Universität Zürich[UZH]

Geographisches Institut

Evaluating Bio-Optical Inversions of Sentinel-3 OLCI for Water Constituent Retrieval in Lake Geneva

GEO 511 Master's Thesis

Author: Samuel Nils Landolt, 21-704-382

Supervised by: Dr. Daniel Odermatt (daniel.odermatt@eawag.ch), Loé Maire

Faculty representative: Prof. Dr. Alexander Damm

27.08.2026

Acknowledgements

I would like to express my gratitude to my supervisors, Dr. Daniel Odermatt and Loé Maire, for their valuable input, feedback, and support throughout this thesis. I genuinely appreciated their guidance along the way. A special thank you goes to Loé, whose help with the colloquium presentation made a real difference.

Abstract

Freshwater ecosystems such as lakes are increasingly threatened by climate change and human activity. In Lake Geneva, algal blooms degrade water quality, posing direct threats to drinking water, humans, and animals through the presence of harmful species.

Bio-optical models such as the Water Colour Simulator (WASI) aim to quantify optically active constituents, namely phytoplankton, non-algal particles (NAP), and colored dissolved organic matter (CDOM). Coupled with satellite data, they can complement *in situ* monitoring with maps of constituent distribution at high spatial and temporal resolution.

This thesis coupled WASI with Sentinel-3 Ocean and Land Color Instrument (OLCI) imagery and validated the results against various *in situ* datasets available for Lake Geneva. To address WASI's computational limitations (notably long processing times), it developed MiniWASI, a standalone Python clone tailored towards operational retrieval of constituents and related inherent optical properties in optically deep Case 2 waters.

Applied to invert Sentinel-3 OLCI remote sensing reflectance for 365 images, it performed well in retrieving bulk chlorophyll *a* (CHL, r = 0.72 – 0.8; for datasets comparable to previous studies) and non-algal particles (r = 0.77 – 0.78) but was unable to reliably retrieve CDOM (r = 0.2) or differentiate between different phytoplankton taxa.

Validation showed accuracy similar to that of empirical, locally calibrated CHL retrieval algorithms. Retrieval errors for phytoplankton taxa and CDOM were attributed to the strong ill-posedness of the inverse problem, driven by low concentrations, limited spectral resolution, and a large number (seven) of unknown variables, as well as to errors in the input remote sensing reflectance. These results support the notion that retrieving individual phytoplankton taxa with multispectral sensors such as Sentinel-3 OLCI is challenging, if not mathematically impossible.

Proposed future research directions include quantifying retrieval uncertainties, applying and improving atmospheric correction and adjacency effect algorithms as well as extending the approach to hyperspectral missions.

