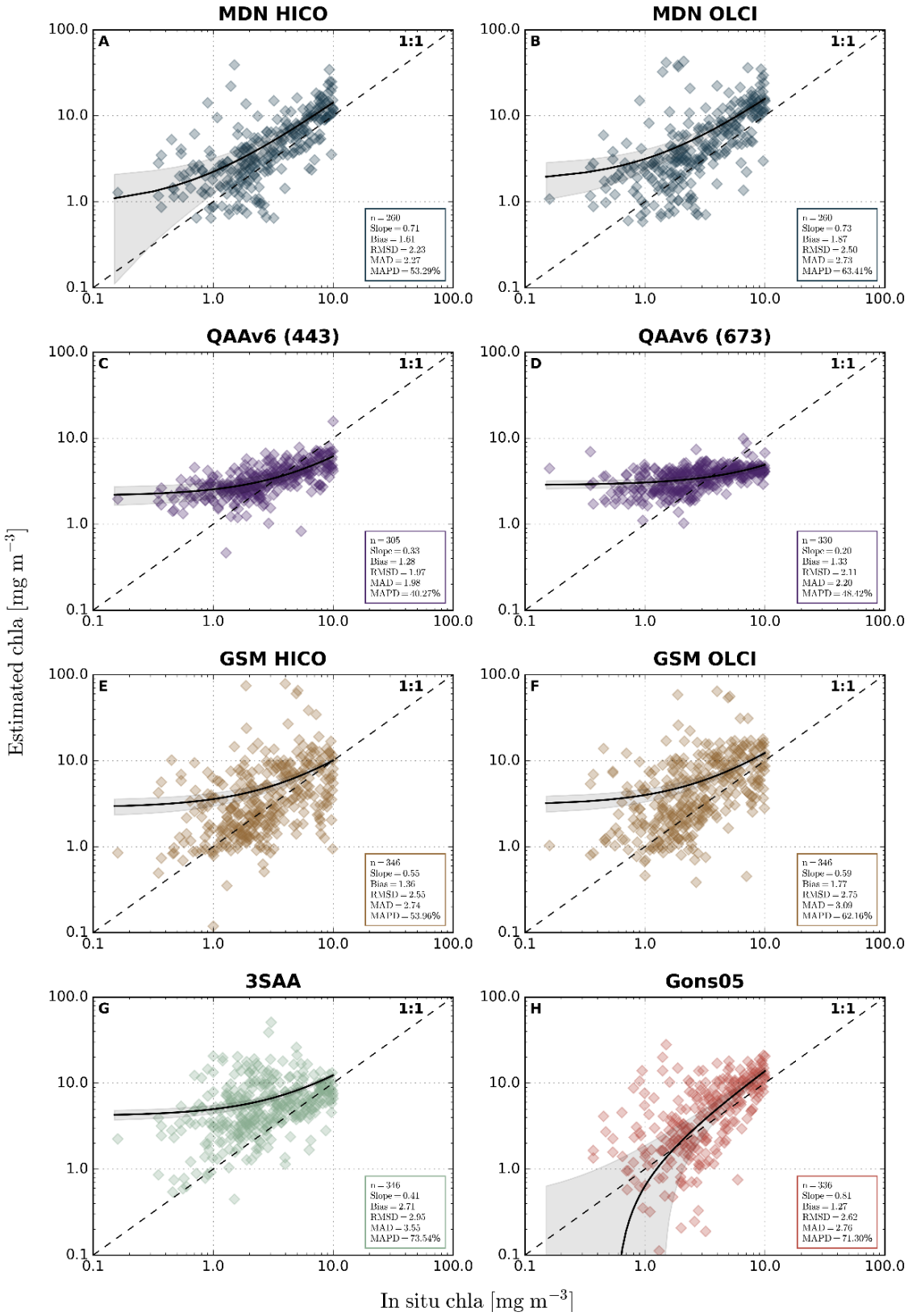


***Figure 2.9:*** *Chla concentration estimates of the MDN (VII - VIII), QAAv6 (IX), GSM (X - XI), 3SAA (XII) and Gons05 (XIII) algorithms. (A, B) MDN HICO and OLCI. (C) QAAv6 (443). Scaling of* $a_\phi(443)$ *to chla concentration via Eq. 2 with dataset coefficients (see Table 2.3 for the coefficients). (D) QAAv6 (673). Scaling of* $a_\phi(673)$ *to chla concentration with dataset coefficients. (E) GSM HICO, (F) GSM OLCI, (G) 3SAA, (H) Gons05 standard chla concentration estimates.*

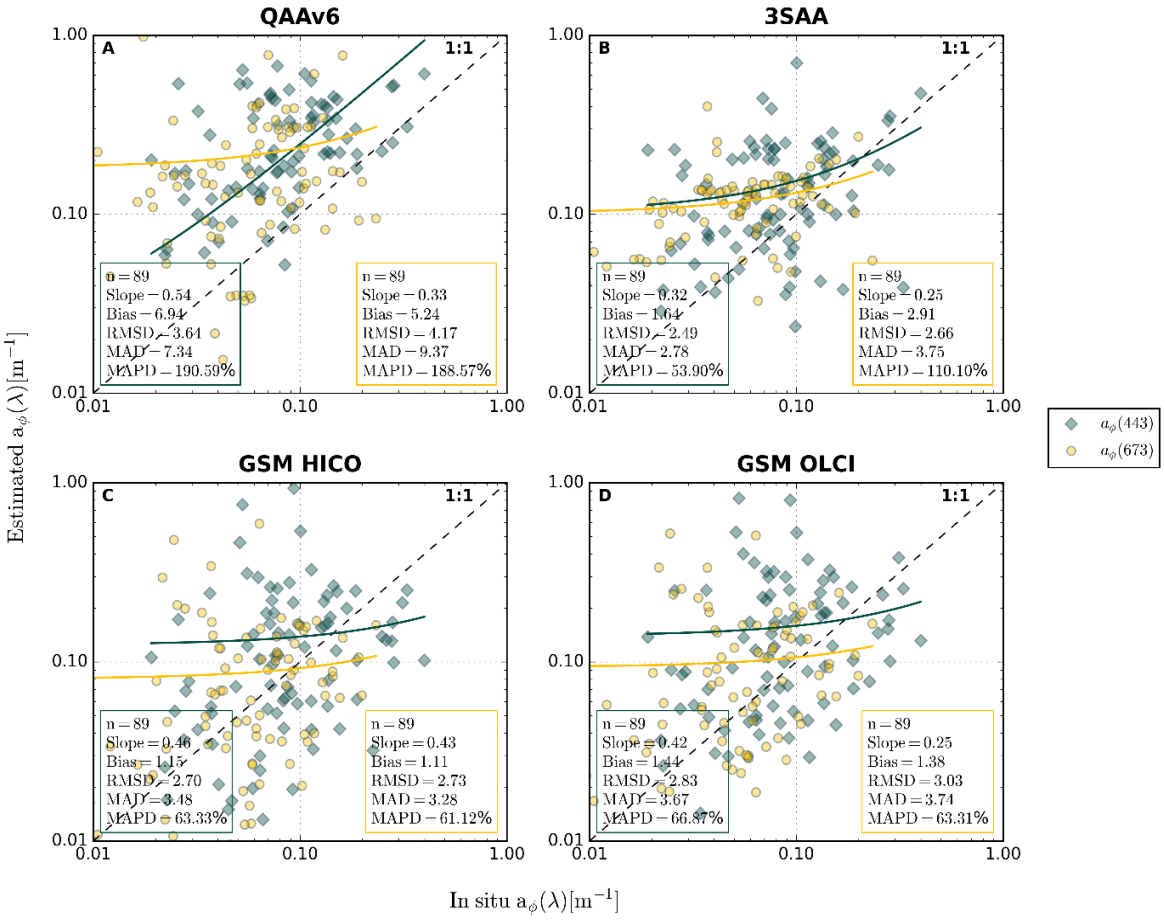


***Figure 2.10:*** *$a_\phi(443)$ and $a_\phi(673)$ estimates (green and yellow, respectively) of (A) QAAv6 (IX), (B) 3SAA (XII), (C, D) GSM HICO and OLCI algorithms (X - XI).*

### 2.4.3 Uncertainty quantification

Uncertainty budgets were derived for all algorithms with some exceptions. OC3 -opt was excluded because of nearly identical performance and a similar RVD as OC4 -opt (not shown). FLH and MPH algorithm estimates had very low retrieval sensitivity over the tested chla concentration range and

***Table 2.5:*** *RF uncertainty quantification accuracies of the RVD from each chla retrieval algorithm included in the uncertainty analysis.*

| **Chla algorithms** | **MdAD** $[mg\ m^{-3}]$ | **MAD** $[mg\ m^{-3}]$ | **RMSE** $[mg\ m^{-3}]$ |
|---|---|---|---|
| **OC4 -opt** | 0.50 | 0.79 | 1.08 |
| **OC6 -opt** | 0.58 | 0.84 | 1.13 |
| **G11 -opt** | 0.68 | 1.05 | 1.51 |
| **MDN HICO** | 0.64 | 0.96 | 1.26 |
| **QAAv6 (443)** | 0.62 | 0.96 | 1.39 |
| **GSM HICO** | 1.56 | 3.08 | 6.02 |

| | | | |
|---|---|---|---|
| **3SAA** | 0.92 | 1.56 | 2.42 |
| **Gons05** | 2.17 | 3.47 | 4.84 |

were therefore omitted. For the remaining algorithm set, the configuration with the lowest retrieval difference in terms of MAD and MAPD was selected. The MDN HICO compared favourably to its OLCI version, and QAAv6 (443) was more precise than QAAv6 (673). The Gons05 algorithm was not optimised and was thus included in its original form. For the GSM algorithms, we used the default GSM HICO chla estimates as they were more accurate than the OLCI configuration. For 3SAA, we used the chla concentrations resulting from addition of the three phytoplankton size classes.

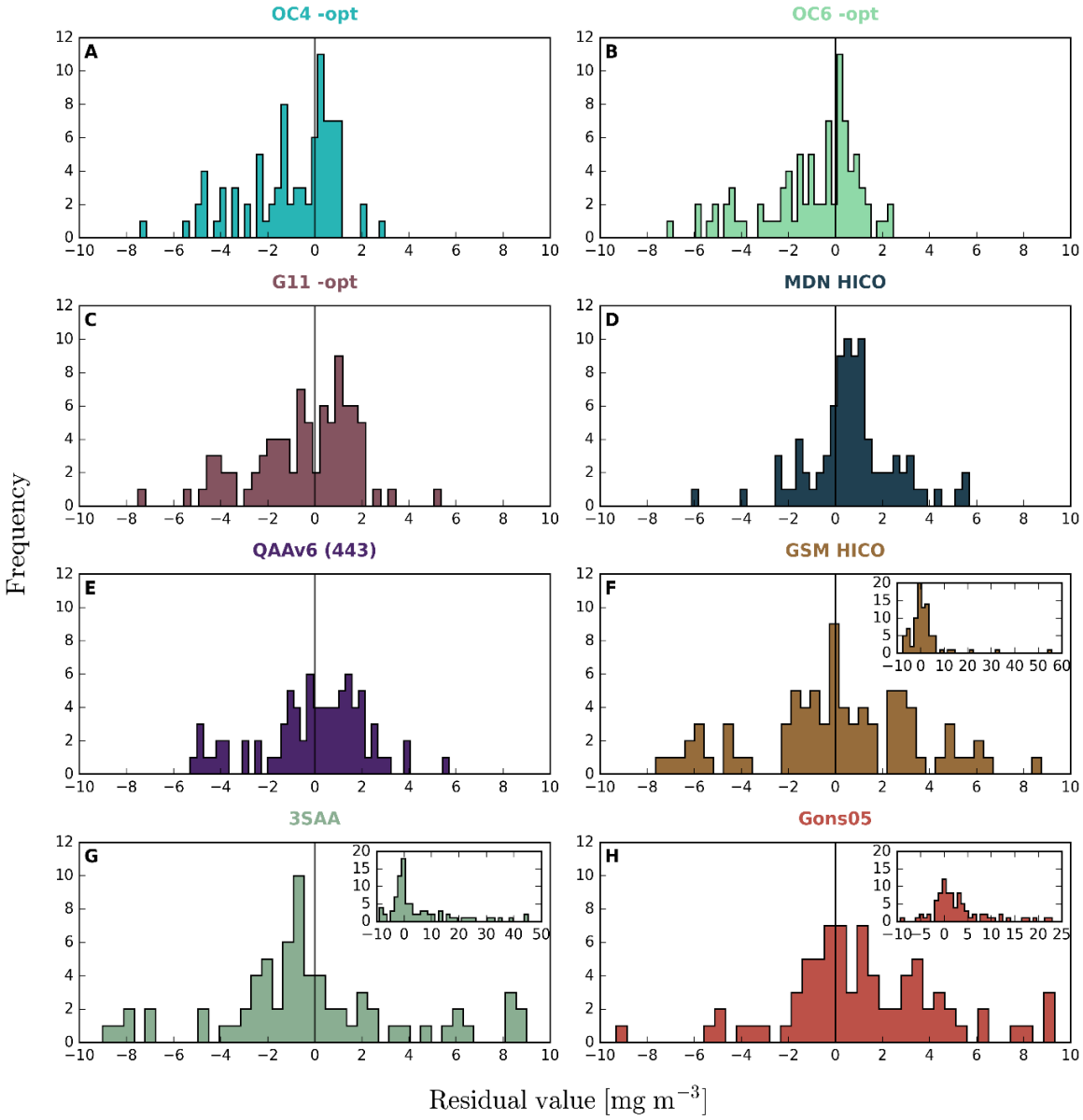


***Figure 2.11:*** *Residual value distributions (RVDs) of the chla algorithms for the observations included in the uncertainty analysis (n = 89). For GSM HICO, 3SAA and Gons05 the RVDs were limited to 10 mg $m^{-3}$ to facilitate a visual comparison of the distributions. The complete RVD of these three algorithms is displayed as an inset (F, G, H) and used in the uncertainty models.*

The VIFs calculated for ISM and TSM (11.75 and 7.64), $a_\phi(443)$ (6.74) and $a_\phi(673)$ (5.43) were higher than any of the other explanatory variables (not shown). The exclusion of ISM and $a_\phi(443)$ reduced the overall VIFs of the variables the most. Thus, seven explanatory variables remained that were used as the input for the uncertainty models.

Results of the RF retrieval uncertainty quantification are depicted in Table 2.5. The RVDs of GSM HICO, 3SAA and Gons05 contain high residual values (see Figure 2.11 F, G, H). Retrieval quantification accuracies of the uncertainty models are thus to be viewed with respect to the RVD of each algorithm. For example, the MdAD of 1.56 mg $m^{-3}$ for the GSM HICO uncertainty model can be considered accurate given the large residual value range (maximum 55.61 mg $m^{-3}$ overestimate). Further, the Gons05 algorithm overestimated 11 observations by more than 9 mg $m^{-3}$, resulting in the least accurate uncertainty model with a MdAD of 2.17 mg $m^{-3}$ (see Figure 2.10 H). The RVDs of the OC4 -opt, OC6 -opt, G11 -opt, MDN HICO and OLCI and QAAv6 (443) algorithms ranged from -8 mg $m^{-3}$ to 6 mg $m^{-3}$. Since the range of their RVDs is smaller than from the GSM HICO, 3SAA and Gons05 algorithms, the uncertainty models were able to quantify them with higher precision (e.g., OC4 MdAD of 0.50 mg $m^{-3}$, MAD of 0.79 mg $m^{-3}$ and RMSE of 1.08 mg $m^{-3}$).

### 2.4.4 Uncertainty characterisation

SHAP values were calculated to assess the explanatory power of each explanatory variable within each RF uncertainty model (Figure 2.12). The SHAP value magnitudes vary between chla algorithms because the associated RVDs of the chla algorithms are unique. An explanatory variable had impact on an uncertainty model if it was used in the estimation of an algorithm's RVD. Consequently, the variable impact is expressed by the corresponding SHAP values as either smaller or larger than 0. We examined each explanatory variable to identify drivers of uncertainty of the tested algorithms.

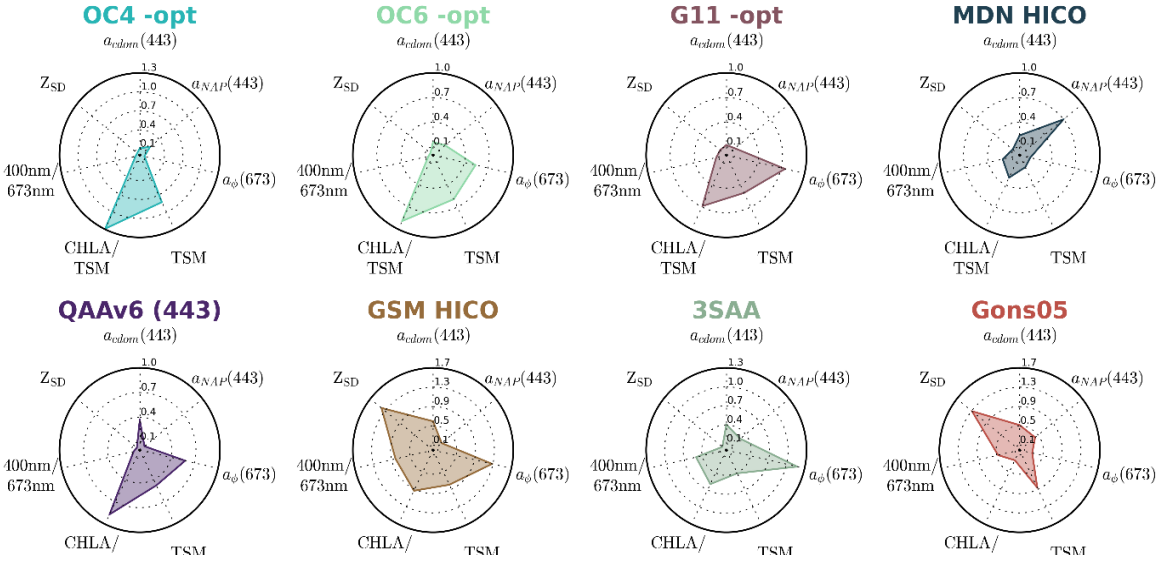


***Figure 2.12**: Average SHAP values of the explanatory variables for each chla algorithm. The average SHAP value was calculated as the average of all the absolute SHAP values for each observation associated with an explanatory variable.*

Except for MDN HICO and Gons05, the chla/TSM ratio was an impactful driver in all algorithms (Figure 2.13). Likewise, $a_\phi(673)$ was a highly impactful model variable for most models except for OC4 -opt, MDN HICO and Gons05. TSM was relevant for the OC4 -opt, OC6 -opt, G11 -opt, QAAv6 (443), GSM HICO, 3SAA and Gons05 uncertainty models. Especially for the OC4 -opt, G11 -opt and QAAv6 (443) algorithms, the only variables with a marked influence were the chla/TSM ratio, TSM and $a_\phi(673)$. In contrast, in the MDN uncertainty model $a_{NAP}(443)$ had significant impact. $a_{NAP}(443)$ had a marginal effect on the OC6 -opt uncertainty model and was not effective in the OC4 -opt and G11 -opt models. In contrast to the other variables of the OC6 -opt uncertainty model, high values of $a_{NAP}(443)$ $a_{CDOM}(443)$ decreased the model output variables. The other explanatory variables did not have a demonstrable effect on the OC6 -opt. $a_{CDOM}(443)$ values also contributed to explain retrieval uncertainty of the MDN, with low absorption values

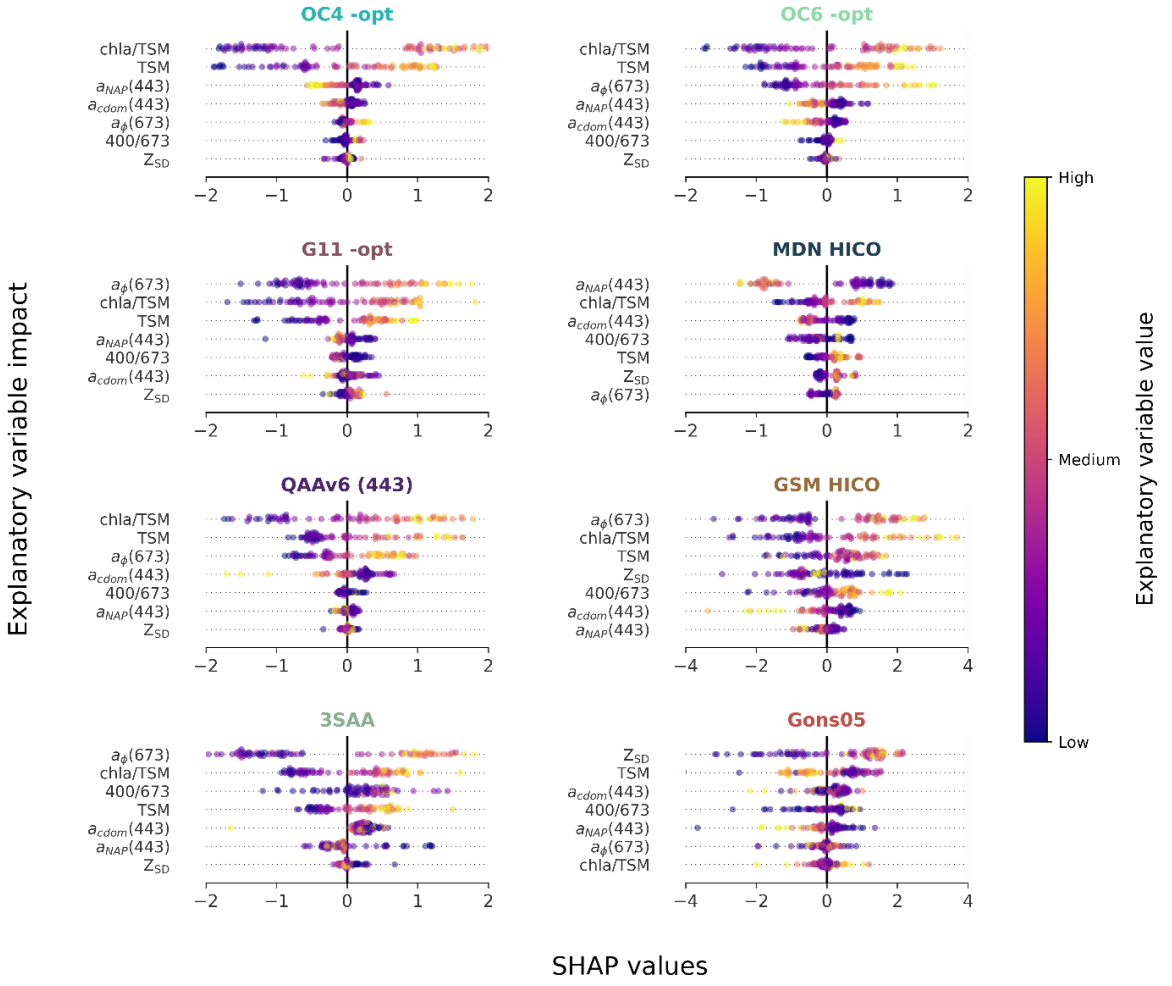


***Figure 2.13:*** *Individual SHAP value impact on the RF uncertainty models. The individual SHAP values associated with each observation of each explanatory variable and algorithm are shown. The impact of an explanatory variable on the RF uncertainty model is plotted in descending order of impact. The colour represents the explanatory variable value (yellow high, blue low). The model impact of a variable can be both positive and negative (see Figure 3A for more conceptual details). For display purposes, the variable* $\frac{R_{rs}(400)}{R_{rs}(673)}$ *is denoted as 400/673.*

having the largest impact on the model. Otherwise, $a_{CDOM}(443)$ was only an impactful predictor for the QAAv6 (443), 3SAA, and Gons05 uncertainty models. $Z_{SD}$ values were the least impactful considering the variable impact order across all models. Yet, for the uncertainty models of GSM HICO and Gons05, $Z_{SD}$ was a relevant variable.

For GSM HICO low $Z_{SD}$ values increased the model output, whereas medium to high values decreased it. The $\frac{R_{rs}(400)}{R_{rs}(673)}$ ratio was unimpactful for most models, except for GSM HICO, 3SAA and Gons05. For GSM HICO, low values of the $\frac{R_{rs}(400)}{R_{rs}(673)}$ ratio, i.e., where $R_{rs}(673) > R_{rs}(400)$, reduced the magnitude output, whereas high values, i.e., where $R_{rs}(400) > R_{rs}(673)$ increased the model output. For the 3SAA model, the ratio $\frac{R_{rs}(400)}{R_{rs}(673)}$ was highly impactful together with other variables related to phytoplankton absorption variability and composition, such as $a_{\phi}(673)$ and the chla/TSM ratio. The impact of $a_{CDOM}(443)$, $a_{NAP}(443)$ and $Z_{SD}$ on the respective 3SAA uncertainty model was low, which relates to the 3SAA algorithm formulation. The combined absorption of CDOM and NAP $\left(a_{dg}(\lambda)\right)$ was estimated in the first step of the algorithm and after removal from $a_{nw}(\lambda)$ the algorithm retrieved $a_{\phi}(\lambda)$. Hence, most of the influence of $a_{CDOM}(443)$ and $a_{NAP}(443)$ was incorporated by the algorithm formulation and the retrieval uncertainty of the algorithm was expectedly not related to these IOPs. $a_{CDOM}(443)$ and $a_{NAP}(443)$ were also irrelevant for the GSM HICO uncertainty model. The two SAA algorithms 3SAA and GSM HICO accounted for $a_{CDOM}(443)$ and $a_{NAP}(443)$ in their inversion procedure, but their retrieval uncertainty was driven by other factors such as phytoplankton absorption in red and variability (see Figure 2.13). For Gons05, $Z_{SD}$ and TSM had the strongest model impact. The values of $Z_{SD}$ and TSM impacted the model in contrary directions: whereas high $Z_{SD}$ values (i.e., higher water transparency) increased the model's output values, high TSM values (i.e., lower water transparency) decreased them. Contrary to the other models, especially G11 -opt, $a_{\phi}(673)$ and chla/TSM were the least impactful variables.

## 2.5 Discussion

While it is widely known that atmospheric disturbance on the water-leaving radiance and its correction contribute a large part to product uncertainty (Pahlevan et al., 2021a; Warren et al., 2019), it is equally important to understand algorithmic uncertainty arising from the translation of above-surface reflectance to the concentration of target substances. The approach presented in this study achieves this analysis using a highly heterogeneous collection of in situ observations and could be extended to inspect end-to-end product uncertainty (including AC effects) by starting from satellite observations.

The RF uncertainty models quantified the total uncertainty budget of a chla algorithm and did not distinguish between observational and algorithmic sources of uncertainty. Although most data used

in this study have been previously published, an unknown degree of observational uncertainty is likely part of the *in situ* reference measurements. This may also include biases between datasets due to use of similar but not identical methodology. The observed systematic and large chla retrieval uncertainties by the algorithms imply that the primary cause of retrieval uncertainty were the algorithmic formulations.

The information represented by the explanatory variables was accurately used by the RF models to quantify chla algorithm uncertainties (see Table 2.5), thereby demonstrating that retrieval uncertainties can be modelled through other *in situ* observations of the same dataset. While we compared the RF to a LASSO in this work and discovered higher uncertainty quantification accuracy (tabulated in Appendix 3), it would be beneficial to cross-validate the results of the RF feature significance using SHAP with other ML models. As a result, the causes of chla estimation uncertainty are more independent of the used ML model. We use the results of our uncertainty characterisation to discuss the algorithmic formulations of the chla algorithms tested in this study.

### 2.5.1 Blue/green algorithms

The reflectance shape and magnitude in the blue-green part of the spectrum in the evaluated dataset was largely influenced by CDOM and NAP. This contrasts the core assumption of the OC algorithms, namely that $R_{rs}(\lambda)$ co-varies with phytoplankton as the dominant optical component. Under- and overestimation of chla was large (residual values ranging from -5 mg $m^{-3}$ to + 4 mg $m^{-3}$). $a_{CDOM}(443)$ and $a_{NAP}(443)$ values were not used by the RF uncertainty models to estimate the RVDs of the OC4 and OC6 algorithms. In fact, the chla/TSM ratio and TSM were the most impactful variables. Thus, under dominating CDOM and NAP absorption, phytoplankton sources that influence the reflectance magnitude and shape variability, including varying phytoplankton cell size or the concentration of accessory photosynthetic and photoprotective pigments, could not be sensed by the OC algorithms.

### 2.5.2 Red/NIR algorithms

We observe that the G11 algorithm produced the most accurate chla values out of all the tested algorithms. G11 underestimated chla for concentrations higher than 4-5 mg $m^{-3}$ but the estimates were relatively accurate when compared to the other algorithms within this concentration range. Conversely, for concentrations lower than 5 mg $m^{-3}$ the G11 algorithm systematically overestimated chla values. This lack of sensitivity for concentrations lower than 5 mg $m^{-3}$ has also been observed in other studies (Pahlevan 2021b, 2020). The red/NIR area of the spectrum is the least affected by CDOM and NAP absorption. The RVD of the G11 algorithm was primarily modelled through the $a_{\phi}(673)$, TSM and chla/TSM ratio variables. The SHAP summary of the individual variables emphasised that high values of $a_{\phi}(673)$ were used to model the residual values of the G11 algorithm. The red/NIR band-ratio, which the G11 is fundamentally constructed from, does not suffice when the 665 and 708 nm bands do not hold sufficient information that can be related to chla concentration.

The Gons05 algorithm was overall less accurate than G11, which was also found by (Neil et al., 2019). The performance of the algorithm in this low – moderate chla concentration range confirms findings described in an earlier application of the algorithm to some of the Laurentian Great Lakes (Gons et al., 2008), where the retrieval differences were large for mesotrophic waters. The Gons05 retrieval uncertainty was most effectively modelled through $Z_{SD}$ and TSM. $a_{\phi}(673)$ and chla/TSM only impacted the Gons05 uncertainty model marginally, however these two variables were substantial for the G11 uncertainty model. Both G11 and Gons05 are based on the 665 and 708 nm band ratio, thus reversed variable importance may seem surprising. For Gons05, $Z_{SD}$ values were effective to estimate low residual values. In contrast, low-medium TSM values increased the uncertainty model output magnitude to model higher residual values (see Figure 2.13). $Z_{SD}$ and TSM are proxies of backscattering, which in turn is correlated to the calculation of the $b_b$ term using 778 nm in the Gons05 algorithm. 778 nm is however not part of the G11 algorithm, thus explaining the difference in uncertainty model variable impact. Consequently, optimisation of the backscatter term in the Gons05 algorithm might lead to performance improvements over the considered concentration range. An alternative route is to investigate whether the backscatter coefficient in the red/NIR range is too sensitive to low signal versus noise at 778 nm in oligo- and mesotrophic water bodies.

### 2.5.3 Machine learning algorithms

The MDN HICO and OLCI algorithms were previously trained with a semi-global dataset of inland and coastal systems. The dataset consisted of primarily eutrophic waters (chla dataset mean 21.7 mg $m^{-3}$, median 8.9 mg $m^{-3}$ and a std. dev of 47.5 mg $m^{-3}$). Seemingly, oligo- and mesotrophic waters with low chla concentrations and variable levels of CDOM and NAP absorption were under-represented during training. Although the algorithm had a positive bias (>1.61), the slope of the regression from the MDN HICO model was moderate (0.71), which implies that chla estimation was fairly consistent across the considered concentration range. This finding is further supported by the algorithm's RVD (Figure 2.11), in which most chla values were over- or underestimated by 1 – 3 mg $m^{-3}$. The uncertainty modelling (MdAD 0.64 mg $m^{-3}$, MAD 0.96 mg $m^{-3}$) of the MDN HICO algorithm revealed that $a_{NAP}(443)$ and $a_{CDOM}(443)$ are significant variables to characterise the retrieval uncertainties. An addition of observations characterised by the conditions of this study to the training measurement pool of the MDN could lead to direct retrieval improvements. Changes of the MDN training dataset would however not demonstrate whether the mixture density network is superior to other ML approaches for these waters, as we did not test other ML algorithms. New ML algorithms should put a larger emphasis on oligo- and mesotrophic datasets to facilitate a broader evaluation of diverse ML approaches to retrieve chla concentration in these conditions.

### 2.5.4 Semi-analytical algorithms

The retrieval uncertainty for the QAAv6, GSM and 3SAA algorithms was characterised by the variables $a_\phi(673)$ and the two ratios chla/TSM and $\frac{R_{rs}(400)}{R_{rs}(673)}$, indicating that a significant amount of phytoplankton absorption and composition variability could not be accounted for in the algorithm retrieval procedures. $a_\phi(\lambda)$ or combined phytoplankton size class absorption must be precisely estimated in the first place to accurately retrieve chla concentration through SAA. $a_\phi(\lambda)$ at 443 nm and 673 nm only has a weak - moderate relationship to chla concentration in this dataset (Figure 2.7). For algorithms that rely on $a_\phi(\lambda)$ to estimate chla, such as SAA (but also empirical fluorescence or IOP neural network algorithms), these weak relationships directly affect their accuracy. Consequently, empirical algorithms are more suited than SAA in scenarios where the uncertainty introduced by $a_{CDOM}(\lambda)$ and $a_{NAP}(\lambda)$ is smaller than by the $a_\phi(\lambda)$/chla ratio. For SAA prior information about $a_\phi(\lambda)$ variability is required. Without calibration, these models default to mass-specific phytoplankton absorption ($a_\phi^*(\lambda)$) values which are likely not appropriate for oligo- and mesotrophic lake conditions. In open ocean and coastal waters, $a_\phi^*(\lambda)$ exhibits a natural variability of a factor of 4 for any given chla value in addition to pigment packaging effects (Bricaud et al., 1995, 2004). This pattern stresses that the retrieval of chla concentration through $a_\phi(\lambda)$ may be significantly altered in the studied inland water conditions where $a_\phi^*(\lambda)$ variability is likely even higher. Thus, if uncalibrated, inappropriate $a_\phi^*(\lambda)$ values are used by the inverse SAA algorithms.

## 2.6 Conclusions

For oligo- and mesotrophic lakes and reservoirs, chla uncertainties were repeatedly shown to be highest when compared to higher biomass waters (Liu et al., 2021; Neil et al., 2019; Werther et al., 2021). The uncertainties persisted for the dataset of this study, as none of the tested algorithms accurately retrieved chla concentration over the considered range (0 – 10 mg $m^{-3}$).

The OC algorithms retrieved chla imprecisely, because the magnitude and shape of the used blue and green bands was mostly driven by CDOM and NAP absorption rather than by phytoplankton pigments. The results of the uncertainty analysis suggest that the OC algorithms are insensitive to phytoplankton absorption and composition variability in oligo- and mesotrophic lakes and reservoirs under the occurrence of CDOM and NAP absorption. The OC algorithms can therefore only be reliably used in scenarios where it is known that CDOM and NAP are not dominating the IOP absorption budget. In contrast, red-band algorithms such as G11 and Gons05 were partially able to retrieve chla accurately because the signal in their employed bands was not significantly influenced by CDOM and NAP absorption. The uncertainty modelling however revealed other causes for their retrieval inaccuracy, most of them related to low signal to noise ratios in these waters. As the uncertainty analysis demonstrated, for G11 the signal of $a_\phi(\lambda)$ in the red bands 665 and 708 nm was

too low to be accurately relatable to chla concentration. For the Gons05 algorithm the explanatory variables were highly related to backscatter which could not be precisely retrieved from the red/NIR bands. Moreover, chla retrieval by red-band peak height algorithms such as FLH and MPH was the least successful of all algorithms because the required sun-induced fluorescence signal was mostly not available or detectable.

The results of the uncertainty modelling for the tested MDN OLCI and HICO ML algorithms revealed that their consistent overestimation of chla concentration was related to CDOM and NAP absorption. The causes however were not comparable to the OC algorithms. Instead, measurements of chla and $R_{rs}(\lambda)$ from oligo- and mesotrophic lakes under high CDOM and NAP absorption were likely underrepresented during MDN model training, as the MDNs were built with a global dataset in which inland waters with low chla concentrations under varying CDOM and NAP absorption were under sampled (Filazzola et al., 2020; Pahlevan et al., 2020).

The SAA algorithms QAAv6, GSM and 3SAA consistently showed higher retrieval uncertainties than empirical algorithms. Our analysis revealed that SAA algorithms are at a disadvantage when the relationship of $a_\phi(\lambda)$ to chla concentration is only weak or moderate, as it was the case for this dataset at 443 and 673 nm for the samples included in the uncertainty analysis. Moreover, SAA make explicit assumptions about $a_\phi(\lambda)$ intensity and variability, which are likely too static when considering the phytoplankton variability and dynamic across multiple oligo- and mesotrophic lakes and reservoirs. It is thus likely that only with known or switching $a_\phi^*(\lambda)$ parametrisations SAA will produce accurate chla concentrations in the studied conditions.

For MDN and GSM, the HICO configuration outperformed OLCI by 10 – 20%, suggesting that higher information content improves the retrieval accuracy. Nevertheless, superior hyperspectral resolution did not overcome the issues faced by the tested algorithms. Possible causes of retrieval inaccuracy over oligo- and mesotrophic waters are still poorly handled by existing algorithms. *In situ* datasets are now starting to be of sufficient size to inform new algorithm development for optically complex inland water bodies with low chla concentrations. As most current chla retrieval algorithms lack confidence intervals, new developments should indicate if their use over the target conditions is appropriate.

## 2.7 Contributions

Concept: MW, ES, DO, SGHS, HL, DSFJ
Data curation: MW, ES, DO, DG, PH, AT
Investigation: MW, DO, SGHS, ES, DSFJ, HL
Methodology: MW
Code: MW
Visualisation: MW

Writing lead: MW

Writing – review & editing: SGHS, DO, ES, DG, DSFJ, HL, PH, AT

## 2.8 Acknowledgements

This study has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No. 776480 (MONOCLE). We thank two anonymous reviewers for their constructive and thorough comments to improve the manuscript. We owe our gratitude to the following data providers of LIMNADES for their measurements used in this study: Stewart Bernard, Caren E. Binding, Mariano Bresciani, Claudia Giardino, Anatoly A. Gitelson, Luis Guanter, Kari Y.O. Kallio, Tiit Kutser, Ciro Manzo, Mark W. Matthews, John F. Schalles and Antonio Ruiz-Verdú. We additionally thank Steven R. Greb for his efforts in collecting the Wisconsin DNR data.

## Appendix A1. Description of Wisconsin DNR *in situ* data

## Radiometric measurements

Radiometric measurements were made with a pair of inter-calibrated Ocean Optics USB series UV-NIR spectrometers (Ocean Insight, Orlando, FL, USA), which measured sub-surface upwelling radiance, $L_u(\lambda)$ and total downwelling irradiance above the surface, $E_d(\lambda)$ at the same time (Gitelson et al., 2008). $R_{rs}(\lambda)$ was derived through:

$$R_{\mathrm{rs}}(\lambda) = \frac{L_{\mathrm{u}}(\lambda)}{E_d(\lambda)} \times \frac{E_{\mathrm{ref}}(\lambda)}{L_{\mathrm{ref}}(\lambda)} \times R_{\mathrm{ref}}(\lambda) \times \frac{1}{100} \times \frac{t}{n_w^2} \times F_0(\lambda) \times \frac{1}{\pi}, \quad (A1)$$

where $E_{ref}(\lambda)$ and $L_{ref}(\lambda)$ are the total downwelling irradiance and upwelling radiance from a reflectance target, $R_{ref}(\lambda)$ is the respective spectral reflectance provided by the manufacturer, $t$ is the water-to-air transmittance (t = 0.98), $n_w$ is the refractive index of water (n = 1.33 at 20 °C), $F_0(\lambda)$ is the spectral immersion factor (Ohde and Siegel, 2003), and $\pi$ is introduced to transform the intermediately calculated water-leaving radiance reflectance $\rho_w(\lambda)$ into $R_{rs}(\lambda)$.

The spectrometers used for the radiometric measurements were customised with a 25-µm slit, a 600-line grating, which was blazed at 500 nm to optimise the instrument response for the bandwidth from 350 nm to 850 nm, and a standard ILX-511B detector, which resulted in a 1.39 nm spectral resolution. One of the spectrometers was connected to a 25° field-of-view optical fibre which was taped to a 4-m long, handheld Unger Opti-Loc extension pole. The pole was pointed away from the boat and the tip of the fibre was kept just beneath the water surface to measure $L_u(\lambda)$ at nadir on the sun-lit side of the boat. The second spectrometer was connected to an optical fibre fitted with a cosine collector to create a 180° field-of-view, taped to a 2-m pole. This pole was attached to a location on the boat free from adjacent influences to measure $E_d(\lambda)$ at zenith concurrently with the upwelling radiance without interference from surface structures. The spectrometers were inter-calibrated at the start of each set of measurements through six calibration scans of a white Spectralon™ reflectance target

(Labsphere, Inc., North Sutton, NH, USA), which was calibrated annually at the manufacturer. The calibration scans were followed by six to twelve scans of optically deep water which were processed to %-reflectance with the CDAP-2 software package (CALMIT, University of Nebraska-Lincoln, USA) and the median spectrum was considered representative for the station. System noise was removed with a smoothing spline function.

## Field measurements and laboratory analysis

Water transparency measurements ($Z_{SD}$) were collected using a standard 20-cm diameter Secchi disk. Surface water samples were collected with a Kemmerer Bottle at a depth of 0.5 m below the surface and stored cold in the dark. Upon completion of the filtering of the water samples shortly after their collection, the membranes for the laboratory analysis of chla were frozen and transferred to the Wisconsin State Laboratory of Hygiene (WSLH) where they were analysed, while the CDOM absorption, particulate absorption ($a_p(\lambda)$), and TSM samples were analysed in-house. The absorption measurements were taken as laboratory triplicates and the final parameter values correspond to the median of the measurements.

$a_{CDOM}(\lambda)$ samples were filtered through acid washed MF-Millipore GSWP mixed cellulose ester membranes while chla samples were filtered through DAWP membranes with a 0.65 µm pore size. For TSM and particulate absorption samples Whatman™ GF/F binder-free glass microfiber filters were used.

Chla was extracted from the concentrated algal samples (U.S. EPA 445.0) in a solution of aqueous 90% acetone aided by bath type sonication. Therefore, the sample racks with tubes were suspended in an ultrasonic bath, covered to exclude light, sonicated for 25 minutes, and stored at a temperature of less than 4°C to complete the extraction overnight. The samples were centrifuged for 30 minutes at 500 XG to clarify the extracts on the next day and the chla concentrations were measured fluorometrically (Welschmeyer, 1994) with a PerkinElmer LS-55 Fluorescence Spectrometer (PerkinElmer, Inc., Waltham, MA). The fluorescence spectrophotometer was calibrated with pure chla standards of a known concentration. TSM and the organic fraction of TSM (OSM) were measured gravimetrically from the dried and combusted residue on pre-combusted and pre-weighed filter pads (APHA 2540D and APHA 2540E).

Laboratory triplicates of $a_{CDOM}(\lambda)$ were transferred to a 0.1 m cuvette and optical densities of the filtrates were measured for a wavelength range from 350 nm to 800 nm and a bandwidth of 2 nm with the UV WinLab software package and a PerkinElmer LAMBDA 35 UV/Vis spectrophotometer. The signal from a de-ionized (DI) water blank was subtracted automatically and $a_{CDOM}(\lambda)$ was calculated as:

$$a_{CDOM}(\lambda) = \frac{l\,n(10)}{l}[OD_s(\lambda) - OD_{null}], \qquad (A1.1)$$

where $OD_S(\lambda)$ is the optical density of the sample, $OD_{null}(\lambda)$ is the average optical density of the sample from 700 to 720 nm (if $OD_S(440) < 0.05$) or 780 to 800 nm (if $OD_S(440) \geq 0.05$) for the null point correction, and l is the pathlength of the cuvette in m. The results represent the averaged laboratory triplicates.

The particulate and NAP absorption coefficients were measured through the quantitative filter technique (Mitchell et al., 2000). Laboratory triplicates were transferred to an optical glass slide mounted against the transmittance port of a Labsphere RSA-PE-20 accessory and the optical densities of the particles retained on the filters were measured over the 400 to 800 nm wavelength range and a bandwidth of 2 nm with the same spectrophotometer. The signal from a previously measured blank filter was subtracted and $a_p(\lambda)$ was calculated as:

$$a_{\mathrm{p}}(\lambda) = \frac{l\,n(10)}{\frac{V_{\mathrm{f}}}{A}}\left[0.3907 \times \left[OD_{\mathrm{fp}}(\lambda) - OD_{\mathrm{null}}\right] + 0.3201 \times \left[OD_{\mathrm{fp}}(\lambda) - OD_{\mathrm{null}}\right]^2\right], \qquad (A1.2)$$

where $OD_{fp}(\lambda)$ is the optical density of the sample and $OD_{fp}(\lambda)$ is the average optical density of the sample from 780 to 800 nm required for the null point correction of the measurements. V and A are the volume of water filtered in $m^3$ and the area of the filter in $m^2$ used to calculate the pathlength in m. The quadratic function for the pathlength amplification correction (Cleveland and Weidemann, 1993) included in the equation was derived for a small set of samples collected in 2015 with the same instrument. The effects of the absorption by pigments were removed through reaction with a sodium hypochlorite solution (Ferrari and Tassan, 1999) and $a_{NAP}(\lambda)$ was measured similarly to $a_p(\lambda)$ after a 20-minute reaction time for the rinsed samples. The subtraction of $a_{NAP}(\lambda)$ from $a_p(\lambda)$ resulted in $a_\phi(\lambda)$, where the results represent the averaged laboratory triplicates.

## Appendix 2. Description of University of Stirling (UoStirling) *in situ* data

Radiometric measurements were made with a ship-mounted set of three hyperspectral sensors (HyperSAS, Seabird Scientific Inc.) installed just above the water surface. The first sensor pointed at the water surface and collected upwelling radiance $L_u(\lambda)$, which comprises both water-leaving radiance $L_w(\lambda)$ and the reflected sky irradiance $\rho_s L_s(\lambda)$. The sky irradiance $L_s(\lambda)$ was measured with a second sensor, while a third sensor measured downwelling irradiance $E_d(\lambda)$. $R_{rs}(\lambda)$ was then derived through Simis and Olsson, (2013):

$$R_{rs}(\lambda) = L_w(\lambda, 0^+)/E_d(\lambda), \qquad (A2)$$

$$L_w(\lambda, 0^+) = L_t(\lambda) - \rho_s L_s(\lambda). \qquad (A2.1)$$

The radiometric measurements were taken under the consideration of the two largest challenges to reduce measurement errors: maintenance of optimal viewing geometry and an accurate value determination of $\rho_s$ under changing illumination conditions (Aas, 2010). Measurements were kept in this dataset if they were taken under cloud free conditions and on calm waters to ensure a correct alignment of the ship at a viewing azimuth angle ($\varphi_v$) > 90° (ideally at 135°) (Hooker and Morel, 2003a).

Water constituents were measured with the same method and protocols as described in Appendix 1 for the U.S. Wisconsin DNR *in situ* data. Chla concentration was derived via spectrophotometric determination (ISO 12060:1992). $a_p(\lambda)$ was measured at Plymouth Marine Laboratory (PML, England). Before 2014 $a_p(\lambda)$ was derived through the method by (Tassan and Ferrari, 1995) and after 2013 via (Röttgers and Gehnke, 2012; Stramski et al., 2015). CDOM absorption for these samples was calculated through (Stedmon et al., 2000). Water transparency ($Z_{SD}$) was measured through a standard Secchi Disk lowered into the water from nadir until invisible to the viewer's eye.

## Appendix 3. Uncertainty models

In this paper, we restricted the modelling of the RVDs to non-linear RF models. Here we compare the RF to a linear model, the so-called Least Absolute Shrinkage and Selection Operator (LASSO). LASSO shrinks parameter coefficients to zero for variables that did not contribute significantly to an uncertainty model (Tibshirani, 1996). LASSO is an extension to the standard linear regression model:

$$y_i = \beta_0 + \sum_{j=1}^{n} \beta_i c_{ij}, \qquad (A3)$$

where $y_i$ is the target variable (here a residual value), $c_{ij}$ an explanatory variable and beta ($\beta_i$) the learned coefficient (or weight) of the i-th observation for the j-th of all explanatory variables n. $\beta_0$ is the intercept.