Table of Contents

1. Introduction 1

1.1 Context 1

1.2 Motivation 1

1.3 Remote sensing as a monitoring tool 2

1.4 Remote sensing of Case 2 waters 2

1.4.1 Optically active constituents 3

1.4.2 Inherent and apparent optical properties 4

1.5 Constituent retrievals: State of research 5

1.6 Objectives 7

2. Data and Methods 8

2.1 Study area 8

2.2 *In situ* data: Sources and sampling stations 8

2.3 *In situ* data: Validation parameters 9

2.3.1 AOPs 9

2.3.2 CHL 10

2.3.3 Phytoplankton 11

2.3.4 TSM and CDOM 11

2.3.5 IOPs 11

2.3.6 Auxiliary data 12

2.4 *In situ* data: Additional quality control 12

2.5 Satellite data: Acquisition and processing 12

2.6 Matchups 13

2.7 Methods: Overview 14

2.8 Bio-optical model MiniWASI 15

2.8.1 Forward model 15

2.8.2 Inverse model 18

2.8.3 Model parametrization 19

2.9 Inversion of *in situ* remote sensing reflectance 21
2.10 Matchup validation 22
2.11 Operational processing chain 23
3. Results 24
3.1 AC performance 24
3.2 SHL2 27
3.2.1 Spectrophotometric CHL 27
3.2.2 Phytoplankton taxa 28
3.3 LéXPLORE 30
3.3.1 Bulk phytoplankton proxies 30
3.3.2 TSM and CDOM 33
3.3.3 Constituent time series and interrelationships 33
3.3.4 IOPs 36
3.4 Eawag field campaigns 40
4. Discussion 45
4.1 Bulk phytoplankton retrieval: Strengths and limitations 45
4.1.1 SHL2 validation and matchup representativeness 45
4.1.2 LéXPLORE: Intercomparison of *in situ* CHL and CPHY references 45
4.1.3 Factors limiting retrieval performance at LéXPLORE 46
4.1.4 Benchmarking against empirical algorithms and literature 47
4.1.5 Pigment composition and SIOP considerations 47
4.2 NAP: Stable retrieval 48
4.3 CDOM: A retrieval failure and its causes 49
4.4 IOPs: Implications for model configuration 50
4.5 Phytoplankton taxa retrieval: Limitations 53
4.6 AC: A limiting factor 53
4.7 Matchup representativeness 54
4.8 Future study directions and general limitations 55

4.8.1 Uncertainty 55
4.8.2 AC and AE advancements 55
4.8.3 Bio-optical modelling: Limitations and possible improvements 56
4.8.4 Hyperspectral missions: PACE 56
4.9 Conclusion 57
Bibliography 58
Appendices 64
Appendix 1: MiniWASI and WASI7 inversion performance 64
Appendix 2: Local relationship between aLH_{676} and CHL_F 68
Appendix 3: Relationship between modeled and measured *in situ* IOPs 69
Appendix 4: aLH_{676} profile on 5.9.2021 71
Appendix 5: Drift of the AC-S instrument 72
AI statement 73
Statement of authorship 74

List of Figures

Figure 1: *Uroglena* sp. bloom in Lake Geneva 2
Figure 2: True-color Level 2 Sentinel-2 A/B example images of optically active constituents in Case-2 waters 3
Figure 3: Overview map of Lake Geneva 9
Figure 4: Schematic overview for the validation of the three basic constituents 14
Figure 5: IOPs and SIOPs as configured with the Lake Constance pre-sets from Table 6. 21
Figure 6: Schematic diagram of the processing chain presented in Section 2. 23
Figure 7: Validation of the atmospheric correction applied to the Sentinel-3 imagery. 25
Figure 8: Validation of the Sentinel-3 remote sensing reflectance after it was re-fitted by the MiniWASI bio-optical inversion model 26
Figure 9: One-to-one validation of the satellite-modeled phytoplankton concentration 27
Figure 10: Time series at the SHL2 monitoring station, 2017 – 2024 28
Figure 11: Relative proportions of the five phytoplankton taxa groups retrieved by the MiniWASI bio-optical inversion model. 29
Figure 12: *In situ* phytoplankton biovolume plotted against the corresponding MiniWASI-modeled pigment concentration retrieved from Sentinel-3 30
Figure 13: Pairwise inter-comparison of four *in situ* estimates of chlorophyll *a* / phytoplankton concentration at LéXPLORE. 31
Figure 14: Validation of the satellite-modeled phytoplankton concentration 32
Figure 15: Validation of the satellite-modeled non-algal particle concentration and colored dissolved organic matter absorption 33
Figure 16: Pairwise Pearson correlation between the three water constituents 34
Figure 17: Time series (LéXPLORE) of satellite and *in situ* water constituent estimates by year. 35
Figure 18: Validation of satellite-modeled constituent absorption 37
Figure 19: Validation of the satellite-modeled constituent backscattering at 620 nm 38
Figure 20: Comparison of satellite-modeled and *in situ* backscattering spectral shapes. 39
Figure 21: Map-based retrievals from a Sentinel-3A image on 20 May 2021 40
Figure 22: Map-based retrievals from a Sentinel-3B image on 11 June 2021. 41
Figure 23: Map-based retrievals from a Sentinel-3A image on 16 June 2021. 43
Figure 24: Combined validation of satellite-modeled phytoplankton concentration and non-algal particle concentration 43
Figure 25: WASI forward-modeled remote sensing reflectance and constituent absorption for Lake Geneva “standard” concentrations 50
Figure 26: Measured particle backscattering coefficient spectra from the MALINA and ICESCAPE cruises 52

List of Tables

Table 1: Literature review of studies that retrieved optically active constituents using bio-optical models WASI and BOMBER 6
Table 2: Morphometric and watershed characteristics of Lake Geneva 8
Table 3: Used measurements obtained from the Thetis profiler. 10
Table 4: Sentinel-3 OLCI Marine User Handbook band description 13
Table 5: *In situ* reference variables used for validation 14
Table 6: References for the MiniWASI equations in the WASI (version 7) manual with lake- and condition-specific parametrization of the equations 20
Table 7: MiniWASI initial, minimum, and maximun values 20
Table 8: Band-by-band validation statistics for Sentinel-3 remote sensing reflectance. 26
Table 9: Validation statistics for the comparison in Figure 9 27
Table 10: Pearson correlation and significance between the MiniWASI-modeled, per-taxon phytoplankton pigment concentrations retrieved from Sentinel-3 and the corresponding *in situ* biovolume and abundance measurements. 29
Table 11: Validation statistics for the six pairwise *in situ* comparisons in Figure 13 31
Table 12: Validation statistics for the comparisons in Figure 14 32
Table 13: Validation statistics for the comparisons in Figure 15 33
Table 14: Validation statistics for satellite-modeled constituent absorption 36
Table 15: Validation statistics for satellite-modeled constituent backscattering from Sentinel-3A and 3B against the *in situ* estimate converted from scattering measurements at four paired wavelengths. 37
Table 16: *In situ* vs. satellite-modeled matchup values at the six sampling points of the 20 May 2021 field campaign. 41
Table 17: *In situ* vs. satellite-modeled matchup values at five of the six sampling points of the 11 June 2021 field campaign 42
Table 18: *In situ* vs. satellite-modeled matchup values at the six sampling points of the 16 June 2021 field campaign 42

Abbreviation Full Term

AC	Atmospheric Correction
AE	Adjacency Effect
AOP(s)	Apparent Optical Properties
BOMBER	Bio-Optical Model Based tool for Estimating water quality and bottom properties from Remote sensing images
CCI	Climate Change Initiative (ESA)
CDOM	Colored Dissolved Organic Matter
$CDOM_R$	Modeled Colored Dissolved Organic Matter from hyperspectral *in situ* measurements
$CDOM_{S3}$	Modeled Colored Dissolved Organic Matter from Sentinel-3
CHL	chlorophyll *a*
CHL_A	Absorption-corrected (fluorometric) chlorophyll *a*
CHL_F	Fluorometric chlorophyll *a*
CHL_{HPLC}	Measured chlorophyll *a* from laboratory HPLC analysis
CHL_S	Spectrophotometric (laboratory-measured) chlorophyll *a*
CIPEL	Commission Internationale pour la Protection des Eaux du Léman (International Commission for the Protection of Lake Geneva)
CPHY	Phytoplankton concentration
$CPHY_R$	Modeled phytoplankton concentration from hyperspectral *in situ* measurements
$CPHY_{HPLC}$	Measured phytoplankton concentration from laboratory HPLC analysis
$CPHY_{S3}$	Modeled phytoplankton concentration from Sentinel-3
Eawag	Swiss Federal Institute of Aquatic Sciences (Eidgenössische Anstalt für Wasserversorgung, Abwasserreinigung und Gewässerschutz)
ESA	European Space Agency
EUMETSAT	European Organisation for the Exploitation of Meteorological Satellites
FWHM	Full Width at Half Maximum
HPLC	High-Performance Liquid Chromatography
IOCCG	International Ocean Colour Coordinating Group
IOP(s)	Inherent Optical Properties
L1	Level 1 (data processing level)
L2	Level 2 (data processing level)
MdSA	Median Symmetric Accuracy
MiniWASI	Minimal WASI deep-water model
MODIS	Moderate Resolution Imaging Spectroradiometer (sensor on board of the Terra and Aqua satellites)
NAP	Non-Algal Particles
NAP_R	Modeled Non-Algal Particles from hyperspectral *in situ* measurements
NAP_{S3}	Modeled Non-Algal Particles from Sentinel-3
NPQ	Non-Photochemical Quenching
OAC(s)	Optically Active Constituent(s)
OCx	Ocean Color (algorithm family, including versions OC2 / OC3 / OC4)
OLCI	Ocean and Land Color Instrument
PACE	Plankton, Aerosol, Cloud, ocean Ecosystem (satellite mission)
PAR	Photosynthetically Active Radiation
POLYMER	POLYnomial-based algorithm applied to MERIS (atmospheric correction algorithm)
PRISMA	PRecursore IperSpettrale della Missione Applicativa (satellite mission)
RMSE	Root Mean Square Error
RS	Remote Sensing
S2(A/B)	Sentinel-2 (A/B) satellites
S3(A/B)	Sentinel-3 (A/B) satellites
SA	Spectral Angle
SHL2	Station Hydrobiologique Lémanique 2 (CIPEL sampling point)
SIOP(s)	Specific Inherent Optical Properties
TSM	Total Suspended Matter
WASI	Water Colour Simulator
WASI-AI	WASI artificial intelligence module
WASI-2D	WASI image module
WASI7	WASI version 7