The coefficients $\beta_0, \beta_1, \ldots, \beta_n$ are estimated through minimising the residual sum of squares:

$$\hat{\beta} = \underset{\beta}{arg\,min} \sum_{i=1}^{m} \left( y_i - \beta_0 - \sum_{j=1}^{n} \beta_j c_{ij} \right)^2 . \qquad (A3.1)$$

To estimate the beta-terms for LASSO, a penalty term $\alpha \sum_{j=1}^{n} |\beta_j|$, also called L1-norm or L1 penalty, is added to Eq. A3.1:

$$\hat{\beta}_{LASSO} = \underset{\beta}{arg\,min} \sum_{i=1}^{m} \left( y_i - \beta_0 - \sum_{j=1}^{n} \beta_j c_{ij} \right)^2 + \alpha \sum_{j=1}^{n} |\beta_j| . \qquad (A3.2)$$

The L1 penalty has the effect of forcing some of the coefficient estimates to be exactly zero when $\alpha$ is sufficiently large. The larger the penalisation, the fewer explanatory variables are present in the model because their coefficients are zero. The L1 penalty performs variable selection, as relevant variables to the model receive non-zero coefficients. In LASSO, $\alpha$ is a model hyper-parameter which we optimised with the same procedure as for the RFs.

LASSO is restricted to model the residual values linearly and would thus require explicit formulations of non-linearities and interaction terms. Since we did not have knowledge about these, we opted for use of RFs.

In linear models, such as LASSO, the $\beta$ terms are sensitive to value magnitude and range, which differ between variables (e.g., TSM in g $m^{-3}$ and $a_{CDOM}(443)$ in $m^{-1}$). The variables therefore need to be normalised for LASSO to fall in the range between 0 - 1:

$$c_{norm} = \frac{c - min(c)}{max(c) - min(c)}. \quad (A3.3)$$

Conversely, RF models are independent of unit scales and do not require prior normalisation.

Table 2.6 shows that differences between RF and LASSO started to become significant with increasing chla retrieval uncertainty. A likely reason for this performance difference was the ability of the RF to estimate outliers of the RVDs better through its underlying non-linear model.

***Table 2.6:*** *LASSO and RF uncertainty quantification accuracies (linear scale) for each chla retrieval algorithm included in the uncertainty analysis. RF outperformed LASSO across all metrics for all algorithms.*

| **Chla algorithm** | **MdAD [mg $m^{-3}$]** | | **MAD [mg $m^{-3}$]** | | **RMSE [mg $m^{-3}$]** | |
|---|---|---|---|---|---|---|
| | **LASSO** | **RF** | **LASSO** | **RF** | **LASSO** | **RF** |
| **OC4 -opt** | 0.78 | 0.50 | 1.15 | 0.79 | 1.84 | 1.08 |
| **OC6 -opt** | 0.73 | 0.58 | 1.08 | 0.84 | 1.72 | 1.13 |
| **G11 -opt** | 0.85 | 0.68 | 1.34 | 1.05 | 2.16 | 1.51 |
| **MDN HICO** | 0.83 | 0.64 | 1.25 | 0.96 | 1.79 | 1.26 |
| **QAAv6 (443)** | 1.09 | 0.62 | 1.68 | 0.96 | 2.61 | 1.39 |
| **GSM HICO** | 2.28 | 1.56 | 4.21 | 3.08 | 7.03 | 6.02 |
| **3SAA** | 1.50 | 0.92 | 2.42 | 1.56 | 3.77 | 2.42 |
| **Gons05** | 2.62 | 2.17 | 3.61 | 3.47 | 5.01 | 4.84 |

An exhaustive dataset with many explanatory variables is highly labour and cost intensive. Hence, we required an uncertainty model that can be constructed with a limited set of explanatory variables (here n = 89). Unlike popular neural networks, LASSO and RFs do not require high amounts of measurements. Other methods, such as support vector regression machines (Boser et al., 1992) or more advanced gradient boosting decision trees (Chen and Guestrin, 2016) may provide similar performance such as the RFs used in this study and can be considered as model alternatives.

## Appendix 4. SHAP and standard variable importance

ML libraries such as Scikit-learn (Pedregosa et al., 2011) provide measures to estimate variable impact on a model (thereby defining its importance). To enable the explanation of the quantified uncertainty by the RFs, the measure to infer variable impact must be consistently defined and available for each individual observation. During the construction process of a RF, each node in a sub-decision tree represents a split using some selection of the input variables. For regression RFs specifically, the aim is to gain information about the variable impact on the RF that resulted in a reduction of variance (or a decrease in error). However, the impact of a variable cannot be

consistently derived for ML models using standard variable importance measures(Altmann et al., 2010; Grömping, 2009; Strobl et al., 2009). In practice, the implementations of the variance reduction criterion vary between the libraries (Zhou and Hooker, 2021). As an example, the popular eXtreme Gradient Boosting (XGBoost) library (Chen and Guestrin, 2016) has five metrics (total gain, gain, total cover, cover, weight) to measure variable importance. The use of these metrics results in different variable rankings when applied to the same dataset. Besides inconsistency drawbacks, standard importance measures share a limitation: the contribution of a variable to the model for a single observation is not discernible from the overall variable impact. For these reasons standard importance measures are insufficient to model the RVD of a chla algorithm. We therefore used SHAP that ensures a consistent measure of variable impact (Lundberg and Lee, 2017) and enables to obtain the impact of a variable for an individual observations (Du et al., 2020; Lundberg et al., 2020) recent review on the practices of variable importance measures can be found in (Belle and Papantonis, 2020).

Although SHAP has advantages over the usual measures of variable importance, it also has disadvantages, such as the computational effort required. To calculate a Shapley value for a variable, the impact of each variable (called the marginal contribution to the model) is calculated relative to all other variables in the model, resulting in an exponential increase in calculations. Therefore, in practise, exact Shapley values are not calculated, but are approximated by algorithms such as TreeSHAP (Lundberg et al., 2020).

# Chapter 3: A Bayesian approach for remote sensing of chlorophyll-*a* and associated retrieval uncertainty in oligotrophic and mesotrophic lakes

*This chapter was published as: Werther, M., Odermatt, D., Simis S.G.H., Gurlin, D., Lehmann, M., K., Kutser, T., Gupana, R., Varley, A., Hunter, P.D., Tyler, A.N., Spyrakos, E., 2022: A Bayesian approach for remote sensing of chlorophyll-a and associated retrieval uncertainty in oligotrophic and mesotrophic lakes. Remote Sensing of Environment 283, 113295. This chapter's version differs from the journal paper.*


## Abstract

Satellite remote sensing of chlorophyll-a concentration (chla) in oligotrophic and mesotrophic lakes faces uncertainties from sources such as atmospheric correction, complex inherent optical property compositions, and imperfect algorithmic retrieval. To improve chla estimation in oligo- and mesotrophic lakes and reservoirs, we developed Bayesian probabilistic neural networks (BNNs) for the Sentinel-3 Ocean and Land Cover Instrument (OLCI) and Sentinel-2 MultiSpectral Imager (MSI). The BNNs were designed with an *in situ* dataset of oligo- and mesotrophic water bodies (1755 observations from 178 systems; median chla: 5.11 mg $m^{-3}$, standard deviation: 10.76 mg $m^{-3}$) and provide a per-pixel uncertainty percentage associated with retrieved chla. The BNNs largely outperformed existing chla algorithms when applied to various oligo- and mesotrophic lakes. Over the entire chla range considered in this study (0 – 68 mg $m^{-3}$), accuracy gains compared to reference chla algorithms exceeded 15% by the OLCI and MSI BNNs. Accuracy gains > 25% were found in oligotrophic lakes (0 – 8 mg $m^{-3}$). To measure the quality of the BNN uncertainty estimate, we calculated the percentage interval coverage probability (PICP), Sharpness and mean absolute calibration difference (MACD) metrics. The associated BNN chla uncertainty estimate included the reference in situ chla values for most observations (PICP 80%) across different performance assessments. Further analysis showed that the BNN chla uncertainty estimate was not constantly well-calibrated across different evaluation strategies (Sharpness 1.7 – 6, MACD 0.04 – 0.25). BNN uncertainties were used to exercise two strategies: 1) identification and filtering of uncertain estimates through scene-specific thresholds and 2) selection of the most accurate prior atmospheric correction algorithm per individual satellite observation to retain chla associated with the lowest BNN uncertainty. Both strategies improved chla product quality and demonstrate the importance of uncertainty estimation. This manuscript serves as a study on Bayesian machine learning for the estimation and visualisation of chla and associated uncertainty to generate harmonised products across OLCI and MSI for small and large oligo- and mesotrophic lakes.

## 3.1 Introduction

The phytoplankton pigment chlorophyll-a concentration (chla) is widely used as a proxy of phytoplankton biomass and net primary production (Carlson, 1977; Y Huot et al., 2007; Poikane et al., 2010; Vörös and Padisák, 1991). Chla is thus an important indicator for the ecological integrity of aquatic ecosystems (Boyer et al., 2009). Optical sensors such as the MultiSpectral Imager (MSI) and Ocean and Land Cover Instrument (OLCI) aboard the Sentinel-2 and 3 satellites, respectively, are used to estimate chla and other optically active constituents (OACs) in lakes (Odermatt et al., 2018; Pahlevan et al., 2020; Toming et al., 2016). The majority of inland water quality studies in the last decades have focussed on eutrophic water bodies (Clark et al., 2017; Coffer et al., 2021; Dekker and Peters, 1993; Gons, 1999; Simis et al., 2005; Urquhart et al., 2017). Compared to eutrophic lakes, oligo- and mesotrophic water bodies are under-sampled (Filazzola et al., 2020), which impedes the development and validation of chla estimation methods for oligo- and mesotrophic lakes.

The remote-sensing reflectance $R_{rs}(\lambda)$ contains information about water column properties and components (Mobley, 1999; O'Reilly et al., 1998). Several approaches exist that related optical features in $R_{rs}(\lambda)$ to chla. The contribution from phytoplankton backscattering to the reflectance spectrum forms distinct reflectance peaks near 560 and 700 nm (Suits, 1975). A phytoplankton pigment feature at 675 nm can be offset against the near-infrared (NIR) scattering peak around 700 nm (Gitelson, 1992; Mittenzwey et al., 1992). A variety of two, three and four-band combination algorithms are based on the 700 nm and 665 nm band ratio to capture the red absorption signature of chla (Dall'Olmo et al., 2003; Gilerson et al., 2010; Gons, 1999; Gurlin et al., 2011; Mishra and Mishra, 2012; Moses et al., 2009; Pepe et al., 2001). Chla absorption and fluorescence peaks in the 665 - 685 nm region are also at the basis of line height algorithms such as fluorescence line height (FLH) (Gower, 1980; Gower et al., 2004; Gupana et al., 2021), maximum peak height (MPH) (Matthews et al., 2012; Matthews and Odermatt, 2015) and maximum chlorophyll index (MCI) (Binding et al., 2013; Gower et al., 2005).

Chla estimation in oligo- and mesotrophic systems is still associated with larger uncertainties when compared to eutrophic waters (Liu et al., 2021; Mouw et al., 2013; Neil et al., 2019). The causes are manifold. Absorption by coloured dissolved organic matter ($a_{CDOM}(\lambda)$) and non-pigmented particulate absorption ($a_{NAP}(\lambda)$) may be larger than by phytoplankton pigments ($a_{\phi}(\lambda)$), particularly at short wavelengths where band arithmetic algorithm specificity then breaks down (Härmä et al., 2001; Kutser et al., 2016; Mouw et al., 2013; Neil et al., 2019; Schalles, 2006). For analytical algorithms that invert a reflectance spectrum into inherent optical properties (IOPs) (Werdell et al., 2018), $a_{\phi}(\lambda)$ must be precisely estimated to relate it to chla. Effects such as pigment packaging impact the linearity of the relationship between chla and $a_{\phi}(\lambda)$(Grzymski et al., 1997; Kirk, 1994; McKee et al., 2014). The scaling of $a_{\phi}(\lambda)$ to chla depends on phytoplankton type and particle size (Johnsen et al., 1994; Lutz, 2001; Simis et al., 2007). Therefore, for analytical algorithm

application to estimate chla, prior information about $a_\phi(\lambda)$ variability is required. Without system-specific calibration, the underlying bio-optical models use default pigment mass-specific phytoplankton absorption ($a_\phi^*(\lambda)$) values. $a_\phi^*(\lambda)$ natural variability spans an order of 4 magnitudes in the ocean (Bricaud et al., 2004b, 1998b, 1995b), and its variability is unknown for global lakes, yet subject to measurement uncertainty (McKee et al., 2014). For algorithms that focus on the red/NIR spectral area, the phytoplankton absorption peak at 675 nm in lakes may be undetectable in conditions where CDOM and NAP absorption is high (Kutser et al., 2016). Moreover, non-photochemical quenching across large euphotic depths affects the water-leaving fluorescence emission peak near 681 nm (Gupana et al., pers. comm.). For oligotrophic lakes, the red-NIR area of the spectrum has low signal to noise ratios, causing chla-related peaks to be below the detection limit. In recent years, machine learning (ML) methods have been developed to estimate chla in inland water bodies. Neural networks (NNs) developed for lakes overcome some of the issues encountered with band ratio methods and analytical approaches when assessed over large chla ranges (Pahlevan et al., 2022, 2020; Schaeffer et al., 2022; Smith et al., 2021). For a ML approach to handle unseen observations, large training datasets of *in situ* and/or simulated measurements are necessary. ML approaches are prone to produce estimation errors when a so-called dataset shift occurs, whereby the IOPs and OACs of the provided, unknown input substantially differ from the measurements used during ML model training (Moreno-Torres et al., 2012; Ovadia et al., 2019). Chla estimation uncertainty introduced through a dataset shift or other sources such as prior AC and adjacent land influence on the radiance distribution are common in lake remote sensing. However, current ML methods do not inform about the uncertainty associated with estimated chla. Through provision of model confidence highly uncertain estimates could be identified.

Bayesian probability theory offers a mathematical tool to reason about uncertainty (Ghahramani, 2015b). Bayesian probabilistic reasoning applied to NNs usually comes at a prohibitive computational cost. NNs with distributions placed over the weights have long been studied as Bayesian Neural Networks (BNNs) (MacKay, 1992; Neal, 1995) commonly through variational inference (VI) (Jordan and Saul, 1999; Saul et al., 1996), but with limited success (Gal and Ghaharamni, 2016; Graves, 2011). Approaches based on VI bring higher model complexity, which for the field of aquatic remote sensing is prohibitive: to represent uncertainty, the number of parameters in BNNs based on VI is doubled when compared to the same non-Bayesian NN size. To train a network with millions of parameters large input datasets are required. However, for inland aquatic remote sensing large datasets are sparse and only just mature through community-wide efforts to collate datasets ready for use with recent ML approaches (Cao et al., 2020; Pahlevan et al., 2022; Werther et al., 2021).

Here, we investigate a potential solution in BNNs developed through Monte Carlo Dropout (Gal and Ghaharamni, 2016) to obtain chla and a well-calibrated uncertainty estimate from a limited set of input training samples. Monte Carlo Dropout refers to the process of randomly turning off all outgoing connections from each neuron in a NN with a previously determined probability $p$

(Srivastava et al., 2014). Dropout can be activated during the application to unknown observations, which results in an ensemble of NNs to estimate a final chla value: for a single input spectrum, a number ($S$) of chla estimates are made, whereby each chla estimate results from another dropout NN variant.

Based on the largest available *in situ* dataset covering typical oligo-, meso- and eutrophic lake optical properties, we aim to: 1) improve the chla estimation accuracy via OLCI and MSI sensors over both small and large oligo- and mesotrophic lakes through Bayesian probabilistic machine learning, 2) provide a well-calibrated, per-pixel uncertainty percentage associated with estimated chla, and 3) demonstrate how the provided uncertainties can be used to deal with dataset shifts and AC errors commonly affecting satellite remote sensing of chla.

## 3.2 Datasets

### 3.2.1 Development dataset

A dataset of 1755 *in situ* observations from 178 lakes and reservoirs was compiled from six sources to develop and evaluate the BNNs: (i) LIMNADES (Lake Bio-optical Measurements and Matchup Data for Remote Sensing (https://limnades.stir.ac.uk/)), (ii) Wisconsin DNR (U.S.), (iii) Earth and Planetary Observation Sciences (EPOS) (Scotland, UK), (iv) the University of Tartu (Estonia), (v) the University of Waikato (New Zealand (NZ)) and the (vi) LéXPLORE Platform (Lake Geneva, Switzerland). Only samples containing above-water $R_{rs}(\lambda)$ and chla measurements were considered for this study. Where available, total suspended matter (TSM) dry weight and the absorption by

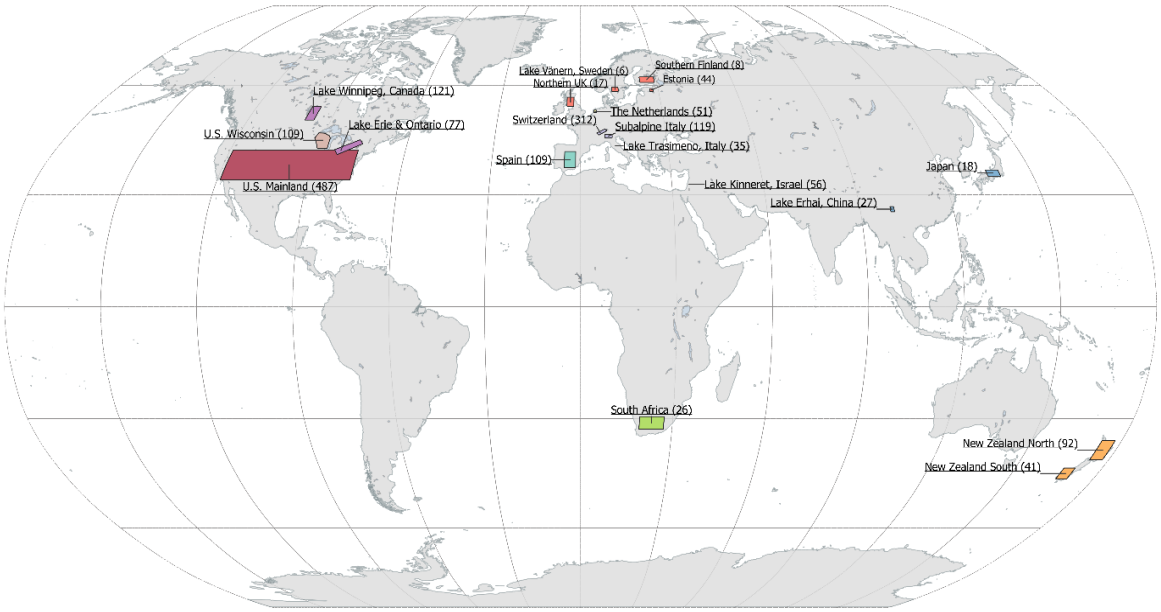


***Figure 3.1:*** *Spatial distribution of the 10 in situ measurement regions constituting the dataset of this study. Grouped regions share the same colour. Number of samples taken in the indicated polygon is shown in brackets. See Appendix 1 for a detailed description of the dataset regions.*

CDOM at 443 nm ($a_{CDOM}(443)$) were included in the dataset. For algorithm development and validation we separated the measurements from these datasets into 10 regions (Figure 2.1) to evaluate the BNNs on independent geographical regions, as previously demonstrated for ML algorithm development (O'Shea et al., 2021; Pahlevan et al., 2022). Details about the regions and included water bodies are provided in Appendix 1. Datasets in LIMNADES were described in (Spyrakos et al., 2018) and references therein. Measurements from Wisconsin DNR were taken across the U.S. State of Wisconsin between spring 2014 and autumn 2016. The Wisconsin DNR dataset was used in recent algorithm development and inter-comparison studies (Pahlevan et al., 2021b, 2020). Collection methods and details are provided in Werther et al. (in revision). Measurements provided by the University of Tartu were previously detailed in (Kutser et al., 2013; Soomets et al., 2020). The measurements from lakes in NZ refer collectively to two datasets. The first one was collected by the University of Waikato between 2015 and 2019 and has been previously published (Pahlevan et al., 2022, 2020). The second dataset refers to data collected jointly by the University of Waikato and University of Stirling in 2020 (referred to as NZ 2020). The NZ 2020 dataset is unpublished and described in Appendix 2. Surface chla obtained through calibration of in vivo fluorescence measurements and $R_{rs}(\lambda)$ from the LéXPLORE platform (https://lexplore.info/) in Lake Geneva were measured between 2018/10 and 2021/09. Measurement details are provided in Appendix 3 and in Irani Rahaghi et al. (pers. comm.).

Optical water types (OWTs) were used to define the development and application range of the BNNs (Moore et al., 2014, 2001; Spyrakos et al., 2018). OWTs membership was derived through estimation of the spectral angle (Kruse et al., 1993; Liu et al., 2021). For all $R_{rs}(\lambda)$ an OWT membership score was calculated and the OWT with the highest similarity score assigned. Observations assigned to OWT 5 were small (n = 40) in this dataset. The corresponding systems were high in CDOM absorption and low in chla, known to be challenging for retrieval algorithms (Kallio et al., 2015; Kutser et al., 2016; Toming et al., 2016).

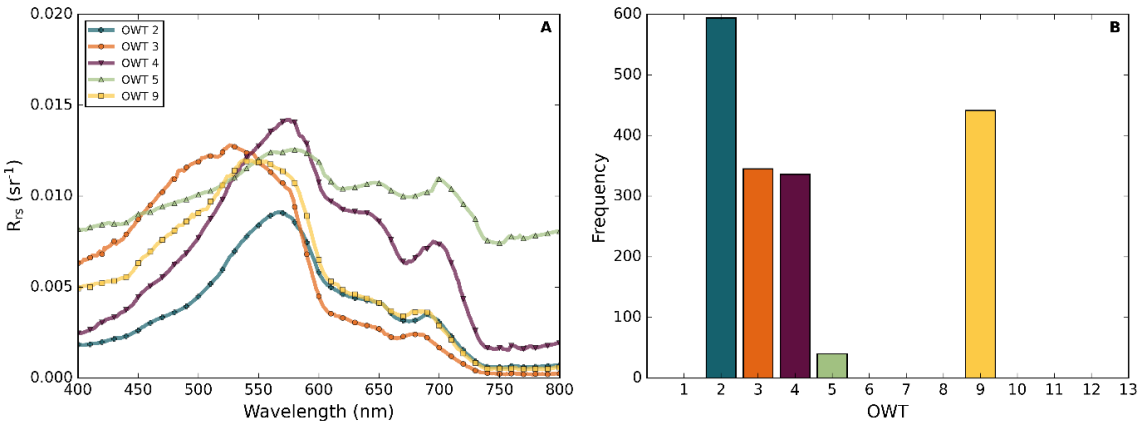


***Figure 3.2:*** *OWTs 2, 3, 4, 5 and 9 used for the development and application of the BNNs. (A) Spectral medians. (B) Frequency per OWT.*

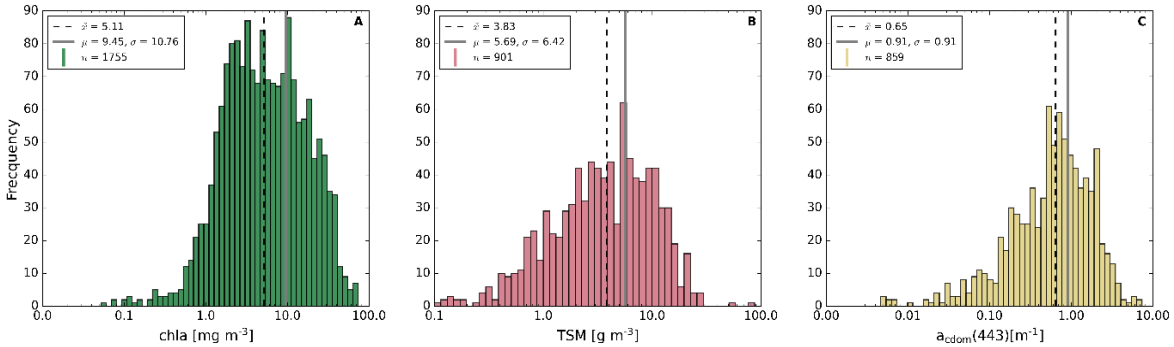


***Figure 3.3:*** *Log-distributions of (A) chla, (B) TSM and (C) $a_{CDOM}(443)$ for the entire dataset. Denoted are median ($\hat{x}$), mean ($\mu$) and standard deviation ($\sigma$) of the respective parameter.*

The chla range of the dataset represents typically encountered oligo-, meso- and eutrophic conditions (median 5.11 mg $m^{-3}$, mean 9.45 mg $m^{-3}$, standard deviation: 10.76 mg $m^{-3}$, maximum 68 mg $m^{-3}$) (Figure 3.3). 167 observations (9.5%) in the dataset exhibit chla > 25 mg $m^{-3}$ with a maximum of 68 mg $m^{-3}$ to account for seasonal shifts in phytoplankton abundance and community composition that may occasionally exceed mesotrophic trophic status (TS) in lakes. The BNN development and assessments in this study are thus focussed on oligo- and mesotrophic lakes, as these water types are under-represented in lake remote sensing of chla.

### 3.2.2 Satellite data processing

Match-ups were generated for the period 2018/10 to 2021/09 between coinciding *in situ* measurements taken on the LéXPLORE platform and satellite observations from Sentinel-3A/B OLCI 300m resolution and Sentinel-2A/B MSI 20m resolution. The closest OLCI and MSI sensor overpass within +- 2 hours of the *in situ* measurement was selected. For OLCI and MSI (in their respective resolution) the closest valid pixel south of the platform surrounded by at least 3x3 valid pixels was used as the match-up location. Valid pixels were identified through the IdePix (Identification of Pixel features) algorithm (Skakun et al., 2022; Wevers et al., 2021). Data flagged as invalid, cloud (including ambiguous, sure, and a 2-pixel buffer) or cloud shadow, snow/ice, bright, coastline land, white and glint risk were excluded.

For AC of OLCI and MSI data, the POLYMER (Steinmetz et al., 2011) and Case 2 Region CoastColour (C2RCC) (Brockmann et al., 2016) algorithms were selected. POLYMER and C2RCC AC performances were previously compared for OLCI and MSI (Liu et al., 2021; Pahlevan et al., 2021; Warren et al., 2021, 2019) and a detailed validation of POLYMER for Lake Geneva is presented in Irani Rahaghi et al. (pers. comm.). Both ACs are widely used in combination with algorithms for the retrieval of chla in lakes (Kravitz et al., 2020; Liu et al., 2021; O'Shea et al., 2021; Pahlevan et al., 2020, 2022; Pereira-Sandoval et al., 2019; Smith et al., 2021).

## 3.3 Methodology

### 3.3.1 Bayesian Neural Networks

At the heart of all BNNs is the Bayesian theorem:

$$p(\theta|D) = \frac{p(D|\theta)p(\theta)}{p(D)}, \tag{3.1}$$

where $p(\theta|D)$ is the posterior, i.e., the probability of a certain value of a neural network parameter $\theta$ given the data $D$, $p(D|\theta)$ the inverse, $p(\theta)$ the prior and $p(D)$ the quantity (also known as the marginal likelihood) (Bolstad and Curran, 2016). In Bayesian statistics, $\theta$ is not defined by one value but has an uncertainty described by a probability distribution $p(\theta)$. This distribution, $p(\theta)$, defines the probability value of each parameter value $\theta$. For a Bayesian model to obtain a predictive chla distribution, the weight distribution of the neural network is averaged:

$$p(y|x_{test}, D) = \sum_i (p(y|x_{test}, \theta_i) \cdot p(\theta_i|D), \tag{3.2}$$

where $x_{test}$ is an input test $R_{rs}(\lambda)$ observation and $\theta_i$ denote the weights $w_i$ of the NN.
Replacement of NN weights with distributions is computationally costly and requires large amounts of training data, which is scarce in lake remote sensing. In this study we therefore used Monte Carlo (MC) dropout applied to NN layers (Gal and Ghaharamni, 2016). MC dropout replaces each fixed $\theta_i$ of a deterministic NN with a binary distribution, whereby either zero or the value $\theta_c$ for a NN connection are obtained. Using MC dropout, our NNs predicted S-times chla for a single reflectance input spectrum. S was set to 50 in this study. Each of the S predictions originated from a different NN variant that corresponded to a sampled network constellation. Combining the S dropout estimates resulted in a chla probability distribution:

$$p(y|x_{test}, D) = \frac{1}{S}\sum_{i=1}^{S} p(y|x_{test}, \theta_i)\,, \tag{3.3}$$

which is an empirical approximation to Eq. 3.3 that captured both the NN model (epistemic) and the intrinsic aleatoric uncertainty (Abdar et al., 2021). We then sampled from all determined Gaussian distributions: $y \sim N(\mu_{x,\theta}, \sigma_{x,\theta})$, where $\mu_\theta$ is the chla mean and $\sigma_\theta$ the standard deviation. Because $\sigma$ was a distribution, for each chla estimate $y$ we calculated the 95% confidence interval (CI) and then estimated the width of the CI ($CI_w$) to get an absolute number. $CI_w$ was divided by two to obtain uncertainty $\pm$ of $y$ and the entire term was multiplied by 100 to obtain a BNN percentage uncertainty for estimated chla:

$$\varepsilon(\%) = \left(\frac{\frac{CI_w}{2}}{y}\right) \cdot 100. \tag{3.4}$$

### 3.3.2 BNN processing

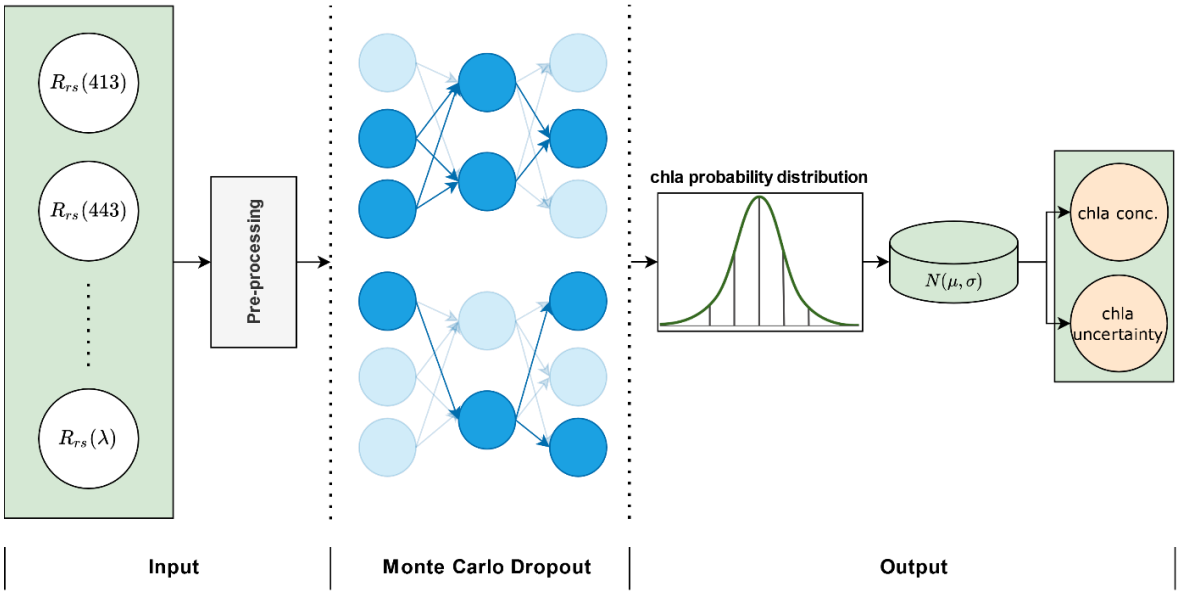


***Figure 3.4:*** *Processing scheme of the BNNs based on Monte Carlo dropout. For MSI the first band is 443 nm. Pre-processing includes the spectral convolution of the training data, normalisation of both training and unknown (test) observations and treatment of negative* $R_{rs}(\lambda)$ *values.*

*In situ* $R_{rs}(\lambda)$ were convolved to match the relative spectral response (RSR) of the OLCI instruments on Sentinel-3A and 3B and the MSI instruments on Sentinel-2A and B, and were used as inputs to the BNNs. Distinct BNNs were developed for OLCI S3A/B and MSI S2A/B. For the OLCI BNNs we selected the 12 bands from 413 through 778 nm (omitting the oxygen bands at 761, 764 and 767 nm).

The waveband centred at 400 nm was excluded because the majority of *in situ* $R_{rs}(\lambda)$, did not extend to < 400 nm and therefore could not be convolved to match the OLCI band. For MSI, all 7 bands from 443 to 783 nm were selected. For clear oligotrophic waters, negative $R_{rs}(\lambda)$ in one or more red-NIR wavebands are likely to occur when using POLYMER, but not when using C2RCC. Negative values can cause unpredictable behaviour of the BNN which may lead to high chla uncertainty. All negative values $R_{rs}(\lambda)$ > 665 nm in the *in situ* dataset were set to zero for BNN training and during pre-processing of test observations. $R_{rs}(\lambda)$ input features were then normalised to fall between the 0 – 1 range. Each feature was treated individually such that it was in the given range of the dataset between 0 and 1. Normalisation reduced the required training time and smoothed the process of minimising the loss function of the network, the negative log-likelihood (NLL) in this study. NLL is a standard measure of the quality of estimate of a probabilistic model (Hastie et al., 2001).

Since the BNNs are based on MC dropout, a stochastic process is involved. To obtain a representative BNN for the different evaluation strategies (see section 3.3 below), 10 BNNs were trained for each sensor and applied to the respective test observations once. The median of the 10 different networks

was used as the final estimate we report, compared to recent practice of NNs with stochastic elements (O'Shea et al., 2021; Pahlevan et al., 2022, 2020). For the evaluation on *in situ* data, the S3A and S2A BNNs versions were used. For match-ups with MSI 2A/B and OLCI 3A/B the BNN version corresponding to the respective instrument was selected.

### 3.3.3 BNN performance assessment

Five different assessments were conducted to evaluate the quality of the chla and uncertainty BNN products under varying conditions: (i) a randomised 50/50 training/test data split, (ii) a leave-one-out exercise (LOO) using *in situ* observations, (iii) BNN OLCI and MSI LéXPLORE match-ups, (iv) a LéXPLORE 3 ½ year time-series (including the match-ups) and lastly (v) single-day product visualisations over both large and small lakes in Europe, New Zealand, Africa and Canada. The BNNs were compared to state-of-the-art chla algorithms. Details of these assessments are provided in the following sections.

#### 3.3.3.1 50/50 training/test data split

For an assessment of overall BNN performance on *in situ* data, the development dataset (n = 1755) was split into 50% training (n = 878) and 50% test (n = 877) sets, following recent ML assessment practices (O'Shea et al., 2021; Pahlevan et al., 2020). The observations in the two sets were randomly drawn from all regions of the entire dataset. The 50% training data were further split into a training (60%; 526 observations) and validation set (40%; 352 observations). An initial OLCI and MSI BNN was constructed using this training subset (n = 526), and the performance of these BNN versions was evaluated on the validation set (n = 352) to find an optimal architecture and to tune the hyper-parameters of the BNNs through Bayesian optimisation implemented in the Python package Weights & Biases (Biewald, 2020; Werther et al., 2021a). BNNs were continually improved by tracking the loss on the validation data. Once the validation loss plateaued, the BNNs were re-trained with the entire training set (878 / 1755 observations) and then applied to the test set. The OLCI and MSI BNN performances on the test set across OWTs and TS classes using the definition by OECD (1982) were recorded.

A randomised 50/50 split of training and test data has two major limitations. First, training and test datasets share the same distribution of OACs (assured through prior randomisation of the overall dataset), which may not occur in a situation outside of model development. Performance under dataset shifts can therefore not be analysed. Second, the performance of the BNN OLCI and MSI versions in this assessment does not represent their final models, as these versions were only trained with 50% of the training data. The final BNN versions made publicly available with this article were re-trained with the entire dataset (n = 1755). The final hyper-parameter configuration of the OLCI and MSI BNNs is listed in Appendix 4.

#### 3.3.3.2 Leave-one-out

We simulated dataset shifts with all available *in situ* measurements through a leave-one-out (LOO) strategy (O'Shea et al., 2021; Pahlevan et al., 2022). In LOO, a BNN was trained with measurements from all regions except for one region. The BNN was then applied to the left-out region and the entire process was repeated until each region was left-out. Through the LOO strategy we assess the generalisation performance of the BNN for independent systems (represented by a region) that were not part of the training process, and which may thus not share the same OAC distribution. We note that the performance of the BNNs for different regions may underestimate overall model performance. Whilst different *in situ* measurement techniques between the datasets were used, the measurement consistency between regions is unknown. Individual regions may thus carry varying degrees of measurement uncertainty.

#### 3.3.3.3 Lake Geneva match-ups and time series assessments

To evaluate the BNNs with independent satellite match-ups, a training set excluding observations from Lake Geneva was created. OLCI and MSI BNNs were trained with the exact same set of bands and architecture as in the 50/50 training/test split and LOO assessments. The match-up evaluation on Lake Geneva facilitated to assess how well the BNNs trained with a semi-global *in situ* dataset of lake properties transfer to an individual system measured through satellite sensors, thereby representing a common use-case of the newly developed algorithm.

#### 3.3.3.4 Comparison with other chla algorithms

Using the 50/50 split, LOO and match-up performance assessments we evaluated the accuracy of the OLCI and MSI BNNs versus five published chla algorithms representing a variety of techniques (hereafter reference algorithms). These included the OLCI and MSI Mixture Density Networks (MDNs)((Pahlevan et al., 2020), the red/NIR Gons05 (Gons et al., 2005), the red/NIR band ratio G11 (Gurlin et al., 2011), the blue/green band ratio OC3 (O'Reilly et al., 1998; O'Reilly and Werdell, 2019), and the Blend algorithm (Smith et al., 2018) which switches between OCI (Hu et al., 2012) and the red-NIR approach by Gilerson et al. (2010) and blends their estimates using empirically derived thresholds.

To reduce bias resulting from calibration of the published algorithms on different datasets, the G11, OC3 and Blend model coefficients were optimised (denoted as -opt) through non-linear least squares fitting against the same training datasets used in the evaluation strategies (identical to the observations for training the BNNs). For Lake Geneva match-ups, all measurements were provided for optimisation, except for the observations measured in Lake Geneva. Algorithms using OLCI wavebands were included in their originally published configuration (denoted as -org) and with optimised coefficients (denoted as -opt). For Blend in the OLCI configuration only the OC4 algorithm was optimised, and for Blend in the MSI configuration the OC3 algorithm was used instead

of OC4 because the 510 nm band in OC4 is not available on MSI. MSI band positions of Gons05, G11 and OC3 (also used in Blend) were slightly shifted, namely from 490, 708 and 778 nm to 492, 704 and 783 nm, which was successfully demonstrated previously in Warren et al. (2021). Optimised model coefficients are tabulated in Appendix 5. For the OLCI and MSI MDNs specifically, in the LOO assessment regions 5 and 9 were excluded from the evaluation because these were already used in full as part of the training set of said algorithms. Their inclusion would thus not represent an independent comparison per region.

### 3.3.4 Accuracy and uncertainty calibration metrics

Performance metrics used to assess the chla retrieval accuracy included median symmetric accuracy (MdSA, in %) and symmetric signed percentage bias (Bias, in %) (Morley et al., 2018). In addition, the mean absolute difference (MAD, in mg $m^{-3}$) and median absolute percentage difference (MAPD, in %) are provided (Seegers et al., 2018). These metrics were calculated in log space.

BNN uncertainty estimates may not capture the true data distribution (Lakshminarayanan et al., 2017). For example, a 90% confidence interval (CI) may not contain the reference *in situ* chla concentration in 9/10 scenarios. The BNN uncertainty estimate is miscalibrated when the CI does not include the reference *in situ* chla. To assess the quality of the BNN chla uncertainty calibration we calculated three metrics between the BNN model estimated (e) and observed (o) *in situ* chla:

Percentage interval coverage probability (PICP; denoted $\rho$). The percentage of observations for which *in situ* chla lies within the 95% confidence interval of the BNN chla estimate:

$$\rho = \frac{1}{N}\sum_{n=1}^{N} \mathbb{1}_{o_n \le e_n^{high}} \cdot \mathbb{1}_{o_n \ge e_n^{low}}, \tag{3.5}$$

where $e_n^{high}$ is the 97.5% percentile and $e_n^{low}$ is the 2.5% percentile of the BNN estimated chla for an input $x_n$ (Yao et al., 2019). The higher the $\rho$ [%] value, the more *in situ* chla values were covered by the BNN uncertainty estimate.

Sharpness (denoted $\sigma$). The standard deviation (var) of a chla estimate whose cumulative distribution function ($F_N$) should be small:

$$\sigma = \sqrt{\frac{1}{N}\sum_{n=1}^{N} var(F_N)^2}\,. \tag{3.6}$$

Sharpness represents the average of the BNN estimated chla standard deviations (Tran et al., 2020).

Mean absolute calibration difference (MACD; denoted $\tau$). MACD is a statistic to measure calibration relative to an ideal reliability diagram. Reliability diagrams show expected observation accuracy as a function of confidence (Degroot and Fienberg, 1983; Niculescu-Mizil and Caruana, 2005). To compute MACD, chla estimates were sorted and divided into b equally-spaced bins (b = 100 in this study) with an approximately equal number of BNN chla estimates in each bin:

$$\tau = \sum_{b=1}^{B} \frac{n_b}{N} \cdot |acc(b) - conf(b)|, \tag{3.7}$$

where $n_b$ is the number of chla estimates in bin $b$, $N$ is the total number of observations, and $acc(b)$ and $conf(b)$ are the accuracy and confidence in bin $b$ (Guo et al., 2017; Nixon et al., 2019). The smaller the values of Sharpness and MACD, the more accurate the uncertainty calibration.

## 3.4 Results

### 3.4.1 50/50 dataset split

OLCI and MSI BNNs outperformed the reference chla algorithms (MdSA 12 – 21% lower) (Figures 3.5, 3.6). The OLCI BNN was slightly more accurate than its MSI version (2 - 3% difference in MdSA). Lower accuracy in the estimation of chla from MSI compared to OLCI was also observed when the reference chla algorithms were employed (Figure 3.7). For OLCI, optimisation of model coefficients from G11, OC3 and Blend improved the estimates by 8 – 25% (MdSA). Thus, the optimisation process of the coefficients is considered transferable to MSI.

Large over- and underestimates of chla were observed for concentrations between 0 and 1 mg $m^{-3}$ across all the algorithms. Only 65 data points (13.5 % of the total) were available in this concentration range in the test set. Chla estimates from the algorithms used for comparison became markedly more accurate for concentrations greater than 10 mg $m^{-3}$. Varying retrieval accuracies became more apparent when the assessment was tabulated per TS class (Table 3.1). The OLCI BNN was > 25% and the MSI BNN > 35% more accurate for oligotrophic waters (0 – 8 mg $m^{-3}$) than the reference algorithms. Red/NIR algorithms started to become highly accurate in mesotrophic waters (8 – 25 mg

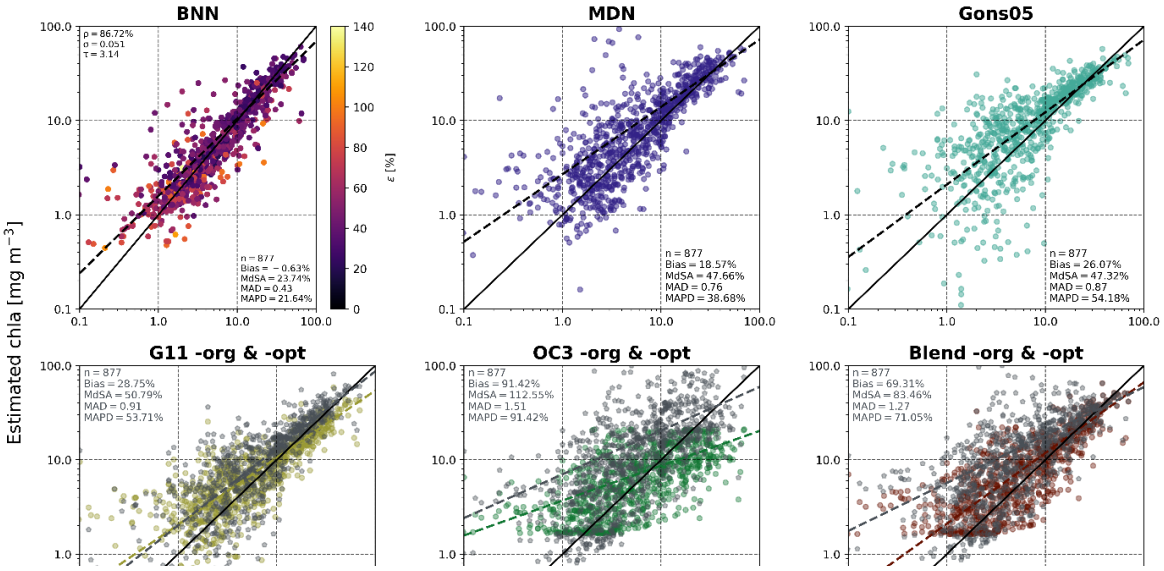


***Figure 3.5:*** *Chla retrieval results by the OLCI BNN and reference algorithms in the 50/50 training/test split assessment. For G11, OC3 and Blend chla retrievals through original coefficients (-org) are shown in grey, and optimised coefficients (-opt) in colour.*

$m^{-3}$) (Table 3.1). For eutrophic waters (> 25 mg $m^{-3}$) the red/NIR OLCI Gons05 algorithm was most accurate (MdSA 13.98%), but its performance did not transfer to MSI. Across OWTs, the BNNs consistently outperformed the reference algorithms (Table 3.2). For OWT 5 it is to note that only 15 observations were assigned to it, thus the evaluation is not representative. The low number in this OWT originated from the randomisation of the entire dataset prior to training and evaluation of the algorithms. Uncertainty calibration by the OLCI and MSI BNNs was accurate (MACD 0.051 and

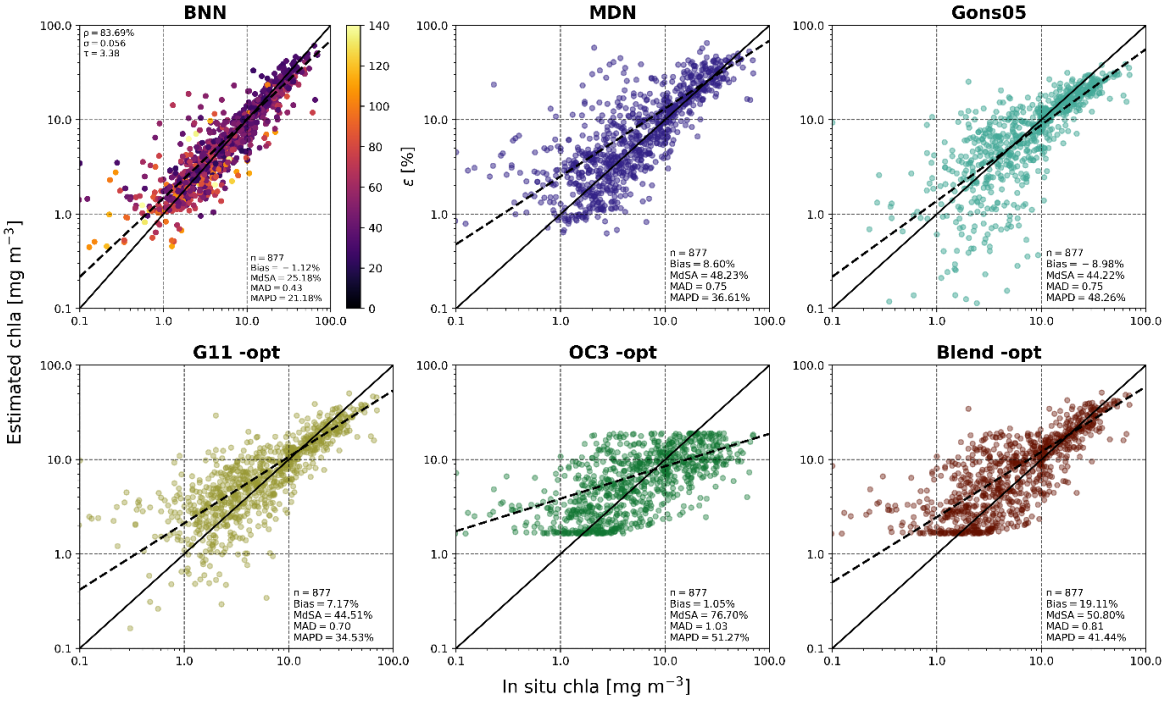


***Figure 3.6:*** *Chla retrieval results by the MSI BNN and reference algorithms in the 50/50 training/test split assessment.*

0.056, respectively) and PICP included > 83% of the *in situ* chla measurements. Sharpness of 3.18 (OLCI) and 3.38 (MSI) across the large concentration range was moderate - low. Highly uncertain estimates were identified by both BNN configurations.