1. Introduction

1.1 Context

Aquatic ecosystems provide essential services to wildlife and people, including drinking water, food, and recreational opportunities. Monitoring the state of these ecosystems is important for assessing their health, understanding biogeochemical processes, and protecting human and animal welfare. For example, monitoring can enable the early detection of harmful algal blooms (IOCCG, 2000).

This is particularly urgent for inland waters, such as lakes, which are under increasing threat from overfishing, habitat loss, the invasion of exotic species, micropollutants, and water quality degradation due to climate change and increased human activity (Jacquet et al., 2026; Moisset, 2017; Niroumand-Jadidi et al., 2021). Spaceborne remote sensing (RS) allows for frequent, large-scale observations of water quality through the identification of different constituents, which is valuable given their high spatial and temporal variability (Giardino et al., 2019).

1.2 Motivation

In Lake Geneva, poor water quality is primarily linked to algal blooms. Since the 1950s, around 23 significant algal blooms have been recorded (Jacquet et al., 2026). Globally, algal blooms have increased due to climate change (see e.g. Hou et al., 2022) and yet despite decreased phosphorous levels, the concentration of chlorophyll *a* (CHL) in Lake Geneva has remained largely constant since the 1980s. While this phenomenon is not fully understood, it is suspected to be linked to climate change (Moisset, 2017). Frequent algal blooms can lead to a "decrease in water transparency and the formation of hypoxic and anoxic zones (Moisset, 2017, p. 35)", which disrupts the equilibrium of lakes. This can lead to an increase in toxic phytoplankton species and a reduction in the abundance of fish and benthic species (Moisset, 2017).

Beyond their serious impact on ecosystem health, such blooms are directly relevant to Lake Geneva residents as they negatively impact ecosystem services. For example, a *Rhodomonas minuta* bloom in March 1990 negatively impacted drinking water quality (Jacquet et al., 2026), with the "La Suisse" newspaper proclaiming "Genève recrache" ("Geneva spits out") on the frontpage of the Geneva section, showing an image of a man spitting out tap water (Roth, 1990).

While largely harmless, a bloom of *Uroglena* sp. in September 2021 led to unpleasant odors and gave the water an unappealing brown color (Figure 1), impacting recreational activities around the lake (Irani Rahaghi et al., 2024; Jacquet et al., 2026). However, although the extent of the bloom was evident from satellite imagery, it could hardly be observed at the SHL2 (Station Hydrobiologique Lémanique 2) monitoring station, which is situated in an area where the water was relatively clear (Jacquet et al., 2026; see Figure 1).



Figure 1: Uroglena sp. bloom in Lake Geneva. At the top: true-color Sentinel-2 image from 6 September 2021, downloaded from the Copernicus browser (©Copernicus), showing sampling station SHL2. Scaling was adjusted to match the images in Figure 2. At the bottom: photographs (©sjacquet) taken in Thonon-les-Bains on the same day. From Jacquet et al. (2026), Figure 6, reproduced with permission from Springer Nature.

While water constituents such as phytoplankton have wide-ranging ecological and biogeochemical significance, the purpose of this thesis is not to evaluate their impact on ecosystem health or services or their connection to various biogeochemical processes. Instead, the goal is to contribute novel insights on how spaceborne RS can quantify constituents in lakes.

1.3 Remote sensing as a monitoring tool

Spaceborne RS is a valuable tool for monitoring water quality, complementing *in situ* measurements. While *in situ* data provides high-quality, point-based measurements, RS provides coarser estimates with broader spatial and temporal context (Giardino et al., 2019; Kiefer et al., 2015).

The Sentinel-3 (S3) Ocean and Land Color Instrument (OLCI) offers promising capabilities for aquatic remote sensing. It has a near-daily revisit frequency, 21 spectral bands (400 – 1020 nm), a spatial resolution of 300 m, a radiometric depth of 12 bits, and improved signal-to-noise performance (EUMETSAT, 2018).

1.4 Remote sensing of Case 2 waters

The RS of optically complex ("Case 2") waters has gained increasing attention in recent years as sensors specifically designed for water applications such as S3 OLCI have become available (Giardino et al., 2019).

Compared to marine ("Case 1") waters, the RS of Case 2 waters, which includes inland water bodies and coastal areas, is more difficult since constituents are not closely correlated and

highly diverse (IOCCG, 2000). This diversity is visually apparent across globally distributed Case 2 waters (Figure 2), from sediment-dominated deltas to phytoplankton-dominated lakes.

Giardino et al. (2019) list three major difficulties for multispectral spaceborne RS of (optically deep) Case 2 waters:

1. Ill-Posedness: The number of retrieved variables, which implicitly includes atmospheric parameters, is often higher than the number of available bands. Additionally, constituent groups (see Section 1.4.1) can vary greatly in their optical properties.
2. Water bodies are very dark, with only about 1% of light being reflected. Therefore, instrument noise and atmospheric correction (AC) errors disproportionately affect the retrieved spectra compared to land applications.
3. Masking effect: If one constituent dominates the others, it becomes difficult to quantify them.