***Table 3.1:*** *Retrieval accuracies the chla algorithms per TS class for the test set, compared through the MdSA (in %) metric. Lowest MdSA achieved by an algorithm for each TS in bold.*

**OLCI**

| **Chla algorithm** | **Oligotrophic** (chla <= 8 [mg/m$^3$]; n = 548) | **Mesotrophic** (chla >= 8 and <= 25 [mg/m$^3$]; n = 253) | **Eutrophic** (chla >= 25 [mg/m$^3$]; n = 76) |
|---|---|---|---|
| BNN | **28.78** | **18.95** | 17.28 |
| MDN | 76.43 | 25.33 | 19.37 |
| Gons05 | 54.60 | 30.00 | **13.98** |
| G11 -opt | 75.09 | 19.26 | 34.85 |
| OC3 -opt | 78.95 | 44.97 | 191.88 |
| Blend -opt | 83.21 | 24.74 | 19.84 |

**MSI**

| | | | |
|---|---|---|---|
| BNN | **32.42** | 19.14 | **19.77** |
| MDN | 71.13 | 26.43 | 20.94 |
| Gons05 | 81.57 | 20.28 | 46.41 |
| G11 -opt | 72.58 | **17.52** | 35.98 |
| OC3 -opt | 79.54 | 47.20 | 207.57 |
| Blend -opt | 81.97 | 27.15 | 21.27 |

***Table 3.2:*** *Retrieval accuracies of the chla algorithms for OLCI (top) and MSI (bottom) for the test set per OWT class, compared through the MdSA (in %) metric. Lowest MdSA achieved by an algorithm for each OWT in bold.*

**OLCI**

| **Chla algorithm** | **OWT 2 (n = 261)** | **OWT 3 (n = 257)** | **OWT 4 (n = 149)** | **OWT 5 (n = 13)** | **OWT 9 (n = 197)** |
|---|---|---|---|---|---|
| BNN | **23.21** | **27.73** | **20.82** | 42.51 | **23.67** |
| MDN | 36.87 | 64.72 | 47.53 | 109.67 | 44.73 |
| Gons05 | 37.48 | 136.90 | 31.90 | 60.51 | 49.07 |
| G11 -opt | 32.71 | 82.80 | 34.47 | 47.48 | 39.66 |
| OC3 -opt | 65.25 | 66.49 | 134.44 | 443.42 | 57.21 |
| Blend -opt | 41.11 | 69.44 | 41.31 | **24.82** | 55.09 |

**MSI**

| | | | | | |
|---|---|---|---|---|---|
| BNN | **24.27** | **23.14** | **18.22** | **19.67** | **28.49** |
| MDN | 30.16 | 67.32 | 51.76 | 83.30 | 41.10 |
| Gons05 | 32.02 | 107.92 | 40.09 | 63.58 | 66.14 |
| G11 -opt | 29.72 | 63.43 | 39.23 | 36.52 | 46.04 |
| OC3 -opt | 72.57 | 59.26 | 135.18 | 433.24 | 57.65 |
| Blend -opt | 44.06 | 102.98 | 36.62 | 37.20 | 51.26 |

### 3.4.2 Leave-one-out: BNN generalisation ability

In the LOO assessment, the OLCI BNN MdSA was expectedly higher for some regions (MdSA between 19 – 59%) than for the 50/50 training/test split assessment. In comparison the OLCI BNN outperformed

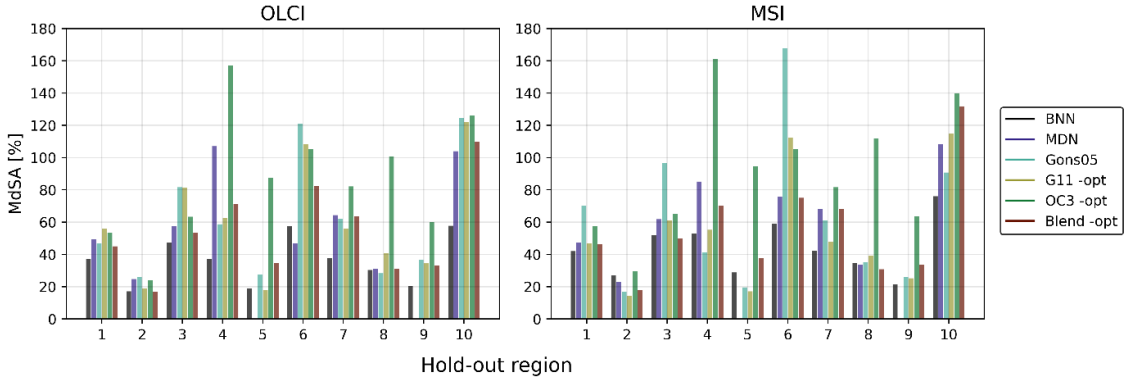


***Figure 3.7:*** *Change in MdSA of the chla algorithms for each region as part of the LOO assessment.*

the reference algorithms in 7/10 regions and remained competitive for the remaining regions. Highest gains in retrieval accuracy by the OLCI BNN were made for region 4 consisting of the UK, Sweden, Finland and Estonia (MdSA 19% lower), the South African reservoirs representing region 7 (MdSA 18% lower) and Lake Erie, Ontario, and Winnipeg of region 10 (MdSA 40% lower). For most regions, MSI MdSA was only slightly higher than for OLCI. However, the largest difference in retrieval accuracy between the MSI and OLCI BNN configurations was linked to regions 4 and 10, whereby the MSI config uration was about 20% less accurate. The results of the independent region evaluation emphasise that measurements included in this dataset are representative for other regions of the globe. Consequently the dataset enabled a robust generalisation of the BNNs to systems that shared the same OWTs, but that were from regions not included in the respective training dataset.

Through LOO, the effectiveness of the uncertainty calibration can be analysed (Figure 3.8). MACD and Sharpness did not co-vary linearly with chla accuracy. For example, OLCI MdSA in regions 2, 5 and 9 was the smallest (see Figure 3.7), but the corresponding uncertainty metrics were not consistently small, too. In region 9, chla accuracy with MdSA < 20% was achieved, but MACD (>

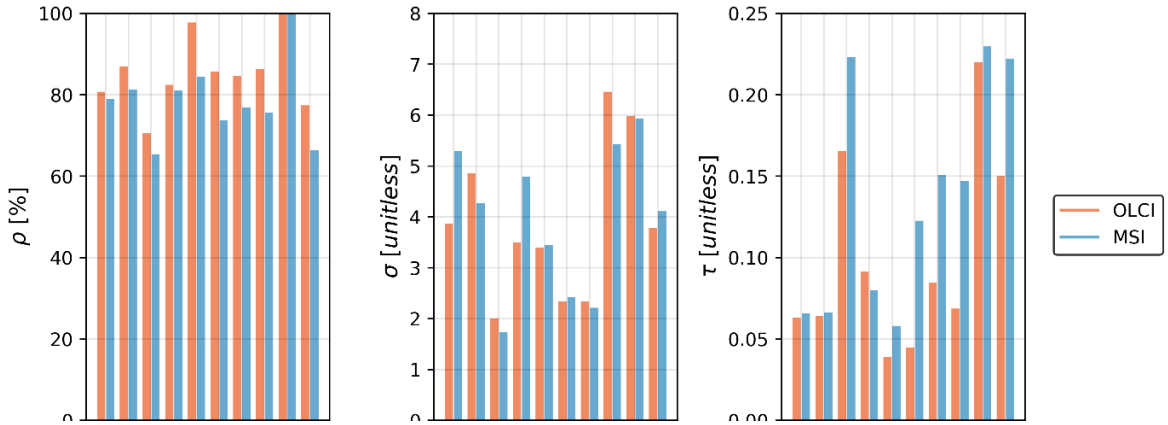


***Figure 3.8:*** *BNN uncertainty calibration metrics PICP ($\rho$), Sharpness ($\sigma$) and MACD ($\tau$) per region as part of the LOO assessment.*

0.20) and Sharpness (> 5) were high in comparison to other regions. Nevertheless, region 9 PICP was 100%, indicating that all BNN chla estimates fell within the estimated confidence intervals covering the reference *in situ* chla value. Overall, OLCI PICP was on average 7% higher than MSI (85.24% and 78.36%, respectively). Although the MSI BNN was only slightly less accurate in the retrieval of chla, the algorithm was less certain about its estimates resulting in a lower PICP percentage and higher MACD.

### 3.4.3 Lake Geneva match-ups

Satellite-derived $R_{rs}(\lambda)$ for *in situ* match-ups over Lake Geneva were generated through POLYMER and C2RCC applied to OLCI and MSI products. POLYMER calculated negative values in several red bands of the spectrum for 21 OLCI and 12 MSI match-ups. To assure a better match-up comparison of the chla algorithm performances, the 21 OLCI and 12 MSI observations obtained through POLYMER were excluded. Further, Gons05 and G11 were excluded because these red/NIR algorithms showed the largest retrieval uncertainties in the 50/50 split and LOO assessments for chla < 10 mg m$^3$. Chla retrievals from OLCI Polymer and C2RCC were more accurate than for MSI from the evaluated algorithms. OLCI POLYMER OC3 -org and Blend -org configurations were most accurate (MdSA 48.41% and 36.61%, respectively), closely followed by the BNN (MdSA 49.75%). Optimising the OC3 and Blend algorithms coefficients did not improve the chla estimation for the match-ups, unlike in the 50/50 split and LOO assessments. The optimisation of the OC3 and Blend algorithm coefficients systematically increased the underestimation of chla (Bias > -70%). The reason for this difference lays in the large value range of the training set used for the algorithm coefficient optimisation, which was naturally higher than the variation in Lake Geneva.

Algorithms using OLCI imagery corrected through POLYMER underestimated chla (Bias between -21% for BNN and -174% for MDN). For OLCI C2RCC match-ups the BNN was four times more accurate than the OC3 and Blend configurations. The reference algorithms performed worse on OLCI

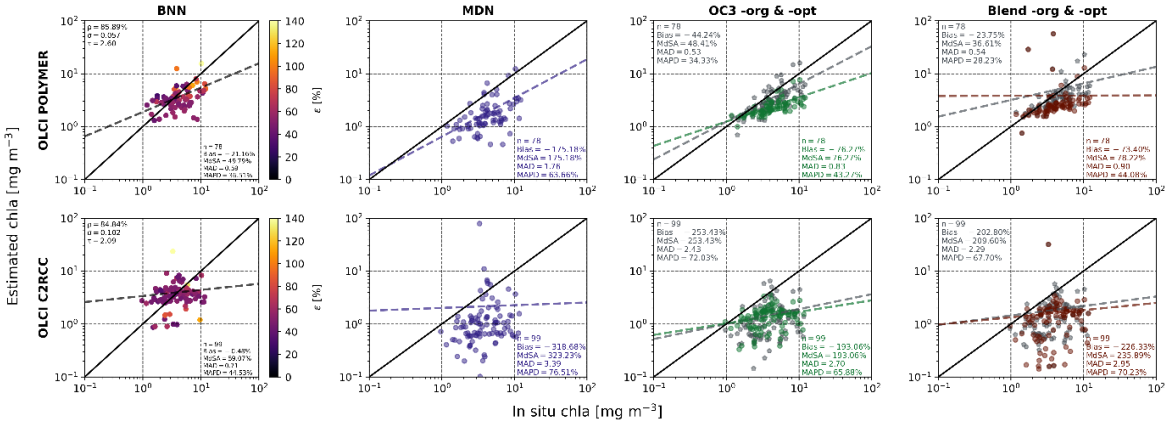


***Figure 3.9:*** *LéXPLORE platform (Lake Geneva) OLCI match-up chla estimates from BNN, MDN, OC3 -org & -opt and Blend -org and -opt algorithms. For OC3 and Blend, grey markers indicate the original and coloured markers the optimised algorithm coefficients.*

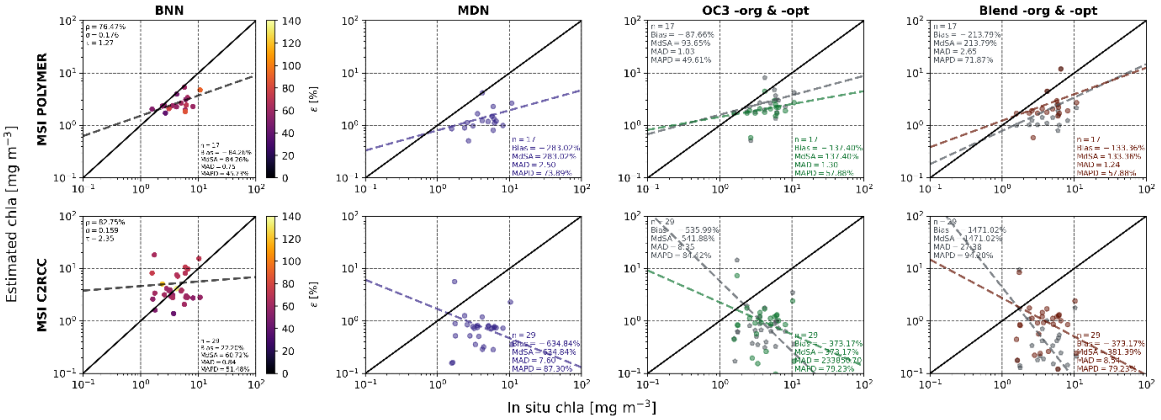


***Figure 3.10**: LéXPLORE platform (Lake Geneva) MSI match-up chla estimates from BNN, MDN, OC3 -org, OC3 -opt, Blend -org and Blend -opt algorithms. For OC3 and Blend, grey markers indicate the original and coloured markers the optimised algorithm coefficients.*

observations compared to the *in situ* 50/50 training/test set split and LOO assessments, which indicates that the reference algorithms were more affected by AC uncertainties than the BNN. Contrary to POLYMER, C2RCC OLCI and MSI BNNs overestimated chla, resulting in less variability of estimated chla. PICP was high (> 75%) across all four sensor and AC combinations, and MACD values were similar to the previous LOO analysis on independent region data. Sharpness values were lower than in the 50/50 split and LOO assessments, indicating small average standard deviations and thus a better uncertainty calibration. OLCI and MSI MDNs systematically underestimated chla across all four sensor and AC configurations (MdSA > 176%, Bias > 174).

### 4.4 Lake Geneva time series

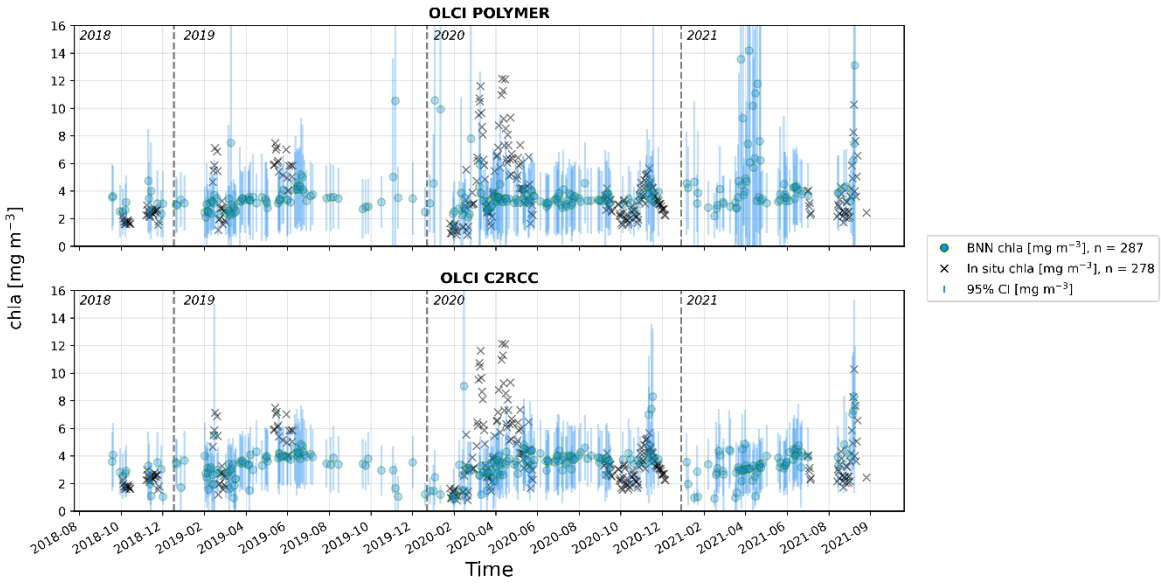


***Figure 3.11**: BNN OLCI time series of* LéXPLORE *(Lake Geneva), including all match-ups.*

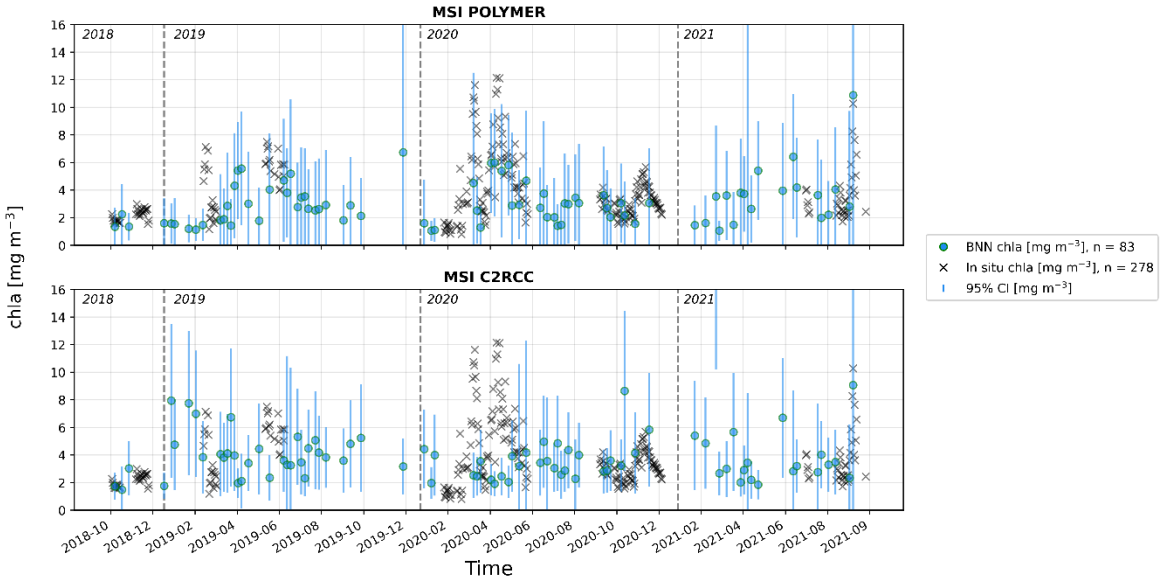


***Figure 3.12**: BNN MSI time series of LéXPLORE (Lake Geneva), including all match-ups.*

Phytoplankton vertical variability is seasonal in Lake Geneva (Minaudo et al., 2021) and a major factor driving optical properties (Nouchi et al., 2018). The impacts of phytoplankton vertical variability and its effect on remote sensing retrievals is still poorly understood. Seasonal variations overs the low to moderate chla range (0 – 8 mg $m^{-3}$) were accurately estimated by the OLCI BNNs during the four-year period, independent of the prior AC algorithm. The confidence intervals in most cases (> 80% PICP) captured the corresponding in situ chla value (see Figures 3.9, 3.10). For a spring bloom in 2020, moderate to high chla (8 – 12 mg $m^{-3}$) was not estimated with similar precision when compared to low chla across the four years. The confidence intervals were too small and did not contain most of the high peaks during the bloom. Conversely, *in situ* estimates in August – September 2021 with chla > 5 mg $m^{-3}$ were closely matched by the OLCI and MSI BNNs.

The MSI BNNs showed a larger spread in the low to moderate chla range compared to the OLCI BNNs, with wider confidence intervals. Larger and less precise intervals (higher Sharpness) via MSI agree with the results of the 50/50 split and the match-up assessments. The uncertainty estimates by the OLCI BNNs were throughout the exercises more accurately calibrated. Across the time series, $R_{rs}(\lambda)$ derived through POLYMER were associated with higher BNN chla uncertainty than through C2RCC. Between April and May 2021, POLYMER estimated negative $R_{rs}(\lambda)$ values in red bands (not shown) which led to higher uncertainty in chla estimates compared to more stable and lower chla estimates following from C2RCC for the same period.

### 3.4.4 OLCI and MSI BNN chla and uncertainty products

The BNNs were applied to several OLCI and MSI images over lakes with varying optical properties to assess horizontal consistency and visualise chla estimates alongside retrieval uncertainties. Single-

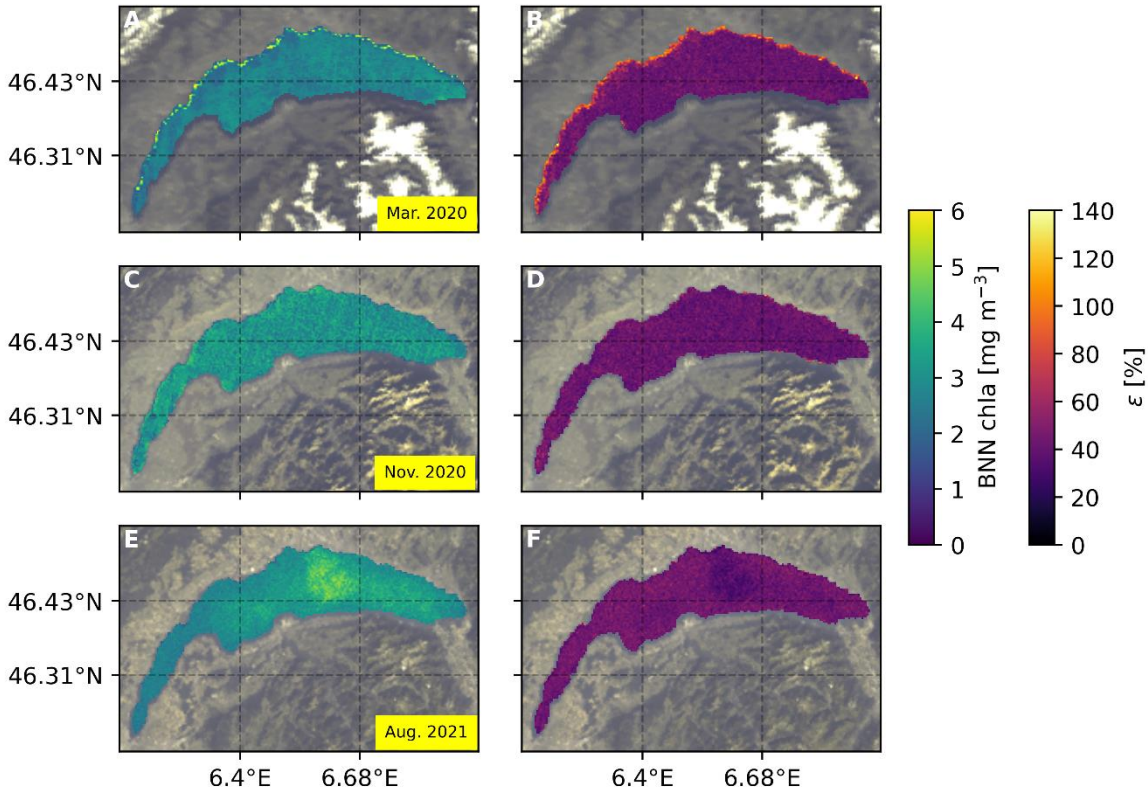


***Figure 3.13**: OLCI BNN chla and uncertainty products obtained through C2RCC AC over Lake Geneva on 13th of March 2020 (A, B), 11th of November 2020 (C, D) and 21st of August 2021 (E, F).*

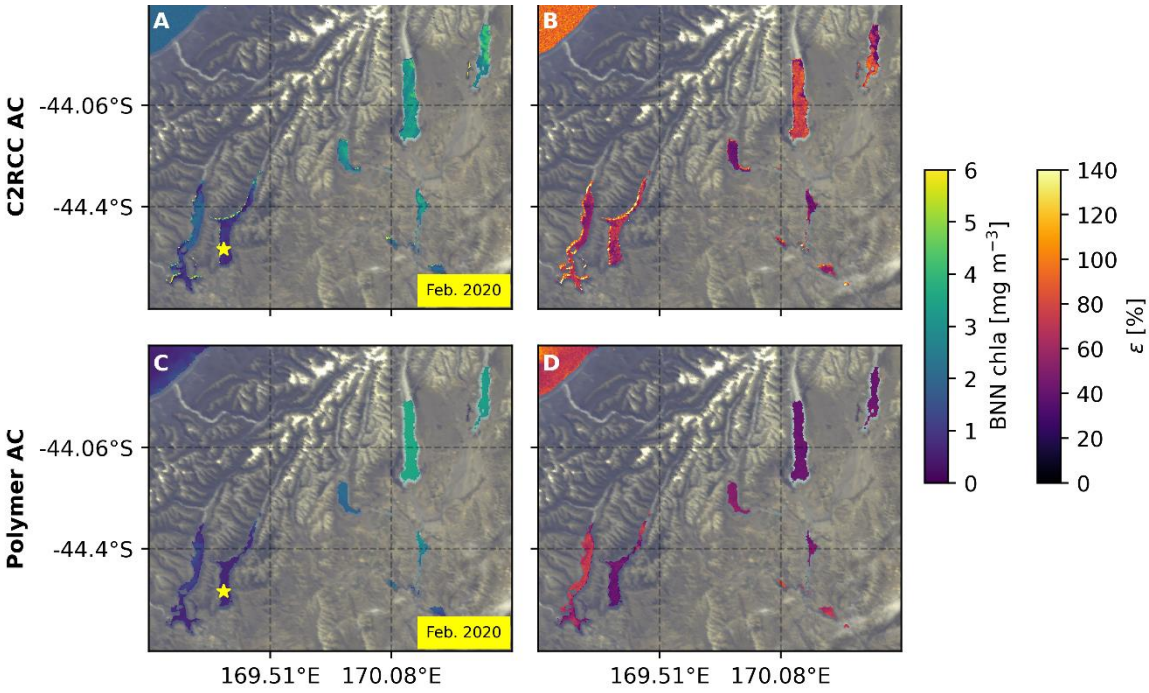


***Figure 3.14:** OLCI BNN chla and uncertainty products over southern New Zealand, 9th of February 2020. Chla was measured in situ at the location of Lake Hawea (star symbol) as 0.89 $mg\ m^{-3}$ during the overpass.*

day products over Lake Geneva, the Southern Island in New Zealand (SNZ), Lake Turkana in Kenya and boreal areas of Sweden and Canada were processed. Figure 3.13 illustrates OLCI BNN chla

through prior C2RCC AC and associated uncertainty products over Lake Geneva from March 2020, November 2020, and August 2021. The maps using the BNN match the time series chla patterns and ranges for Lake Geneva from the oligotrophic states in March 2020 to mesotrophic levels in August 2021 (Soulignac et al., 2018).

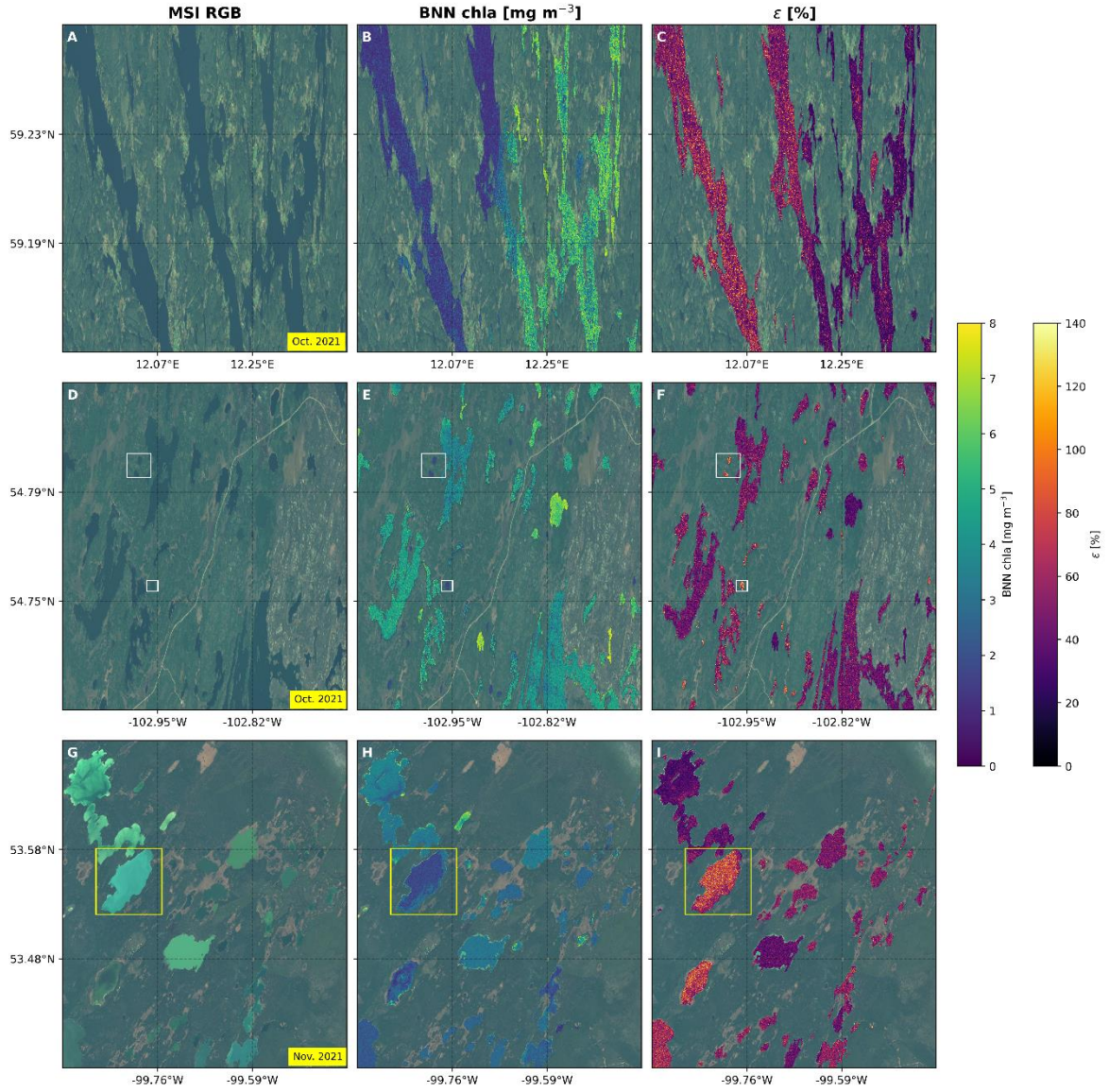


***Figure 3.15:*** *BNN-C2RCC chla and uncertainty products for MSI (20 m) over Sweden (A, B, C) on 11$^{th}$ of October 2021, and Canada on 16$^{th}$ of October (D, E, F) and 4$^{th}$ of November 2021 (G, H, I). A, D, G represent MSI L1 products with true water colours obtained from top-of-atmosphere sensor radiance, while B, E, H depict produced chla. C, F, I represent associated chla uncertainty. White and yellow boxes surround small water bodies with high BNN uncertainty. See explanations in this section for the boxes.*

The retrieval uncertainties ranged between 20 – 60% in Lake Geneva similar to the results of the previous match-up and time series assessments. The results over the LéXPLORE location can thus be considered transferable to other regions of the lake. During March 2020, areas with high uncertainty are observed near the north shore of Lake Geneva, likely associated with contamination of water-leaving radiance by bottom reflectance, adjacent land, or both.

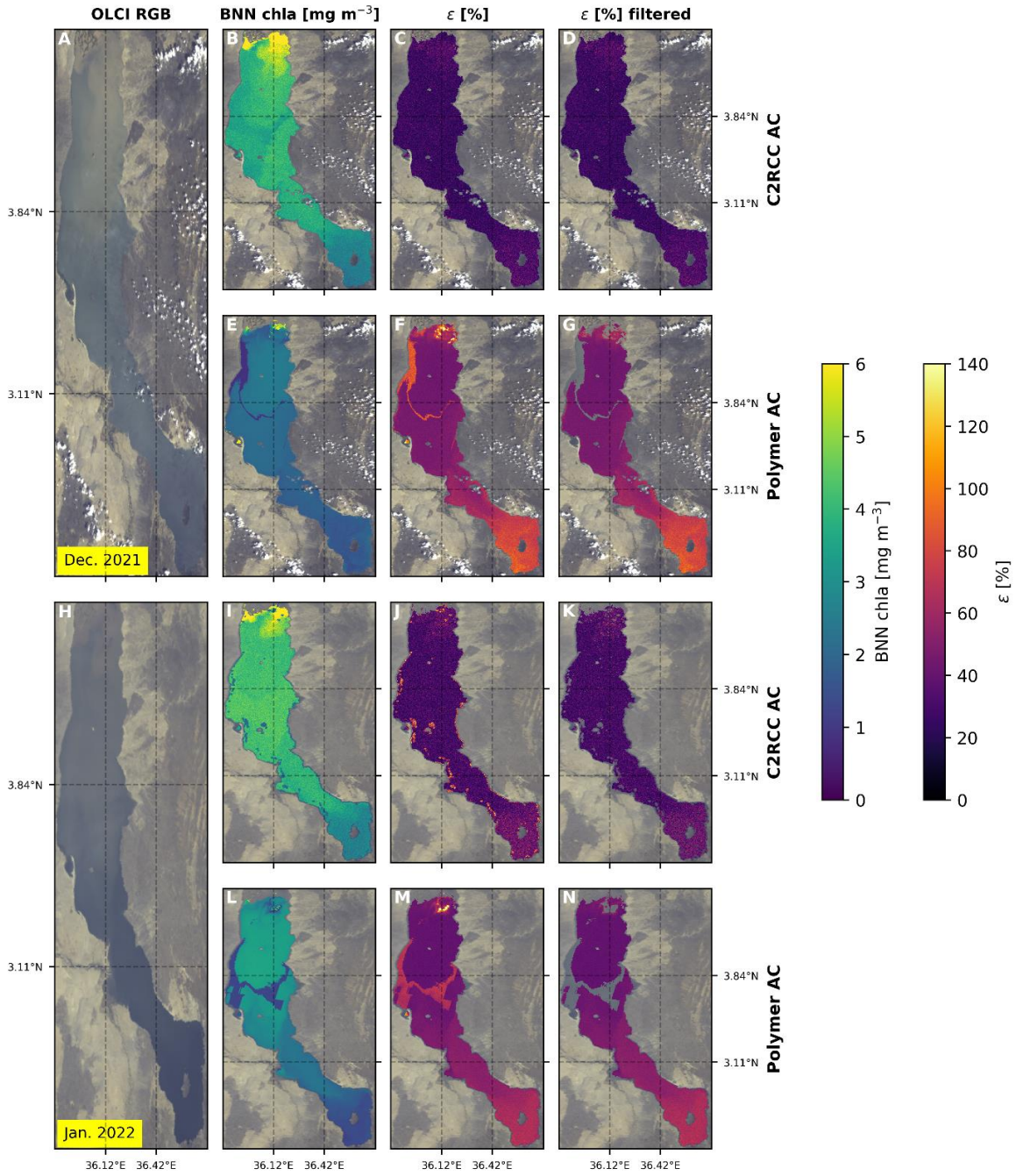


***Figure 3.16:*** *BNN chla and uncertainty products for Lake Turkana (Kenya) on 13$^{th}$ of December 2021 and 14$^{th}$ of January 2022 using C2RCC and POLYMER ACs. A, H: OLCI L1 RGB. B, E, I, L: BNN chla. C, F, J, M: Associated uncertainty. D, G, K, N: Filtered pixels based on uncertainty higher than 85% for Dec. 2021 and 65% for Jan. 2022.*

OLCI products processed with C2RCC and POLYMER for SNZ are shown in Figure 3.14. BNN chla products varied spatially between the lakes and AC algorithms. Moreover, uncertainties of C2RCC derived-chla were significantly higher (between 50 - 100%) than for POLYMER, as shown in Figure 3.14. Unlike for Lake Geneva, POLYMER did not produce negative $R_{rs}(\lambda)$ in any wavebands over this area. A surface measurement of *in situ* chla in Lake Hawea (stars on Figure 3.14) was 0.89 mg m$^{-3}$ on the 9$^{th}$ of February during the Sentinel-3 overpass, which closely matched the BNN chla of this scene. Overall, highest uncertainties were generated for chla < 1 mg m$^{-3}$, a range

for which training data were scarce. The scarcity became apparent in the top left corner of the imagery in the clear coastal waters. The uncertainties for chla < 1 mg $m^{-3}$ were similar to the results of the initial 50/50 split analysis. Highest chla retrieval uncertainties from the BNN and the reference algorithms are observed in this concentration range.

The MSI BNN generates high-quality chla products for small water bodies (SWBs). Figure 3.15 contains BNN MSI products over SWBs of Canada and Sweden. Satellite-derived $R_{rs}(\lambda)$ from SWBs are prone to adjacency effects which degrade the quality of the retrievals (Paulino et al., 2022; Sterckx et al., 2011b). In Figure 3.15 D-F, small lakes contained water pixels (white squares) for which the MSI BNN produced high uncertainties. Associated reflectance spectra of small lakes in this scene used to retrieve chla were influenced by the proximity to land, resulting in high uncertainty. The MSI BNN products demonstrate sensitivity to changes in chla between the lakes captured on this scene. Consistent with OLCI BNN retrievals, lake areas depicted in MSI BNN products exhibited higher uncertainties where chla was lowest (Figure 3.15 A, B, C). In November 2021 (Figure 3.15

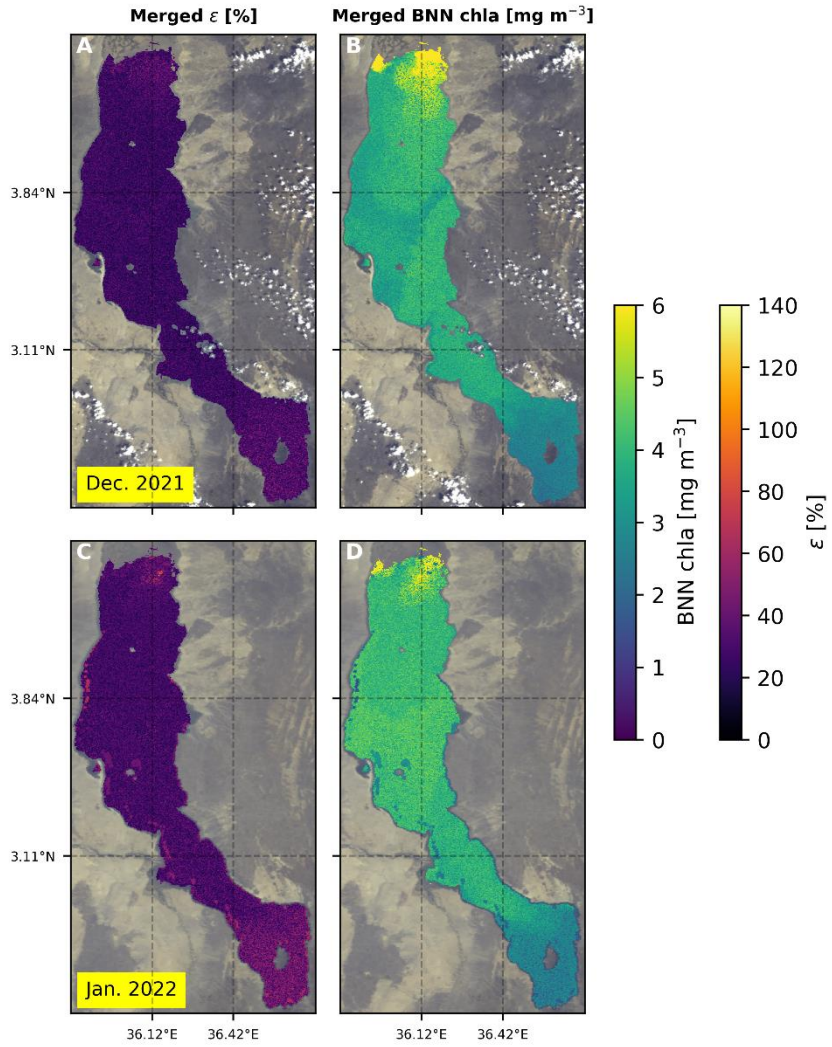


***Figure 3.17:*** *BNN OLCI merged uncertainty and chla products from 13th of December 2021 and 14th of January 2022 over Lake Turkana. See Figure 3.16 for the products used for the merge.*

G-I) the RGB of “Clearwater Lake” appears extremely bright (see yellow square), which the BNN associated with high chla uncertainty.

OLCI BNN products from December 2021 and January 2022 (Figure 3.16) cover the largest desert lake in the world, Lake Turkana (Kenya), known for its spatially variable optical properties (Liu et al., 2021d; Tebbs et al., 2020). OLCI BNN products over Lake Turkana captured a gradient in chla of the more turbid waters in the north of the lake than towards the southern end. BNN chla following C2RCC (Figure 3.16 B, J) was higher (2 – 6 mg $m^{-3}$) than through POLYMER (2 – 3.5 mg $m^{-3}$; F, N). Chla uncertainties from C2RCC were approximately 50% lower than by POLYMER for both days. The product uncertainties can be used to filter retrievals and exclude areas of water bodies for which the BNN assigned high uncertainty. $R_{rs}(\lambda)$ from POLYMER caused the BNN to produce unrealistic patches of chla around 1 mg $m^{-3}$ (Figure 3.16 F, N). These patches do not resemble the optical gradient apparent in the area and were not produced through prior C2RCC AC. In contrast, for the Jan. 2022 product (Figure 3.16 K), C2RCC reflectance spectra caused the BNN to generate unrealistically low chla in the middle of the lake and nearshore when compared to neighboring pixels (Figure 16 J). The unusual BNN chla patches and areas caused by the POLYMER and C2RCC ACs were consistently associated with 40 – 100% higher uncertainties than surrounding pixels (Figure 3.16 G, K, O). These highly uncertain areas over Lake Turkana were treated and resulted in new products. First, the corresponding pixels were removed by imposing BNN uncertainty thresholds (higher than 85% for Dec. 2021 and 65% for Jan. 2021) on the entire scene (Figure 3.16 D, H, L, P). The procedure eliminated almost all observations with

high uncertainty and resulted in a more homogenous product. Second, the BNN chla uncertainties associated with each AC algorithm were used to generate merged chla and uncertainty products (Figure 3.17) by selecting the result per scene with the lowest uncertainty from either AC. The merged product replaced the highly uncertain areas of either AC processor (see Figure 3.16) and led to an improved product quality both in terms of chla and uncertainty.

## 3.5 Discussion

Chla estimation uncertainties have repeatedly been shown to be highest in oligotrophic and mesotrophic conditions (Liu et al., 2021; Neil et al., 2019; Werther et al., 2021). The retrieval results over the chla range considered here confirmed these findings. Large gains in retrieval accuracy by the BNNs were made across the five assessments of this study, affirming that a probabilistic ensemble of NNs based on MC dropout trained with a representative *in situ* dataset can lead to estimation improvements over oligo- and mesotrophic lakes. Uncertainty is common in lake remote sensing and unlike the reference chla algorithms, the BNNs indicated when retrieved chla was uncertain. Here we discuss the BNN chla and uncertainty estimation, the uncertainty calibration and the handling of uncertain estimates as exemplified for Lake Turkana.

### 3.5.1 BNN chla estimation

The uncertainty associated with BNN chla varied between the five performance assessments. Lowest BNN chla retrieval uncertainty was found for the 50/50 split assessment over established chla algorithms. The 50/50 split performance of the BNN resembles results reported in the MDN publications (O'Shea et al., 2021; Pahlevan et al., 2020). In this assessment strategy the systems in both training and test sets are not strictly independent. Observations from the same system can occur in both sets due to the prior randomisation of the entire dataset, thereby introducing knowledge about similar optical properties in the algorithm training stage. Controlled randomisation prevents dataset shifts and the 50/50 dataset split results therefore represent an idealistic evaluation scenario. However, dataset shifts are common during satellite application of a developed ML algorithm. Therefore, the region-wise LOO assessment was conducted, and the results provided further insight into the BNN capabilities. The LOO assessment corroborated the 50/50 split results, whereby for 7 out of 10 regions the BNN outperformed reference algorithms. Expectedly, overall accuracy for LOO was worse than in the 50/50 split assessment. Regional chla accuracy varied from 19 % to nearly 60 % MdSA for the BNN but was consistently higher for the reference chla algorithms. Moreover, the reference chla algorithm performance varied strongly between regions. LOO results highlight that the training dataset of this study is representative for measurements sharing the same OWTs, even if from different regions. The LOO results also suggest that varying degrees of uncertainty between the observations of the dataset exist that propagated into the BNN chla estimation. This source of observational uncertainty is irreducible for the BNNs and sets a realistic baseline for expected performance. Compared to the *in situ* assessments, overall BNN accuracy was expectedly lower for the match-ups over Lake Geneva due additional sources of uncertainty such as prior AC. Retrieval accuracy varied between 40 – 60% for OLCI and was 10 – 20% higher for MSI over Lake Geneva (Figure 3.11).

BNN chla uncertainties were highest for chla < 1 mg $m^{-3}$ across larger and smaller water bodies captured on OLCI and MSI products over SNZ, Canada and Lake Turkana. Because measurement availability is globally scarce in the clearest lakes, observation coverage was low in the training dataset. The low amount of training data and resulting high chla uncertainty over this range clearly indicate a measurement gap. Radiative transfer simulations or coastal data sets may be able to fill this training observation gap to further reduce chla uncertainties. In addition to measurement scarcity, systematic measurement error is prone to lowest chla and phytoplankton absorption coefficients (McKee et al., 2014). With regards to higher biomass conditions, only 9.5% of the dataset consisted of chla > 25 mg $m^{-3}$ (maximum 68 mg $m^{-3}$), therefore chla estimation uncertainty will rise with increasing chla. Oligo- and mesotrophic waters most likely do not exceed this threshold even when experiencing seasonal, short-term phytoplankton blooms such as those observed in Lake Geneva. The BNNs were specifically developed for oligo- and mesotrophic waters and the results of this study show that algorithms should be specifically developed for these lake types.

### 3.5.2 Calibration of uncertainty

The quality of the uncertainty calibration was measured through the PICP, Sharpness and MACD metrics. Across the different assessments, the confidence intervals covered the *in situ* reference chla value in 3/4 cases (PICP > 75%). However, the provided uncertainty estimate was not consistently well-calibrated across the assessments. The Sharpness metric specifically measured the range of the uncertainty estimate per observation and small Sharpness values were not achieved in all performance assessments. Moreover, even though OLCI performance did not differ significantly from MSI, the OLCI uncertainties were consistently better calibrated. More research is needed to understand under which conditions the BNN uncertainty calibration was systematically inaccurate and how it correlated with optical properties and training data availability. The source of uncertainty can be the underlying BNN statistical model (the epistemic uncertainty) or an external source (the aleatoric uncertainty), such as prior AC, an AE or other unknown error as part of the measurements contained in the dataset (IOCCG, 2019). Whereas epistemic uncertainty can be improved through better data coverage and incremental improvements to the underlying statistical model, aleatoric uncertainty sources are often more difficult to identify and correct as they are de-coupled from the BNN. Moreover, it is desirable to distinguish between the aleatoric and epistemic uncertainty contributors altogether to provide specific pathways for BNN improvements.

Besides the investigation into observation conditions and optical properties that affect BNN uncertainty, computational methods exist to improve uncertainty calibration as a post-processing step. The approaches usually require a separate dataset, which in practice can be a subset of the overall data set aside that is then not used for training or evaluation. Example techniques include histogram binning (Zadrozny and Elkan, 2001), isotonic regression (Kuleshov et al., 2018; Zadrozny and Elkan, 2002) and Platt scaling (Platt, 1999).

### 3.5.3 AC selection through BNN uncertainty

Atmospheric correction is the greatest source of uncertainty in satellite remote sensing of OACs in lakes (Pahlevan et al., 2021b; Pereira-Sandoval et al., 2019; Wang et al., 2019; Warren et al., 2019). Observations associated with high uncertainty differed between the BNN estimates for the same products over Lake Geneva, SNZ and Lake Turkana. To deal with varying degrees of uncertainty introduced by AC algorithms, the BNN chla uncertainties were used as a selection mechanism by imposing thresholds (as undertaken in Figure 3.16) or by retaining the observation associated with the lowest BNN uncertainty (Figure 3.17). The former approach omits observations and is desirable when AC quality is unknown, during unsupervised processing of imagery or when a comparison to another AC algorithm is not suitable. When two or more ACs for the same product are available, the latter approach to obtain a merged product is the favoured option. Moreover, the merging exemplified in Figure 3.17 provides a pathway to decide between different ACs. Since AC performance largely varies with faced atmospheric properties and IOPs, it is a priori impossible to know which AC

performs most accurate. A more thorough analysis with multiple products, areas and atmospheric conditions should be undertaken to determine whether AC selection through BNN uncertainty consistently improves the chla product quality.

## 3.6 Conclusion

The BNNs developed in this study represent a novel approach to derive chla and associated uncertainty estimates from $R_{rs}(\lambda)$ from Sentinel-2 (MSI) and Sentinel-3 (OLCI) over oligo- and mesotrophic lakes with the ability to account for phytoplankton abundance shifts into the eutrophic regime. Using the largest *in situ* dataset over oligo- and mesotrophic lakes and reservoirs available to date, the BNNs were trained, validated, and compared against state-of-the-art and community-established algorithms. In this study we demonstrated that the developed BNN algorithms work similarly for both MSI and OLCI observations (Bias < 1.5%, MdSA < 26%, MAPD < 22%), suggesting their potential to produce high-quality chla products across multiple missions. We found that chla uncertainties for MSI were on average 10-20% higher than for OLCI reaffirming that additional bands on OLCI not only benefit the chla retrieval accuracy, but effectively reduce the associated uncertainty. Lower uncertainties through an increased number of available bands may be of significant relevance for the rising number of hyperspectral missions such as PRISMA and PACE. Throughout the performance assessments, chla uncertainties ranged between 10 – 65%. As evidenced through long-term match-ups over Lake Geneva, the BNNs were able to improve upon the chla retrieval of established methods. Across all assessment strategies, the largest (> 25%) performance improvements were made over oligotrophic water bodies (<= 8 mg $m^{-3}$), whereas for chla > 10 mg $m^{-3}$ established algorithm approaches performed similarly. As for any other chla algorithm relying on satellite-derived $R_{rs}(\lambda)$, BNN chla accuracy varied with the prior AC algorithm. The BNN uncertainty estimate enabled to handle high uncertainty introduced by AC algorithms and other sources prior to chla retrieval. We exemplified common issues that are caused by ACs over oligo- and mesotrophic lakes: negative $R_{rs}(\lambda)$ in red bands, uncertainty introduced by the adjacency effect and unreasonable $R_{rs}(\lambda)$ values stemming from extreme variation in bio-optical lake properties. To deal with inconsistent results obtained by the different AC algorithms, we suggest further research into the selection of AC products through BNN uncertainties to retain accurate chla retrievals. Moreover, research into strategies to improve the uncertainty calibration is recommended. Further downstream products depending on accurate chla, such as trophic status and net primary production, may impose uncertainty quality thresholds to retain high-quality chla retrievals. The BNN architecture presented in this study can be extended to other multi- and hyperspectral missions such as Landsat, PRISMA and PACE and supports the retrieval of other OACs and IOPs such as TSM and CDOM.

## 3.7 Code Availability

The OLCI and MSI BNNs will be made available under https://github.com/mowerther/BNN_2022

## 3.8 Contributions

Concept: MW, ES, DO, SGHS

Data curation: MW, ES, DO, ML, DG, TK, PH, AT, RG, AV

Investigation: MW, ES, DO, SGHS, DG

Methodology: MW

Code: MW

Visualisation: MW

Writing lead: MW

Writing – review & editing: SGHS, DO, ES, DG, TK, ML, PH, AT, RG, AV

## 3.9 Acknowledgements

This study has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No. 776480 (MONOCLE) and the EUMETSAT Copernicus Collaborative Exchange Award. We would like to thank providers of data to LIMNADES for their measurements used in this study: Stewart Bernard, Caren E. Binding, Mariano Bresciani, Claudia Giardino, Anatoly A. Gitelson, Luis Guanter, Kari Y.O. Kallio, Tiit Kutser, Ciro Manzo, Mark W. Matthews, John F. Schalles and Antonio Ruiz-Verdú. We thank Camille Minaudo for assistance with processing chla and $R_{rs}(\lambda)$ measurements obtained on LéXPLORE.

**Appendix 1. Regions and inland water bodies constituting the dataset of this study. Colour shading matches Figure 3.1.**

| Region | Countries | Dataset Measure-ments (n = 1755) | Inland water bodies (n = 178) | References |
|---|---|---|---|---|
| 1 | Spain | 109 | Aguilar reservoir, Alarcon reservoir, Lake Albufera, Alcantara reservoir, Alcorlo reservoir, Almendra reservoir, Bornos reservoir, Burguillo reservoir, Canelles reservoir, Cernadilla reservoir, Cijara reservoir, Contreras reservoir, Cortes de Pallas reservoir, Cuerda del Pozo reservoir, El Atazar reservoir, Giribaile reservoir, Guadalcacin reservoir, Guadalen reservoir, Guadalteba reservoirs, Iznajar reservoir, La-Serena reservoir, Negratin reservoir, Pinilla reservoir, Rialb reservoir, Riano | (Ruiz-Verdú et al., 2005, 2008; Simis et al., 2007) |

| | | | | |
|---|---|---|---|---|
| | | | reservoir, Ricobayo reservoir, San Juan reservoir, Lake Sanabria, Santa Teresa reservoir, Terradets reservoir, Pantano de Tremp reservoir, Uribarri Ganboako urtegia reservoir, Valdecanas reservoir, Valmayor reservoir, Valparaiso reservoir, Valuengo reservoir | |
| 2 | The Nether-lands, Is-rael | 107 | IJsselmeer, Lake Kinneret | (Simis et al., 2005; 2007; Yacobi et al., 2011) |
| 3 | Switzer-land, Italy | 466 | Lake Biel, Lake Garda, Lake Idro, Lake Como, Lake Iseo, Lake Geneva, Lake Maggiore, Lake Trasimeno | (Bresciani et al., 2011; Giardino et al., 2005, 2013, 2014, 2015; Guanter et al., 2010; Manzo et al., 2015) |
| 4 | United Kingdom, Sweden, Finland, Estonia | 75 | Lake Antu, Lake Bassenthwaite, Lake Coniston, Loch Katrine, Loch Lomond, Lake Maleren, Loch Ness, Lake Peijänne, Lake Peipsi, Lake Pyhäjärvi, Lake Ullswater, Lake Vänern, Lake Vesijärvi | (Kallio et al., 2015; Kutser et al., 2013) |
| 5 | China, Japan | 45 | Lake Biwa, Lake Erhai, Lake Suwa | (Matsushita et al., 2015) |
| 6 | New Zealand | 133 | Lake Brunner, Lake Hauroko, Lake Hawea, Lake Ellesmere, Lake Humuhumu, Lake Kahuparere, Lake Kai-iwi, Lake Kanono, Lake Kegonsa, Lake Mokeno, Lake Okareka, Lake Okaro, Lake Okataina, Lake Ototoa, Lake Pearson, Lake Pupuke, Lake Rotoehu, Lake Rotoiti, Lake Rotokakahi, Lake Rotokare, Lake Rotokawau, Lake Rotoma, Lake Rotonuiaha, Lake Rotoroa, Lake Rotorua, Lake Rototuna, Lake Sheppard, Lake Tarawera, Lake Taupo, Lake Tutuira, Lake Waikaremoana, Lake Waikere, Lake Mapourika, Lake Moeraki, Lake Monowai, Lake Paringa, Lake Te Anau, Lake Wahapo, Lake Wakatipu | (Pahlevan et al., 2020, 2022; Wang et al., 2018) |
| 7 | South Africa | 26 | Loskop reservoir, Theewaterskloof reservoir | (Matthews, 2014; Matthews and Bernard, 2013) |
| 8 | U.S. Mainland | 487 | Lake Amos, Lake BranchedOak, Lake Collins, Lake CorpusCristi, Lake Crescent, Lake DiamondPond, Lake EagleCreek, Lake EastTwin, Lake Ewell, Lake Fivemile, Flatriver reservoir, Lake Forest, Lake Fremond, Lake Fresmond, Lake Geist, Lake GingerCove, Lake GoosePond, Lake Granite, Great Salt | (Bradt, 2012; Dall'Olmo et al., 2003, 2005; Gitelson et al., 2007; Gitelson et al., 2008; Li et al., 2013, 2015; Moore et |

| | | | | |
|---|---|---|---|---|
| | | | Lake, Green Lake, Groton Lake, Hickleys Lake, Kettlebrook reservoir, Lake Attitash, Lake Champlain, Lake Hayward, Lake Manawa, Lake Mattawa, Lake Sunapee, Lake Wentworth, Lake LittleBigWood, Long Lake, Long Pond, Lowell Lake, Lake Lowerbranch, Lake Mantova, Lake Morse, Norton reservoir, Lake Okoboji, Perch Pond, Pine Lake, Pleasant Lake, Quacumquasit Pond, Quinebaug Lake, Sebago Lake, Sebasticook Lake, Silver Lake, Squam Lake, Tilden Pond, Tyler Lake, Lake Whalum, Wallum Lake | al., 2014; Schalles, 2006; Schalles and Hladik, 2012) |
| 9 | U.S. Wisconsin | 109 | Big Saint Germain Lake, Big Sand Lake, Butternut Lake, Fence Lake, Fox Lake, Geneva Lake, Lac Courte Oreilles, Lac Vieux Desert, Lake Chippewa, Lake Mendota, Lake Monona, Lake Waubesa, Lake Winnebago, Metonga Lake, Pelican Lake, Rock Lake, Round Lake, Shawano Lake, Trout Lake | (Pahlevan et al., 2022, 2021, 2020) |
| 10 | U.S., Canada | 198 | Lake Erie, Lake Ontario, Shadow Lake, Lake Winnipeg | (Binding et al., 2008, 2010; 2011; 2013) |

## Appendix 2. *In situ* measurements NZ 2020

Chla water samples were filtered (using a 0.45 $\mu m$ GFF filter) and then extracted using 95% methanol (spiked with Tocopheryl acetate as an internal standard) before being sonicated and frozen to complete extraction of all chla (Hooker et al., 2009). Finally, the extract was filtered through a 0.45 $\mu m$ Teflon filter. Resultant extracts were then run on an Agillent Infinity 1290 uHPLC using the method by Van Heukelem and Thomas (2001). TSM was measured gravimetrically from the dried and combusted residue on pre-combusted and pre-weighed filter pads (APHA 2540D and APHA 2540E). Laboratory triplicates of $a_{CDOM}(\lambda)$ were transferred to a 0.1 m cuvette and optical densities of the filtrates were measured for a wavelength range from 350 to 800 nm using the method described in Stedmon et al. (2000).

Radiometric measurements were made with a ship-mounted set of three hyperspectral sensors (TriOS RAMSES) installed just above the water surface. The first sensor pointed at the water surface and collected upwelling radiance $L_u(\lambda)$, which comprises both water-leaving radiance $L_w(\lambda)$ and the reflected sky irradiance $\rho_s L_s(\lambda)$. The sky irradiance $L_s(\lambda)$ was measured with a second sensor, while a third sensor measured downwelling irradiance $E_d(\lambda)$. $R_{rs}(\lambda)$ was then calculated through Simis and Olsson (2013):

$$R_{rs}(\lambda) = L_w(\lambda, 0^+)/E_d(\lambda), \quad (A2)$$

$$L_w(\lambda, 0^+) = L_t(\lambda) - \rho_s L_s(\lambda). \quad (A2.1)$$

Measurements were kept in this dataset if they were taken under cloud free conditions and on calm waters to ensure a correct alignment of the ship at a viewing azimuth angle ($\varphi_v$) > 90° (ideally at 135°) (Hooker and Morel, 2003b). All $R_{rs}(\lambda)$ were corrected for residual reflected skylight through the method described in (Jiang et al., 2020).

## Appendix 3. LéXPLORE platform measurements

The autonomous Thetis profiler (Seabird Scientific Inc.) is deployed at the LéXPLORE platform in Lake Geneva. The profiler is comprised of hyperspectral radiometers, absorption, backscattering and attenuation sensors and a fluorescence probe for chla measurements. The top 50 m of the water column are measured by the profiler at 3-hour intervals and a vertical resolution of 10 cm. A total of 278 profiles taken between October 2018 and September 2021 were used in this study. *In situ* measurements from the Thetis profiler are available in the Datalakes portal (https://www.datalakes-eawag.ch/).

Fluorescence probe-derived chla ($CHL_F$) from the Thetis profiler were measured using the WetLabs ECO Triplets BBFL2W, which has excitation and emission wavelengths at 470 and 695 nm, respectively. $CHL_F$ estimates are impacted by non-photochemical quenching (NPQ) of fluorescence leading to an underestimation of chla in high irradiance conditions Gupana et al. (pers. comm.). To compensate for the underestimation, we applied the chla correction procedure developed by (Roesler and Barnard, 2013). We obtained a linear relationship between $CHL_F$ and absorption line height (aLH) based chla estimates ($CHL_A$) during low light conditions which ensured minimal impact of NPQ to $CHL_F$. This linear relationship was used for daytime aLH measurements to correct NPQ-impacted $CHL_F$ estimates. aLH was calculated using the following equations:

$$a_{BL}(\lambda_{ref}) = \frac{a(715) - a(650)}{715 - 650} * (\lambda_{ref} - 650) + a(650), \quad (A3.1)$$

$$a_{LH}(676) = a(676) - a_{BL}(676).$$

where $a_{BL}$ is the absorption measured from the AC-S at the reference wavelength $\lambda_{ref}$ of 676 nm. $CHL_A$ was calculated for the upper 50 m of the water column and the top 5 m were averaged to represent *in situ* surface chla.

The below water upwelling radiance ($L_u$) and downwelling irradiance ($E_d$) measurements from the Thetis profiler were measured at nadir using the Satlantic HOCR RO8W and ICSW, respectively. Both radiometers have 180 spectral channels spanning 300 – 1200 nm. The underwater reflectance ($R_m$) was calculated using the following equation:

$$R_m(\lambda, z) = \pi \frac{L_u\,(\lambda, z)}{E_d\,(\lambda, z)} [\text{dl}], \quad (A3.2)$$

where $\lambda$ is wavelength region from 400 to 800 nm and $z$ is depth in meters. The $R_m$ closest to the water surface was converted to above-water reflectance following (Steinmetz et al., 2011).

$$R_{rs}(\lambda, z) = 0.544 * R_m(\lambda, z). \quad (A3.3)$$

## Appendix 4. BNN hyper-parameters

| Hyper-parameter | Value |
|---|---|
| Training epochs | 9000 |
| Hidden layers | 6 |
| Neurons in hidden layers 1-6 | 200, 500, 500, 500, 200, 2 |
| Batch-size | 512 |
| Activation function | Rectified Linear Unit (ReLU) |
| Dropout rate | 0.25% |
| Learning rate | $1.5 \times 10^{-4}$ |
| Optimiser | Adam |
| Loss function | Negative-log likelihood (NLL) |

## Appendix 5. Optimised coefficients of reference chla algorithms

**OLCI**

| Algorithm | Coefficient 1 | Coefficient 2 | Coefficient 3 | Coefficient 4 | Coefficient 5 |
|---|---|---|---|---|---|
| OC3 -opt | 0.07406589 | -2.02001282 | 0.16124464 | -17.55852465 | -33.62782537 |
| OC4 -opt | 0.12359851 | -2.80056003 | 3.94756167 | -23.76880782 | -92.02561384 |
| G11 -opt | -7.47220546 | 47.51051798 | -22.15274869 | | |

**MSI**

| | | | | | |
|---|---|---|---|---|---|
| OC3 -opt | 0.07090613 | -2.15824786 | 1.41248107 | -25.01903886 | -60.82627189 |
| G11 -opt | 2.22937993 | 38.80134518 | -22.82347274 | | |

# Chapter 4: Meta-classification of remote sensing reflectance to estimate trophic status of inland and nearshore waters

*This chapter was published as: Werther, M., Spyrakos, E., Simis S.G.H., Odermatt, D., Stelzer, K., Krawczyk, H., Berlage, O., Hunter, P., Tyler, a., 2021. Meta-classification of remote sensing reflectance to estimate trophic status of inland and nearshore waters. ISPRS J. Photogramm. Remote Sens. 176, 109-126.*

## Abstract

Common aquatic remote sensing algorithms estimate the trophic state (TS) of inland and nearshore waters through the inversion of remote sensing reflectance (Rrs($\lambda$)) into chlorophyll-a (chla) concentration. In this study we present a novel method that directly inverts Rrs($\lambda$) into TS without prior chla retrieval. To cope with the optical diversity of inland and nearshore waters, the proposed method stacks supervised classification algorithms and combines them through meta-learning. We demonstrate the developed methodology using the waveband configuration of the Sentinel-3 Ocean and Land Colour Instrument on 40 globally distributed inland and nearshore waters (567 observations). To assess the performance of the developed approach, we compare the results with TS derived through optical water type (OWT) switching of chla retrieval algorithms. Meta-classification of TS was on average 6.75% more accurate than TS derived via OWT switching of chla algorithms. The presented method achieved >90% classification accuracies for eutrophic and hypereutrophic waters and was >12% more accurate for oligotrophic waters than derived through OWT chla retrieval. However, mesotrophic waters were estimated with lower accuracy from both our developed method and through OWT chla retrieval (52.17% and 46.34%, respectively), highlighting the need for improved base algorithms for low - moderate biomass waters. Misclassified observations were characterised by highly absorbing and/or scattering optical properties for which we propose adaptations to our classification strategy.

## 4.1 Introduction

Eutrophication is the process whereby nutrient enrichment leads to excessive primary production of phytoplankton (cyanobacteria and algae) in water bodies (Conley et al., 2009; Smith et al., 2006). While eutrophication can occur as a natural process over long time scales, anthropogenic activities have increased eutrophication rates (Schindler, 2012; Sinha et al., 2017). The main causes are non-point pollution from agricultural practices, urban development and energy production and consumption (Glibert et al., 2005; Mainstone and Parr, 2002).

Therefore, eutrophication of inland and nearshore waters has become a leading cause of poor water quality (Smith, 2003; Smith et al., 2014; Wang, 2006). Increasing frequency and extent of phytoplankton blooms can have implications for ecosystem services and health (Heisler et al., 2008; Lewis et al., 2011; Nixon, 1995). In affected waters, cyanobacteria may produce cyanotoxins that adversely affect human and animal health (Codd, 2000; Merel et al., 2013). Naturally, lentic waters such as lakes are significant emitters of the greenhouse gases carbon dioxide ($CO_2$), nitrous oxide ($N_2O$) and methane ($CH_4$)(Cole et al., 2007; DelSontro et al., 2018). Enhanced eutrophication due to anthropogenic climate change is expected to increase aquatic $CH_4$ emissions from lentic waters by 30 – 90% over the next century (Beaulieu et al., 2019; Tranvik et al., 2009).

Over the last decades, several frameworks have been developed to assess and manage eutrophication. (Carlson 1977) proposed a Trophic State Index (TSI) linking transparency (Secchi disk depth (zSD [m])), surface phosphorus (P [mg/l]) and phytoplankton chlorophyll-a (chla [mg/m$^3$]) concentrations to the trophic state (TS) of lakes. The index partitioned TS into three classes: oligo-, meso- and eutrophic. In later work, (Carlson and Simpson, 1996) introduced an additional TS class (hypereutrophic) to include extreme biomass scenarios. More recently, other parameters linked to water optical properties, such as turbidity (NTU) and colour scales, were employed for the retrieval of TS (Binding et al., 2007; Lehmann et al., 2018; Wang et al., 2018).

Of the aforementioned parameters, *in situ* measurements of chla are most frequently used to estimate TS. Chla is a reliable proxy directly for phytoplankton biomass and indirectly for primary production (Carlson, 1977; Huot et al., 2007; Kasprzak et al., 2008). *In situ* derived chla is a core indicator in monitoring programs such as the European Water Framework Directive or the U.S. Clean Water Act (Carvalho et al., 2008; Keller and Cavallaro, 2008; Søndergaard et al., 2005). While the extraction of chla from *in situ* collected water samples has few, and likely low, associated uncertainties, this monitoring approach cannot be scaled up to include remote sites and short-lived phytoplankton bloom phenomena (Schaeffer et al., 2013; Tyler et al., 2016). Aquatic remote sensing complements *in situ* measurements by providing a spatial and temporal observation advantage (Mouw et al., 2015). Remote sensing methods only estimate surface water concentrations approximately up to the first optical depth, whereas *in situ* measurements can characterise the whole water column (Gordon and Clark, 1980; Zaneveld et al., 2005).

In aquatic remote sensing, the inherent optical properties (IOPs, i.e. absorption, backscatter and fluorescence) of water and the optically active constituents (OACs), namely phytoplankton pigments ($\phi(\lambda)$[1/m]), non-pigmented particles (nap($\lambda$)[1/m]) and the absorption by the chromophoric fraction of dissolved organic matter ($a_{\mathrm{cdom}}(\lambda)$[1/m]), impact the remote sensing reflectance (Rrs($\lambda$, sr$^{-1}$)) vector. Rrs($\lambda$, sr$^{-1}$) is defined as the ratio of water-leaving radiance $L_w(\mu\mathrm{W\ cm^{-2}\ sr^{-1}\ nm^{-1}})$ to total downwelling irradiance $E_d(\mu\mathrm{W\ cm^{-2}\ nm^{-1}})$:

$$\mathrm{Rrs}(\lambda,\ \mathrm{sr}^{-1}) = L_w/E_d. \tag{4.1}$$

Rrs($\lambda$) is thus the critical optical property to derive information from a water body about OACs dispersed in the water column. The retrieval of phytoplankton chla concentration, or the phytoplankton absorption component, $a_\phi(\lambda)$[1/m], can be expressed as a function estimation problem that requires inversion of Rrs($\lambda$):

$$x = f^{-1}[\text{Rrs}(\lambda)], \tag{4.2}$$

whereby $x$ is the quantity to invert Rrs($\lambda$) for. The inversion of Rrs($\lambda$) is known to be mathematically ill-posed, as multiple combinations of IOPs can result in the same Rrs($\lambda$) vector and may thus cause ambiguity in the inversion (Defoin-Platel and Chami, 2007; Sydor et al., 2004).

OAC compositions and concentrations strongly vary across inland and nearshore waters, thus accurate modelling of Equation 4.2 has led to the development of numerous chla retrieval algorithms over the past decades (see reviews by (Blondeau-Patissier et al., 2014; Matthews, 2011; Odermatt et al., 2012; Tyler et al., 2016). Chla retrieval algorithms may be divided into two categories: empirical and semi-analytical. As the name implies, algorithms of the former category are based on empiricism, in which a functional relationship between an OAC and the optical Rrs($\lambda$) vector is established from field observations, domain knowledge and/or experiments. Popular examples are the Fluorescence Line Height (FLH) (Gower et al., 1999), the Maximum Peak-Height (MPH) (Matthews et al., 2012) and Maximum Chlorophyll Index (MCI) (Gower et al., 2005) algorithms, which use band arithmetic to relate spectral phenomena associated with phytoplankton to the concentration of chla. Machine learning (ML) algorithms are also empirical, as the underlying statistical models are developed with datasets consisting of field and/or simulated observations. The fundament of ML algorithms are typically non-linear regression models trained with the available dataset. Non-linear regression models have the ability to approximate any type of function, which may be necessary to accurately solve for $x$ in Equation 4.2 (Hieronymi et al., 2017). Compared to band arithmetic algorithms, ML approaches can include any number of bands in the model development phase, which enables to exhaust the information contained within current multispectral sensors and opens up new possibilities for the exploration of hyperspectral resolutions (Pahlevan et al., 2021b). However, ML algorithms are often limited by available and representative datasets to train a model that generalises well to unseen observations. Regression approaches (not limited to non-linear ML models) can also invert Rrs($\lambda$) for IOPs such as $a_\phi(\lambda)$. Retrieved $a_\phi(\lambda)$ is then scaled to chla concentration.

Semi-analytical solution algorithms (SAA) invert Rrs($\lambda$) for IOPs (Werdell et al., 2018). SAA base the retrieval on physical reasoning, but partly employ statistical methods (hence the term 'semi'). In the inversion for $a_\phi(\lambda)$, SAA show many variants and differ in their definition of the $a_\phi(\lambda)$ spectral shape, the method to calculate the magnitude of $a_\phi(\lambda)$ and the defined relationship between Rrs($\lambda$) and $a_\phi(\lambda)$. The scaling of $a_\phi(\lambda)$ to chla can be significantly confounded in optically complex inland and nearshore waters due to pigment packaging and the contribution of accessory pigments to

absorption (Bricaud et al., 1995; Simis et al., 2007). Unless this variability is accounted for, non-linear effects in the relationship between $a_\phi(\lambda)$ and chla will also affect TS estimation.

Empirical and SAA approaches estimate the TS of a water body indirectly by inverting Rrs($\lambda$) for chla or by scaling $a_\phi(\lambda)$ to chla. The retrieved concentration indicates a TS class. In a recent study, (Shi et al., 2019) outlined that significant uncertainties may propagate into TS estimation due to the limited precision associated with inversion for chla. To overcome intermediate chla retrieval when TS information is ultimately required, (Shi et al., 2019) developed an approach that directly relates the light absorption coefficient of OACs to TS using the quasi-analytical algorithm (QAA) by (Lee et al., 2002).

To reduce associated retrieval uncertainties, another approach is to assign Rrs($\lambda$) into previously defined, distinct optical water types (OWTs) (Moore et al., 2014; Spyrakos et al., 2018). OWTs are then utilised to guide the retrieval, since a single chla algorithm in practice often shows limited accuracy across a range of OWTs. OWT switching and blending of several algorithms following prior classification into known OWTs has become established practice (Eleveld et al., 2017; Neil et al., 2019). These schemes exploit the complementary retrieval ranges of individual algorithms.

In this study, we develop a methodology to overcome issues associated with indirect TS derivation through inversion of Rrs($\lambda$) into chla (or $a_\phi(\lambda)$). To accomplish this, our method inverts for TS classes directly through modelling of the TSI system as a classification task. Retrieving TS classes instead of a continuous chla concentration value requires to develop a ML classification algorithm. To enable the classification, we establish a relationship between the Rrs($\lambda$) vector and TS through an *in situ* dataset (n= 2184) of co-located chla and Rrs($\lambda$) measurements. We further recognise the limited validity of a single algorithm for the information retrieval across many OWTs. However, instead of common switching or blending of algorithms through OWT schemes, we stack multiple classification models and combine their TS class predictions through a higher-level classifier.

To illustrate classification of TS, we define our dataset in this study as $D = \{(y_i,\ x_i), i = 1, \dots, N\}$, where $y_i$ is the TS class and $x_i$ is a vector representing the Rrs($\lambda$) values of the i-th instance. Examples of vector classification algorithms (classifiers) are decision trees, support vector machines, neural networks and k-nearest neighbours (Ham et al., 2005; Mou et al., 2017). The aim of vector classifiers is to learn statistically meaningful patterns of observations through the minimisation of a defined loss criterion (Vapnik, 1999). In practice, different statistical approaches have similarities because they learn the same properties for a given Rrs($\lambda$) vector. Each classifier model is the result of a statistical learning process generating unique class decision boundaries. Therefore, while one algorithm may fail to correctly predict the true TS class, another may succeed (Polley and van der Laan, 2011; Ting and Witten, 1999). The combination of individual learners is the baseline for ensemble learning. The idea explored here is to construct a strong single learner from several weak learners.

Two of the most popular ensemble methods are bagging (Breiman, 1996) and boosting (Freund and Schapire, 1996) combines the predictions of weak learners using different bootstrap samples of the training set. Boosting sequentially trains a series of weak learners with weighted versions of the training set based on the performance of previously constructed learners. Wolpert (1992) proposed a linear combination of individual models to form an ensemble and named it "stacked generalisation" also known as "stacking." van der Laan et al. (2007) have extended the original stacking approach with a cross-validation framework and coined it "Super Learner."

The classification framework presented here is based on the concept by van der Laan et al. (2007). In the context of this research, we call this higher-level classification algorithm the "meta-classifier". The meta-classifier acts as a meta instance, learning from the decisions of each individual model to predict TS classes for Rrs($\lambda$). Statistically, the meta-classifier generates classification predictions based on the optimal combination of learning algorithms that minimise the cross-validated loss and provide the highest prediction confidence (Ting and Witten, 1999; van der Laan et al., 2007). Here we exemplify direct classification of Rrs into TS. To cope with the optical diversity of inland and nearshore waters, we explore the concept of stacking classifiers in a meta-learning scheme. We evaluate the method on the multispectral resolution of the Sentinel-3A Ocean and Land Colour Instrument (OLCI) for 40 inland and nearshore sites. An established practice is to estimate TS through inversion of Rrs into chla, whereby multiple chla algorithms may be combined in an OWT scheme to address in-water optical complexity. The developed meta-classifier involves multiple classification algorithms, thus we assess the utility of our approach through comparison with TS derivation via OWT switching of several chla retrieval algorithms.

## 4.2 Methods

*In situ* bio-optical data were sourced from LIMNADES (Lake Bio-optical Measurements and Matchup Data for Remote Sensing: https://limnades.stir.ac.uk/). We assembled a subset of LIMNADES with 2751 samples of co-located *in situ* chla and hyperspectral Rrs($\lambda$) measurements. The datasets and measurement

***Table 4.1:*** *Inland and nearshore waters included in the training and test datasets.*

| Dataset name (according to | Number of observations (n) | | Systems | References |
|---|---|---|---|---|
| Training set (n = 2184) | | | | |
| CEDEX | 107 | | 26 Spanish reservoirs | (Ruiz-Verdú et al., 2008, 2005; Simis et al., 2007) |

| | | | | |
|---|---|---|---|---|
| CU, EC, IU, UNL-A, UNH | 1121 | | 90 U.S. waters | (Binding et al., 2008, 2010; 2013; Bradt, 2012b Dall'Olmo et al., 2005, 2003; Dall'Olmo and Gitelson, 2006; Gitelson et al., 2008; 2007; Gurlin et al., 2011; Li et al., 2015, 2013; Schalles, 2006; Schalles and Hladik, 2012) |
| UNL-B | 52 | | Lake Kinneret (Israel) | (Yacobi et al., 2011) |
| SYKE | 10 | | Three Finnish lakes | (Kallio et al., 2015) |
| CNR | 215 | | Five Italian lakes | (Bresciani et al., 2011; Giardino et al., 2015, 2014, 2013, 2005; Guanter et al., 2010; Manzo et al., 2015) |
| UCT | 56 | | Three South African reservoirs | (Matthews and Bernard, 2013) |
| CAS | 243 | | Lake Taihu (China) | (Zhang et al., 2010, 2007) |
| UT | 38 | | Lake Peipsi (Estonia) | (Kutser et al., 2013, 2012) |
| NIOO-KNAW | 198 | | Two Dutch lakes | (Guanter et al., 2010; Simis et al., 2005; 2007) |
| UTSU | 144 | | Six Japanese and Chinese lakes | (Jaelani et al., 2013; Matsushita |

| | | | | |
|---|---|---|---|---|
| | | | | et al., 2015; Yang et al., 2013) |
| Test set (n = 567) | | | | |
| CEDEX | 76 | | 17 Spanish reservoirs and two lakes | (Ruiz-Verdú et al., 2008, 2005; Simis et al., 2007) |
| UNH, UNL-A, UI, CU | 448 | | 28 U.S. waters | (Bradt, 2012; Dall'Olmo et al., 2005, 2003; Gitelson et al., 2008; Li et al., 2013; Moore et al., 2014; Schalles, 2006) |
| USTIR | 29 | | Lake Balaton (Hungary) | (Riddick et al., 2015) |
| UL | 14 | | Lake Bogoria (Kenya) | (Tebbs et al., 2013) |

protocols of OACs and the derivation of Rrs($\lambda$) are described in (Spyrakos et al., 2018)(see Table 4.1 herein). Rrs($\lambda$) in all datasets were measured just above the water surface ($0^+$). We did not apply an additional correction to Rrs($\lambda$), assuming the measurements were made under optimal viewing angles and quality controlled during the original study. For brevity, we omit the wavelength-dependency for the remainder of the paper.

For the development of the classification method two independent datasets were created: one for training the classification algorithms and one for evaluating their performance. A priority in the development process was to avoid the allocation of observations from the same water body to both training and test datasets. Mixing or randomising the *in situ* measurements across both datasets would introduce knowledge about the water bodies included in the test set to the classification algorithms in the training stage and prevent an independent evaluation of the proposed method. We therefore split the entire dataset using the first letter of the water bodies names: A-D (n = 567, 20%) for testing and E-Y (n = 2184, 80%) for training.

### 4.2.1 Radiometric data pre-processing

The *in situ* hyperspectral Rrs measurements from the training and test datasets were spectrally resampled to the multispectral band configuration of Sentinel-3A OLCI, normalised and subsequently assigned a TS class (Figure 4.1). These steps are detailed below. Resampling of the hyperspectral data was necessary to combine Rrs from multiple sources and sensors. The spectral data were convolved to the spectral response function (SRF) of OLCI because the observation

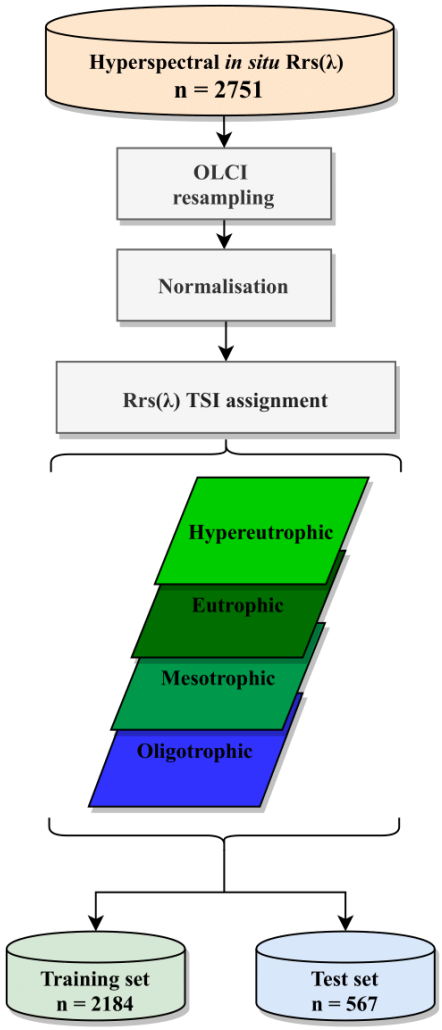


***Figure 4.1:*** *Pre-processing scheme of the entire dataset (n = 2751) resulting in independent training (n = 2184) and test (n = 567) sets.*

frequency, spectral sensitivity and resolution of this sensor are relevant to observe the TS of inland and nearshore waters (Kravitz et al. 2020). To convolve Rrs for each OLCI band *i*, we calculated the values of the spectral albedo for the bands:

$$\bar{R}(\lambda_i) = \frac{\int_{\lambda_1}^{\lambda_2} R(\lambda)\phi_i(\lambda)\,\mathrm{d}(\lambda)}{\int_{\lambda_1}^{\lambda_2} \phi_i(\lambda)\,\mathrm{d}(\lambda)}, \tag{4.3}$$

where ($\lambda$) is the wavelength, ($\lambda_i$) is the center wavelength in the *i*-th spectral band, $\phi_i(\lambda)$ is the SRF of the i-th spectral band, ($\lambda_1$, $\lambda_2$) are the boundary wavelengths of the considered spectral range, R is the spectral albedo and $\bar{R}(\lambda_i)$ is the mean spectral albedo in the *i*-th spectral band. The mean albedo values within the bands represent the spectral albedo seen by the sensor and are often also called the spectral signature of the viewed surface. We note that the convolution was performed on Rrs rather than the mathematically correct $L_w$ and $E_d$ measurements. Any effects are considered negligible at the 10-nm bandwidth of OLCI (Burggraaff, 2020).

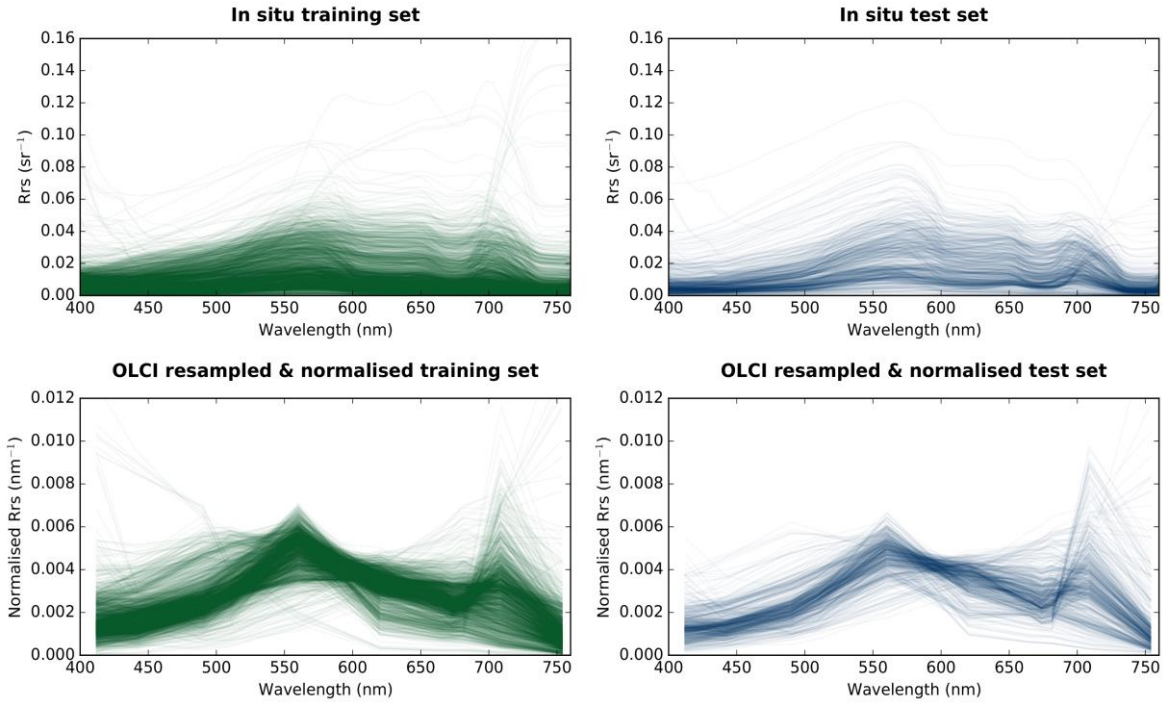


***Figure 4.2:*** *Hyperspectral in situ data (top row) and the resulting Sentinel-3A OLCI resampled and normalised multispectral reflectance spectra (bottom row) for both training (green) and test measurements (blue).*

Several previous optical classification studies classified untreated Rrs, which is sensitive to both amplitude and spectral shape (Lee et al., 2010; Moore et al., 2009, 2001). More recently, and particularly relevant to optically complex water bodies, the reflectance vector has been normalised prior to the classification(Hieronymi et al., 2017; Mélin et al., 2011; Spyrakos et al., 2018; Xi et al., 2017, 2015). The amplitude of Rrs is strongly influenced by particulate (back-) scattering, while the shape is primarily affected by $a_\phi(\lambda)$, $a_{\mathrm{cdom}}(\lambda)$ and the absorption by non-pigmented particles ($a_{\mathrm{nap}}(\lambda)$[1/m]) (Roesler et al., 1989).s In this study we followed the prior normalisation approach of Rrs to emphasise the shape:

$$r_n(\lambda) = \frac{Rrs}{\int_{\lambda_1}^{\lambda_2} Rrs(\lambda)\,\mathrm{d}(\lambda)}, \tag{4.4}$$

where $r_n$ (in units of $\mathrm{nm}^{-1}$) indicates the normalised spectrum obtained through trapezoidal integration between $\lambda_1$ (412 nm) and $\lambda_2$ (753 nm) (Mélin and Vantrepotte, 2015).

The resampled and normalised reflectance datasets are displayed in Figure 4.2. To enable the inversion for TS we established a relationship between Rrs and the TS classes. We used the TSI definition by Carlson (1977) and the extension made for hypereutrophic waters by Carlson and Simpson (1996). The TSI definition provides a logarithmic model to interpret chla concentrations as indicators for TS classes. We separated our Rrs dataset into four reflectance TS classes (oligo-, meso-, eutro- and hypereutrophic) based on the chla concentrations measured from co-located *in situ* water samples (Table 4.2). The compiled dataset covers a wide range of bio-optical conditions and trophic states found across inland and nearshore waters. Figure 4.3 displays the logarithmic distribution of chla [mg/m$^3$], total suspended matter (TSM [g/m$^3$]), and $a_{cdom}(443)$[1/m] for the training and test sets. The minimum and maximum values of the OACs are given in Table 4.3. The training set was used multiple times throughout the classification scheme, whereas the resultant classification algorithms were applied once to the test set for evaluation.

***Table 4.2:*** *TSI classification after Carlson and Simpson (1996) and assigned reflectance spectra per class in our training and test sets. n is the number of observations.*

| Class | TSI | Chla range [mg/m$^3$] | Water quality characteristics | Training (n = 2184) | Test (n = 567) |
|---|---|---|---|---|---|
| 1 | Oligotrophic | 0 – 2.6 | Low primary productivity and high oxygen in hypolimnion. | 356 | 36 |
| 2 | Mesotrophic | > 2.6 – 7.3 | Intermediate levels of productivity. | 328 | 92 |
| 3 | Eutrophic | > 7.3 – 56 | High biological productivity. | 1164 | 332 |
| 4 | Hypereutrophic | > 56 | Excessive biological productivity, algal blooms, low transparency and oxygen levels. | 336 | 107 |

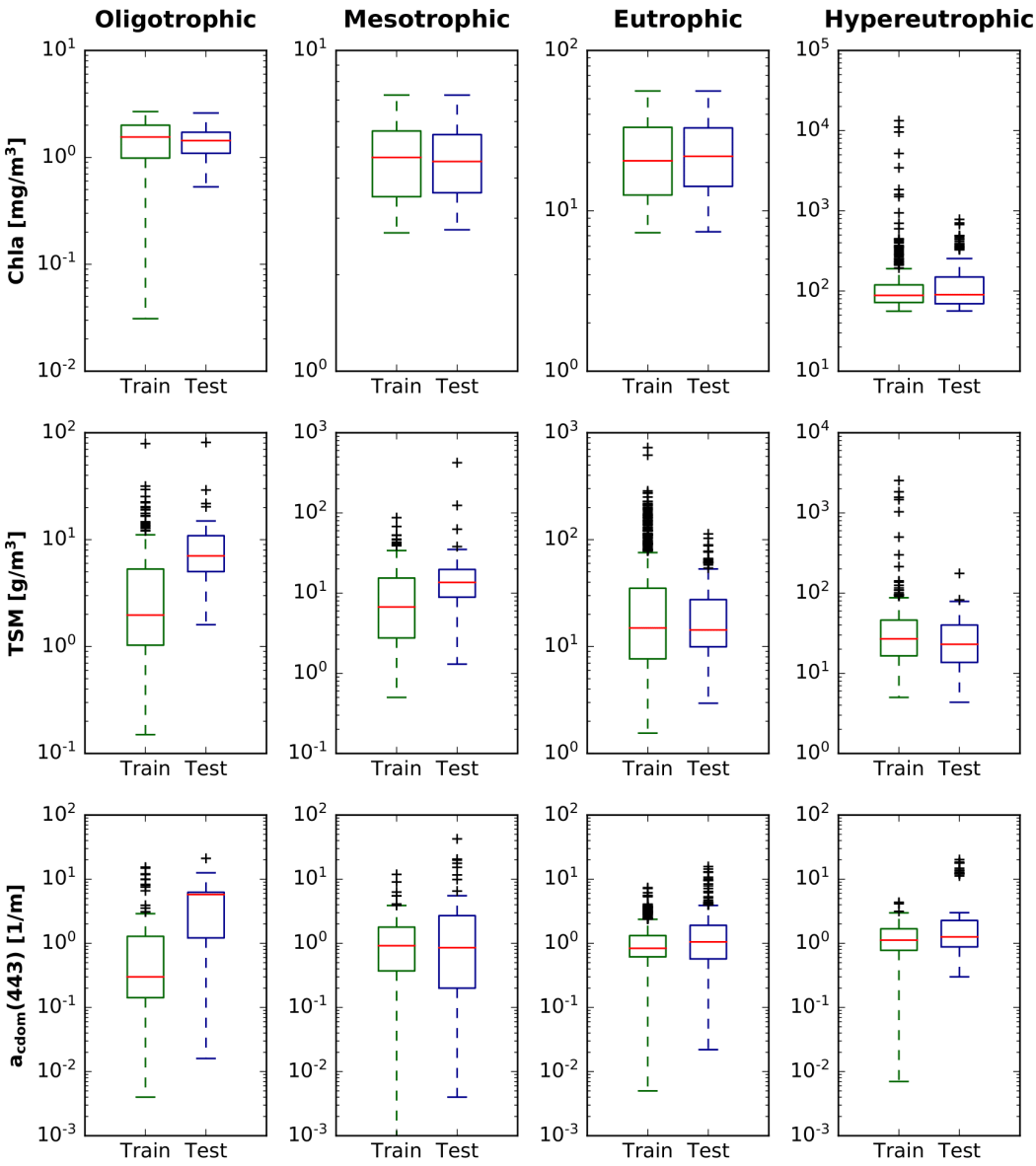


***Figure 4.3**: Logarithmic distribution of available Chla [mg/m$^3$] (n = 2751), TSM [g/m$^3$] (n = 1758) and $a_{cdom}$(443)[1/m] (n = 1754) samples in our dataset per TS class (green training, blue test).*

***Table 4.3:** Minimum and maximum values of the training and test set constituents.*

| | Chla train [mg/m$^3$] | Chla test [mg/m$^3$] | TSM train [g/m$^3$] | TSM test [g/m$^3$] | $a_{CDOM}(443)$ train [1/m] | $a_{CDOM}(443)$ test [1/m] |
|---|---|---|---|---|---|---|
| Minimum | 0.0031 | 0.53 | 0.15 | 1.3 | 0.001 | 0.004 |
| Minimum | 13296.7 | 782 | 2533.3 | 423 | 15.38 | 42.46 |

### 4.2.2 Meta-classification of remote sensing reflectance

In our dataset $D = \{(y_i,\ x_i), i = 1, \dots, N\}$, $y_i$ represents the trophic class values and $x_i$ the reflectance vector values of the i-th instance. To classify an instance, a library L with $k = 4$ base classification algorithms (base-classifiers), i.e. $L = \{k_1, k_2, \dots, k_4\}$, was created. The library is a collection of

vector classifiers. We invoked the k-th base-classifier in L to predict the class for each instance $x_i$, along with its true TS classification $y_i$. Combining these predictions along with the true trophic class vector led to a new dataset, the level-zero data. The level-zero dataset was treated as the training ground for a new learning problem subsequently solved by an additional classification algorithm, the meta-classifier.