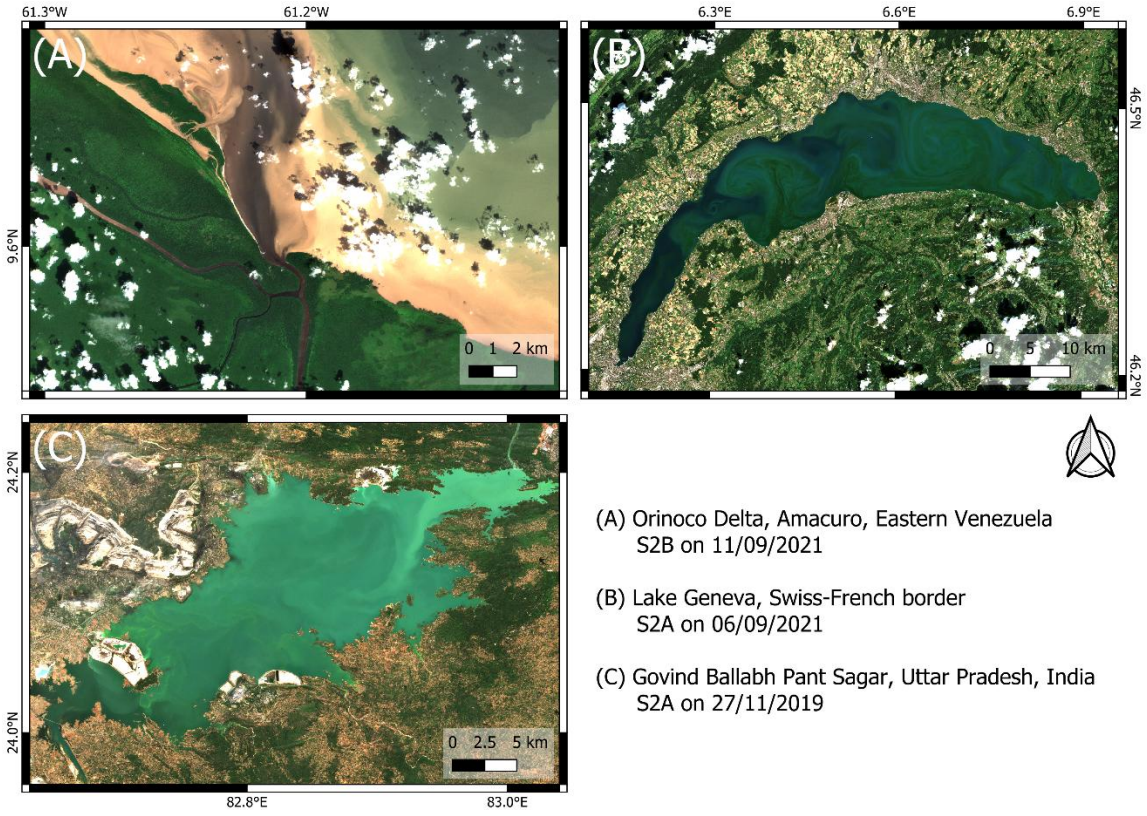

Figure 2: True-color Level 2 Sentinel-2 A/B (S2A/S2B) example images of optically active constituents in Case 2 waters, downloaded from the Copernicus browser (©Copernicus). The same scaling of the 490, 560, and 665 nm bands was applied to all images. (A) Large quantities of suspended and dissolved matter in Orinoco Delta, Eastern Venezuela (B) Bloom of Uroglena sp. phytoplankton species in Lake Geneva (C) Suspended ash particles originating from coal plants in Govind Ballabh Pant Sagar (Rihand Dam), Uttar Pradesh, India.

1.4.1 Optically active constituents

RS can detect three general types of optically active constituents (OACs, constituents that absorb or scatter light), namely phytoplankton pigments, CDOM, and NAP (Bi et al., 2023; Gege, 2004; Giardino et al., 2012). While CDOM represents dissolved matter, NAP and phytoplankton are suspended and are therefore referred to collectively as total suspended matter

(TSM) or suspended particulate matter (IOCCG, 2000). As the concentration of NAP is typically orders of magnitude higher than that of phytoplankton pigments (see Bi et al., 2023, Table 2), NAP and TSM are sometimes used interchangeably.

The main phytoplankton pigment is CHL. It is used as a proxy for detecting algal blooms and calculating primary production (IOCCG, 2000). Optically, CHL is mainly distinguished by two absorption peaks in the blue and red, respectively, which give it its greenish color (Gege, 2004). Accessory pigments, such as chlorophyll *b* and phycocyanin, can be used to distinguish between different phytoplankton species (Babin et al., 2003b; Detoni et al., 2026; Maire et al., 2025).

NAP includes both mineral particles, such as sediments and carbonate, and organic particles, such as dead and decaying phytoplankton cells or degraded organic matter from terrestrial sources (Roesler, 2021). These particles scatter light back to the sensor. At high concentrations, this results in relatively bright water color (see Figure 2, C).

CDOM can be of diverse chemical composition, impacting its optical properties (Griffin et al., 2018). Allochthonous CDOM originates from outside the Lake, i.e. "from terrestrial vegetation and soil sources (Brandão et al., 2018, p. 2932)", while autochthonous CDOM is produced *in situ* by phytoplankton. CDOM is typically characterized by strong absorption in the ultraviolet and blue wavelength ranges (Brandão et al., 2018; Griffin et al., 2018), resulting in a brownish water color (Figure 2, A).

Other constituents besides phytoplankton may also be of interest when assessing water quality. In Lake Geneva, for example, spaceborne retrievals of TSM have been used to detect whiting events (see Nouchi, 2018). Whiting events can occur due to calcite precipitation, which is thought to be connected to the net CO_2 outgassing of many large lakes worldwide (Many et al., 2024). Furthermore, TSM can originate from particulate pollution, such as ash from coal-fired power stations (see Figure 2, C). On a global scale, decreased turbidity levels in typically turbid lakes during the COVID19 lockdown could be attributed to reduced anthropogenic inputs (Wu et al., 2026). High CDOM concentrations can lead to water treatment issues and have been associated with metal mobilization (Griffin et al., 2018).

1.4.2 Inherent and apparent optical properties

Along with the signal of natural water, OACs determine the optical properties of a water body (Odermatt and Gege, 2022). These properties can be apparent (AOPs) or inherent (IOPs). AOPs depend on the ambient light field in a medium and are therefore influenced by the sun and observation geometry, whereas IOPs are not (Ansari et al., 2025; Bi et al., 2023). AOPs include various radiance, irradiance, and reflectance terms and can be related to IOPs via physical modelling (see Albert & Mobley, 2003).