#### 4.2.2.1 Base-classifiers

We trained the meta-classifier with the predictions of four base-classifiers characterised by different statistical assumptions: eXtreme Gradient Boosting (XGBoost), LightGBM (LGBM), Naïve Bayes (NB) and a neural network (NN). The classifier assumptions and the training procedure, involving the stacking of base-classifiers to fit the meta-classifier, are as follows.

The first two classifiers are XGBoost and LGBM (Chen and Guestrin 2016; Ke et al. 2017). Both classifiers have their statistical origin in gradient boosting machines (GBM) combined with decision trees as base-learners (GBDT) (Freund and Schapire, 1997; Friedman et al., 1998). GBDTs create new models sequentially to provide more accurate estimates of the target variable. The principle is to construct new learners that focus on weak areas already learnt, or in statistical terms, construct learners correlated with the negative gradient of the used loss function (Natekin and Knoll, 2013) (Natekin and Knoll 2013). For reviews on boosting algorithms see (Bühlmann and Hothorn, 2007; Schapire, 2003). The prediction of the XGBoost algorithm at each iteration t is based on the defined objective function $\hat{f}$:

$$\hat{f}^{(t)} = L(\theta) + \Omega(\theta), \tag{4.5}$$

where

$$L(\theta) = \sum_{i=1}^{n} \ell\left[y_i, \hat{y}_i^{(t)}\right], \tag{4.6}$$

and

$$\Omega(\theta) = \gamma T + \frac{1}{2}\lambda \sum_{j=1}^{T} w_j^2. \tag{4.7}$$

The objective function $\hat{f}^{(t)}$ consists of two parts: $L(\theta)$ and $\Omega(\theta)$. $\theta$ describes the parameters in the equation. $L(\theta)$ is a differentiable convex loss function that measures the difference (residual) between a class prediction $\hat{y}_i$ and $y_i$ at the t-th iteration. The goal of the optimisation process is to construct a tree structure that minimises a loss function in each iteration. The updated tree structure in each iteration learns from the previous tree's model decision and uses the residuals to fit a new residual tree. To construct models that generalise and avoid overfitting, Equation 4.5 denotes a regularisation term $\Omega(\theta)$. T in Equation 4.7 is the number of terminal nodes in a tree and $\gamma$ the learning rate (between 0 – 1). $\gamma$ is multiplied by T to enable tree pruning. Terminal nodes and the learning rate are hyper-parameters that we optimised in separate steps (see section 2.2.3). Compared

to classic GBM algorithms, XGBoost introduces the term $\frac{1}{2}\lambda\sum_{j=1}^{T} w_j^2$, where $\lambda$ is an additional regularisation parameter, and $w_j$ enables to control the weights of the tree leaves (Goodfellow et al., 2016). $\Omega(\theta)$ prevents overfitting and allows a better generalisation of the constructed model.

In this study we used the multi-class logarithmic loss function (mlogloss) for both XGBoost and LGBM. Mlogloss measures the performance of the models with an output probability value between 0 and 1 and increases when the predicted probability diverges from the actual class label:

$$\text{mlogloss} = -\frac{1}{N}\sum_{i}^{N}\sum_{j}^{M} y_{ij}\ln(p_{ij}), \tag{4.8}$$

where N is the number of observations, M the number of TS class labels, $y_{ij}$ a variable with the predicted class label and $p_{ij}$ is the classification probability output by the classifier for the i-th instance and the j-th label (Bishop, 1995; Hsieh, 2009). Solving Equation 4.8 becomes challenging and computationally demanding. Therefore, Equation 4.8 is transformed using a second-order Taylor expansion (Bishop, 2006). The transformation allows the objective function to depend only on the first and second order derivatives of a point in the loss function, also speeding up the process. The main difference between XGBoost and LGBM is the tree construction process. Both classifiers can capture highly non-linear feature-target class relationships. The models can be precisely controlled by tuning a set of hyper-parameters. In addition, each classifier can be trained on both small and large datasets making them suitable for application to any given classification task.

The third classifier in our ensemble is a Naïve Bayes (NB) probabilistic model based on the Bayes' Theorem:

$$P(C = y_i|X = x) = \frac{P(x|y_i)P(y_i)}{P(x)}, \tag{4.9}$$

where $P(y_i|x)$ is the conditional probability that a reflectance spectrum x belongs to a trophic class $y_i$. The Bayes' rule specifies how this conditional probability can be calculated from the features (wavelengths) of the reflectance vectors of each trophic class, and likewise the unconditional probability (Lewis, 1998). The NB classifier calculated the probability of each trophic class for a given reflectance and output the trophic class with the highest one. We specifically wanted to include the assumption that for some reflectance spectra independent, single wavelengths are dominant, and hence strongly influence the class assignment. Since the wavelengths of multispectral reflectance vectors are at least partly correlated, the NB assumption is naive. In our ensemble of classifiers, NB is one of several base-classifiers generating a TS prediction. Following the theory of stacked generalisation, the meta-classifier should recognise when the NB assumptions apply through evaluation of the predictions for each test reflectance. In case the NB assumptions hold, high prediction accuracies are expected and thus the meta-classifier could prioritise the NB predictions in the decision-making to generate a final TS estimate. In case the NB assumptions do not apply, the

accuracies will be low and lead to higher influence on the meta-classifier of other, more accurate base-classifiers.

As the fourth base-classifier we used a NN. NNs have shown success across a diverse set of waters due to their aptitude to approximate non-linear input-output functions (Brockmann et al., 2016; Doerffer and Schiller, 2007; Hieronymi et al., 2017; Ioannou et al., 2013; Krasnopolsky et al., 201, 2002). In this study, a NN with one layer and multiple hidden neurons $h_j$ was trained. The output of the NN for the test dataset was given by:

$$y = \hat{a} + \sum_{j=1}^{n} \widetilde{w}_j \cdot h_j \,, \tag{4.10}$$

with

$$h_j = \phi\big(a_j + w_j \cdot x_i\big), \tag{4.11}$$

where $x_i$ and $y$ are input and output vectors, respectively; $w_j$ and $\widetilde{w}_j$ are weights, $a_j$ and $\hat{a}$ are fitting parameters and $\phi$ is the non-linear hyperbolic tangent activation function (Hsieh, 2009). In Equation 4.10, *n* is the number of the non-linear activation function, $\phi$ in Equation 4.11. The defined objective function $\hat{f}^{(t)}$(Equation 4.5) for the NN minimised the mlogloss function (Equation 4.8) and the regularisation term $\Omega(\theta) = \sum_k w_k^2$, also known as weight decay. Mathematically, in each iteration k of the network, weights are pushed towards zero. The decay enables to avoid large weights that would cause the network to overfit on a given part of the input data. The NN was trained using backpropagation (Goodfellow et al., 2016). For the multi-class output layer, we used the standard softmax function (Bishop, 2006).

It is worth noting that the NN with one layer can be considered shallow, whereas it is becoming more common to use "deeper" NNs characterised by more layers. We could have added additional layers or used a more advanced architecture such as a mixture density network (MDN), as recently demonstrated for inland and coastal waters in (Pahlevan et al., 2020). However, the intent of this paper is to present meta-classification, and not to showcase various NN architectures. Further, a deeper NN increases training time and requires more *in situ* measurements which are naturally limited. For this application, we therefore opted to keep the NN architecture as basic as possible. Our library *L* consists of a diverse set of base-classifiers and underlying statistical models that forwarded the unique information learnt about the reflectance spectra, constituting each trophic class, to the meta-classifier.

#### 4.2.2.2 Meta-classifier

The meta-classification training and prediction procedures were multi-step processes, as we illustrate schematically in Figure 4.4. All classifiers were trained using a v-fold cross-validation scheme with v = 5. Cross-validation enables performance assessment of the classification algorithms during the training process (Schaffer, 1993). The process of training the meta-classifier on the predictions of the base-classifiers in our library $L = \{k_1, k_2, \dots, k_4\}$ was as following:

We split the reflectance training set into 5 exclusive folds of $n/v = 2184/5 \sim 437$.

For each fold $v = \{(v_i) = 1, \ldots, 5\}$:

a) Reflectance spectra in fold $v_i$ were validation data (hold-out set), while the remaining observations (80% of the reflectance spectra) constituted the training set. Each base-classifier was fit on the training set.

b) With each base-classifier we predicted the outcome $\hat{y}_i$ for each reflectance instance $x_i$ in a validation set $v_i$. The resulting loss of each base-classifier was estimated between the true outcome $y_i$ and its prediction $\hat{y}_i$ for all reflectance spectra.

c) For each classifier, the v-estimated loss rates over the v-validation sets were averaged resulting in the cross-validated loss. For each reflectance the model of the respective base-classifier with the smallest cross-validated loss was selected for subsequent use by the meta-classifier.

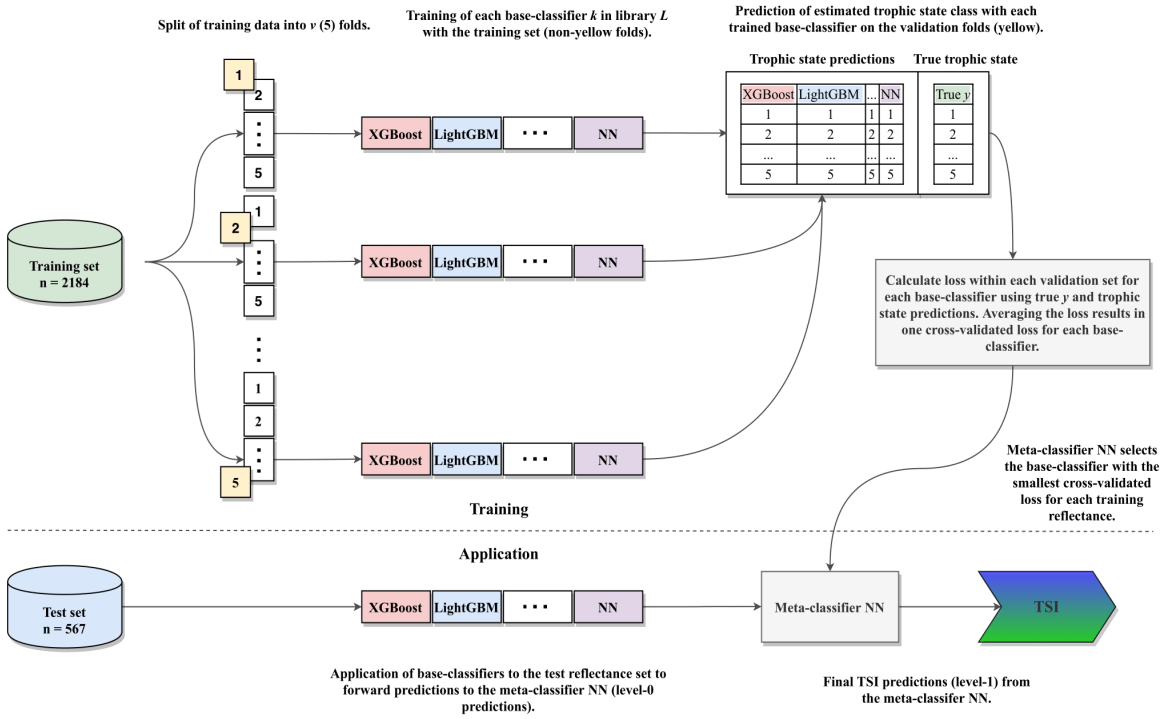


***Figure 4.4:*** *Schematic diagram of the training and application processes included in the meta-classification framework.*

Combined with the true trophic class $y_i$, we stacked the cross-validated predictions made on the training set of each base learner to generate a vector of level-zero predictions: $\text{pzero}_{\text{train}} = \{(p_i, x_i) = 1, \ldots, N\}$. This important step constituted the training set of the meta-classifier, where each feature in $\text{pzero}_{\text{train}}$ was a single prediction of the base-classifiers. For each prediction, we knew the true outcome $y_i$ and hence provided the meta-classifier the necessary training data. The meta-classifier learnt which of the base-classifiers predicted the true trophic class $y_i$ for each training reflectance. We used a separate NN as the meta-classifier. We selected a NN because of its high approximation capability to learn the non-linear decision boundaries necessary to distinguish between the base-

classifier predictions. The task of the meta-classifier NN was to select the most accurate base-classifier for each reflectance and assign a final trophic class. The training procedure of the meta-classifier required the level-zero predictions $\text{pzero}_{\text{train}}$. Overfitting of the meta-classifier on the training data was avoided through weight decay, that was applied in the same manner as for the base-learner NN. In the application, each base-classifier in *L* predicted a trophic class for each reflectance in our test set, resulting in a vector with level-zero test predictions: $\text{pzero}_{\text{test}} = \{(p_i, x_i) = 1, \dots, N\}$. These predictions were then stacked and provided to the meta-learner NN to estimate a final trophic class for each test reflectance. In this study we utilised the open-source ML-ensemble Python library that interfaces with Skicit-Learn (Flennerhag, 2017; Pedregosa et al., 2011).

### 4.2.2.3 Hyper-parameter optimisation

We optimised the learning process of the considered classifiers through hyper-parameter optimisation (HPO). Given a learner A of any of the base-classifiers k in our library L, hyper-parameters were exposed $\lambda\ x\ \Lambda$. Tuning hyper-parameters changed the way model A learnt to correctly classify training reflectance spectra in the dataset D. For example, a hyper-parameter of the base-classifiers XGBoost and LGBM limits the maximum depth of the constructed tree. Further, the NNs require a selection of neurons in a layer (as opposed to weights that are model parameters learnt during training). Mathematically, HPO can be represented as:

$$x^* = {}^{arg\ min}_{x} f(x), \tag{4.12}$$

where $f(x)$ is the objective function to minimise (or maximise) - such as the mlogloss - and $x^*$ is the lowest (or highest) value of a function for a set of hyper-parameters we have drawn from a domain $\mathcal{X}$. In practice, $\mathcal{X}$ was a previously defined grid of hyper-parameters. $f$ is a black-box function and has a large set of hyper-parameter combinations that are computationally costly to evaluate. The search that optimises $f$ is often either manually performed or accomplished by selecting randomly from a set of hyper-parameters. Another option is to search through a grid consisting of a substantial combination of all possible hyper-parameter configurations (Bergstra and Bengio, 2012; Bergstra et al., 2011; Thornton et al., 2012). Because our meta-classification approach involved the training of classifiers with several hyper-parameters, a manual, random or grid search approach was considered unpractical. These search approaches are time intensive and susceptible to missing an optimal hyper-parameter configuration. In this study, we instead followed the concept of Bayesian optimization (Jones et al., 1998; Streltsov and Vakili, 1999). As in the Naïve Bayes classifier, Bayesian optimisation is based on the Bayes' Theorem stating that the posterior probability (or hypothesis) *M* of a learner (or model) *A* given data points *D* is proportional to the likelihood of *D* given *M* multiplied by the prior probability of *M*:

$$P(M|D) = \frac{P(D|M)P(M)}{P(D)}. \tag{4.13}$$

Bayesian optimisation methods can be understood as a sequence process that builds a probabilistic model by keeping track of past evaluation results (Brochu et al., 2010). A probabilistic model is build by mapping hyper-parameters to a probability of a score on the objective function $f$. One can represent this model as $P(C|x)$, where $C$ is the score for each hyper-parameter $x$. In literature, the model $P(C|x)$ is called utility (or surrogate) function $u$. In this study, we built the model with a Gaussian Process (GP) as the prior probability model on $f$ (Rasmussen and Williams, 2005). GPs have become a standard in Bayesian optimization (Martinez-Cantin, 2014; Snoek et al., 2012). The surrogate function $u$ was then optimised for and based on the posterior distribution for sampling the next set of hyper-parameters ($x_{t+1}$):

$$x^* = {}^{arg\,min}_{x} f(x), \tag{4.14}$$

To find the next best point to sample $f$ from, a point was chosen that maximised an acquisition (or selection) function, here the expected improvement (EI):

$$\mathrm{EI}(x) = \mathbb{E}[\max\{0, f(x) - f(\hat{x}\}], \tag{4.15}$$

where $\hat{x}$ is the current optimal set of hyper-parameters. Maximising $\hat{x}$ was the objective to improve upon $f$ most. $f$ was continually evaluated against the true objective function until a defined maximum of iterations was reached. By continually updating the surrogate probability model, Bayesian reasoning led to reliable results. The next set of hyper-parameters was selected based on the previous performance history instead of a costly grid search through all possible hyper-parameter combinations. Several libraries exist that implement Bayesian optimisation. Here we used Scikit-Optimize, as it was built on top of Scikit-Learn (Head et al., 2018).

#### 4.2.2.4 Optical water type switching to derive trophic status

To assess the performance of the developed meta-classifier, we compare it against derivation of TS through OWT switching of chla retrieval algorithms (see Figure 4.5).

Our OWT switching approach is based on three previous studies. First, the OWTs for all Rrs in our training and test datasets are available from (Spyrakos et al., 2018) (Figure 4.6). Second, our dataset is almost identical (98% common) to the one used in the study by Neil et al. (2019) (n = 2807, compared to n = 2751 herein). Neil et al. (2019) assessed the performance of 48 chla algorithms on their dataset, resulting in one best-performing algorithm per OWT (see Table 5 therein).

Since the datasets of the two studies are nearly identical, the performance results of Neil et al., (2019) are considered valid for the dataset of the present study. Third, Neil et al. (2019) recommended chla algorithms for groups of OWTs when applied to independent, unknown data (such as the test set herein). We slightly modified the selection of algorithms, based on recent performance evaluations from the European Copernicus Global Land Service (Simis et al., 2020). Four chla algorithms were thus assigned to groups of OWTs (Table 4.4). Restriction of the OWT scheme to four chla algorithms increases the quality of the exercised comparison, as the meta-classifier is equally based on four base-learners.

For the retrieval of chla through OWTs two approaches exist. The first approach uses the most dominant / highest OWT membership score a Rrs received in the clustering process performed in (Spyrakos et al., 2018). Chla is then retrieved with an assigned algorithm per OWT. In the present study we refer to this approach as OWT switching of chla algorithms. The second approach utilises the highest n OWT membership scores per Rrs to retrieve chla with n algorithms. The n retrieved chla values are then weighted to reflect the OWT membership scoring, resulting in a blended chla value (Moore et al., 2014).

The blending procedure varies depending on the value of n and the definition of the weighting function. Since the largest impact on the chla retrieval originates from the algorithm chosen for each OWT, we simplified the process and utilised the highest OWT membership score per Rrs. The meta-classifier was trained with 80% observations of the overall dataset. The coefficients of the chla algorithms included in the switching scheme were thus re-calibrated solely with measurements included in the respective OWT group of the training data set. For example, in our OWT switching scheme, the OC2 algorithm was assigned for OWTs 3, 9, 10, 13, which combined constitute 511 observations in the training set. Using these OWT group measurements of the training set, the

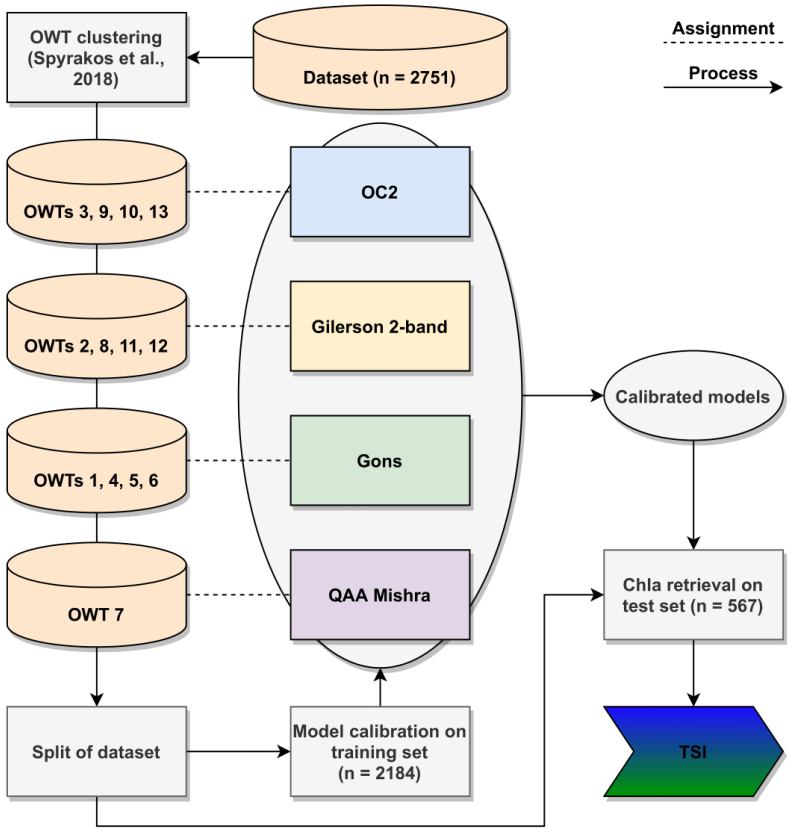


***Figure 4.5:*** *OWT switching scheme of chla algorithms to derive TS. OWT clustering of the dataset was performed in Spyrakos et al. (2018). Chla algorithm selection was based on benchmark results from Neil et al. (2019) for this dataset and modifications undertaken in Simis et al. (2020). Each group of OWTs was assigned one chla algorithm. Algorithm coefficient calibration was performed on the respective OWT group training data and the re-calibrated algorithms were applied to the test observations of the respective OWT group. TS was derived from the retrieved chla value based on the TS class ranges defined in Table 4.2.*

coefficients of the OC2 fourth-order polynomial were estimated using non-linear least squares fitting. We did not re-calibrate the coefficients of the Gons and Gilerson 2-band algorithms, as the number of required chla-specific absorption ($a_\phi^*(\lambda)[\mathrm{m}^2\ \mathrm{g}^{-1}]$ measurements included in the respective OWT training groups was low and thus insufficient for this purpose. Chla from $a_\phi(665)[1/\mathrm{m}]$ retrieved by QAA Mishra was estimated using the equation by (Bricaud et al. 1998):

$$Chla\_QAA\_Mishra = \left(\frac{a_\phi(665)}{\boldsymbol{a}}\right)^{\frac{1}{\boldsymbol{b}}}, \tag{4.16}$$

where a and b were calibrated with training data included in OWT group 7. Unlike the meta-classifier that operated on normalised Rrs (see section 2.1), the chla retrieval algorithms were applied to non-normalised Rrs at corresponding OLCI wavelengths. Each algorithm retrieved chla for the test Rrs measurements contained in the assigned group of OWTs. TS was subsequently derived from the retrieved chla concentration according to the TS class ranges depicted in Table 4.2.

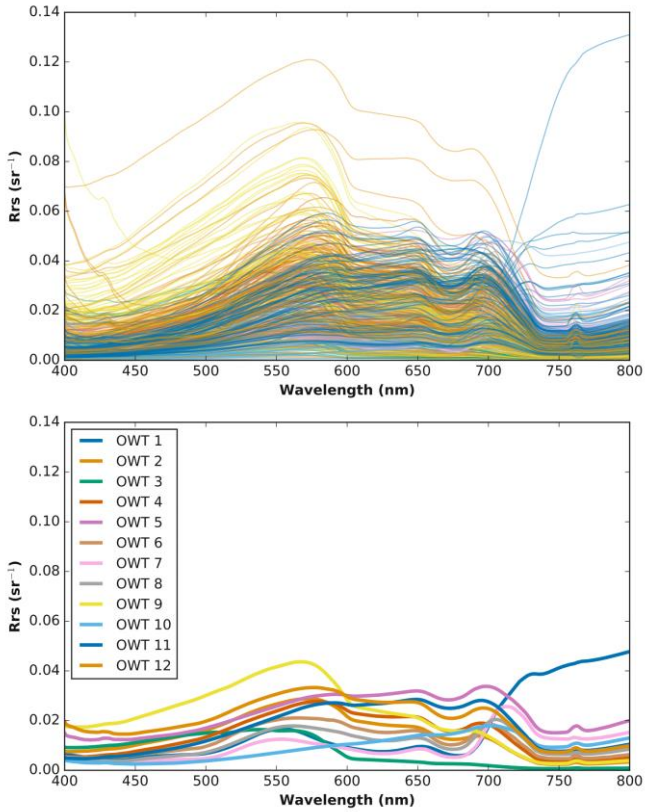


***Figure 4.6:*** *Hyperspectral Rrs (top, n = 567) and clustered OWTs (bottom, n = 12) of the test dataset.*

### 4.2.2.5 Performance evaluation

To evaluate the base-classifiers independently of the meta-classifier and vice versa, we compared them to a separate support vector machine (SVM) classifier (Cortes and Vapnik, 1995). Identical to the base-classifiers, we used the same training and test sets and procedures to train the SVM.

For the evaluation of TS classifications either through meta-classification or derived via conventional chla retrieval, we calculated the following metrics:

Overall Accuracy (OA). Accuracy of all reflectance instances $x$ correctly classified into each of the four TS classes $M$, divided by the total number of test samples (n = 567):

$$OA = \sum_{i=1}^{M} x\,/n. \tag{4.17}$$

Average Accuracy (AA). Average classification accuracy across all four trophic classes:

$$AA = \sum_{i=1}^{M} x\,/4. \tag{4.18}$$

Kappa Coefficient (Kappa). Percentage agreement corrected by the level of agreement that could be expected due to chance alone:

$$\kappa = (p_0 - p_e)/(1 - p_e), \tag{4.19}$$

where $p_0$ was the accuracy and $p_e$ was the probability of agreement by chance (Cohen, 1960; Congalton, 1991).

OA and AA are not equal because each trophic class holds a different number of samples. Because of the dataset split procedure, the dataset suffers from an imbalance between the classes (see Table 4.2). Using OA alone lacks precision because the smaller number of samples in the trophic classes 1 and 2 have less impact on the final accuracy score than class 3 (eutrophic). Hence, we calculated AA for all classification models. Because the eutrophic class has the highest amount of samples, large differences between OA and AA may indicate a biased classification model. We included the Kappa coefficient to estimate the probability of a correct class assignment by chance alone. In the comparison of the meta-classifier against TS derived via OWT switching, we evaluated the results of the regression of chla retrieved from an algorithm (estimated (E)) versus the *in situ* chla values (observation (O)). We assessed the residuals of $E_i - O_i$ with log-transformed metrics, as they enable a robust assessment of the algorithms over large concentration ranges of chla (O'Reilly and Werdell, 2019; Seegers et al., 2018).

Bias, which quantifies the average difference between estimated chla and the observed *in situ* value and is robust to systematic errors produced by an algorithm:

$$\text{Bias} = 10^{Y} \text{ where } Y = \left[\frac{\sum_{i=1}^{\mathrm{n}} \log_{10}(E_i) - \log_{10}(O_i)}{\mathrm{n}}\right]. \tag{4.20}$$

Mean Absolute Error (MAE), which captures the error magnitude accurately but can be sensitive to outliers:

$$\text{MAE} = 10^{Z} \text{ where } Z = \left[\frac{\sum_{i=1}^{\mathrm{n}} |\log_{10}(E_i) - \log_{10}(O_i)|}{\mathrm{n}}\right]. \tag{4.21}$$

Median Absolute Percentage Error (MdAPE), which is outlier-resistant. For each sample $(i)$:

$$\text{MdAPE} = 100 \times \tilde{\mathrm{x}}$$

$$\text{where } \tilde{x} \text{ is the median of } \left[\frac{|E_i - O_i|}{O_i}\right]. \quad (4.22)$$

Additionally, we provide the slope of the linear regression fit to enable comparisons with previously published results. We omit the coefficient of determination ($r^2$) as it lacks a response to bias and is sensitive to outliers and thus subject to false interpretation (Seegers et al., 2018).

***Table 4.4:*** *Chlorophyll-a algorithms included in the OWT switching scheme. Calibration coefficients for each model highlighted in bold.*

| OWTs | Algorithm | References | Equation | a | b | c | d | e |
|---|---|---|---|---|---|---|---|---|
| 3, 9, 10, 13 | OC2 | (O'Reilly et al., 1998; O'Reilly and Werdell, 2019) | $CHLA\_OC = 10^{(a+bX+cX^2+dX^2+eX^2)}$ X = $log_{10}\left(\frac{R_{rs}490}{R_{rs}560}\right)$ | -0.0087 | -1.9803 | 5.0867 | 1.0043 | -15.7660 |
| 2, 8, 11, 12 | Gilerson 2-band | (Gilerson et al., 2010) | $Chla\_Gil = [35.75\ x\ \frac{R_{rs}708}{R_{rs}665} - 19.30]^{1.124}$ | | | | | |
| 1, 4, 5, 6 | Gons | (Gons et al., 2008, 2005, 2002) | Chla_Gons = $[(\frac{R_{rs}708}{R_{rs}665})\ x\ (0.7 + b_b) - 0.4 - b_b^a]\ /\ b$ | 0.016 | | | | |
| 7 | QAA Mishra | (Mishra and Mishra, 2014) | $a_\phi(665) = a(665) - a_w(665) - a_{cdom}(665)$ $a_{cdom}(665) = a_{cdom}(443)^{(-a(665-443))}$ | 0.0138 | | | | |

## 4.3 Results

### 4.3.1 Meta-classification

For the meta-classifier LGBM and XGBoost were base-learners. LGBM marginally outperformed XGBoost for all TS classes (Table 4.5). The overall accuracies (OA) were 80.56% and 79.15% for LGBM and XGBoost, respectively (Figure 4.7). Both classified a high proportion of the eutro- and hypereutrophic systems (classes 3 and 4) correctly (86.14% and 91.59% for LGBM and 85.54% and 89.72% for XGBoost).

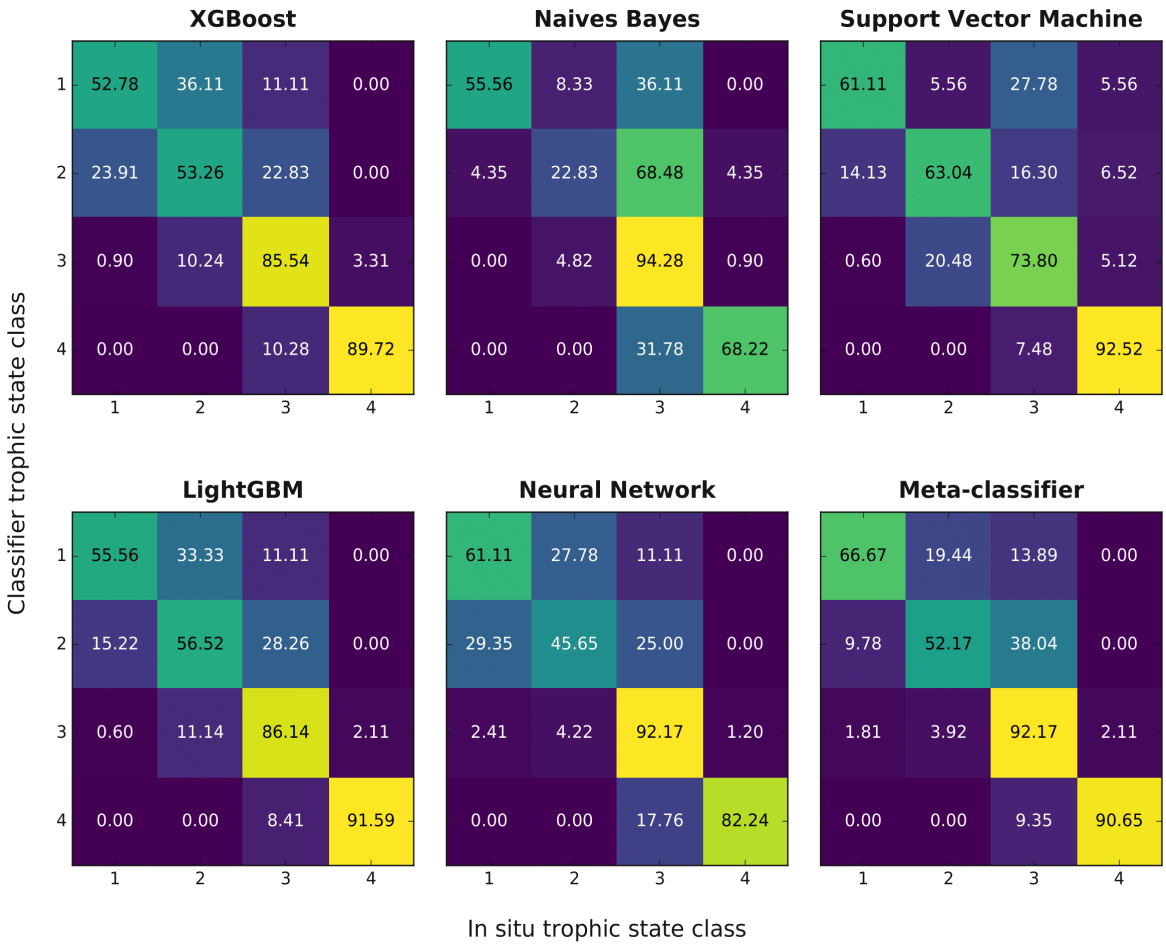


***Figure 4.7**: Classification matrices for predictions made by all classifiers on the independent test set (n = 567). The percentage of reflectance spectra assigned per TS class is shown. Yellow colours indicate high, purple colours low percentages per classifier. TS classes are denoted as 1 = Oligotrophic, 2 = Mesotrophic, 3 = Eutrophic, 4 = Hypereutrophic.*

***Table 4.5:** Classification accuracies of the different classifiers for the test set shown in percentages. The highest accuracy in each row is shown in bold.*

| Class | Class Name | XGBoost | LightGBM | NN | NB | SVM | Meta-classifier |
|---|---|---|---|---|---|---|---|
| 1 | Oligotrophic | 52.77 | 55.55 | 61.11 | 55.55 | 61.11 | **66.66** |
| 2 | Mesotrophic | 53.26 | 56.52 | 45.65 | 22.82 | **63.04** | 52.17 |
| 3 | Eutrophic | 85.54 | 86.14 | 92.16 | **94.27** | 73.79 | 92.16 |
| 4 | Hypereutrophic | 89.71 | 91.58 | 82.24 | 68.22 | **92.52** | 90.65 |
| OA | - | 79.15 | 80.56 | 80.91 | 75.44 | 74.91 | **83.92** |
| AA | - | 70.32 | 72.45 | 70.29 | 60.22 | 72.61 | **75.41** |
| Kappa | - | 65.10 | 67.18 | 66.88 | 52.73 | 60.02 | **71.72** |

Both classifiers achieved moderate classification accuracies for the oligo- and mesotrophic classes (classes 1 and 2). For the oligotrophic class, 33.33% and 36.11% misclassifications occurred (assigned mesotrophic) for LGBM and XGBoost, respectively. For mesotrophic systems, LGBM misclassified 28.26% and XGBoost 22.83% of observations as eutrophic. LGBM and XGBoost performed similarly, with slightly higher prediction accuracies by LGBM across all classes. Differences were primarily due their distinct approaches to build the decision trees. Both models constituted balanced base-learners, without major prediction failures for any of the TS classes.

Out of all the base-classifiers, NB classifications had the highest variance. While 55.56% of observations were correctly assigned oligotrophic (similar to the performances of XGBoost and LGBM), NB performed poorly (22.83%) for the mesotrophic class, predicting most water systems as eutrophic (68.48%). In contrast, NB classified 94.28% of the eutrophic observations accurately, outperforming all classifiers for this TS class. Although the other classes were not predicted with high precision, NB reached an OA score of 75.44%. This high OA score can be explained by the higher number of test samples in the eutrophic class 3 (n = 332) and the disproportionate impact on the overall accuracy metric. Consequently, AA becomes a more relevant metric because it incorporates the imbalanced dataset distribution. The NB AA score was approximately 15% lower at 60.22%. NB assumptions only applied to eutrophic and partially hypereutrophic waters (68.22%). For oligotrophic waters, the NB classifier performed comparable to the other classifiers.

The base-classifier NN achieved the highest accuracies for oligotrophic systems (61.11%). Compared to LGBM and XGBoost, the results were inferior for mesotrophic (45.65%) and hypereutrophic waters (82.24%). However, as for the other classifiers, the NN scored high accuracies for eutrophic waters (92.17%). Whereas for oligotrophic and eutrophic waters the prediction accuracies by the NN were competitive, the model's predictions were not balanced across the entire set of TS classes. It is to observe that higher accuracies for the oligotrophic class were accompanied by lower precision for mesotrophic waters. Similarly, the eutrophic class was retrieved with high precision, whereas less accurate predictions for hypereutrophic waters were made. Because the NN in this study is considered shallow, adding depth to the architecture may stabilise the predictions made across the TS classes. Therefore, more thorough experiments with different NN architectures need to be undertaken than they were exercised in this study. For exemplification of the meta-classifier concept, the NN sufficed to add meaningful information to the ensemble of base-learners.

The non-base SVM classifier scored the highest accuracy for mesotrophic waters (63.04%, 6.52% more accurate than the highest base-learner prediction by LGBM (56.52%)) and hypereutrophic waters (0.94% compared to LGBM predictions). SVM predictions were 10.87% and 1.87% more accurate than from the meta-classifier for these two classes, which in sum represents a significant performance gain. However, the SVM misclassified a large proportion of the eutrophic class (73.79% compared to 92.16% by the meta-classifier), reducing all performance metrics significantly.

In the present study, the SVM functioned as a standalone comparison model which therefore could not be incorporated into the ensemble of base-learners. However, given the performance gains of the SVM over other base-learners for mesotrophic and hypereutrophic waters, the addition of the SVM to the ensemble should be investigated. Before adding the SVM, it needs to be clarified whether the eutrophic misclassifications are a primary consequence of the model's more accurate mesotrophic and hypereutrophic classifications. If included, it is furthermore important to validate if the lower accuracies for eutrophic waters can be accurately handled by the meta-classifier without an overall performance loss for this class. Only if the meta-classifier can discard misclassifications accurately, performance gains of the SVM for mesotrophic and hypereutrophic waters would improve overall meta-classifier accuracies. Inclusion of the SVM into the ensemble of base-learners would also require to re-train the meta-classifier NN.

The meta-classifier achieved the highest classification accuracies across all performance metrics (OA: 83.92%, AA: 75.41%, Kappa: 71.72%) and the oligotrophic class 1 (66.67%). In comparison, the base-classifiers' average accuracy for oligotrophic waters was 56.25%. The meta-classifier improved on this score by 10.42%. Compared to the oligotrophic class, the meta-classifier did not improve over the most accurate base-classifiers for mesotrophic waters. The decision-making process of the meta-classifier to prioritise a reliable base-classifier became increasingly complex in the case of strongly differing base-classifier predictions. For mesotrophic waters, the meta-classifier had to discard the poor performing base-classifiers NB (22.83%) and NN (45.65%) in favour of the more accurate XGBoost (53.26%) and LGBM (56.52%). The meta-classifier was not able to entirely dismiss the NB and NN classifiers compared to the most reliable performance achieved by the base-learner LGBM. Despite the poor performances by NB and the NN, the meta-classifier scored 52.17% prediction accuracy for mesotrophic waters. Since the selection of a base-classifier for a reflectance was learnt using the training data, choosing incorrect classifiers for a reflectance of the test set was an expected outcome in heterogeneous classification scenarios.

In contrast, the meta-classifier generated highly accurate results for eutrophic and hypereutrophic waters (92.16% and 90.65%, respectively), which were significantly higher than for the oligo- and mesotrophic classes. Confusion by the meta-classifier between these two classes was below 10%.

### 4.3.2 Optical water type switching of chla algorithms

Figure 4.8 illustrates the performances of the chla retrieval algorithms included in the OWT switching scheme for the assigned group of OWTs on the test dataset. The observations contained in the OWT groups of the Gons and Gilerson 2-band algorithms represent 78% of the test set (442/567). The Gons algorithm retrieved chla with low MAE $< 0.33$ [mg/m$^3$], MdAPE $< 20\%$ and a small negative bias (-0.04). Higher residual errors (MAE of 0.61 [mg/m$^3$] and MdAPE of 37.21%) were produced by the Gilerson 2-band algorithm for the respective observations included in the OWT group test set.

QAA Mishra overestimated the values of extremely high chla, turbid waters (positive bias (0.52)). Nonetheless, the overestimation had no negative impact on the TS class assignment, as all retrieved chla values were > 100 [mg/m$^3$] and thus the TS class assignment was 100% accurate (chla values > 56 [mg/m$^3$] are hypereutrophic). The impact of QAA Mishra on the overall hypereutrophic class was low, as there were only 24 measurements included in the OWT 7 test set.

Out of the four chla algorithms, the OC2 algorithm performed accurately for low chla concentrations, but struggled to predict higher chla waters (MAE of 0.79 [mg/m$^3$], negative bias (-0.19)). The OC2 stagnated at approximately 14 mg/m$^3$ of chla, which can be explained by the limit of the polynomial (expressed by its coefficients) to calculate higher chla concentrations. The same stagnation can be observed in the algorithm's application to the training data it was calibrated with (grey hexagons in Figure 4.8, metrics not shown).

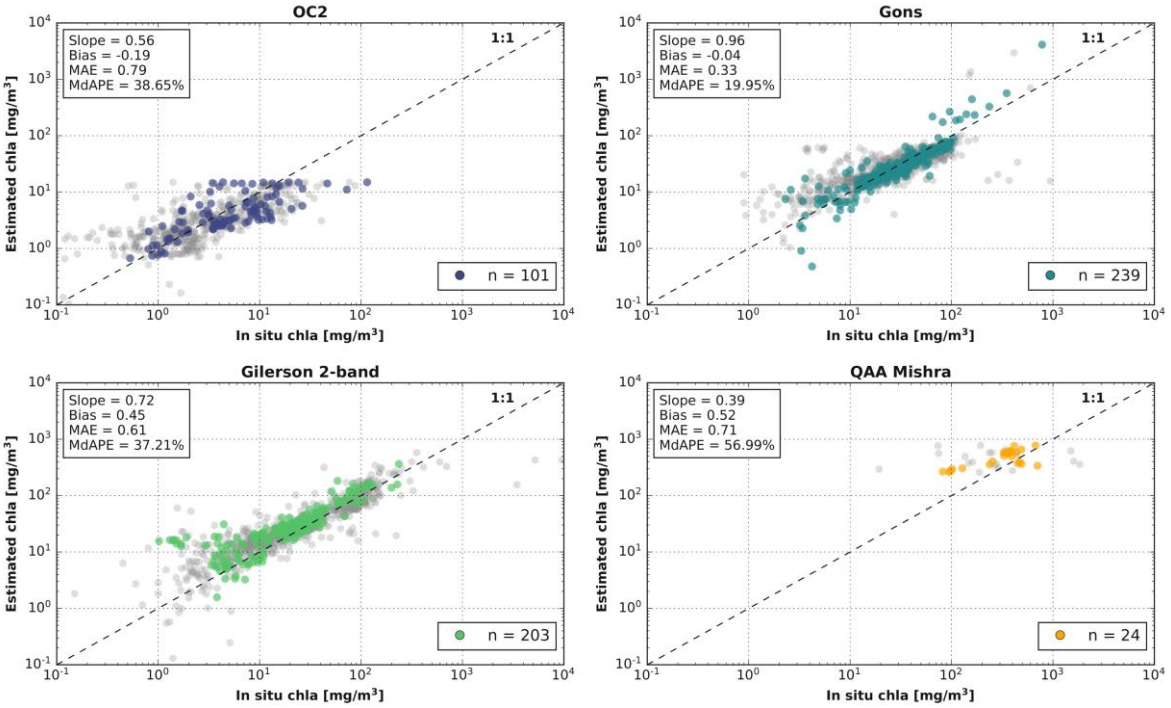


***Figure 4.8:*** *Performance evaluation of chla retrieval algorithms included in the OWT switching scheme: OC2, Gons, Gilerson 2-band and QAA Mishra. Coloured circles represent algorithm retrievals for measurements included in the respective OWT test groups (n = 567). For illustrative purposes, grey hexagons represent algorithm retrievals for the respective OWT training groups, that the algorithms were calibrated with. Metrics are shown for test data.*

The failure of the OC2 algorithm to retrieve chla accurately for higher concentrations is an expected outcome, since the OC2 algorithm was designed for clear waters, where phytoplankton dominates the optical properties. The retrieval result of OC2 indicates that the blue and green areas of the test Rrs were not only changing as a function of phytoplankton. Thus, the blue/green ratio of the OC2 algorithm led to inaccurate retrievals.

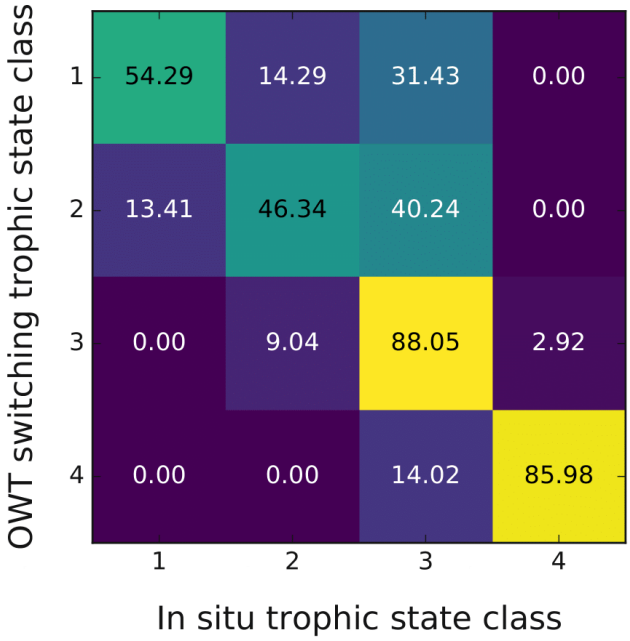


***Figure 4.9:*** *Classification matrix for TS predictions on the independent test set (n = 567) derived from OWT switching of chla algorithms. The percentage of reflectance spectra assigned per TS class is shown. Yellow colours indicate high, purple colours low percentages. TS classes are denoted as 1 = Oligotrophic, 2 = Mesotrophic, 3 = Eutrophic, 4 = Hypereutrophic.*

TS derivation following the chla retrieval through OWT switching is shown in Figure 4.9. The accuracy of the OWT chla algorithm switching approach to derive TS for the test dataset was 79.54% (OA), 68.66% (AA) and 63.38 % (Kappa). As for the meta-classifier, the largest errors occurred for oligo- and mesotrophic waters, whereas the retrieval was highly accurate for eutro- and hypereutrophic waters (> 85% accuracy). For oligotrophic waters, the meta-classifier was 12.38% more accurate (66.67% versus 54.29%, respectively) and 5.83% more precise in the classification of mesotrophic waters than derived through OWT switching of chla algorithms (52.17% and 46.34%, respectively). Our classification approach was slightly more accurate for eutrophic (4.12%) and hypereutrophic waters (4.67%). Using the AA metric that incorporates the imbalance of samples per TS class, the meta-classifier was on average 6.75% more accurate than the OWT switching of chla algorithms (75.41% and 68.66%, respectively).

### 4.3.3 Misclassifications of oligo- and mesotrophic classes

Both the meta-classifier and OWT switching scheme misclassified a high percentage of oligo- and mesotrophic reflectance spectra. Here we investigate the misclassifications of the meta-classifier that are higher for both classes when derived through OWT switching of chla algorithms. The meta-classifier misclassified 19.44% reflectance spectra of the oligotrophic class as mesotrophic and 38.04% reflectance spectra of the mesotrophic class were falsely classified as eutrophic (see Figure 4.7). None of the oligotrophic and mesotrophic test waters were misclassified as hypereutrophic. To investigate the misclassifications, we plotted the distributions of the OACs per TS class of the training and test sets (Figure 4.10). Based on the TS definition and the split of

measurements into training and test sets after each Rrs was assigned a TS class, the two datasets showed almost identical chla concentrations within each class. Greater variation occurred only in the hypereutrophic class for which a maximum chla [mg/m$^3$] concentration was not defined. In contrast, TSM [g/m$^3$] concentrations and $a_{\mathrm{cdom}}(443)$[1/m] strongly varied between the oligo- and mesotrophic classes. Since chla concentrations were low for both the oligo- and mesotrophic classes, TSM was dominated by inorganic particle loads, leading to highly turbid and strongly scattering water properties.

Based on the constituent medians of the OACs, the optical properties of the oligotrophic class in the training set were mostly dominated by phytoplankton chla, as $a_{\mathrm{cdom}}(443)$[1/m] (0.30) and TSM [g/m$^3$] (1.97) induced scattering concentrations were low. In contrast, the oligotrophic test set was characterised by high $a_{\mathrm{cdom}}(443)$[1/m] (5.77) and turbid waters with high TSM [g/m$^3$] concentrations (7.04).

For the mesotrophic class, the meta-classifier assigned 35 out of 92 test reflectance spectra to the eutrophic class (38.04%). The medians of chla [mg/m$^3$] and $a_{\mathrm{cdom}}(443)$[1/m] for all reflectance spectra in this class were comparable between the training and test datasets, however the test TSM concentration was twice as high as the training dataset counterpart (13.6 [g/m$^3$] in the test set compared to 6.71 [g/m$^3$] in the training set). For the 35 misclassified reflectance spectra this difference in TSM persisted with a median of 12.78 [g/m$^3$].

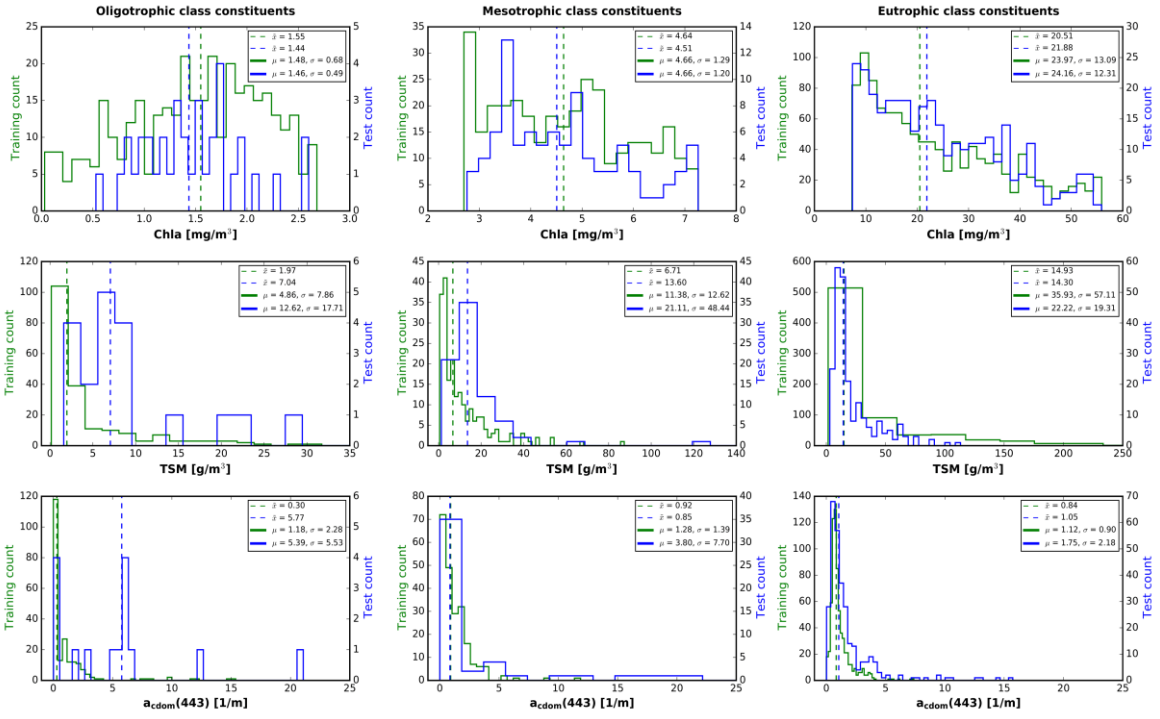


***Figure 4.10:*** *Histograms of Chla [mg/m$^3$], TSM [g/m$^3$] and $a_{cdom}(443)$[1/m] measurements included in the training (green) and test (blue) sets of the oligotrophic, mesotrophic and eutrophic classes. Dashed lines indicate the class median ($\tilde{x}$) of the parameter, $\mu$ and $\sigma$ the mean and standard deviation, respectively.*

The misclassified reflectance spectra of both the oligotrophic and mesotrophic waters reflect the influence of high sediment loads (Figure 4.11). The Rrs vectors of misclassified oligotrophic instances (19.44% as mesotrophic and 13.89% as eutrophic) do not reflect a significant reduction in Rrs values at 560 nm to 620 nm that characterises correctly assigned oligotrophic class observations. Moreover, misclassifications show high reflectance values in the red to near-infrared part of the spectrum. The reflectance spectra are similar in shape and magnitude compared to the training data of the mesotrophic and eutrophic waters. A comparable pattern can be observed for the 35 misclassifications of the mesotrophic class (classified as eutrophic), wherein both shape and magnitude are similar to the training vectors of the eutrophic class.

The reflectance spectra contained in the test sets of the two lowest TS classes were influenced by higher concentrations of absorbing $a_{\mathrm{cdom}}(\lambda)$ and/or concentrations of scattering particles than represented in the provided training data. Consequently, the corresponding Rrs vectors were substantially less present in the training sets, which influenced the learning of the classifier. Without appropriate representation of these waters, the classifies were unable to adjust their class decision boundaries accordingly. For the classifiers in the training stage, the corresponding Rrs vectors were more similar to those abundant in higher trophic classes, which consequently led to incorrect TS predictions on the test set.

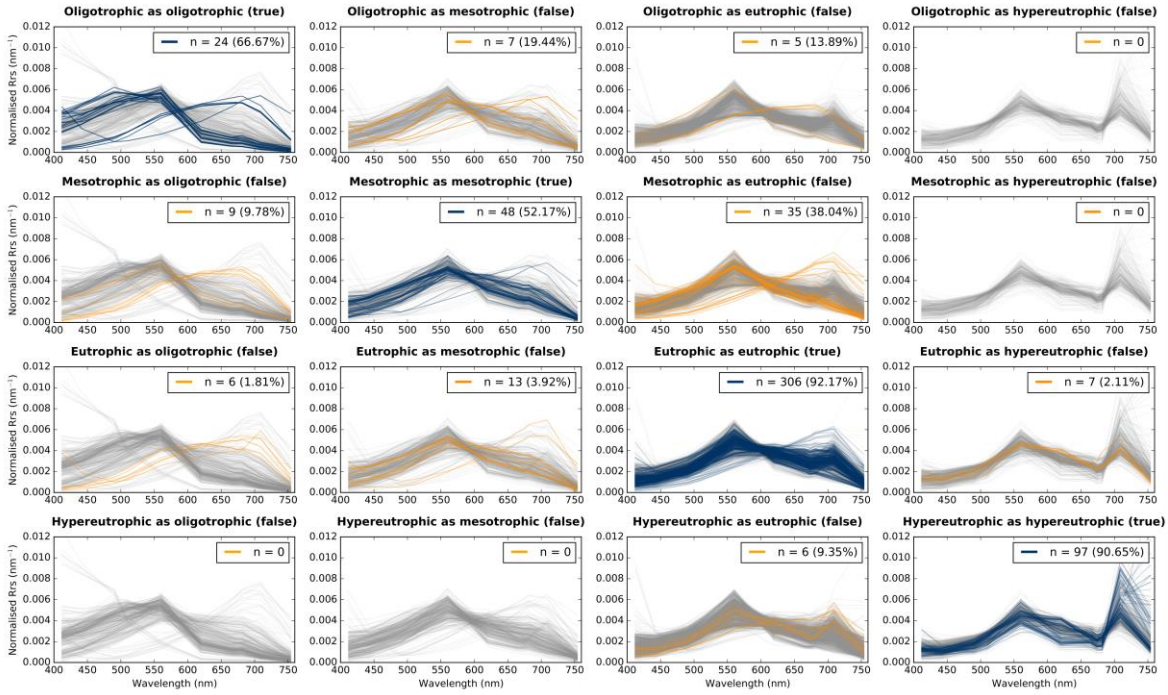


***Figure 4.11:*** *Confusion matrix of remote sensing reflectance spectra. Shown are the classification results of the meta-classifier. Colours of the Rrs vectors correspond to training observations (grey), correctly classified (blue) and misclassified observations (orange). The percentages are identical to those in Figure 4.7.*

## 4.4 Discussion

### 4.4.1 Meta-learning

The prediction accuracies of the base-classifiers and the SVM strongly varied across the four TS classes. With high precision, the meta-classifier was able to identify correct and incorrect TS class predictions made by the base-classifiers, resulting in accurate TS predictions overall. The performance gains of the meta-classifier over individual base-classifiers and the separate SVM classifier validate the stacking theory. Training an additional classification algorithm on the predictions of base-learners can result in significant prediction improvements and reduces the dependency on individual algorithms. The procedure of meta-learning also decreases required knowledge about the performance of a single retrieval algorithm prior to its application to unseen observations. Inherently, the meta-learner has access to the prediction performance of each base-learner through the provided level-zero predictions. Independent of the encountered water type, the meta-learner can thus decide when to employ a specific base-learner during the application to unseen observations. This feature of the presented learning scheme might hold great potential to overcome single algorithm application limitations.

### 4.4.2 Direct trophic status classification

For the training of the base-classifiers, Rrs with previously assigned TS classes were utilised. Since the TS class ranges in this study were a function of chla, the base-classifiers learnt this functional relationship indirectly. Notably, the chla values used to define the class ranges were unknown to the classifiers during the training process. Corresponding reflectance vectors were treated by the classifiers without knowledge about the concentrations of OACs. Consequently, the classifiers learnt to represent a TS class in their feature space through the provided Rrs vectors, whereby the input features corresponded to the values at the band positions of OLCI. Additional bands may be added to the Rrs vectors in the training stage to improve the optical separability of the TS classes.

Through direct TS classification, the presented method avoided some of the issues inherent to TS derivation via chla. For example, the meta-classifier was not confronted with the task of scaling $a_{\phi}(\lambda)$ to chla correctly. Naturally, phytoplankton is part of TSM and produces dissolved organic matter. Indirectly, certain water bodies favour specific phytoplankton groups, and also cluster by turbidity or dissolved organic matter loads (e.g. due to riverine influence). When higher or lower phytoplankton absorption efficiency is correlated to changes in $a_{\mathrm{cdom}}(\lambda)$ or $a_{\mathrm{nap}}(\lambda)$, the classifiers can incorporate the resulting influence on the absorption budget in their decision-making through the varying contribution of these IOPs to the Rrs vector.

### 4.4.3 Adaptation of the classification framework

In misclassified, turbid oligotrophic and mesotrophic waters, high inorganic particle loads dominated the optical properties of the water bodies. Turbid water properties weakened the established relationship between chla and the Rrs vectors that defined the TS class assignment. However, scenarios where biological productivity is light-limited due to high suspended sediment loads are common in natural waterbodies (e.g. rivers) and must thus be better incorporated into the presented classification scheme. To adapt the classification method to turbid waters, one may employ other optical parameters for the TS class assignment. For example, the TSI definition by Carlson (1977) enables to relate transparency (in the form of zSD measurements) to TS, which is inversely related to turbidity. The TS class assignment would then be based on the relationship of transparency to Rrs. However turbid, oligotrophic and mesotrophic waters only constitute a small proportion of our dataset. Therefore, the adaptation requires to define an optical criterion that enables to switch between chla and zSD in the TS class assignment. The use of other optical or water colour indicator parameters to classify Rrs directly into TS might require a different TSI definition than used in the present study. For example, $a_{cdom}(\lambda)$ could also be included in the TS assignment, however the relationship of this IOP to a TS class is not as clearly established as it is the case for chla and zSD. In addition, $a_{cdom}(\lambda)$ is not as commonly measured as chla or zSD.

## 4.5 Conclusion

The present study serves as the first demonstration of direct Rrs classification into TS. We enabled the presented classification scheme through stacking of unique base-classifiers in a meta-learning scheme, trained with a large *in situ* dataset of co-located Rrs and chla measurements (n = 2184). Application of the developed approach to test observations (n = 567) demonstrated that direct meta-classification of TS can significantly outperform indirect TS derivation via OWT switching of chla algorithms. The meta-classifier estimated eutrophic and hypereutrophic waters with $>$ 90% prediction accuracy, making the proposed method a reliable tool to assess and monitor eutrophication of inland and nearshore waters. Although the meta-classifier was able to improve retrieval accuracies for oligo- and mesotrophic waters over OWT switching of chla algorithms by 5 - 12%, accurate classification of TS from low - moderate biomass waters influenced by high TSM concentrations and/or $a_{cdom}(\lambda)$ remains a primary challenge to solve.

The classifiers of the present study were trained with 80% of the dataset. A future version of the developed approach can be based on the inclusion of additional base-learners such as the SVM and the re-training of the classifiers with the entire dataset. In addition, the TS class assignment may be based on other optical TS indicators. Performance improvements for the oligo- and mesotrophic waters are therefore likely. In this work we exemplified the algorithm on the multispectral resolution of Sentinel-3A OLCI. After resampling of the training dataset, the algorithm can, however, be

applied to other sensors such as the Multispectral Instrument (MSI) of the Sentinel-2 satellite to enable cross-mission retrievals of TS with the same method.

## 4.6 Contributions

Mortimer Werther, Harald Krawczyk, Kerstin Stelzer, Evangelos Spyrakos, Daniel Odermatt, Stefan G.H. Simis, Oberon Berlage: Conceptualisation; Mortimer Werther, Evangelos Spyrakos: Data compilation; Mortimer Werther, Evangelos Spyrakos, Stefan G.H. Simis, Daniel Odermatt, Kerstin Stelzer, Harald Krawczyk, Oberon Berlage: Investigation; Mortimer Werther, Harald Krawczyk, Oberon Berlage: Methodology; Mortimer Werther: Visualisation; Mortimer Werther: Writing - original draft; Mortimer Werther, Stefan G.H. Simis, Evangelos Spyrakos, Daniel Odermatt, Kerstin Stelzer, Harald Krawczyk, Peter Hunter, Andrew Tyler: Writing - review & editing.

## 4.7 Acknowledgments

The authors wish to thank Carsten Brockmann for his support and fruitful discussions during algorithm development, Jan Clevers for guidance and supervision during early development stages and Olivier Burggraaff for helpful suggestions on the manuscript. We are also grateful to two anonymous reviewers for their thorough and constructive reviews. We appreciate all the LIMNADES data providers that supported the study through direct or indirect supply of *in situ* measurements: Caren E. Binding, Shane Bradt, Mariano Bresciani, Giorgio Dall'Olmo, Claudia Giardino, Anatoly A. Gitelson, Steve Greb, Daniela Gurlin, Kari Y.O. Kallio, Tiit Kutser, Lin Li, Bunkei Matsushita, Victor Martinez-Vicente, Mark W. Matthews, Timothy S. Moore, Igor Ogashawara, Caitlin Riddick, Antonio Ruiz-Verdú, John F. Schalles, Emma Tebbs, Yosef Z. Yacobi, Yunlin Zhang. This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No. 776480, MONOCLE and from the ESA Sentinel Application Platform (SNAP) contract held by Brockmann Consult.

# Chapter 5: Synopsis

The synopsis features the main findings of this dissertation, a conclusion on phytoplankton remote sensing in oligo- and mesotrophic lakes and a topical outlook.

## 5.1 Main findings

The main findings are structured according to the three chapters and the investigated research questions in Chapter 1.6.2.

### 5.1.1 Characterising retrieval uncertainty of chlorophyll-*a* algorithms in oligotrophic and mesotrophic lakes and reservoirs

***Chapter 2:*** *Werther, M., Odermatt, D., Simis, S.G.H., Gurlin, D., Jorge, D.S.F, Loisel, H., Hunter, P.D., Tyler, A.N., Spyrakos, E. (2022): Characterising retrieval uncertainty of chlorophyll-a algorithms in oligotrophic and mesotrophic lakes and reservoirs. ISPRS J. Photogramm. Remote Sens. 190, 279 – 300.*

In this chapter, 13 recently developed and well-established algorithm configurations were tested on a large *in situ* dataset (n = 346) compiled from oligo- and mesotrophic lakes. While the results confirmed previous studies reporting high uncertainties over these water types, this study identified and measured the effect of different variables on the observed retrieval uncertainties.

This overarching research question of this chapter was:

- What are the reasons for the chla retrieval uncertainty of established and recently developed algorithms in low – moderate biomass lakes?

We found that retrieval uncertainty in low biomass lakes is algorithm-specific. For blue/green band ratio approaches, dominating CDOM and NAP absorption was related to low retrieval accuracy. Since most of the reflectance magnitude and shape variability originated from these two IOPs in the blue-green, the OC algorithms were unable to distinguish between the contribution by phytoplankton and the other substances present in the water column. Similarly, the uncertainty of the MDN HICO and OLCI algorithms was also related to CDOM and NAP absorption. The retrieval uncertainty of algorithms using mainly red and NIR bands (G11, Gons05, MPH) was expectedly not related to CDOM and NAP. However, the uncertainty of red-NIR algorithms was related to phytoplankton absorption at 673 nm, as well as low phytoplankton peaks in $R_{rs}(\lambda)$ that could be effectively related to chla. While SAA approaches overall accurately accounted for CDOM and NAP absorption in their inversion, these algorithms revealed issues prone to their parameterisation. Current SAA parameterisations are static, whereby $a_{\phi}^{*}$, the slope of CDOM *S* and other parameters are pre-defined or derived from a dataset. Current SAA implementations are based on generalised IOP

parameterisations, which can only be adapted to a system through *in situ* measurements. This feature of SAA seems to be difficult to overcome, as IOP datasets of low biomass lakes are even more sparse than OAC measurements accompanied by reflectance. The uncertainty analysis in chapter 2 revealed that SAA chla estimation issues were related to phytoplankton absorption variability and intensity in the red, a spectral area in low biomass waters prone to low signal-to-noise ratios. Furthermore, SAA algorithms only estimate $a_\phi(\lambda)$ indirectly through scaling to chla. Accurate retrieval of chla through SAA therefore relies on a clear correlation between the two variables in a specific system. Chla and $a_\phi(\lambda)$ at 443 and 673 nm in the compiled *in situ* dataset of chapter 2 was weak - moderate (see Figure 2.7). Thus, even if $a_\phi(\lambda)$ is accurately estimated by the algorithms, the relationship to chla is at most moderate. The relationship reported for this dataset is different to datasets collected in the ocean where $a_\phi(\lambda)$ was found to correlate well to chla (Bricaud et al., 2004, 1998, 1995).

### 5.1.2 A Bayesian approach for remote sensing of chlorophyll-*a* and associated retrieval uncertainty in oligotrophic and mesotrophic lakes

***Chapter 3:*** *Werther, M., Odermatt, D., Simis, S.G.H., Gurlin, D., Lehmann, M., Kutser, T., Gupana, R., Varley, A., Hunter, P.D., Tyler, A.N., Spyrakos, E. (2022): A Bayesian approach for remote sensing of chlorophyll-a and associated retrieval uncertainty in oligotrophic and mesotrophic inland waters. Accepted for publication in Remote Sensing of Environment.*

The two following research questions guided the third chapter:

- Do Bayesian neural networks (BNNs) improve the estimation of chla in oligo- and mesotrophic inland waters over reference algorithms?
- How can the provided uncertainty be used to improve chla products?

BNNs were specifically developed for Sentinel-3 OLCI and Sentinel-2 MSI to enable improved chla quantification over both large and small lakes. Five different assessment strategies were carried out to evaluate the newly developed BNNs. Accuracy gains by the OLCI and MSI were found > 25% over established chla algorithms for oligotrophic waters on randomised and regional observation assessments of *in situ* data. The accuracy gains decreased with increasing biomass, whereby for eutrophic waters other algorithms such as Gons05, MDN and G11 were competitive or more accurate. The algorithms were then evaluated using different AC algorithms combinations over Lake Geneva. OLCI AC through Polymer was accompanied by the most accurate retrievals by all tested algorithms. For C2RCC the chla estimation accuracy decreased for all algorithms. Yet, POLYMER produced negative $R_{rs}(\lambda)$ in wavelengths longer than 665 nm for 21 OLCI and 12 MSI match-ups. Consequently, C2RCC was favoured to produce chla and uncertainty products. For MSI the overall accuracy over Lake Geneva was worse by all algorithms compared to our BNN, independent of the AC. The produced time series over Lake Geneva showed seasonal variations in *in situ* measured phytoplankton abundance. While chla less than 5 mg $m^{-3}$ were estimated accurately by the BNNs

throughout the years, a spring bloom in 2020 was not captured well by the OLCI BNN. In contrast, the MSI BNN through POLYMER captured the rise and fall of this bloom. A recent study by Gupana et al. (pers. comm.) found that the peak at 681 nm creates ambiguity in this band due to non-photochemical quenching (NPQ), whereby the same peak height can correspond to a large range of chla. The 681 nm is not available on MSI but included on OLCI, which could explain this phenomenon. It is to note that Lake Geneva *in situ* data was not part of the training dataset, thus enhanced retrieval for these conditions is expected during re-training of the BNNs with the entire dataset.

Besides estimation improvements, the BNNs produced an uncertainty estimate alongside chla which covered the reference *in situ* chla value in > 80% of the cases. This chapter introduced a set of metrics to measure the uncertainty calibration in remote sensing of lakes. Further analysis of the uncertainty calibration revealed that while the provided chla uncertainty covered most reference *in situ* chla values (PICP > 75%), the uncertainty was not consistently well-calibrated (Sharpness > 3). Finally, a range of chla and uncertainty OLCI and MSI products were generated for lakes in Europe, Canada, Africa and New Zealand. BNN products of chla and associated uncertainty were spatially consistent, and examples from Canada and Lake Turkana (Kenya) were used to demonstrate the added value of providing an uncertainty estimate alongside chla. Specifically over Lake Turkana, AC was challenging due to complex atmospheric and IOP compositions. Both C2RCC and POLYMER ACs independently produced erroneous $R_{rs}(\lambda)$ for January 2020 over Lake Turkana. Using the BNN uncertainty estimate, the products were successfully merged by retaining the observation with the lowest chla uncertainty. The resulting product enhanced the overall product quality and opens new pathways to switch between ACs depending on estimated chla uncertainty.

### 5.1.3 Meta-classification of remote sensing reflectance to estimate trophic status of inland and nearshore waters

***Chapter 4:*** *Werther, M., Spyrakos, E., Simis, S.G.H., Odermatt, D., Stelzer, K., Krawczyk, H., Hunter, P., Tyler, A. (2021): Meta-classification of remote sensing reflectance to estimate trophic status of inland nearshore waters. ISPRS J. Photogramm. Remote Sens. 176, 109 – 126.*

For this chapter two research questions were posed:

- Is it possible to estimate trophic status of a water body directly from $R_{rs}(\lambda)$?
- Can an ensemble of classification algorithms improve trophic status estimation over conventional switching of multiple chla algorithms per OWT?

Phytoplankton contains the chla pigment, its concentration is conventionally used as an indicator for TS. However, as the two previous chapters showed, accurate remote sensing retrieval of chla concentration is a challenging endeavor in lakes. Therefore, the question explored in this chapter was whether TS, a high-level indicator of phytoplankton biomass, can be estimated from $R_{rs}(\lambda)$ directly. Since any TS is a classification system, the task can be framed as a supervised classification problem.

To enable the classification, the entire LIMNADES dataset of lakes and nearshore waters was used (at the state of 2020). Each $R_{rs}(\lambda)$ in LIMNADES was assigned a TS class depending on recorded chla concentration and the class definitions from Carlson and Simpson (1996). Thereby, chla concentration retrieval could be circumvented, and a relationship between $R_{rs}(\lambda)$ and TS established. Several studies have previously shown that a combination of algorithms may overcome the limitations of individual candidates (Eleveld et al., 2017; Moore et al., 2014; Neil et al., 2019; Smith et al., 2018). This chapter explored the theory of stacking individual algorithms to form a strong single learner (herein called 'meta-learner'). The study had several outcomes. First, the classification results on an independent test set (567 observations from 47 systems) confirmed that $R_{rs}(\lambda)$ can be directly classified into TS by framing the task as a classification problem. Second, stacking of base-learners outperformed the individual base-learners > 5% across all TS classes. More importantly, classification through the meta-learner was more balanced across the classes than from any of the individual base-learners and a separate SVM used for comparison. Third, the classification accuracy for eutrophic and hypereutrophic classes was much more accurate (approximately 40%) than for oligotrophic and mesotrophic waters. The design of the training and test sets in this study was undertaken through prior separation of the included water bodies by their name. This strategy differed from the employed assessments in Chapter 3, as the dataset was neither randomised nor specifically split per region. For oligo- and mesotrophic waters a dataset shift occurred, whereby the oligotrophic class in the training set was low in CDOM (mean 0.30 $m^{-1}$) and TSM (mean 1.97 g $m^{-3}$), but high in both CDOM (mean 5.77 $m^{-1}$) and TSM (mean 7.04 g $m^{-3}$). Therefore, the waters in the test set can be considered turbid and rich in humic substances, whereas the training set consisted mostly of clear lakes. A similar dataset shift occurred for mesotrophic waters, whereby TSM concentration was doubled in the test set. It is thus to conclude that the results reported in the study for the two classes may be viewed as the lower end of the performance range, as for eutrophic and hypereutrophic waters the training conditions were equal to those in during testing. Moreover, the results suggest that an estimate of uncertainty is crucial to detect limited performance when dataset shifts occur. Lastly, the study showed that the meta-learner outperformed switching of chla algorithms through OWTs by 5 – 12% depending on the considered TS class.

## 5.2 Conclusions

This thesis demonstrated improved phytoplankton remote sensing in low biomass lakes through machine learning. Prior remote sensing research on the estimation of low biomass conditions was focussed on individual regions such as the perialpine Italian and Swiss lakes (Bresciani et al., 2011; Giardino et al., 2014, 2005, 2001; Odermatt et al., 2010; Pepe et al., 2001), the U.S. Great Lakes (Binding et al., 2013; Gons et al., 2008; Sayers et al., 2015), northern European (Kutser et al., 2016, 2012), South African (Kravitz et al., 2020; Matthews and Bernard, 2013) or New Zealand lakes

(Lehmann et al., 2018). For this dissertation, multiple available *in situ* datasets were gathered and compiled into larger collections to enable a representative analysis of uncertainties and algorithm development across multiple nations. Through the uncertainty analysis in Chapter 2, the results clearly demonstrated that estimation uncertainty is not specific to a certain region or type of low biomass lakes. Whether clear alpine systems with high euphotic depths or boreal lakes where CDOM and NAP dominate the optical properties, estimation of low phytoplankton biomass was the most uncertain for chla concentrations below 5 mg $m^{-3}$ and became steadily more accurate towards 10 mg $m^{-3}$. The results of the uncertainty characterisation furthermore revealed that chla retrieval uncertainty is algorithm-specific. Chapter 2 showed that neither SAA nor the empirical band ratio approaches tested, using all kinds of band combinations, were able to accurately estimate chla. This finding allows to draw two main conclusions.

First, a specific set of multispectral bands or their combination favourably suited for the chla retrieval is currently unknown. The results of Chapters 2 and 3 show that hyperspectral resolution can lead to improved chla estimation. In Chapter 2, the MDN and GSM algorithms in the hyperspectral HICO configuration resulted in an increase in accuracy between 10 - 20%. In Chapter 3, the OLCI BNN (12 bands) were only slightly more accurate (MdSA 1.44 - 20% depending on the assessment strategy) than their MSN configuration (7 bands), but their uncertainty was better calibrated. These results suggest that hyperspectral resolution has the potential to significantly improve the accuracy of chla estimation in lakes. However, MDN and GSM in the HICO configuration were not the most accurate algorithms for estimating chla. The uncertainty characterisation in Chapter 2 revealed problems of these algorithms that cannot be solved by solely increasing the information dimensionality.

Moreover, the growth of the feature space of the input data must be accurately covered by increasing the available training observations to overcome the curse of dimensionality. Otherwise, the generalisation of the ML algorithm is limited and a large estimation uncertainty is to be expected. The available *in situ* training data is sparse and likely to remain so. Therefore, an important aspect of future work will be to experiment with new approaches and synthetic datasets to increase the available observations for ML algorithms developed for hyperspectral sensors.

Second, while physics-based SAA accurately accounted for CDOM and NAP absorption in their inversion, these algorithms revealed issues prone to their generalised IOP parameterisation that is not system-specific enough. SAA are suboptimal for satellite application over low – moderate biomass lakes for which *in situ* measurements are unavailable or the optical properties strongly vary. Large-scale *in situ* datasets that covered a wide variety of optical properties and geographical regions enabled to develop the OLCI and MSI BNNs and meta-classifier. The underlying statistical models were constructed with a full set of OLCI and MSI bands, thereby overcoming the restriction to specific bands typically characterising band ratios. In addition, the ML models did not require system-specific IOP parametrisations. While the developed approaches demonstrated that for highly complex low biomass lakes the independence of IOP parametrisations was beneficial, dataset shifts

partially affected the ML approaches. These shifts are similarly affecting band ratio and SAA approaches. In a dataset shift scenario, the underlying statistical model is forced to extrapolate into unknown feature-target dimensions. At best the result is comparable to a scenario included in the training data, but most often estimation uncertainty will rise. To deal with dataset dependency, the BNNs and meta-classifier incorporated retrieval uncertainty. The BNNs expressed uncertainty associated with chla estimation. In dataset shift scenarios and when uncertainty was introduced through AC or a dataset shift, high uncertainty was observable. The BNN uncertainty estimate was successfully used for the removal of uncertain observations and a merging of AC-dependent chla products. Through the meta-learner, uncertain base-learner estimates could be identified and discarded, thereby resulting in accuracy gains. Whereas meta-learning does not decrease the model dependency on data, it did overcome limited applicability of a single algorithm.

## 5.3 Outlook

*In situ* data collection is becoming more automated and through citizen science projects also more accessible. New low-cost sensors are investigated, developed, or already deployed. *In situ* and satellite observations of lake water quality, both in terms of volume and their spectral, spatial, and temporal characteristics, are expanding at unparalleled pace. It is therefore expected that ML developments, depending on large datasets, will attract continuous scientific interest.

The word uncertainty is negatively connotated, but it is a very powerful indicator. Uncertainty may hold information about previously unconsidered phenomena that influence the retrieval algorithm performance. Algorithm development should therefore put larger emphasis on the provision, visualisation, and characterisation of uncertainties. It is crucial for any scientific research to understand how an approach works and where limitations occur. Approaches such as SHAP or the provision of retrieval uncertainty should become standard practice.

Water quality remote sensing, especially over low - moderate biomass lakes, has seen great advancements over the last decade. However, significant challenges for the development of accurate remote sensing products remain. Firstly, there is a need for further characterisation of the bio-geo-optical properties, with respect to seasonal and spatial differences between and within low - moderate biomass water bodies. The full range of (specific)-IOP values should be better incorporated in the inversion of phytoplankton biomass algorithms in order to avoid significant uncertainties in estimated chla. The uncertainty analysis in chapter 2 revealed that retrieval uncertainty is related to limited incorporation of phytoplankton variability and the influence of CDOM and NAP. Therefore, besides exploration of new algorithmic approaches, greater emphasis should be on the understanding of underlying phytoplankton absorption and backscatter variability, IOP heterogeneity both on vertical and horizontal scales as well as the variation in the relationship between phytoplankton absorption and chla in low – moderate biomass lakes. Improved remote sensing of primary

production, phytoplankton types and structure will be key elements to fundamentally improve the retrieval of phytoplankton estimates in oligotrophic and mesotrophic lakes. Moreover, whereas prior studies demonstrated optical ambiguity of ocean colour inversion (Defoin-Platel and Chami, 2007; Sydor et al., 2004), the ambiguity in CDOM and NAP dominated systems and its effect on OAC retrieval is still not well-understood.

The meta-learner in this thesis was solely based on ML as base-learners. Future advancements should combine physics-based and ML algorithms to explore synergies between them. The meta-learner in this study was designed for a classification problem, therefore analytical approaches could not be incorporated. It is *a priori* not possible to know which algorithm is best performing for an observation. However, OWT switching or blending schemes of algorithms require knowledge about the best-performing algorithm per OWT. This is a cumbersome task because many algorithms need to be evaluated before their inclusion in OWT schemes (Neil et al., 2019). Moreover, the prior evaluation of algorithms depends on the dataset used to derive the OWTs. A regression meta-learner does not require an OWT pre-classification. Instead the meta-learner can decide per individual observation which base-algorithm to employ. Meta-learning overcomes the problem of algorithm evaluation and selection through a decision on the base-algorithm per observation.

Oligo- and mesotrophic lakes are most precious water resources but at the same time highly threatened through global warming and anthropogenic activities. Only through timely and recurrent monitoring activities, changes in lake ecosystems caused by environmental pollution and global warming can be addressed. Fundamentally, the provision of uncertainty is critical for such efforts in order to inspire faith in the offered remote sensing products among lake managers, environmental authorities, and policymakers. When the constraints of remote sensing as an information tool are clearly stated, it is an invaluable information source. It is therefore critical that uncertainty provision, visualisation and use-cases of it become a bigger focus in the remote sensing community focused on lakes and other optically complex water environments. As a result, remote sensing products of lakes will be a dependable data source that can be used in combination with *in situ* observations and hydrodynamic modeling. Combining many data sources will allow us to gain a better understanding of the complicated picture of water component fluctuations on horizontal, vertical, and temporal scales.

# References


Aas, E., 2010. Estimates of radiance reflected towards the Zenith at the surface of the sea. Ocean Science 6, 861–876. https://doi.org/10.5194/os-6-861-2010

Aas, E., 1996. Refractive index of phytoplankton derived from its metabolite composition. Journal of Plankton Research 18, 2249.

Abdar, M., Pourpanah, F., Hussain, S., Rezazadegan, D., Liu, L., Ghavamzadeh, M., Fieguth, P., Cao, X., Khosravi, A., Acharya, U.R., Makarenkov, V., Nahavandi, S., 2021. A review of uncertainty quantification in deep learning: Techniques, applications and challenges. Information Fusion 76, 243–297. https://doi.org/10.1016/j.inffus.2021.05.008

Altmann, A., Toloşi, L., Sander, O., Lengauer, T., 2010. Permutation importance: a corrected feature importance measure. Bioinformatics 26, 1340–1347. https://doi.org/10.1093/bioinformatics/btq134

Anderson, D.M., Glibert, P.M., Burkholder, J.M., 2002. Harmful algal blooms and eutrophication: Nutrient sources, composition, and consequences. Estuaries 25, 704–726. https://doi.org/10.1007/BF02804901

Babin, M., 2003. Variations in the light absorption coefficients of phytoplankton, nonalgal particles, and dissolved organic matter in coastal waters around Europe. Journal of Geophysical Research 108, 3211. https://doi.org/10.1029/2001JC000882

Barbour, J.B., 1989. The Discovery of Dynamics: A Study From a Machian Point of View of the Discovery and the Structure of Dynamical Theories. Cambridge, England. Oxford University Press, USA.

Beaulieu, J.J., DelSontro, T., Downing, J.A., 2019. Eutrophication will increase methane emissions from lakes and impoundments during the 21st century. Nature Communications 10, 1375. https://doi.org/10.1038/s41467-019-09100-5

Belle, V., Papantonis, I., 2020. Principles and Practice of Explainable Machine Learning. arXiv.

Bergstra, J., Bengio, Y., 2012. Random Search for Hyper-Parameter Optimization. J. Mach. Learn. Res. 13, 281–305.

Bergstra, J., Bengio, Y., 2012. Random Search for Hyper-Parameter Optimization. Journal of Machine Learning Research 13, 281-305.

Bergstra, J.S., Bardenet, R., Bengio, Y., Kégl, B., 2011. Algorithms for Hyper-Parameter Optimization, in: Shawe-Taylor, J., Zemel, R.S., Bartlett, P.L., Pereira, F., Weinberger, K.Q. (Eds.), Advances in Neural Information Processing Systems 24. Curran Associates, Inc., 2546–2554.

Biewald, L., 2020. Experiment Tracking with Weights and Biases [WWW Document]. URL https://www.wandb.com/

Binding, C., Jerome, J., Bukata, R., Booty, W., 2008. Spectral absorption properties of dissolved and particulate matter in Lake Erie. Remote Sensing of Environment - REMOTE SENS ENVIRON 112, 1702–1711. https://doi.org/10.1016/j.rse.2007.08.017

Binding, C. E., Greenberg, T.A., Bukata, R.P., 2013. The MERIS Maximum Chlorophyll Index; its merits and limitations for inland water algal bloom monitoring. Journal of Great Lakes Research 39, 100–107. https://doi.org/10.1016/j.jglr.2013.04.005

Binding, C.E., Greenberg, T.A., Jerome, J.H., Bukata, R.P., Letourneau, G., 2011. An assessment of MERIS algal products during an intense bloom in Lake of the Woods. Journal of Plankton Research 33, 793–806. https://doi.org/10.1093/plankt/fbq133

Binding, C.E., Jerome, J.H., Bukata, R.P., Booty, W.G., 2010. Suspended particulate matter in Lake Erie derived from MODIS aquatic colour imagery. International Journal of Remote Sensing 31, 5239–5255. https://doi.org/10.1080/01431160903302973

Binding, C.E., Jerome, J.H., Bukata, R.P., Booty, W.G., 2007. Trends in Water Clarity of the Lower Great Lakes from Remotely Sensed Aquatic Color. Journal of Great Lakes Research 33, 828–841.

Bishop, C.M., 2006. Pattern Recognition and Machine Learning (Information Science and Statistics). Springer-Verlag, Berlin, Heidelberg.

Bishop, C.M., 1995. Neural Networks for Pattern Recognition. Oxford University Press, Inc., USA.

Blois, J.L., Zarnetske, P.L., Fitzpatrick, M.C., Finnegan, S., 2013. Climate Change and the Past, Present, and Future of Biotic Interactions. Science 341, 499–504. https://doi.org/10.1126/science.1237184

Blondeau-Patissier, D., Gower, J.F.R., Dekker, A.G., Phinn, S.R., Brando, V.E., 2014. A review of ocean color remote sensing methods and statistical techniques for the detection, mapping and analysis of phytoplankton blooms in coastal and open oceans. Progress in Oceanography 123, 123–144. https://doi.org/10.1016/j.pocean.2013.12.008

Bolstad, W.M., Curran, J.M., 2016. Introduction to Bayesian Statistics, Third Edition, Introduction to Bayesian Statistics, Third Edition. John Wiley & Sons, Inc., Hoboken, NJ, USA. https://doi.org/10.1002/9781118593165

Boser, B.E., Boser, B.E., Guyon, I.M., Vapnik, V.N., 1992. A Training Algorithm for Optimal Margin Classifiers. Proceedings of the 5th Annual ACM Workshop on computational learning theory, 144-152.

Boyer, J.N., Kelble, C.R., Ortner, P.B., Rudnick, D.T., 2009. Phytoplankton bloom status: Chlorophyll a biomass as an indicator of water quality condition in the southern estuaries of Florida, USA. Ecological Indicators 9, S56–S67. https://doi.org/10.1016/j.ecolind.2008.11.013

Bradt, S.R., 2012. Development of bio-optical algorithms to estimate chlorophyll in the Great Salt Lake and New England lakes using in situ hyperspectral measurements. The Uni. of New Hampshire.

Breiman, L., 2004. Population theory for boosting ensembles. The Annals of Statistics 32, 1–11. https://doi.org/10.1214/aos/1079120126

Breiman, L., 2001. Random forests. Machine Learning 45, 5–32. https://doi.org/10.1023/A:1010933404324

Breiman, L., 1996. Bagging Predictors. Machine Learning 24, 123–140. https://doi.org/10.1023/A:1018054314350

Bresciani, M., Cazzaniga, I., Austoni, M., Sforzi, T., Buzzi, F., Morabito, G., Giardino, C., 2018. Mapping phytoplankton blooms in deep subalpine lakes from Sentinel-2A and Landsat-8. Hydrobiologia 824, 197–214. https://doi.org/10.1007/s10750-017-3462-2

Bresciani, M., Stroppiana, D., Odermatt, D., Morabito, G., Giardino, C., 2011. Assessing remotely sensed chlorophyll-a for the implementation of the Water Framework Directive in European perialpine lakes. Science of The Total Environment 409, 3083–3091. https://doi.org/10.1016/j.scitotenv.2011.05.001

Brewin, R.J.W., Sathyendranath, S., Müller, D., Brockmann, C., Deschamps, P.Y., Devred, E., Doerffer, R., Fomferra, N., Franz, B., Grant, M., Groom, S., Horseman, A., Hu, C., Krasemann, H., Lee, Z.P., Maritorena, S., Mélin, F., Peters, M., Platt, T., Regner, P., Smyth, T., Steinmetz, F., Swinton, J., Werdell, J., White, G.N., 2015. The Ocean Colour Climate Change Initiative: III. A round-robin comparison on in-water bio-optical algorithms. Remote Sensing of Environment 162, 271–294. https://doi.org/10.1016/j.rse.2013.09.016

Bricaud, A., Babin, M., Morel, A., Claustre, H., 1995. Variability in the chlorophyll-specific absorption coefficients of natural phytoplankton: Analysis and parameterization. Journal of Geophysical Research: Oceans 100, 13321–13332. https://doi.org/10.1029/95JC00463

Bricaud, A., Claustre, H., Ras, J., Oubelkheir, K., 2004. Natural variability of phytoplanktonic absorption in oceanic waters: Influence of the size structure of algal populations. Journal of Geophysical Research: Oceans 109. https://doi.org/10.1029/2004JC002419

Bricaud, A., Morel, A., Babin, M., Allali, K., Claustre, H., 1998. Variations of light absorption by suspended particles with chlorophyll a concentration in oceanic (case 1) waters : Analysis and implications for bio-optical models. Journal of Geophysical Research 103, 31033–31044.

Brochu, E., Cora, V., Freitas, N., 2010. A Tutorial on Bayesian Optimization of Expensive Cost Functions, with Application to Active User Modeling and Hierarchical Reinforcement Learning. CoRR abs/1012.2.

Brockmann, C., Doerffer, R., Peters, M., Kerstin, S., Embacher, S., Ruescas, A., 2016. 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: Living Planet Symposium, ESA Special Publication. p. 54.

Buckton, D., O'Mongain, E., Danaher, S., 1999. The use of Neural Networks for the estimation of oceanic constituents based on the MERIS instrument. International Journal of Remote Sensing 20, 1841–1851. https://doi.org/10.1080/014311699212515

Bühlmann, P., Hothorn, T., 2007. Boosting Algorithms: Regularization, Prediction and Model Fitting. Statist. Sci. 22, 477–505. https://doi.org/10.1214/07-STS242

Bukata, R.P., 1995. Optical properties and remote sensing of inland and coastal waters. CRC Press.

Bulgarelli, B., Kiselev, V., Zibordi, G., 2014. Simulation and analysis of adjacency effects in coastal waters: a case study. Appl Opt 53, 1523–1545. https://doi.org/10.1364/AO.53.001523

Bulgarelli, B., Zibordi, G., 2018. On the detectability of adjacency effects in ocean color remote sensing of mid-latitude coastal environments by SeaWiFS, MODIS-A, MERIS, OLCI, OLI and MSI. Remote Sens Environ 209, 423–438. https://doi.org/10.1016/j.rse.2017.12.021

Burggraaff, O., 2020. Biases from incorrect reflectance convolution. Opt. Express 28, 13801–13816. https://doi.org/10.1364/OE.391470

Calbet, A., Landry, M.R., 2004. Phytoplankton growth, microzooplankton grazing, and carbon cycling in marine systems. Limnology and Oceanography 49, 51–57. https://doi.org/10.4319/lo.2004.49.1.0051

Cao, Z., Ma, R., Duan, H., Pahlevan, N., Melack, J., Shen, M., Xue, K., 2020. A machine learning approach to estimate chlorophyll-a from Landsat-8 measurements in inland lakes. Remote Sensing of Environment 248, 111974. https://doi.org/10.1016/j.rse.2020.111974

Carlson, R.E., 1977. A trophic state index for lakes. Limnology and Oceanography 22, 361–369. https://doi.org/10.4319/lo.1977.22.2.0361

Carlson, R.E., Simpson, J., 1996. A coordinator's guide to volunteer lake monitoring methods. USA: North American Lake Management Society Madison 96.

Carvalho, L., Solimini, A., Phillips, G., van den Berg, M., Pietiläinen, O.-P., Lyche Solheim, A., Poikane, S., Mischke, U., 2008. Chlorophyll reference conditions for European lake types used for intercalibration of ecological status. Aquatic Ecology 42, 203–211. https://doi.org/10.1007/s10452-008-9189-4

Chen, Q., Zhang, Y., Ekroos, A., Hallikainen, M., 2004. The role of remote sensing technology in the EU water framework directive (WFD). Environmental Science & Policy 7, 267–276. https://doi.org/10.1016/j.envsci.2004.05.002

Chen, T., Guestrin, C., 2016. XGBoost, in: Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. ACM. https://doi.org/10.1145/2939672.2939785

Clark, J.M., Schaeffer, B.A., Darling, J.A., Urquhart, E.A., Johnston, J.M., Ignatius, A.R., Myer, M.H., Loftin, K.A., Werdell, P.J., Stumpf, R.P., 2017. Satellite monitoring of cyanobacterial harmful algal bloom frequency in recreational waters and drinking water sources. Ecological Indicators 80, 84–95. https://doi.org/10.1016/j.ecolind.2017.04.046

Claustre, H., Hooker, S.B., Van Heukelem, L., Berthon, J.F., Barlow, R., Ras, J., Sessions, H., Targa, C., Thomas, C.S., Van Der Linde, D., Marty, J.C., 2004. An intercomparison of HPLC phytoplankton pigment methods using in situ samples: Application to remote sensing and database activities. Marine Chemistry 85, 41–61. https://doi.org/10.1016/j.marchem.2003.09.002

Cleveland, J.S., Weidemann, A.D., 1993. Quantifying absorption by aquatic particles: A multiple scattering correction for glass-fiber filters. Limnology and Oceanography 38, 1321–1327. https://doi.org/10.4319/lo.1993.38.6.1321

Codd, G.A., 2000. Cyanobacterial toxins, the perception of water quality, and the prioritisation of eutrophication control. Ecological Engineering 16, 51–60. https://doi.org/doi.org/10.1016/S0925-8574(00)00089-6

Coffer, M.M., Schaeffer, B.A., Salls, W.B., Urquhart, E., Loftin, K.A., Stumpf, R.P., Werdell, P.J., Darling, J.A., 2021. Satellite remote sensing to assess cyanobacterial bloom frequency across the United States at multiple spatial scales. Ecological Indicators 128, 107822. https://doi.org/10.1016/j.ecolind.2021.107822

Cohen, J., 1960. A Coefficient of Agreement for Nominal Scales. Educational and Psychological Measurement 20, 37–46. https://doi.org/10.1177/001316446002000104

Cole, J.J., Prairie, Y.T., Caraco, N.F., McDowell, W.H., Tranvik, L.J., Striegl, R.G., Duarte, C.M., Kortelainen, P., Downing, J.A., Middelburg, J.J., Melack, J., 2007. Plumbing the Global Carbon Cycle: Integrating Inland Waters into the Terrestrial Carbon Budget. Ecosystems 10, 172–185. https://doi.org/10.1007/s10021-006-9013-8

Comte, L., Olden, J.D., 2017. Climatic vulnerability of the world's freshwater and marine fishes. Nature Clim Change 7, 718–722. https://doi.org/10.1038/nclimate3382

Congalton, R.G., 1991. A review of assessing the accuracy of classifications of remotely sensed data. Remote Sensing of Environment 37, 35–46. https://doi.org/10.1016/0034-4257(91)90048-B

Conley, D., Paerl, H., Howarth, R., Boesch, D., Seitzinger, S., Havens, K., Lancelot, C., Likens, G., 2009. Controlling Eutrophication: Nitrogen and Phosphorus. Science 323, 1014–1015.

Cortes, C., Vapnik, V., 1995. Support-vector networks. Machine Learning 20, 273–297. https://doi.org/10.1007/BF00994018

Dall'Olmo, G., Gitelson, A.A., 2006. Absorption Properties of Dissolved and Particulate Matter in Turbid Productive Inland Lakes, in: Proceedings of Ocean Optics XVIII. Ocean Optics Conference, pp. 1–15.

Dall'Olmo, G., Gitelson, A.A., Rundquist, D.C., 2003. Towards a unified approach for remote estimation of chlorophyll-a in both terrestrial vegetation and turbid productive waters. Geophysical Research Letters 30. https://doi.org/10.1029/2003GL018065

Dall'Olmo, G., Gitelson, A.A., Rundquist, D.C., Leavitt, B., Barrow, T., Holz, J.C., 2005. Assessing the potential of SeaWiFS and MODIS for estimating chlorophyll concentration in turbid productive waters using red and near-infrared bands. Remote Sensing of Environment 96, 176–187. https://doi.org/10.1016/j.rse.2005.02.007

De Keukelaere, L., Sterckx, S., Adriaensen, S., Knaeps, E., Reusen, I., Giardino, C., Bresciani, M., Hunter, P., Neil, C., Van der Zande, D., Vaiciute, D., 2018. Atmospheric correction of Landsat-8/OLI and Sentinel-2/MSI data using iCOR algorithm: validation for coastal and inland waters. European Journal of Remote Sensing 51, 525–542. https://doi.org/10.1080/22797254.2018.1457937

Defoin-Platel, M., Chami, M., 2007. How ambiguous is the inverse problem of ocean color in coastal waters? Journal of Geophysical Research: Oceans 112, 1–16. https://doi.org/10.1029/2006JC003847

Degroot, M.D., Fienberg, S.E., 1983. The comparison and evaluation of forecasters., in: Proceedings of the 1982 .I.O.S. Annual Conference on Practical Bayesian Statistics. Carnegie-Mellon University, Pittsburgh, PA, Department of Statistics; TR-244, pp. 12–22. https://doi.org/10.2307/2987588

Dekker, A.G., Peters, S.W.M., 1993. The use of the thematic mapper for the analysis of eutrophic lakes: A case study in the netherlands. International Journal of Remote Sensing 14, 799–821. https://doi.org/10.1080/01431169308904379

DelSontro, T., Beaulieu, J.J., Downing, J.A., 2018. Greenhouse gas emissions from lakes and impoundments: Upscaling in the face of global change. Limnology and Oceanography Letters 3, 64–75. https://doi.org/10.1002/lol2.10073

Demars, B.O.L., Gíslason, G.M., Ólafsson, J.S., Manson, J.R., Friberg, N., Hood, J.M., Thompson, J.J.D., Freitag, T.E., 2016. Impact of warming on CO2 emissions from streams countered by aquatic photosynthesis. Nature Geosci 9, 758–761. https://doi.org/10.1038/ngeo2807

Doerffer, R., Schiller, H., 2007. The MERIS case 2 water algorithm. International Journal of Remote Sensing - INT J REMOTE SENS 28, 517–535. https://doi.org/10.1080/01431160600821127

Downing, J.A., Prairie, Y.T., Cole, J.J., Duarte, C.M., Tranvik, L.J., Striegl, R.G., McDowell, W.H., Kortelainen, P., Caraco, N.F., Melack, J.M., Middelburg, J.J., 2006. The global abundance and size distribution of lakes, ponds, and impoundments. Limnology and Oceanography 51, 2388–2397. https://doi.org/10.4319/lo.2006.51.5.2388

Du, M., Liu, N., Hu, X., 2020. Techniques for interpretable machine learning. Communications of the ACM 63, 68–77. https://doi.org/10.1145/3359786

Eleveld, M.A., Ruescas, A.B., Hommersom, A., Moore, T.S., Peters, S.W.M., Brockmann, C., 2017. An Optical Classification Tool for Global Lake Waters. Remote Sensing 9. https://doi.org/10.3390/rs9050420

Fan, Y., Li, W., Chen, N., Ahn, J.-H., Park, Y.-J., Kratzer, S., Schroeder, T., Ishizaka, J., Chang, R., Stamnes, K., 2021. OC-SMART: A machine learning based data analysis platform for

satellite ocean color sensors. Remote Sensing of Environment 253, 112236. https://doi.org/10.1016/j.rse.2020.112236
Fan, Y., Li, W., Gatebe, C.K., Jamet, C., Zibordi, G., Schroeder, T., Stamnes, K., 2017. Atmospheric correction over coastal waters using multilayer neural networks. Remote Sensing of Environment 199, 218–240. https://doi.org/10.1016/j.rse.2017.07.016
Ferrari, G.M., Tassan, S., 1999. A method using chemical oxidation to remove light absorption by phytoplankton pigments. Journal of Phycology 35, 1090–1098. https://doi.org/10.1046/j.1529-8817.1999.3551090.x
Filazzola, A., Mahdiyan, O., Shuvo, A., Ewins, C., Moslenko, L., Sadid, T., Blagrave, K., Imrit, M.A., Gray, D.K., Quinlan, R., O'Reilly, C.M., Sharma, S., 2020. A database of chlorophyll and water chemistry in freshwater lakes. Scientific Data 7, 1–10. https://doi.org/10.1038/s41597-020-00648-2
Flennerhag, S., 2017. ML-Ensemble. https://github.com/flennerhag/mlens.
Foley, B., Jones, I.D., Maberly, S.C., Rippey, B., 2012. Long-term changes in oxygen depletion in a small temperate lake: effects of climate change and eutrophication. Freshwater Biology 57, 278–289. https://doi.org/10.1111/j.1365-2427.2011.02662.x
Franz, B.A., Bailey, S.W., Kuring, N., Werdell, P.J., 2015. Ocean color measurements with the Operational Land Imager on Landsat-8: implementation and evaluation in SeaDAS. J. Appl. Remote Sens 9, 096070. https://doi.org/10.1117/1.JRS.9.096070
Franz, B.A., Bailey, S.W., Werdell, P.J., McClain, C.R., 2007. Sensor-independent approach to the vicarious calibration of satellite ocean color radiometry. Appl. Opt. 46, 5068. https://doi.org/10.1364/AO.46.005068
Freund, Y., Schapire, R.E., 1997. A Decision-Theoretic Generalization of On-Line Learning and an Application to Boosting. Journal of Computer and System Sciences 55, 119–139. https://doi.org/10.1006/jcss.1997.1504
Freund, Y., Schapire, R.E., 1996. Experiments with a new boosting algorithm, in: Machine Learning: Proceedings of the Thirteenth International Conference. pp. 148–156.
Friedman, J., Hastie, T., Tibshirani, R., 1998. Additive Logistic Regression: a Statistical View of Boosting. Annals of Statistics 28, 2000.
Frouin, R.J., Franz, B.A., Ibrahim, A., Knobelspiesse, K., Ahmad, Z., Cairns, B., Chowdhary, J., Dierssen, H.M., Tan, J., Dubovik, O., Huang, X., Davis, A.B., Kalashnikova, O., Thompson, D.R., Remer, L.A., Boss, E., Coddington, O., Deschamps, P.-Y., Gao, B.-C., Gross, L., Hasekamp, O., Omar, A., Pelletier, B., Ramon, D., Steinmetz, F., Zhai, P.-W., 2019. Atmospheric Correction of Satellite Ocean-Color Imagery During the PACE Era. Frontiers in Earth Science 7.
Gal, Y., Ghaharamni, Z., 2016. Dropout as a Bayesian Approximation: Representing Model Uncertainty in Deep Learning Zoubin Ghahramani. PMLR, pp. 1050–1059.
Gao, B.-C., Montes, M.J., Ahmad, Z., Davis, C.O., 2000. Atmospheric correction algorithm for hyperspectral remote sensing of ocean color from space. Appl. Opt., AO 39, 887–896. https://doi.org/10.1364/AO.39.000887
Gege, P., 2014. WASI-2D: A software tool for regionally optimized analysis of imaging spectrometer data from deep and shallow waters. Computers & Geosciences 62, 208–215. https://doi.org/10.1016/j.cageo.2013.07.022
Ghahramani, Z., 2015. Probabilistic machine learning and artificial intelligence. Nature 521, 452–459. https://doi.org/10.1038/nature14541
Giardino, C., Bresciani, M., Cazzaniga, I., Schenk, K., Rieger, P., Braga, F., Matta, E., Brando, V.E., 2014. Evaluation of Multi-Resolution Satellite Sensors for Assessing Water Quality and Bottom Depth of Lake Garda. Sensors 14, 24116–24131. https://doi.org/10.3390/s141224116
Giardino, C., Bresciani, M., Stroppiana, D., Oggioni, A., Morabito, G., 2014b. Optical remote sensing of lakes: An overview on Lake Maggiore. Journal of Limnology 73, 201–214. https://doi.org/10.4081/jlimnol.2014.817
Giardino, C., Bresciani, M., Valentini, E., Gasperini, L., Bolpagni, R., Brando, V.E., 2015. Airborne hyperspectral data to assess suspended particulate matter and aquatic vegetation in a shallow

and turbid lake. Remote Sensing of Environment 157, 48–57. https://doi.org/10.1016/j.rse.2014.04.034
Giardino, C., Candiani, G., Bresciani, M., Lee, Z., Gagliano, S., Pepe, M., 2012. BOMBER: A tool for estimating water quality and bottom properties from remote sensing images. Comput. Geosci. 45, 313–318. https://doi.org/10.1016/j.cageo.2011.11.022
Giardino, C., Candiani, G., Zilioli, E., 2005. Detecting chlorophyll-a in Lake Garda using TOA MERIS radiances. Photogrammetric Engineering & Remote Sensing 71, 1045–1051. https://doi.org/10.14358/PERS.71.9.1045
Giardino, C., Pepe, M., Brivio, P.A., Ghezzi, P., Zilioli, E., 2001. Detecting chlorophyll, Secchi disk depth and surface temperature in a sub-alpine lake using Landsat imagery. Science of the Total Environment 268, 19–29. https://doi.org/10.1016/S0048-9697(00)00692-6
Gilerson, A., Zhou, J., Hlaing, S., Ioannou, I., Schalles, J., Gross, B., Moshary, F., Ahmed, S., 2007. Fluorescence component in the reflectance spectra from coastal waters. Dependence on water composition. Optics Express 15, 15702. https://doi.org/10.1364/oe.15.015702
Gilerson, Alexander A., Gitelson, A.A., Zhou, J., Gurlin, D., Moses, W., Ioannou, I., Ahmed, S.A., 2010. Algorithms for remote estimation of chlorophyll-a in coastal and inland waters using red and near infrared bands. Optics Express 18, 24109. https://doi.org/10.1364/oe.18.024109
Gitelson, A., Garbuzov, G., Szilagyi, F., Mittenzwey, K.H., Karnieli, A., Kaiser, A., 1993. Quantitative remote sensing methods for real-time monitoring of inland waters quality. International Journal of Remote Sensing 14, 1269–1295. https://doi.org/10.1080/01431169308953956
Gitelson, A.A., 1992. The peak near 700 nm on radiance spectra of algae and water: Relationships of its magnitude and position with chlorophyll. International Journal of Remote Sensing 13, 3367–3373. https://doi.org/10.1080/01431169208904125
Gitelson, Anatoly A., Dall'Olmo, G., Moses, W., Rundquist, D.C., Barrow, T., Fisher, T.R., Gurlin, D., Holz, J., 2008. A simple semi-analytical model for remote estimation of chlorophyll-a in turbid waters: Validation. Remote Sensing of Environment 112, 3582–3593. https://doi.org/10.1016/j.rse.2008.04.015
Gitelson, A.A., Schalles, J.F., Hladik, C.M., 2007. Remote chlorophyll-a retrieval in turbid, productive estuaries: Chesapeake Bay case study. Remote Sensing of Environment 109, 464–472. https://doi.org/10.1016/j.rse.2007.01.016
Gitelson, A.A., Schalles, J.F., Rundquist, D.C., Schiebe, F.R., Yacobi, Y.Z., 1999. Comparative reflectance properties of algal cultures with manipulated densities. Journal of Applied Phycology 11, 345–354. https://doi.org/10.1023/A:1008143902418
Glibert, P.M., Seitzinger, S., Heil, C.A., Burkholder, J.M., Parrow, M.W., Codispoti, L.A., Kelly, V., 2005. The Role of Eutrophication in the Global Proliferation of Harmful Algal Blooms. Oceanography 18, 198–209.
Gold, A.J., Sims, J.T., 2004. Eutrophication, in: Encyclopedia of Soils in the Environment. Elsevier Inc., pp. 486–494. https://doi.org/10.1016/B0-12-348530-4/00093-X
Gons, H., Auer, M., Effler, S., 2008. MERIS satellite chlorophyll mapping of oligotrophic and eutrophic waters in the Laurentian Great Lakes. Remote Sensing of Environment 112, 4098–4106. https://doi.org/10.1016/j.rse.2007.06.029
Gons, H., Rijkeboer, M., Ruddick, K., 2005. Effect of a waveband shift on chlorophyll retrieval from MERIS imagery of inland and coastal waters. Journal of Plankton Research 27.
Gons, H., Rijkeboer, M., Ruddick, K., 2002. A chlorophyll-retrieval algorithm for satellite imagery (Medium Resolution Imaging Spectrometer) of inland and coastal waters. Journal of Plankton Research 24. https://doi.org/10.1093/plankt/24.9.947
Gons, H.J., 1999. Optical teledetection of chlorophyll a in turbid inland waters. Environmental Science and Technology 33, 1127–1132. https://doi.org/10.1021/es9809657
González Vilas, L., Spyrakos, E., Torres Palenzuela, J.M., 2011. Neural network estimation of chlorophyll a from MERIS full resolution data for the coastal waters of Galician rias (NW Spain). Remote Sensing of Environment 115, 524–535. https://doi.org/10.1016/j.rse.2010.09.021
Goodfellow, I., Bengio, Y., Courville, A., 2016. Deep Learning. MIT Press.
Gordon, H.R., 2019. Physical Principles of Ocean Color Remote Sensing. University of Miami. https://doi.org/10.33596/ppocrs-19

Gordon, H.R., Brown, O.B., Evans, R.H., Brown, J.W., Smith, R.C., Baker, K.S., Clark, D.K., 1988. A semianalytic radiance model of ocean color. Journal of Geophysical Research 93, 10909. https://doi.org/10.1029/JD093iD09p10909

Gordon, H.R., Brown, O.B., Jacobs, M.M., 1975. Computed Relationships Between the Inherent and Apparent Optical Properties of a Flat Homogeneous Ocean. Appl. Opt., AO 14, 417–427. https://doi.org/10.1364/AO.14.000417

Gordon, H.R., Clark, D.K., 1980. Remote sensing optical properties of a stratified ocean: an improved interpretation. Appl. Opt. 19, 3428–3430. https://doi.org/10.1364/AO.19.003428

Gordon, H.R., Wang, M., 1994. Retrieval of water-leaving radiance and aerosol optical thickness over the oceans with SeaWiFS: a preliminary algorithm. Appl. Opt., AO 33, 443–452. https://doi.org/10.1364/AO.33.000443

Gordon, H.R., Wang, M., 1992. Surface-roughness considerations for atmospheric correction of ocean color sensors. 1: The Rayleigh-scattering component. Appl. Opt., AO 31, 4247–4260. https://doi.org/10.1364/AO.31.004247

Gower, J., King, S., Borstad, G., Brown, L., 2005. Detection of intense plankton blooms using the 709 nm band of the MERIS imaging spectrometer. International Journal of Remote Sensing 26, 2005–2012. https://doi.org/10.1080/01431160500075857

Gower, J.F.R., 1980. Observations of in situ fluorescence of chlorophyll-a in Saanich Inlet. Boundary-Layer Meteorology 18, 235–245. https://doi.org/10.1007/BF00122022

Gower, J.F.R., Brown, L., Borstad, G.A., 2004. Observation of chlorophyll fluorescence in west coast waters of Canada using the MODIS satellite sensor. Canadian Journal of Remote Sensing 30, 17–25. https://doi.org/10.5589/m03-048

Gower, J.F.R., Doerffer, R., Borstad, G.A., 1999. Interpretation of the 685nm peak in water-leaving radiance spectra in terms of fluorescence, absorption and scattering, and its observation by MERIS. International Journal of Remote Sensing 20, 1771–1786. https://doi.org/10.1080/014311699212470

Graves, A., 2011. Practical Variational Inference for Neural Networks. Advances in Neural Information Processing Systems 24.

Graziani, F., Trofimov, S., Battistini, S., 2021. 1 - Applied astrodynamics, in: Cappelletti, C., Battistini, S., Malphrus, B.K. (Eds.), Cubesat Handbook. Academic Press, pp. 35–52. https://doi.org/10.1016/B978-0-12-817884-3.00001-1

Greb, S., Dekker, A.G., Binding, C., Bernard, S., Binding, C., Brockmann, C., Dekker, A.G., DiGiacomo, P., Greb, S., Griffith, D., Groom, S., Hestir, E., Hunter, P., Kutser, T., Mannaerts, C.M., Matthews, M., Odermatt, D., Lain, L.R., Schaeffer, B., Simis, S., Spyrakos, E., Stumpf, R.P., Tyler, A., Urquhart, E.A., Wang, M., 2018. Earth observations in support of global water quality monitoring. International Ocean-Colour Coordinating Group.

Grizzetti, B., Lanzanova, D., Liquete, C., Reynaud, A., Cardoso, A.C., 2016. Assessing water ecosystem services for water resource management. Environmental Science & Policy 61, 194–203. https://doi.org/10.1016/j.envsci.2016.04.008

Grömping, U., 2009. Variable importance assessment in regression: Linear regression versus random forest. American Statistician 63, 308–319. https://doi.org/10.1198/tast.2009.08199

Grzymski, J., Johnsen, G., Sakshaug, E., 1997. The significance of intracellular self-shading on the biooptical properties of brown, red, and green macroalgae. Journal of Phycology 33, 408–414. https://doi.org/10.1111/j.0022-3646.1997.00408.x

Guanter, L., Ruiz-Verdú, A., Odermatt, D., Giardino, C., Simis, S., Estellés, V., Heege, T., Domínguez-Gómez, J.A., Moreno, J., 2010a. Atmospheric correction of ENVISAT/MERIS data over inland waters: Validation for European lakes. Remote Sensing of Environment 114, 467–480. https://doi.org/10.1016/j.rse.2009.10.004

Guanter, L., Ruiz-Verdú, A., Odermatt, D., Giardino, C., Simis, S., Estellés, V., Heege, T., Domínguez-Gómez, J.A., Moreno, J., 2010b. Atmospheric correction of ENVISAT/MERIS data over inland waters: Validation for European lakes. Remote Sensing of Environment 114, 467–480. https://doi.org/10.1016/j.rse.2009.10.004

Guo, C., Pleiss, G., Sun, Y., Weinberger, K.Q., 2017. On Calibration of Modern Neural Networks.

Gupana, R.S., Odermatt, D., Cesana, I., Giardino, C., Nedbal, L., Damm, A., 2021. Remote sensing of sun-induced chlorophyll-a fluorescence in inland and coastal waters: Current state and future prospects. Remote Sensing of Environment 262, 112482. https://doi.org/10.1016/j.rse.2021.112482

Gurlin, D., Gitelson, A.A., Moses, W.J., 2011. Remote estimation of chl-a concentration in turbid productive waters — Return to a simple two-band NIR-red model? Remote Sensing of Environment 115, 3479–3490. https://doi.org/10.1016/j.rse.2011.08.011

Ham, J., Yangchi Chen, Crawford, M.M., Ghosh, J., 2005. Investigation of the random forest framework for classification of hyperspectral data. IEEE Transactions on Geoscience and Remote Sensing 43, 492–501. https://doi.org/10.1109/TGRS.2004.842481

Härmä, P., Vepsäläinen, J., Hannonen, T., Pyhälahti, T., Kämäri, J., Kallio, K., Eloheimo, K., Koponen, S., 2001. Detection of water quality using simulated satellite data and semi-empirical algorithms in Finland. Science of the Total Environment 268, 107–121. https://doi.org/10.1016/S0048-9697(00)00688-4

Harmel, T., Chami, M., Tormos, T., Reynaud, N., Danis, P.-A., 2018. Sunglint correction of the Multi-Spectral Instrument (MSI)-SENTINEL-2 imagery over inland and sea waters from SWIR bands. Remote Sensing of Environment 204, 308–321. https://doi.org/10.1016/j.rse.2017.10.022

Hastie, T., Friedman, J., Tibshirani, R., 2001. The Elements of Statistical Learning, Springer Series in Statistics. Springer New York, New York, NY. https://doi.org/10.1007/978-0-387-21606-5

Head, T., MechCoder, Louppe, G., Shcherbatyi, I., fcharras, Vinícius, Z., cmmalone, Schröder, C., nel215, Campos, N., Young, T., Cereda, S., Fan, T., Schwabedal, J., Hvass-Labs, Pak, M., SoManyUsernamesTaken, Callaway, F., Estève, L., Besson, L., Landwehr, P.M., Komarov, P., Cherti, M., Shi, K. (KJ), Pfannschmidt, K., Linzberger, F., Cauet, C., Gut, A., Mueller, A., Fabisch, A., 2018. scikit-optimize/scikit-optimize: High five - v0.5.

Heisler, J., Glibert, P.M., Burkholder, J.M., Anderson, D.M., Cochlan, W., Dennison, W.C., Dortch, Q., Gobler, C.J., Heil, C.A., Humphries, E., Lewitus, A., Magnien, R., Marshall, H.G., Sellner, K., Stockwell, D.A., Stoecker, D.K., Suddleson, M., 2008. Eutrophication and harmful algal blooms: A scientific consensus. Harmful Algae 8, 3–13. https://doi.org/10.1016/j.hal.2008.08.006

Hieronymi, M., Müller, D., Doerffer, R., 2017. The OLCI Neural Network Swarm (ONNS): A Bio-Geo-Optical Algorithm for Open Ocean and Coastal Waters. Frontiers in Marine Science 4, 140. https://doi.org/10.3389/fmars.2017.00140

Hoge, F.E., Lyon, P.E., 1996. Satellite retrieval of inherent optical properties by linear matrix inversion of oceanic radiance models: An analysis of model and radiance measurement errors. Journal of Geophysical Research 101, 16,631-16,648. https://doi.org/10.1029/96JC01414

Hoogenboom, H.J., Dekker, A.G., de Haan, J.F., 1998. Retrieval of Chlorophyll and Suspended Matter from Imaging Spectrometry Data by Matrix Inversion. Canadian Journal of Remote Sensing 24, 144–152. https://doi.org/10.1080/07038992.1998.10855234

Hooker, S.B., Esaias, W.E., 1993. An overview of the SeaWiFS Project. Eos, Transactions American Geophysical Union 74, 241–246. https://doi.org/10.1029/93EO00945

Hooker, S.B., Morel, A., 2003. Platform and environmental effects on above-water determinations of water-leaving radiances. Journal of Atmospheric and Oceanic Technology 20, 187–205. https://doi.org/10.1175/1520-0426(2003)020<0187:PAEEOA>2.0.CO;2

Hooker, S.B., Van Heukelem, L., Thomas, C.S., Claustre, H., Ras, J., Barlow, R., Sessions, H., Schlüter, L., Perl, J., Trees, C., Stuart, V., Head, E., Clementson, L., Fishwick, J., Llewellyn, C., Aiken, J., 2005. The Second SeaWiFS HPLC Analysis Round-Robin Experiment (SeaHARRE-2), NASA Technical Memorandum.

Hooker, S.B., Van Heukelem, L., Thomas, C.S., Claustre, H., Ras, J., Schlüter, L., Clementson, L., Van Der Linde, D., Eker-Develi, E., Berthon, J.-F., Barlow, R., Sessions, H., Ismail, H., Perl, J., 2009. The Third SeaWiFS HPLC Analysis Round-Robin Experiment (SeaHARRE-3).

Hovis, W.A., Clark, D.K., Anderson, F., Austin, R.W., Wilson, W.H., Baker, E.T., Ball, D., Gordon, H.R., Mueller, J.L., El-Sayed, S.Z., Sturm, B., Wrigley, R.C., Yentsch, C.S., 1980. Nimbus-

7 Coastal Zone Color Scanner: System Description and Initial Imagery. Science 210, 60–63. https://doi.org/10.1126/science.210.4465.60
Hsieh, W.W., 2009. Machine Learning Methods in the Environmental Sciences: Neural Networks and Kernels. Cambridge University Press, USA.
Hu, C., Lee, Z., Franz, B., 2012. Chlorophyll *a* algorithms for oligotrophic oceans: A novel approach based on three-band reflectance difference. Journal of Geophysical Research: Oceans 117. https://doi.org/10.1029/2011JC007395
Huot, Y, Babin, M., Bruyant, F., Grob, C., Twardowski, M.S., Claustre, H., 2007. Does chlorophyll a provide the best index of phytoplankton biomass for primary productivity studies? Biogeosciences Discussions 4, 707–745.
Huot, Yannick, Brown, C.A., Cullen, J.J., 2007. Retrieval of phytoplankton biomass from simultaneous inversion of reflectance, the diffuse attenuation coefficient, and Sun-induced fluorescence in coastal waters. Journal of Geophysical Research: Oceans 112, 1–26. https://doi.org/10.1029/2006JC003794
Ioannou, I., Gilerson, A., Gross, B., Moshary, F., Ahmed, S., 2013. Deriving ocean color products using neural networks. Remote Sensing of Environment 134, 78–91. https://doi.org/10.1016/j.rse.2013.02.015
IOCCG, 2019. Uncertainties in Ocean Colour Remote Sensing. International Ocean-Colour Coordinating Group (IOCCG), Dartmouth, Canada.
Jaelani, L.M., Matsushita, B., Yang, W., Fukushima, T., 2013. Evaluation of four MERIS atmospheric correction algorithms in Lake Kasumigaura, Japan. International Journal of Remote Sensing 34, 8967–8985. https://doi.org/10.1080/01431161.2013.860660
Jerlov, N.G., 1968. Optical oceanography. Limnology and Oceanography 13, 731–732. https://doi.org/10.4319/lo.1968.13.4.0731
Jiang, D., Matsushita, B., Yang, W., 2020. A simple and effective method for removing residual reflected skylight in above-water remote sensing reflectance measurements. ISPRS Journal of Photogrammetry and Remote Sensing 165, 16–27. https://doi.org/10.1016/j.isprsjprs.2020.05.003
Johnsen, G., Nelson, N.B., Jovine, R.V.M., Prezelin, B.B., 1994. Chromoprotein-and pigment-dependent modeling of spectral light absorption in two dinoflagellates, Prorocentrum minimum and Heterocapsa pygmaea.
Jones, D.R., Schonlau, M., Welch, W.J., 1998. Efficient Global Optimization of Expensive Black-Box Functions. Journal of Global Optimization 13, 455–492. https://doi.org/10.1023/A:1008306431147
Jonson, J.E., Borken-Kleefeld, J., Simpson, D., Nyíri, A., Posch, M., Heyes, C., 2017. Impact of excess NO $\less$sub$\greater$x$\less$/sub$\greater$ emissions from diesel cars on air quality, public health and eutrophication in Europe. Environ. Res. Lett. 12, 094017. https://doi.org/10.1088/1748-9326/aa8850
Jordan, M.I., Saul, L.K., 1999. An Introduction to Variational Methods for Graphical Models.
Jorge, D.S.F., Loisel, H., Jamet, C., Dessailly, D., Demaria, J., Bricaud, A., Maritorena, S., Zhang, X., Antoine, D., Kutser, T., Bélanger, S., Brando, V.O., Werdell, J., Kwiatkowska, E., Mangin, A., d'Andon, O.F., 2021. A three-step semi analytical algorithm (3SAA) for estimating inherent optical properties over oceanic, coastal, and inland waters from remote sensing reflectance. Remote Sensing of Environment 263, 112537. https://doi.org/10.1016/j.rse.2021.112537
Kallio, K., Koponen, S., Ylöstalo, P., Kervinen, M., Pyhälahti, T., Attila, J., 2015. Validation of MERIS spectral inversion processors using reflectance, IOP and water quality measurements in boreal lakes. Remote Sensing of Environment 157, 147–157. https://doi.org/10.1016/j.rse.2014.06.016
Kasprzak, P., Padisák, J., Koschel, R., Krienitz, L., Gervais, F., 2008. Chlorophyll a concentration across a trophic gradient of lakes: An estimator of phytoplankton biomass? Limnologica 38, 327–338. https://doi.org/10.1016/j.limno.2008.07.002
Ke, G., Meng, Q., Finley, T., Wang, T., Chen, W., Ma, W., Ye, Q., Liu, T.-Y., 2017. LightGBM: A Highly Efficient Gradient Boosting Decision Tree, in: Guyon, I., Luxburg, U.V., Bengio, S.,

Wallach, H., Fergus, R., Vishwanathan, S., Garnett, R. (Eds.), Advances in Neural Information Processing Systems 30. Curran Associates, Inc., pp. 3146–3154.
Keiner, L.E., Yan, X.-H., 1998. A Neural Network Model for Estimating Sea Surface Chlorophyll and Sediments from Thematic Mapper Imagery. Remote Sensing of Environment 66, 153–165. https://doi.org/10.1016/S0034-4257(98)00054-6
Keller, A.A., Cavallaro, L., 2008. Assessing the US Clean Water Act 303(d) listing process for determining impairment of a waterbody. Journal of Environmental Management 86, 699–711. https://doi.org/10.1016/j.jenvman.2006.12.013
Keller, S., Maier, P.M., Riese, F.M., Norra, S., Holbach, A., Börsig, N., Wilhelms, A., Moldaenke, C., Zaake, A., Hinz, S., 2018. Hyperspectral Data and Machine Learning for Estimating CDOM, Chlorophyll a, Diatoms, Green Algae and Turbidity. International Journal of Environmental Research and Public Health 15, 1881. https://doi.org/10.3390/ijerph15091881
Kiefer, I., Odermatt, D., Anneville, O., Wüest, A., Bouffard, D., 2015. Application of remote sensing for the optimization of in-situ sampling for monitoring of phytoplankton abundance in a large lake. Science of the Total Environment 527–528, 493–506. https://doi.org/10.1016/j.scitotenv.2015.05.011
Kirk, J.T.O., 1994. Light and photosynthesis in aquatic ecosystems, third edition, Light and Photosynthesis in Aquatic Ecosystems, third edition. Cambridge University Press. https://doi.org/10.1017/CBO9781139168212
Kirk, J.T.O., 1976. Yellow substance (gelbstoff) and its contribution to the attenuation of photosynthetically active radiation in some inland and coastal south-eastern Australian waters. Mar. Freshwater Res. 27, 61–71. https://doi.org/10.1071/mf9760061
Krasnopolsky, V., Chalikov, D., Tolman, H., 2002. A neural network technique to improve computational efficiency of numerical oceanic models. Ocean Modelling 4, 363–383. https://doi.org/10.1016/S1463-5003(02)00010-0
Krasnopolsky, V., Nadiga, S., Mehra, A., Bayler, E., 2018. Adjusting Neural Network to a Particular Problem: Neural Network-Based Empirical Biological Model for Chlorophyll Concentration in the Upper Ocean. Applied Computational Intelligence and Soft Computing 2018, 1–10. https://doi.org/10.1155/2018/7057363
Krasnopolsky, V.M., Schiller, H., 2003. Some neural network applications in environmental sciences. Part I: Forward and inverse problems in geophysical remote measurements. Neural Networks 16, 321–334. https://doi.org/10.1016/S0893-6080(03)00027-3
Kravitz, J., Matthews, M., Bernard, S., Griffith, D., 2020. Application of Sentinel 3 OLCI for chl-a retrieval over small inland water targets: Successes and challenges. Remote Sensing of Environment 237, 111562. https://doi.org/10.1016/j.rse.2019.111562
Kruse, F.A., Lefkoff, A.B., Boardman, J.W., Heidebrecht, K.B., Shapiro, A.T., Barloon, P.J., Goetz, A.F.H., 1993. The spectral image processing system (SIPS)-interactive visualization and analysis of imaging spectrometer data, in: AIP Conference Proceedings. AIP Publishing, pp. 192–201. https://doi.org/10.1063/1.44433
Kuchinke, C.P., Gordon, H.R., Franz, B.A., 2009. Spectral optimization for constituent retrieval in Case 2 waters I: Implementation and performance. Remote Sensing of Environment 113, 571–587. https://doi.org/10.1016/j.rse.2008.11.001
Kuleshov, V., Fenner, N., Ermon, S., 2018. Accurate uncertainties for deep learning using calibrated regression. 35th International Conference on Machine Learning, ICML 2018 6, 4369–4377.
Kutser, T., Paavel, B., Verpoorter, C., Kauer, T., Vahtmäe, E., 2012. Remote Sensing of Water Quality in Optically Complex Lakes. The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences XXXIX-B8, 165–169. https://doi.org/10.5194/isprsarchives-xxxix-b8-165-2012
Kutser, T., Paavel, B., Verpoorter, C., Ligi, M., Soomets, T., Toming, K., Casal, G., 2016. Remote sensing of black lakes and using 810 nm reflectance peak for retrieving water quality parameters of optically complex waters. Remote Sensing 8. https://doi.org/10.3390/rs8060497
Kutser, T., Vahtmäe, E., Paavel, B., Kauer, T., 2013. Removing glint effects from field radiometry data measured in optically complex coastal and inland waters. Remote Sensing of Environment 133, 85–89. https://doi.org/10.1016/j.rse.2013.02.011

Lakshminarayanan, B., Pritzel, A., Blundell, C., 2017. Simple and Scalable Predictive Uncertainty Estimation using Deep Ensembles, in: Proceedings of the 31st International Conference on Neural Information Processing Systems. pp. 6405–6416.

Lee, Z., Ahn, Y.-H., Mobley, C., Arnone, R., 2010. Removal of surface-reflected light for the measurement of remote-sensing reflectance from an above-surface platform. Opt. Express 18, 26313–26324. https://doi.org/10.1364/OE.18.026313

Lee, Z., Carder, K.L., Arnone, R.A., 2002. Deriving inherent optical properties from water color: a multiband quasi-analytical algorithm for optically deep waters. Appl. Opt. 41, 5755–5772. https://doi.org/10.1364/AO.41.005755

Lee, Z., Pahlevan, N., Ahn, Y.-H., Greb, S., O'Donnell, D., 2013. Robust approach to directly measuring water-leaving radiance in the field. Appl. Opt., AO 52, 1693–1701. https://doi.org/10.1364/AO.52.001693

Lehmann, M.K., Nguyen, U., Allan, M., der Woerd, H.J., 2018. Colour Classification of 1486 Lakes across a Wide Range of Optical Water Types. Remote Sensing 10(8), 1-22. https://doi.org/10.3390/rs10081273

Leung, H., Haykin, S., 1991. The complex backpropagation algorithm. IEEE Transactions on Signal Processing 3(9), 2101-2104.

Lewis, D.D., 1998. Naive (Bayes) at forty: The independence assumption in information retrieval, in: Nédellec, C., Rouveirol, C. (Eds.), Machine Learning: ECML-98. Springer Berlin Heidelberg, Berlin, Heidelberg, pp. 4–15.

Lewis, W.M., Wurtsbaugh, W.A., Paerl, H.W., 2011. Rationale for Control of Anthropogenic Nitrogen and Phosphorus to Reduce Eutrophication of Inland Waters. Environmental Science & Technology 45, 10300–10305. https://doi.org/10.1021/es202401p

Li, Linhai, Li, Lin, Song, K., 2015. Remote sensing of freshwater cyanobacteria: An extended IOP Inversion Model of Inland Waters (IIMIW) for partitioning absorption coefficient and estimating phycocyanin. Remote Sensing of Environment 157, 9–23. https://doi.org/10.1016/j.rse.2014.06.009

Li, Linhai, Li, Lin, Song, K., Li, Y., Tedesco, L.P., Shi, K., Li, Z., 2013. An inversion model for deriving inherent optical properties of inland waters: Establishment, validation and application. Remote Sensing of Environment 135, 150–166. https://doi.org/10.1016/j.rse.2013.03.031

Liashchynskyi, Petro, Liashchynskyi, Pavlo, 2019. Grid Search, Random Search, Genetic Algorithm: A Big Comparison for NAS. arXiv:1912.06059.

Liu, X., Steele, C., Simis, S., Warren, M., Tyler, A., Spyrakos, E., Selmes, N., Hunter, P., 2021. Retrieval of Chlorophyll-a concentration and associated product uncertainty in optically diverse lakes and reservoirs. Remote Sensing of Environment 267, 112710. https://doi.org/10.1016/j.rse.2021.112710

Loisel, H., Stramski, D., 2000. Estimation of the inherent optical properties of natural waters from the irradiance attenuation coefficient and reflectance in the presence of Raman scattering. Appl. Opt., AO 39, 3001–3011. https://doi.org/10.1364/AO.39.003001

Longhi, M.L., Beisner, B.E., 2009. Environmental factors controlling the vertical distribution of phytoplankton in lakes. Journal of Plankton Research 31, 1195–1207. https://doi.org/10.1093/plankt/fbp065

Lundberg, S., Lee, S.-I., 2017. A Unified Approach to Interpreting Model Predictions. Advances in Neural Information Processing Systems 2017-December, 4766–4775.

Lundberg, S., Lee, S.-I., 2016. An unexpected unity among methods for interpreting model predictions.

Lundberg, S.M., Erion, G., Chen, H., DeGrave, A., Prutkin, J.M., Nair, B., Katz, R., Himmelfarb, J., Bansal, N., Lee, S.-I., 2020. From local explanations to global understanding with explainable AI for trees. Nature Machine Intelligence 2, 56–67. https://doi.org/10.1038/s42256-019-0138-9

Lundberg, S.M., Erion, G.G., Lee, S.-I., 2018. Consistent Individualized Feature Attribution for Tree Ensembles. arXiv.

Lutz, V.A., 2001. Changes in the In Vivo Absorption and Fluorescence Excitation Spectra with Growth Irradiance in Three Species of Phytoplankton. Journal of Plankton Research 23, 555–569. https://doi.org/10.1093/plankt/23.6.555

Maberly, S.C., O'Donnell, R.A., Woolway, R.I., Cutler, M.E.J., Gong, M., Jones, I.D., Merchant, C.J., Miller, C.A., Politi, E., Scott, E.M., Thackeray, S.J., Tyler, A.N., 2020. Global lake thermal regions shift under climate change. Nat Commun 11, 1232. https://doi.org/10.1038/s41467-020-15108-z

MacKay, D.J.C., 1992. Bayesian Interpolation. Neural Computation 4, 415–447. https://doi.org/10.1162/neco.1992.4.3.415

Maier, P.M., Keller, S., Hinz, S., 2021. Deep Learning with WASI Simulation Data for Estimating Chlorophyll a Concentration of Inland Water Bodies. Remote Sensing 13, 718. https://doi.org/10.3390/rs13040718

Mainstone, C.P., Parr, W., 2002. Phosphorus in rivers — ecology and management. Science of The Total Environment 282–283, 25–47. https://doi.org/10.1016/S0048-9697(01)00937-8

Manzo, C., Bresciani, M., Giardino, C., Braga, F., Bassani, C., 2015. Sensitivity analysis of a bio-optical model for Italian lakes focused on Landsat-8, Sentinel-2 and Sentinel-3. European Journal of Remote Sensing 48, 17–32. https://doi.org/10.5721/EuJRS20154802

Martinez-Cantin, R., 2014. BayesOpt: A Bayesian Optimization Library for Nonlinear Optimization, Experimental Design and Bandits. Journal of Machine Learning Research 15, 3735–3739.

Matarrese, R., Morea, A., Tijani, K., DePasquale, V., Chiaradia, M.T., Pasquariello, G., 2008. A specialized support vector machine for coastal water chlorophyll retrieval from water leaving reflectances. International Geoscience and Remote Sensing Symposium (IGARSS) 4, IV-910-IV–913. https://doi.org/10.1109/IGARSS.2008.4779871

Matsushita, B., Yang, W., Yu, G., Oyama, Y., Yoshimura, K., Fukushima, T., 2015. A hybrid algorithm for estimating the chlorophyll-a concentration across different trophic states in Asian inland waters. ISPRS Journal of Photogrammetry and Remote Sensing 102, 28–37. https://doi.org/10.1016/j.isprsjprs.2014.12.022

Matthews, M.W., 2014. Eutrophication and cyanobacterial blooms in South African inland waters: 10years of MERIS observations. Remote Sensing of Environment 155, 161–177. https://doi.org/10.1016/j.rse.2014.08.010

Matthews, M.W., 2011. A current review of empirical procedures of remote sensing in inland and near-coastal transitional waters. International Journal of Remote Sensing 32, 6855–6899. https://doi.org/10.1080/01431161.2010.512947

Matthews, M.W., Bernard, S., 2013. Characterizing the absorption properties for remote sensing of three small optically-diverse south african reservoirs. Remote Sensing 5, 4370–4404. https://doi.org/10.3390/rs5094370

Matthews, M.W., Bernard, S., Robertson, L., 2012. An algorithm for detecting trophic status (chlorophyll-a), cyanobacterial-dominance, surface scums and floating vegetation in inland and coastal waters. Remote Sensing of Environment 124, 637–652. https://doi.org/10.1016/j.rse.2012.05.032

Matthews, M.W., Odermatt, D., 2015. Improved algorithm for routine monitoring of cyanobacteria and eutrophication in inland and near-coastal waters. Remote Sensing of Environment 156, 374–382. https://doi.org/10.1016/j.rse.2014.10.010

McKee, D., Röttgers, R., Neukermans, G., Calzado, V.S., Trees, C., Ampolo-Rella, M., Neil, C., Cunningham, A., 2014. Impact of measurement uncertainties on determination of chlorophyll-specific absorption coefficient for marine phytoplankton. Journal of Geophysical Research: Oceans 119, 9013–9025. https://doi.org/10.1002/2014JC009909

McKinna, L.I.W., Cetinić, I., Werdell, P.J., 2021. Development and Validation of an Empirical Ocean Color Algorithm with Uncertainties: A Case Study with the Particulate Backscattering Coefficient. Journal of Geophysical Research: Oceans 126, e2021JC017231. https://doi.org/10.1029/2021JC017231

Mélin, F., Franz, B.A., 2014. Assessment of satellite ocean colour radiometry and derived geophysical products, in: Experimental Methods in the Physical Sciences. Academic Press, pp. 609–638. https://doi.org/10.1016/B978-0-12-417011-7.00020-9

Mélin, F., Sclep, G., Jackson, T., Sathyendranath, S., 2016. Uncertainty estimates of remote sensing reflectance derived from comparison of ocean color satellite data sets. Remote Sensing of Environment 177, 107–124. https://doi.org/10.1016/j.rse.2016.02.014

Mélin, F., Vantrepotte, V., 2015. How optically diverse is the coastal ocean? Remote Sensing of Environment 160, 235–251. https://doi.org/10.1016/j.rse.2015.01.023

Mélin, F., Vantrepotte, V., Clerici, M., D'Alimonte, D., Zibordi, G., Berthon, J.-F., Canuti, E., 2011. Multi-sensor satellite time series of optical properties and chlorophyll-a concentration in the Adriatic Sea. Progress in Oceanography 91, 229–244. https://doi.org/10.1016/j.pocean.2010.12.001

Merchant, C.J., Paul, F., Popp, T., Ablain, M., Bontemps, S., Defourny, P., Hollmann, R., Lavergne, T., Laeng, A., De Leeuw, G., Mittaz, J., Poulsen, C., Povey, A.C., Reuter, M., Sathyendranath, S., Sandven, S., Sofieva, V.F., Wagner, W., 2017. Uncertainty information in climate data records from Earth observation. Earth Syst. Sci. Data 9, 511–527. https://doi.org/10.5194/essd-9-511-2017

Merel, S., Walker, D., Chicana, R., Snyder, S., Baurès, E., Thomas, O., 2013. State of knowledge and concerns on cyanobacterial blooms and cyanotoxins. Environment International 59, 303–327. https://doi.org/10.1016/j.envint.2013.06.013

Michelutti, N., Blais, J.M., Cumming, B.F., Paterson, A.M., Rühland, K., Wolfe, A.P., Smol, J.P., 2010. Do spectrally inferred determinations of chlorophyll a reflect trends in lake trophic status? J Paleolimnol 43, 205–217. https://doi.org/10.1007/s10933-009-9325-8

Minaudo, C., Odermatt, D., Bouffard, D., Rahaghi, A.I., Lavanchy, S., Wüest, A., 2021. The Imprint of Primary Production on High-Frequency Profiles of Lake Optical Properties. Environ. Sci. Technol. 55, 14234–14244. https://doi.org/10.1021/acs.est.1c02585

Mishchenko, M.I., 2014. Directional radiometry and radiative transfer: The convoluted path from centuries-old phenomenology to physical optics. Journal of Quantitative Spectroscopy and Radiative Transfer, Electromagnetic and Light Scattering by Nonspherical Particles XIV 146, 4–33. https://doi.org/10.1016/j.jqsrt.2014.02.033

Mishra, S., Mishra, D.R., 2014. A novel remote sensing algorithm to quantify phycocyanin in cyanobacterial algal blooms. Environmental Research Letters 9, 114003. https://doi.org/10.1088/1748-9326/9/11/114003

Mishra, S., Mishra, D.R., 2012. Normalized difference chlorophyll index: A novel model for remote estimation of chlorophyll-a concentration in turbid productive waters. Remote Sensing of Environment 117, 394–406. https://doi.org/10.1016/j.rse.2011.10.016

Mitchell, T., 1997. Machine Learning. McGraw-Hill, New York.

Mittenzwey, K.-H., Ullrich, S., Gitelson, A.A., Kondratiev, K.Y., 1992. Determination of chlorophyll a of inland waters on the basis of spectral reflectance. Limnology and Oceanography 37, 147–149. https://doi.org/10.4319/lo.1992.37.1.0147

Mobley, C.D., 2022. The Oceanic Optics Book. International Ocean Colour Coordinating Group (IOCCG).

Mobley, C.D., Stramski, D., Bissett, W.P., Boss, E., S., 2004. Is the Case 1 – Case 2 classification still useful? Oceanography 17(2), 61-67.

Mobley, C.D., 1999. Estimation of the remote-sensing reflectance from above-surface measurements. Applied Optics 38, 7442. https://doi.org/10.1364/ao.38.007442

Mobley, C.D., 1994. Light and Water: Radiative Transfer in Natural Waters. Academic Press Inc, San Diego.

Mobley, C.D., Sundman, L.K., Bissett, W.P., Cahill, B., 2009. Fast and accurate irradiance calculations for ecosystem models. Biogeosciences Discussions 6, 10625–10662. https://doi.org/10.5194/bgd-6-10625-2009

Moore, T.S., Campbell, J.W., Dowell, M.D., 2009. A class-based approach to characterizing and mapping the uncertainty of the MODIS ocean chlorophyll product. Remote Sensing of Environment 113, 2424–2430. https://doi.org/10.1016/j.rse.2009.07.016

Moore, T.S., Campbell, J.W., Hui Feng, 2001. A fuzzy logic classification scheme for selecting and blending satellite ocean color algorithms. IEEE Transactions on Geoscience and Remote Sensing 39, 1764–1776. https://doi.org/10.1109/36.942555

Moore, T.S., Dowell, M.D., Bradt, S., Verdú, A.R., 2014. An optical water type framework for selecting and blending retrievals from bio-optical algorithms in lakes and coastal waters. Remote sensing of environment 143, 97—111. https://doi.org/10.1016/j.rse.2013.11.021

Morel, A., 1988. Optical modeling of the upper ocean in relation to its biogenous matter content (case I waters). Journal of Geophysical Research: Oceans 93, 10749–10768. https://doi.org/10.1029/JC093iC09p10749

Morel, A., Bricaud, A., 1981. Theoretical results concerning light absorption in a discrete medium, and application to specific absorption of phytoplankton. Deep Sea Research Part A. Oceanographic Research Papers 28, 1375–1393. https://doi.org/10.1016/0198-0149(81)90039-X

Morel, A., Prieur, L., 1977. Analysis of variations in ocean color1. Limnology and Oceanography 22, 709–722. https://doi.org/10.4319/lo.1977.22.4.0709

Moreno-Torres, J.G., Raeder, T., Alaiz-Rodríguez, R., Chawla, N. V., Herrera, F., 2012. A unifying view on dataset shift in classification. Pattern Recognition 45, 521–530. https://doi.org/10.1016/j.patcog.2011.06.019

Morley, S.K., Brito, T.V., Welling, D.T., 2018. Measures of Model Performance Based On the Log Accuracy Ratio. Space Weather 16, 69–88. https://doi.org/10.1002/2017SW001669

Moses, W.J., Gitelson, A.A., Berdnikov, S., Povazhnyy, V., 2009. Estimation of chlorophyll-a concentration in case II waters using MODIS and MERIS data - Successes and challenges. Environmental Research Letters 4, 45005. https://doi.org/10.1088/1748-9326/4/4/045005

Moses, W.J., Sterckx, S., Montes, M.J., De Keukelaere, L., Knaeps, E., 2017. Atmospheric Correction for Inland Waters, in: Bio-Optical Modeling and Remote Sensing of Inland Waters. Elsevier Inc., pp. 69–100. https://doi.org/10.1016/B978-0-12-804644-9.00003-3

Mou, L., Ghamisi, P., Zhu, X.X., 2017. Deep Recurrent Neural Networks for Hyperspectral Image Classification. IEEE Transactions on Geoscience and Remote Sensing 55, 3639–3655. https://doi.org/10.1109/TGRS.2016.2636241

Mouw, C.B., Chen, H., McKinley, G.A., Effler, S., O'Donnell, D., Perkins, M.G., Strait, C., 2013. Evaluation and optimization of bio-optical inversion algorithms for remote sensing of Lake Superior's optical properties. Journal of Geophysical Research: Oceans 118, 1696–1714. https://doi.org/10.1002/jgrc.20139

Mouw, C.B., Greb, S., Aurin, D., DiGiacomo, P.M., Lee, Z., Twardowski, M., Binding, C., Hu, C., Ma, R., Moore, T., Moses, W., Craig, S.E., 2015. Aquatic color radiometry remote sensing of coastal and inland waters: Challenges and recommendations for future satellite missions. Remote Sensing of Environment 160, 15–30.

Natekin, A., Knoll, A., 2013. Gradient boosting machines, a tutorial. Frontiers in Neurorobotics 7, 21. https://doi.org/10.3389/fnbot.2013.00021

Neal, R.M., 1996. Bayesian learning for neural networks. University of Toronto.

Neil, C., Spyrakos, E., Hunter, P.D., Tyler, A.N., 2019. A global approach for chlorophyll-a retrieval across optically complex inland waters based on optical water types. Remote Sensing of Environment 229, 159–178. https://doi.org/10.1016/j.rse.2019.04.027

Neville, R.A., Gower, J.F.R., 1977. Passive remote sensing of phytoplankton via chlorophyll α fluorescence. Journal of Geophysical Research 82, 3487–3493. https://doi.org/10.1029/JC082i024p03487

Niculescu-Mizil, A., Caruana, R., 2005. Predicting good probabilities with supervised learning, in: ICML 2005 - Proceedings of the 22nd International Conference on Machine Learning. ACM Press, New York, New York, USA, pp. 625–632. https://doi.org/10.1145/1102351.1102430

Nixon, J., Dusenberry, M., Jerfel, G., Nguyen, T., Liu, J., Zhang, L., Tran, D., 2019. Measuring Calibration in Deep Learning.

Nixon, S.W., 1995. Coastal marine eutrophication: A definition, social causes, and future concerns. Ophelia 41, 199–219. https://doi.org/10.1080/00785236.1995.10422044

Nouchi, V., Odermatt, D., Wüest, A., Bouffard, D., 2018. Effects of non-uniform vertical constituent profiles on remote sensing reflectance of oligo- to mesotrophic lakes. European Journal of Remote Sensing 51, 808–821. https://doi.org/10.1080/22797254.2018.1493360

Odermatt, D., Danne, O., Philipson, P., Brockmann, C., 2018. Diversity II water quality parameters from ENVISAT (2002-2012): A new global information source for lakes. Earth System Science Data 10, 1527–1549. https://doi.org/10.5194/essd-10-1527-2018

Odermatt, D., Giardino, C., Heege, T., 2010. Chlorophyll retrieval with MERIS Case-2-Regional in perialpine lakes. Remote Sensing of Environment 114, 607–617. https://doi.org/10.1016/j.rse.2009.10.016

Odermatt, D., Gitelson, A., Brando, V.E., Schaepman, M., 2012. Review of constituent retrieval in optically deep and complex waters from satellite imagery. Remote Sensing of Environment 118, 116–126. https://doi.org/10.1016/j.rse.2011.11.013

Odermatt, D., Heege, T., Nieke, J., Kneubühler, M., Itten, K., 2008. Water Quality Monitoring for Lake Constance with a Physically Based Algorithm for MERIS Data. Sensors 8, 4582–4599. https://doi.org/10.3390/s8084582

OECD, 1982. Eutrophication of Waters. Monitoring, Assessment and Control. Final Report of the OECD Cooperative Programme on Monitoring of Inland Waters (Eutrophication Control).

O'Reilly, J.E., Werdell, P.J., 2019. Chlorophyll algorithms for ocean color sensors - OC4, OC5 & OC6. Remote Sensing of Environment 229, 32–47. https://doi.org/10.1016/j.rse.2019.04.021

O'Reilly, Maritorena, S., Mitchell, B.G., Siegel, D.A., Carder, K.L., Garver, S.A., Kahru, M., McClain, C., 1998. Ocean color chlorophyll algorithms for SeaWiFS. Journal of Geophysical Research: Oceans 103, 24937–24953. https://doi.org/10.1029/98JC02160

O'Shea, R.E., Pahlevan, N., Smith, B., Bresciani, M., Egerton, T., Giardino, C., Li, L., Moore, T., Ruiz-Verdu, A., Ruberg, S., Simis, S.G.H., Stumpf, R., Vaičiūtė, D., 2021. Advancing cyanobacteria biomass estimation from hyperspectral observations: Demonstrations with HICO and PRISMA imagery. Remote Sensing of Environment 266, 112693. https://doi.org/10.1016/j.rse.2021.112693

Ovadia, Y., Research, G., Fertig, E., Ren, J., Nado Google Research, Z., Nowozin Google Research, S., Dillon Google Research, J. V, Lakshminarayanan, B., Snoek, J., 2019. Can You Trust Your Model's Uncertainty? Evaluating Predictive Uncertainty Under Dataset Shift, in: NIPS'19: Proceedings of the 33rd International Conference on Neural Information Processing Systems. pp. 14003–14014.

Pagano, T.S., Durham, R.M., 1993. Moderate Resolution Imaging Spectroradiometer (MODIS), in: Sensor Systems for the Early Earth Observing System Platforms. SPIE, pp. 2–17. https://doi.org/10.1117/12.152835

Pahlevan, N., Mangin, A., Balasubramanian, S. V., Smith, B., Alikas, K., Arai, K., Barbosa, C., Bélanger, S., Binding, C., Bresciani, M., Giardino, C., Gurlin, D., Fan, Y., Harmel, T., Hunter, P., Ishikaza, J., Kratzer, S., Lehmann, M.K., Ligi, M., Ma, R., Martin-Lauzer, F.R., Olmanson, L., Oppelt, N., Pan, Y., Peters, S., Reynaud, N., Sander de Carvalho, L.A., Simis, S., Spyrakos, E., Steinmetz, F., Stelzer, K., Sterckx, S., Tormos, T., Tyler, A., Vanhellemont, Q., Warren, M., 2021b. ACIX-Aqua: A global assessment of atmospheric correction methods for Landsat-8 and Sentinel-2 over lakes, rivers, and coastal waters. Remote Sensing of Environment 258, 112366. https://doi.org/10.1016/j.rse.2021.112366

Pahlevan, N., Smith, B., Alikas, K., Anstee, J., Barbosa, C., Binding, C., Bresciani, M., Cremella, B., Giardino, C., Gurlin, D., Fernandez, V., Jamet, C., Kangro, K., Lehmann, M.K., Loisel, H., Matsushita, B., Hà, N., Olmanson, L., Potvin, G., Simis, S.G.H., VanderWoude, A., Vantrepotte, V., Ruiz-Verdù, A., 2022. Simultaneous retrieval of selected optical water quality indicators from Landsat-8, Sentinel-2, and Sentinel-3. Remote Sensing of Environment 270, 112860. https://doi.org/10.1016/j.rse.2021.112860

Pahlevan, N., Smith, B., Binding, C., Gurlin, D., Li, L., Bresciani, M., Giardino, C., 2021a. Hyperspectral retrievals of phytoplankton absorption and chlorophyll-a in inland and nearshore coastal waters. Remote Sensing of Environment 253, 112200. https://doi.org/10.1016/j.rse.2020.112200

Pahlevan, N., Smith, B., Schalles, J., Binding, C., Cao, Z., Ma, R., Alikas, K., Kangro, K., Gurlin, D., Hà, N., Matsushita, B., Moses, W., Greb, S., Lehmann, M.K., Ondrusek, M., Oppelt, N., Stumpf, R., 2020. Seamless retrievals of chlorophyll-a from Sentinel-2 (MSI) and Sentinel-3 (OLCI) in inland and coastal waters: A machine-learning approach. Remote Sensing of Environment 240, 111604. https://doi.org/10.1016/j.rse.2019.111604

Palmer, S.C.J., Kutser, T., Hunter, P.D., 2015. Remote sensing of inland waters: Challenges, progress and future directions. Remote Sensing of Environment 157, 1–8. https://doi.org/10.1016/j.rse.2014.09.021

Paulino, R.S., Martins, V.S., Novo, E.M.L.M., Barbosa, C.C.F., de Carvalho, L.A.S., Begliomini, F.N., 2022. Assessment of Adjacency Correction over Inland Waters Using Sentinel-2 MSI Images. Remote Sensing 14, 1829. https://doi.org/10.3390/rs14081829

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M., Duchesnay, E., 2011. Scikit-learn: Machine Learning in Python. Journal of Machine Learning Research 12, 2825–2830.

Pepe, M., Giardino, C., Borsani, G., Cardoso, A.C., Chiaudani, G., Premazzi, G., Rodari, E., Zilioli, E., 2001. Relationship between apparent optical properties and photosynthetic pigments in the sub-alpine Lake Iseo. Science of the Total Environment 268, 31–45. https://doi.org/10.1016/S0048-9697(00)00691-4

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

Philipson, P., Kratzer, S., Ben Mustapha, S., Strömbeck, N., Stelzer, K., 2016. Satellite-based water quality monitoring in Lake Vänern, Sweden. International Journal of Remote Sensing 37, 3938–3960. https://doi.org/10.1080/01431161.2016.1204480

Platt, J.C., 1999. Probabilistic Outputs for Support Vector Machines and Comparisons to Regularized Likelihood Methods, in: Advances in Large Margin Classifiers. MIT Press, pp. 61--74.

Poikane, S., Alves, M.H., Argillier, C., Van Den Berg, M., Buzzi, F., Hoehn, E., De Hoyos, C., Karottki, I., Laplace-Treyture, C., Solheim, A.L., Ortiz-Casas, J., Ott, I., Phillips, G., Pilke, A., Pádua, J., Remec-Rekar, S., Riedmüller, U., Schaumburg, J., Serrano, M.L., Soszka, H., Tierney, D., Urbanic, G., Wolfram, G., 2010. Defining chlorophyll-a reference conditions in European Lakes. Environmental Management 45, 1286–1298. https://doi.org/10.1007/s00267-010-9484-4

Polley, E.C., van der Laan, M.J., 2011. Super Learning, in: van der Laan, M.J., Rose, S. (Eds.), . Springer, pp. 43–66. https://doi.org/10.1007/978-1-4419-9782-1

Rasmussen, C.E., Williams, C.K.I., 2005. Gaussian Processes for Machine Learning (Adaptive Computation and Machine Learning). The MIT Press.

Rast, M., Bezy, J.L., Bruzzi, S., 1999. The ESA Medium Resolution Imaging Spectrometer MERIS a review of the instrument and its mission. International Journal of Remote Sensing 20, 1681–1702. https://doi.org/10.1080/014311699212416

Reinersman, P.N., Carder, K.L., 1995. Monte Carlo simulation of the atmospheric point-spread function with an application to correction for the adjacency effect. Appl. Opt., AO 34, 4453–4471. https://doi.org/10.1364/AO.34.004453

Reynolds, C.S., 1984. The Ecology of Freshwater Phytoplankton. Cambridge University Press.

Riddick, C.A.L., Hunter, P.D., Tyler, A.N., Martinez-Vicente, V., Horváth, H., Kovács, A.W., Vörös, L., Preston, T., Présing, M., 2015. Spatial variability of absorption coefficients over a biogeochemical gradient in a large and optically complex shallow lake. Journal of Geophysical Research: Oceans 120, 7040–7066. https://doi.org/10.1002/2015JC011202

Ritchie, J.C., Cooper, C.M., Schiebe, F.R., 1990. The relationship of MSS and TM digital data with suspended sediments, chlorophyll, and temperature in Moon Lake, Mississippi. Remote Sensing of Environment 33, 137–148. https://doi.org/10.1016/0034-4257(90)90039-O

Roesler, C.S., Barnard, A.H., 2013. Optical proxy for phytoplankton biomass in the absence of photophysiology: Rethinking the absorption line height. Methods in Oceanography 7, 79–94. https://doi.org/10.1016/j.mio.2013.12.003

Roesler, C.S., Perry, M.J., Carder, K.L., 1989. Modeling in situ phytoplankton absorption from total absorption spectra in productive inland marine waters. Limnology and Oceanography 34, 1510–1523. https://doi.org/10.4319/lo.1989.34.8.1510

Röttgers, R., Gehnke, S., 2012. Measurement of light absorption by aquatic particles: Improvement of the quantitative filter technique by use of an integrating sphere approach. Applied Optics 51, 1336–1351. https://doi.org/10.1364/AO.51.001336

Ruddick, K.G., Gons, H.J., Rijkeboer, M., Tilstone, G., 2001. Optical remote sensing of chlorophyll a in case 2 waters by use of an adaptive two-band algorithm with optimal error properties. Appl. Opt., AO 40, 3575–3585. https://doi.org/10.1364/AO.40.003575

Ruder, S., 2017. An overview of gradient descent optimization algorithms. arXiv:1609.04747 [cs].

Ruiz-Verdú, A., Dominguez-Gomez, J.-A., Pena-Martinez, R., 2005. Use of CHRIS for Monitoring Water Quality in Rosarito Reservoir, in: ESA Special Publication, ESA Special Publication. p. 26.

Ruiz-Verdú, A., Simis, S.G.H., de Hoyos, C., Gons, H.J., Peña-Martínez, R., 2008. An evaluation of algorithms for the remote sensing of cyanobacterial biomass. Remote Sensing of Environment 112, 3996–4008. https://doi.org/10.1016/j.rse.2007.11.019

Russell, J.L., 1964. Kepler's Laws of Planetary Motion: 1609-1666. The British Journal for the History of Science 2, 1–24.

Salama, M.S., Stein, A., 2009. Error decomposition and estimation of inherent optical properties. Applied Optics 48, 4947–4962. https://doi.org/10.1364/AO.48.004947

Santer, R., Schmechtig, C., 2000. Adjacency effects on water surfaces: primary scattering approximation and sensitivity study. Appl. Opt., AO 39, 361–375. https://doi.org/10.1364/AO.39.000361

Saul, L.K., Jaakkola, T., Jordan, M.I., 1996. Mean field theory for sigmoid belief networks. Journal of Artificial Intelligence Research 4, 61–76. https://doi.org/10.1613/jair.251

Saulquin, B., Fablet, R., Bourg, L., Mercier, G., d'Andon, O.F., 2016. MEETC2: Ocean color atmospheric corrections in coastal complex waters using a Bayesian latent class model and potential for the incoming sentinel 3 — OLCI mission. Remote Sensing of Environment 172, 39–49. https://doi.org/10.1016/j.rse.2015.10.035

Sayers, M.J., Grimm, A.G., Shuchman, R.A., Deines, A.M., Bunnell, D.B., Raymer, Z.B., Rogers, M.W., Woelmer, W., Bennion, D.H., Brooks, C.N., Whitley, M.A., Warner, D.M., Mychek-Londer, J., 2015. A new method to generate a high-resolution global distribution map of lake chlorophyll. International Journal of Remote Sensing 36, 1942–1964. https://doi.org/10.1080/01431161.2015.1029099

Schaeffer, B., Salls, W., Coffer, M., Lebreton, C., Werther, M., Stelzer, K., Urquhart, E., Gurlin, D., 2022. Merging of the Case 2 Regional Coast Colour and Maximum-Peak Height chlorophyll-a algorithms: validation and demonstration of satellite-derived retrievals across US lakes. Environmental Monitoring and Assessment 194, 1–25. https://doi.org/10.1007/s10661-021-09684-w

Schaeffer, B.A., Schaeffer, K.G., Keith, D., Lunetta, R.S., Conmy, R., Gould, R.W., 2013. Barriers to adopting satellite remote sensing for water quality management. International Journal of Remote Sensing 34, 7534–7544. https://doi.org/10.1080/01431161.2013.823524

Schaffer, C., 1993. Selecting a classification method by cross-validation. Machine Learning 13, 135–143. https://doi.org/10.1007/BF00993106

Schalles, J.F., 2006. Optical remote sensing techniques to estimate phytoplankton chlorophyll a concentrations in coastal waters with varying suspended matter and cdom concentrations, in: Remote Sensing and Digital Image Processing, Remote Sensing and Digital Image Processing. Springer International Publishing, pp. 27–79.

Schalles, J.F., Hladik, C.M., 2012. Mapping phytoplankton chlorophyll in turbid, Case 2 estuarine and coastal waters. Israel Journal of Plant Sciences 60, 169–191. https://doi.org/10.1560/IJPS.60.1-2.169

Schapire, R.E., 2003. The Boosting Approach to Machine Learning: An Overview, in: Denison, D.D., Hansen, M.H., Holmes, C.C., Mallick, B., Yu, B. (Eds.), Nonlinear Estimation and Classification. Springer New York, New York, NY, pp. 149–171. https://doi.org/10.1007/978-0-387-21579-2_9

Schiller, H., 2007. Model inversion by parameter fit using NN emulating the forward model: evaluation of indirect measurements. Neural Netw 20, 479–483. https://doi.org/10.1016/j.neunet.2007.04.022

Schiller, H., Doerffer, R., 1999. Neural network for emulation of an inverse model operational derivation of Case II water properties from MERIS data. International Journal of Remote Sensing 20, 1735–1746. https://doi.org/10.1080/014311699212443

Schindler, D.W., 2012. The dilemma of controlling cultural eutrophication of lakes. Proceedings of the Royal Society B: Biological Sciences 279, 4322–4333. https://doi.org/10.1098/rspb.2012.1032

Schroeder, Th., Schaale, M., Fischer, J., 2007. Retrieval of atmospheric and oceanic properties from MERIS measurements: A new Case-2 water processor for BEAM. International Journal of Remote Sensing 28, 5627–5632. https://doi.org/10.1080/01431160701601774

Seegers, B.N., Stumpf, R.P., Schaeffer, B.A., Loftin, K.A., Werdell, P.J., 2018. Performance metrics for the assessment of satellite data products: an ocean color case study. Optics Express 26, 7404. https://doi.org/10.1364/oe.26.007404

Sharaf, N., Bresciani, M., Giardino, C., Faour, G., Slim, K., Fadel, A., 2019. Using Landsat and in situ data to map turbidity as a proxy of cyanobacteria in a hypereutrophic Mediterranean reservoir. Ecological Informatics 50, 197–206. https://doi.org/10.1016/j.ecoinf.2019.02.001

Shi, K., Li, Y., Li, L., Lu, H., Song, K., Liu, Z., Xu, Y., Li, Z., 2013. Remote chlorophyll-a estimates for inland waters based on a cluster-based classification. Science of the Total Environment 444, 1–15. https://doi.org/10.1016/j.scitotenv.2012.11.058

Shi, K., Zhang, Yunlin, Song, K., Liu, M., Zhou, Y., Zhang, Yibo, Li, Y., Zhu, G., Qin, B., 2019. A semi-analytical approach for remote sensing of trophic state in inland waters: Bio-optical mechanism and application. Remote Sensing of Environment 232, 111349. https://doi.org/doi.org/10.1016/j.rse.2019.111349

Simis, S.G.H., Olsson, J., 2013. Unattended processing of shipborne hyperspectral reflectance measurements. Remote Sensing of Environment 135, 202–212. https://doi.org/10.1016/j.rse.2013.04.001

Simis, S.G.H., Peters, S.W.M., Gons, H.J., 2005. Remote sensing of the cyanobacterial pigment phycocyanin in turbid inland water. Limnology and Oceanography 50, 237–245. https://doi.org/10.4319/lo.2005.50.1.0237

Simis, S.G.H., Ruiz-Verdú, A., Domínguez-Gómez, J.A., Peña-Martinez, R., Peters, S.W.M., Gons", H.J., 2007. Influence of phytoplankton pigment composition on remote sensing of cyanobacterial biomass. Remote Sensing of Environment 106, 414–427. https://doi.org/10.1016/j.rse.2006.09.008

Simis, S.G.H., Stelzer, K., Mueller, D., Selmes, N., 2020. Copernicus Global Land Service - Lake Water Quality. Algorithm Theoretical Basis Document. (No. 1.3.0-1.4.0). Copernicus.

Sinha, E., Michalak, A.M., Balaji, V., 2017. Eutrophication will increase during the 21st century as a result of precipitation changes. Science 357, 405–408. https://doi.org/10.1126/science.aan2409

Skakun, S., Wevers, J., Brockmann, C., Doxani, G., Aleksandrov, M., Batič, M., Frantz, D., Gascon, F., Gómez-Chova, L., Hagolle, O., López-Puigdollers, D., Louis, J., Lubej, M., Mateo-García, G., Osman, J., Peressutti, D., Pflug, B., Puc, J., Richter, R., Roger, J.-C., Scaramuzza, P., Vermote, E., Vesel, N., Zupanc, A., Žust, L., 2022. Cloud Mask Intercomparison eXercise (CMIX): An evaluation of cloud masking algorithms for Landsat 8 and Sentinel-2. Remote Sensing of Environment 274, 112990. https://doi.org/10.1016/j.rse.2022.112990

Smith, B., Pahlevan, N., Schalles, J., Ruberg, S., Errera, R., Ma, R., Giardino, C., Bresciani, M., Barbosa, C., Moore, T., Fernandez, V., Alikas, K., Kangro, K., 2021. A Chlorophyll-a Algorithm for Landsat-8 Based on Mixture Density Networks. Frontiers in Remote Sensing 1, 5. https://doi.org/10.3389/frsen.2020.623678

Smith, M.E., Robertson Lain, L., Bernard, S., 2018. An optimized Chlorophyll a switching algorithm for MERIS and OLCI in phytoplankton-dominated waters. Remote Sensing of Environment 215, 217–227. https://doi.org/10.1016/j.rse.2018.06.002

Smith, V.H., 2003. Eutrophication of freshwater and coastal marine ecosystems a global problem. Environmental Science and Pollution Research 10, 126–139. https://doi.org/10.1065/espr2002.12.142

Smith, V.H., Dodds, W.K., Havens, K.E., Engstrom, D.R., Paerl, H.W., Moss, B., Likens, G.E., 2014. Comment: Cultural eutrophication of natural lakes in the United States is real and widespread. Limnology and Oceanography 59, 2217–2225. https://doi.org/10.4319/lo.2014.59.6.2217

Smith, V.H., Joye, S.B., Howarth, R.W., 2006. Eutrophication of freshwater and marine ecosystems. Limnology and Oceanography 51, 351–355. https://doi.org/10.4319/lo.2006.51.1_part_2.0351

Snoek, J., Larochelle, H., Adams, R., 2012. Practical Bayesian Optimization of Machine Learning Algorithms. Advances in Neural Information Processing Systems 4.

Søndergaard, M., Jeppesen, E., Peder Jensen, J., Lildal Amsinck, S., 2005. Water Framework Directive: ecological classification of Danish lakes. Journal of Applied Ecology 42, 616–629. https://doi.org/10.1111/j.1365-2664.2005.01040.x

Soomets, T., Uudeberg, K., Jakovels, D., Brauns, A., Zagars, M., Kutser, T., 2020. Validation and Comparison of Water Quality Products in Baltic Lakes Using Sentinel-2 MSI and Sentinel-3 OLCI Data. Sensors 20, 742. https://doi.org/10.3390/s20030742

Sørensen, K., Grung, M., Röttgers, R., 2007. An intercomparison of in vitro chlorophyll a determinations for MERIS level 2 data validation. International Journal of Remote Sensing 28, 537–554. https://doi.org/10.1080/01431160600815533

Sosik, H., 2008. Characterizing seawater constituents from optical properties, op. cit.,UNESCO, pp. 281-32.

Soulignac, F., Danis, P.-A., Bouffard, D., Chanudet, V., Dambrine, E., Guénand, Y., Harmel, T., Ibelings, B.W., Trevisan, D., Uittenbogaard, R., Anneville, O., 2018. Using 3D modeling and remote sensing capabilities for a better understanding of spatio-temporal heterogeneities of phytoplankton abundance in large lakes. Journal of Great Lakes Research 44, 756–764. https://doi.org/10.1016/j.jglr.2018.05.008

Spyrakos, E., O'Donnell, R., Hunter, P.D., Miller, C., Scott, M., Simis, S.G.H., Neil, C., Barbosa, C.C.F., Binding, C.E., Bradt, S., Bresciani, M., Dall'Olmo, G., Giardino, C., Gitelson, A.A., Kutser, T., Li, L., Matsushita, B., Martinez-Vicente, V., Matthews, M.W., Ogashawara, I., Ruiz-Verdú, A., Schalles, J.F., Tebbs, E., Zhang, Y., Tyler, A.N., 2018a. Optical types of inland and coastal waters. Limnology and Oceanography 63, 846–870. https://doi.org/10.1002/lno.10674

Srivastava, N., Hinton, G., Krizhevsky, A., Salakhutdinov, R., 2014. Dropout: A Simple Way to Prevent Neural Networks from Overfitting, Journal of Machine Learning Research.

Stamnes, K., Hamre, B., Stamnes, S., Chen, N., Fan, Y., Li, W., Lin, Z., Stamnes, J., 2018. Progress in Forward-Inverse Modeling Based on Radiative Transfer Tools for Coupled Atmosphere-Snow/Ice-Ocean Systems: A Review and Description of the AccuRT Model. Applied Sciences 8, 2682. https://doi.org/10.3390/app8122682

Stedmon, C.A., Markager, S., Kaas, H., 2000. Optical properties and signatures of chromophoric dissolved organic matter (CDOM) in Danish coastal waters. Estuarine, Coastal and Shelf Science 51, 267–278. https://doi.org/10.1006/ecss.2000.0645

Steinmetz, F., Deschamps, P.-Y., Ramon, D., 2011. Atmospheric correction in presence of sun glint: application to MERIS. Optics Express 19, 9783. https://doi.org/10.1364/oe.19.009783

Sterckx, S., Knaeps, E., Ruddick, K., 2011. Detection and correction of adjacency effects in hyperspectral airborne data of coastal and inland waters: The use of the near infrared similarity spectrum. International Journal of Remote Sensing 32, 6479–6505. https://doi.org/10.1080/01431161.2010.512930

Sterner, R.W., Keeler, B., Polasky, S., Poudel, R., Rhude, K., Rogers, M., 2020. Ecosystem services of Earth's largest freshwater lakes. Ecosystem Services 41, 101046. https://doi.org/10.1016/j.ecoser.2019.101046

Stockwell, J.D., Doubek, J.P., Adrian, R., Anneville, O., Carey, C.C., Carvalho, L., De Senerpont Domis, L.N., Dur, G., Frassl, M.A., Grossart, H.-P., Ibelings, B.W., Lajeunesse, M.J., Lewandowska, A.M., Llames, M.E., Matsuzaki, S.-I.S., Nodine, E.R., Nõges, P., Patil, V.P., Pomati, F., Rinke, K., Rudstam, L.G., Rusak, J.A., Salmaso, N., Seltmann, C.T., Straile, D., Thackeray, S.J., Thiery, W., Urrutia-Cordero, P., Venail, P., Verburg, P., Woolway, R.I., Zohary, T., Andersen, M.R., Bhattacharya, R., Hejzlar, J., Janatian, N., Kpodonu, A.T.N.K., Williamson, T.J., Wilson, H.L., 2020. Storm impacts on phytoplankton community dynamics in lakes. Global Change Biology 26, 2756–2784. https://doi.org/10.1111/gcb.15033

Stramski, D., Boss, E., Bogucki, D., Voss, K.J., 2004. The role of seawater constituents in light backscattering in the ocean. Progress in Oceanography 61, 27–56. https://doi.org/10.1016/j.pocean.2004.07.001

Stramski, D., Kiefer, D.A., 1991. Light-Scattering By Microorganisms in the Open Ocean. Progress in Oceanography 28, 343–383.

Stramski, D., Reynolds, R.A., Kaczmarek, S., Uitz, J., Zheng, G., 2015. Correction of pathlength amplification in the filter-pad technique for measurements of particulate absorption coefficient in the visible spectral region. Applied Optics 54, 6763. https://doi.org/10.1364/ao.54.006763

Streltsov, S., Vakili, P., 1999. A Non-myopic Utility Function for Statistical Global Optimization Algorithms. Journal of Global Optimization 14, 283–298. https://doi.org/10.1023/A:1008284229931

Strobl, C., Boulesteix, A.L., Kneib, T., Augustin, T., Zeileis, A., 2008. Conditional variable importance for random forests. BMC Bioinformatics 9, 1–11. https://doi.org/10.1186/1471-2105-9-307

Strobl, C., Malley, J., Tutz, G., 2009. An Introduction to Recursive Partitioning: Rationale, Application and Characteristics of Classification and Regression Trees, Bagging and Random Forests. Psychol Methods 14, 323–348. https://doi.org/10.1037/a0016973

Suits, G.H., 1975. The nature of electromagnetic radiation., in: Reeves, R.G. (Ed.), Manual of Remote Sensing. American Society of Photogrammetry, Falls Church, Va., pp. 51–73.

Sydor, M., Gould, R.W., Arnone, R.A., Haltrin, V.I., Goode, W., 2004. Uniqueness in remote sensing of the inherent optical properties of ocean water. Applied Optics 43, 2156. https://doi.org/10.1364/AO.43.002156

Talling, J.F., 1986. The seasonality of phytoplankton in African lakes, in: Munawar, M., Talling, J.F. (Eds.), Seasonality of Freshwater Phytoplankton: A Global Perspective, Developments in Hydrobiology. Springer Netherlands, Dordrecht, pp. 139–160. https://doi.org/10.1007/978-94-009-4818-1_10

Tassan, S., Ferrari, G.M., 1995. An alternative approach to absorption measurements of aquatic particles retained on filters. Limnology and Oceanography 40, 1358–1368. https://doi.org/10.4319/lo.1995.40.8.1358

Tebbs, E.J., Avery, S.T., Chadwick, M.A., 2020. Satellite remote sensing reveals impacts from dam-associated hydrological changes on chlorophyll- *a* in the world's largest desert lake. River Research and Applications 36, 211–222. https://doi.org/10.1002/rra.3574

Tebbs, E.J., Remedios, J.J., Harper, D.M., 2013. Remote sensing of chlorophyll-a as a measure of cyanobacterial biomass in Lake Bogoria, a hypertrophic, saline–alkaline, flamingo lake, using Landsat ETM+. Remote Sensing of Environment 135, 92–106. https://doi.org/10.1016/j.rse.2013.03.024

Thackeray, S.J., Henrys, P.A., Hemming, D., Bell, J.R., Botham, M.S., Burthe, S., Helaouet, P., Johns, D.G., Jones, I.D., Leech, D.I., Mackay, E.B., Massimino, D., Atkinson, S., Bacon, P.J., Brereton, T.M., Carvalho, L., Clutton-Brock, T.H., Duck, C., Edwards, M., Elliott, J.M., Hall, S.J.G., Harrington, R., Pearce-Higgins, J.W., Høye, T.T., Kruuk, L.E.B., Pemberton, J.M., Sparks, T.H., Thompson, P.M., White, I., Winfield, I.J., Wanless, S., 2016. Phenological sensitivity to climate across taxa and trophic levels. Nature 535, 241–245. https://doi.org/10.1038/nature18608

Thépaut, J.-N., Dee, D., Engelen, R., Pinty, B., 2018. The Copernicus Programme and its Climate Change Service, in: IGARSS 2018 - 2018 IEEE International Geoscience and Remote Sensing Symposium. pp. 1591–1593. https://doi.org/10.1109/IGARSS.2018.8518067

Thornton, C., Hutter, F., Hoos, H.H., Leyton-Brown, K., 2012. Auto-WEKA: Automated Selection and Hyper-Parameter Optimization of Classification Algorithms. CoRR abs/1208.3719.

Ting, K.M., Witten, I.H., 1999. Issues in Stacked Generalization. Journal of Artificial Intelligence Research 10, 271–289. https://doi.org/10.1613/jair.594

Toming, K., Kutser, T., Laas, A., Sepp, M., Paavel, B., Nõges, T., 2016. First experiences in mapping lakewater quality parameters with sentinel-2 MSI imagery. Remote Sensing 8, 1–14. https://doi.org/10.3390/rs8080640

Tong, Y., Wang, M., Peñuelas, J., Liu, X., Paerl, H.W., Elser, J.J., Sardans, J., Couture, R.-M., Larssen, T., Hu, H., Dong, X., He, W., Zhang, W., Wang, X., Zhang, Y., Liu, Y., Zeng, S.,

Kong, X., Janssen, A.B.G., Lin, Y., 2020. Improvement in municipal wastewater treatment alters lake nitrogen to phosphorus ratios in populated regions. Proceedings of the National Academy of Sciences 117, 11566–11572. https://doi.org/10.1073/pnas.1920759117

Tran, K., Neiswanger, W., Yoon, J., Zhang, Q., Xing, E., Ulissi, Z.W., 2020. Methods for comparing uncertainty quantifications for material property predictions.

Tranvik, L.J., Downing, J.A., Cotner, J.B., Loiselle, S.A., Striegl, R.G., Ballatore, T.J., Dillon, P., Finlay, K., Fortino, K., Knoll, L.B., Kortelainen, P.L., Kutser, T., Larsen, Soren., Laurion, I., Leech, D.M., McCallister, S.L., McKnight, D.M., Melack, J.M., Overholt, E., Porter, J.A., Prairie, Y., Renwick, W.H., Roland, F., Sherman, B.S., Schindler, D.W., Sobek, S., Tremblay, A., Vanni, M.J., Verschoor, A.M., von Wachenfeldt, E., Weyhenmeyer, G.A., 2009. Lakes and reservoirs as regulators of carbon cycling and climate. Limnology and Oceanography 54, 2298–2314. https://doi.org/10.4319/lo.2009.54.6_part_2.2298

Tyler, A.N., Hunter, P.D., Spyrakos, E., Groom, S., Constantinescu, A.M., Kitchen, J., 2016. Developments in Earth observation for the assessment and monitoring of inland, transitional, coastal and shelf-sea waters. Science of the Total Environment 572, 1307–1321. https://doi.org/10.1016/j.scitotenv.2016.01.020

Urquhart, E.A., Schaeffer, B.A., Stumpf, R.P., Loftin, K.A., Werdell, P.J., 2017. A method for examining temporal changes in cyanobacterial harmful algal bloom spatial extent using satellite remote sensing. Harmful Algae 67, 144–152. https://doi.org/10.1016/j.hal.2017.06.001

van der Laan, M.J., Polley, E.C., Hubbard, A.E., 2007. Super Learner. Statistical Applications in Genetics and Molecular Biology 6. https://doi.org/10.2202/1544-6115.1309

Van Heukelem, L., Thomas, C.S., 2001. Computer-assisted high-performance liquid chromatography method development with applications to the isolation and analysis of phytoplankton pigments. Journal of Chromatography A 910, 31–49. https://doi.org/10.1016/S0378-4347(00)00603-4

Vanhellemont, Q., Ruddick, K., 2018. Atmospheric correction of metre-scale optical satellite data for inland and coastal water applications. Remote Sensing of Environment 216, 586–597. https://doi.org/10.1016/j.rse.2018.07.015

Vapnik, V.N., 1999. An overview of statistical learning theory. IEEE Transactions on Neural Networks 10, 988–999. https://doi.org/10.1109/72.788640

Vasistha, P., Ganguly, R., 2020. Water quality assessment of natural lakes and its importance: An overview. Materials Today: Proceedings, 3rd International Conference on Innovative Technologies for Clean and Sustainable Development 32, 544–552. https://doi.org/10.1016/j.matpr.2020.02.092

Vermote, E.F., Tanre, D., Deuze, J.L., Herman, M., Morcette, J.-J., 1997. Second Simulation of the Satellite Signal in the Solar Spectrum, 6S: an overview. IEEE Transactions on Geoscience and Remote Sensing 35, 675–686. https://doi.org/10.1109/36.581987

Verpoorter, C., Kutser, T., Seekell, D.A., Tranvik, L.J., 2014. A global inventory of lakes based on high-resolution satellite imagery. Geophysical Research Letters 41, 6396–6402. https://doi.org/10.1002/2014GL060641

Vörös, L., Padisák, J., 1991. Phytoplankton biomass and chlorophyll-a in some shallow lakes in central Europe. Hydrobiologia 215, 111–119. https://doi.org/10.1007/BF00014715

Wang, B., 2006. Cultural eutrophication in the Changjiang (Yangtze River) plume: History and perspective. Estuarine, Coastal and Shelf Science 69, 471–477. https://doi.org/10.1016/j.ecss.2006.05.010

Wang, D., Ma, R., Xue, K., Loiselle, S.A., 2019. The Assessment of Landsat-8 OLI Atmospheric Correction Algorithms for Inland Waters. Remote Sensing 11, 169. https://doi.org/10.3390/rs11020169

Wang, J., Lee, Z., Wang, D., Shang, S., Wei, J., Gilerson, A., 2021. Atmospheric correction over coastal waters with aerosol properties constrained by multi-pixel observations. Remote Sensing of Environment 265, 112633. https://doi.org/10.1016/j.rse.2021.112633

Wang, M., 1999. Atmospheric correction of ocean color sensors: computing atmospheric diffuse transmittance. Appl. Opt., AO 38, 451–455. https://doi.org/10.1364/AO.38.000451

Wang, S., Li, J., Zhang, B., Spyrakos, E., Tyler, A.N., Shen, Q., Zhang, F., Kuster, T., Lehmann, M.K., Wu, Y., Peng, D., 2018. Trophic state assessment of global inland waters using a MODIS-derived Forel-Ule index. Remote Sensing of Environment 217, 444–460.

Warren, M.A., Simis, S.G.H., Martinez-Vicente, V., Poser, K., Bresciani, M., Alikas, K., Spyrakos, E., Giardino, C., Ansper, A., 2019. Assessment of atmospheric correction algorithms for the Sentinel-2A MultiSpectral Imager over coastal and inland waters. Remote Sensing of Environment 225, 267–289. https://doi.org/10.1016/j.rse.2019.03.018

Warren, M.A., Simis, S.G.H., Selmes, N., 2021. Complementary water quality observations from high and medium resolution Sentinel sensors by aligning chlorophyll-a and turbidity algorithms. Remote Sensing of Environment 265, 112651. https://doi.org/10.1016/j.rse.2021.112651

Welschmeyer, N.A., 1994. Fluorometric analysis of chlorophyll a in the presence of chlorophyll b and pheopigments. Limnology and Oceanography 39, 1985–1992. https://doi.org/10.4319/lo.1994.39.8.1985

Werdell, P.J., McKinna, L.I.W., Boss, E., Ackleson, S.G., Craig, S.E., Gregg, W.W., Lee, Z., Maritorena, S., Roesler, C.S., Rousseaux, C.S., Stramski, D., Sullivan, J.M., Twardowski, M.S., Tzortziou, M., Zhang, X., 2018a. An overview of approaches and challenges for retrieving marine inherent optical properties from ocean color remote sensing. Progress in Oceanography. https://doi.org/10.1016/j.pocean.2018.01.001

Werther, M., Spyrakos, E., Simis, S.G.H., Odermatt, D., Stelzer, K., Krawczyk, H., Berlage, O., Hunter, P., Tyler, A., 2021. Meta-classification of remote sensing reflectance to estimate trophic status of inland and nearshore waters. ISPRS Journal of Photogrammetry and Remote Sensing 176, 109–126. https://doi.org/10.1016/j.isprsjprs.2021.04.003

Werther, M., Odermatt, D., Simis, S.G.H., Gurlin, D., Jorge, D.S.F., Loisel, H., Hunter, P.D., Tyler, A.N., Spyrakos, E., 2022a. Characterising retrieval uncertainty of chlorophyll-a algorithms in oligotrophic and mesotrophic lakes and reservoirs. ISPRS Journal of Photogrammetry and Remote Sensing 190, 279–300. https://doi.org/10.1016/j.isprsjprs.2022.06.015

Werther, M., Odermatt, D., Simis, S.G.H., Gurlin, D., Lehmann, M.K., Kutser, T., Gupana, R., Varley, A., Hunter, P.D., Tyler, A.N., Spyrakos, E., 2022b. A Bayesian approach for remote sensing of chlorophyll-a and associated retrieval uncertainty in oligotrophic and mesotrophic lakes. Remote Sensing of Environment 283, 113295. https://doi.org/10.1016/j.rse.2022.113295

Wevers, J., Müller, D., Scholze, J., Kirches, G., Quast, R., Brockmann, C., 2021. IdePix for Sentinel-2 MSI Algorithm Theoretical Basis Document. https://doi.org/10.5281/ZENODO.5788067

Williamson, C.E., Saros, J.E., Vincent, W.F., Smol, J.P., 2009. Lakes and reservoirs as sentinels, integrators, and regulators of climate change. Limnol. Oceanogr. 54, 2273–2282. https://doi.org/10.4319/lo.2009.54.6_part_2.2273

Wolpert, D.H., 1992. Stacked generalization. Neural Networks 5, 241–259. https://doi.org/10.1016/S0893-6080(05)80023-1

Wolters, E., Toté, C., Sterckx, S., Adriaensen, S., Henocq, C., Bruniquel, J., Scifoni, S., Dransfeld, S., 2021. iCOR Atmospheric Correction on Sentinel-3/OLCI over Land: Intercomparison with AERONET, RadCalNet, and SYN Level-2. Remote Sensing 13, 654. https://doi.org/10.3390/rs13040654

Wu, J., Chen, X.-Y., Zhang, H., Xiong, L.-D., Lei, H., Deng, S.-H., 2019. Hyperparameter Optimization for Machine Learning Models Based on Bayesian Optimizationb. Journal of Electronic Science and Technology 17, 26–40. https://doi.org/10.11989/JEST.1674-862X.80904120

Xi, H., Hieronymi, M., Krasemann, H., Röttgers, R., 2017. Phytoplankton Group Identification Using Simulated and In situ Hyperspectral Remote Sensing Reflectance. Frontiers in Marine Science 4, 272. https://doi.org/10.3389/fmars.2017.00272

Xi, H., Hieronymi, M., Röttgers, R., Krasemann, H., Qiu, Z., 2015. Hyperspectral Differentiation of Phytoplankton Taxonomic Groups: A Comparison between Using Remote Sensing Reflectance and Absorption Spectra. Remote Sensing 7, 14781–14805. https://doi.org/10.3390/rs71114781

Yacobi, Yosef Z, Moses, W.J., Kaganovsky, S., Sulimani, B., Leavitt, B.C., Gitelson, A.A., 2011. NIR-red reflectance-based algorithms for chlorophyll-a estimation in mesotrophic inland

and coastal waters: Lake Kinneret case study. Water Research 45, 2428–2436. https://doi.org/10.1016/j.watres.2011.02.002

Yang, J., Gong, P., Fu, R., Zhang, M., Chen, J., Liang, S., Xu, B., Shi, J., Dickinson, R., 2013. The role of satellite remote sensing in climate change studies. Nature Climate Change 3, 875–883. https://doi.org/10.1038/nclimate1908

Yao, J., Pan, W., Ghosh, S., Doshi-Velez, F., 2019. Quality of Uncertainty Quantification for Bayesian Neural Network Inference.

Yvon-Durocher, G., Jones, J.I., Trimmer, M., Woodward, G., Montoya, J.M., 2010. Warming alters the metabolic balance of ecosystems. Philos Trans R Soc Lond B Biol Sci 365, 2117–2126. https://doi.org/10.1098/rstb.2010.0038

Zadrozny, B., Elkan, C., 2002. Transforming classifier scores into accurate multiclass probability estimates, in: Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. Association for Computing Machinery (ACM), New York, New York, USA, pp. 694–699. https://doi.org/10.1145/775047.775151

Zadrozny, B., Elkan, C., 2001. Obtaining calibrated probability estimates from decision trees and naive Bayesian classifiers, in: Proceedings of the Eighteenth International Conference on Machine Learning. pp. 609–616.

Zaneveld, J.R.V., Barnard, A.H., Boss, E., 2005. Theoretical derivation of the depth average of remotely sensed optical parameters. Opt. Express 13, 9052–9061. https://doi.org/10.1364/OPEX.13.009052

Zhang, Y., Feng, L., Li, J., Luo, L., Yin, Y., Liu, M., Li, Y., 2010. Seasonal–spatial variation and remote sensing of phytoplankton absorption in Lake Taihu, a large eutrophic and shallow lake in China. Journal of Plankton Research 32, 1023–1037. https://doi.org/10.1093/plankt/fbq039

Zhang, Y., Zhang, B., Wang, X., Li, J., Feng, S., Zhao, Q., Liu, M., Qin, B., 2007. A study of absorption characteristics of chromophoric dissolved organic matter and particles in Lake Taihu, China. Hydrobiologia 592, 105–120. https://doi.org/10.1007/s10750-007-0724-4

Zheng, G., Stramski, D., 2013. A model for partitioning the light absorption coefficient of suspended marine particles into phytoplankton and nonalgal components. Journal of Geophysical Research: Oceans 118, 2977–2991. https://doi.org/10.1002/jgrc.20206

Zhou, Z., Hooker, G., 2021. Unbiased measurement of feature importance in tree-based methods. ACM Transactions on Knowledge Discovery from Data 15, 1–21. https://doi.org/10.1145/3429445

Zibordi, G., Voss, K.J., 2014. In situ optical radiometry in the visible and near infrared, in: Experimental Methods in the Physical Sciences. Academic Press, pp. 247–304. https://doi.org/10.1016/B978-0-12-417011-7.00010-6

Zilioli, E., Brivio, P.A., Gomarasca, M.A., 2014. A correlation between optical properties from satellite data and some indicators of eutrophication in Lake Garda (Italy). Science of the total environment 158, 127 - 133.

Zimmermann, G., Neumann, A., Suemnich, K.-H., Schwarzer, H.H., 1993. MOS/PRIRODA: an imaging VIS/NIR spectrometer for ocean remote sensing, in: Imaging Spectrometry of the Terrestrial Environment. SPIE, pp. 201–206. https://doi.org/10.1117/12.157057