In the context of spaceborne RS of water bodies, $R_{rs}(\lambda)$ is the main AOP used for the retrieval of OACs. It is defined as the ratio of upwelling radiance to downwelling irradiance. IOPs include scattering, backscattering, absorption, and attenuation as well as terms derived from these quantities such as the single scattering albedo. IOPs are directly related to the optical properties of a water body, i.e. the concentrations of OACs (Albert and Mobley, 2003; Bi et al., 2023; Gege, 2004).

Bio-optical models use physical modelling to relate OACs, IOPs, and AOPs. To establish these relationships, bio-optical models use specific inherent optical properties (SIOPs), which link IOP spectra to constituent concentrations (Gege, 2004; Giardino et al., 2012). The most well-known SIOP is the "Bricaud curve" (Bricaud et al., 1995), which relates CHL to phytoplankton absorption and remains widely used. However, since the concentration-specific absorption of CDOM varies considerably, CDOM abundance is typically expressed as CDOM absorption at 440 nm (Giardino et al., 2019). The conversion of IOPs to AOPs such as $R_{rs}(\lambda)$ is largely based on the work of Albert & Mobley (2003).

1.5 Constituent retrievals: State of research

For constituent retrievals in Case 2 waters using spaceborne RS, mostly (semi-) empirical or semi-analytical models are applied, approximating bio-optical relationships between AOPs, IOPs, and OACs (Giardino et al., 2019). Typically, these models retrieve one constituent at a time (Odermatt et al., 2012). The most notable of these models for CHL are the Ocean Color (OCx) algorithms, which quantify CHL based on blue-to-green reflectance ratios (O'Reilly et al., 1998). Similar empirical and semi-analytical algorithms exist for the retrieval of TSM (e.g. the widely used "Nechad algorithm"; see Nechad et al., (2010)) and CDOM.

While such TSM and CHL algorithms have been successfully applied for a variety of water types, retrievals of CDOM have been less successful and are typically constrained to highly concentrated waters (Odermatt et al., 2012). Current OAC retrieval research has increasingly focused on machine learning techniques (see, for example, Werther et al., 2022). As this thesis does not focus on empirical and semi-analytical algorithms, readers are directed to the extensive reviews of different kinds of retrieval algorithms by Giardino et al. (2019) and Odermatt et al. (2012).

While empirical and semi-analytical models are appealing due to their relative simplicity, they are constrained in their application to Case 2 waters due to the varying background signals of other constituents (Giardino et al., 2019). Location-specific calibration is needed to provide good results (Bonnier et al., 2024).

For constituent retrievals, bio-optical models invert AOPs, such as $R_{rs}(\lambda)$. The two main bio-optical models available for inverting OACs from satellite data are WASI and the Bio-Optical Model Based tool for Estimating water quality and bottom properties from Remote sensing images (BOMBER) (Gege, 2004; Giardino et al., 2012). As BOMBER is an add-on to the subscription-based ENVI software (Giardino et al., 2012) and WASI is a free, stand-alone software, this thesis focuses on WASI. However, it should be noted that both share the same core of bio-optical equations (see Gege, 2004). Bio-optical models promise more global applicability and the ability to distinguish signals originating from different constituents. Recently, the inversion of *in situ* spectrometric $R_{rs}(\lambda)$ measurements has revealed promising capabilities for the retrieval of phytoplankton taxa, extending beyond the simple retrieval of CHL, NAP, and CDOM (see Detoni et al., 2026; Maire et al., 2025).

The large-scale application of WASI and BOMBER to satellite data has not been sufficiently explored, as such an exercise is computationally expensive. For example, inverting a single

Sentinel-2 image using the WASI-2D image module (see Gege (2014); hereafter referred to as "WASI" for brevity) takes over 24 hours for Lake Geneva (own assessment) and approximately 20 hours for Lake Garda (Niroumand-Jadidi and Gege, 2025).

Furthermore, these models require appropriate input parameters to ensure the convergence of the minimization algorithm, especially when dealing with ill-posed conditions (Gege, 2004; Giardino et al., 2019). The WASI-AI module (Niroumand-Jadidi & Gege, 2025) was developed to reduce lengthy processing. This is achieved by training a small, image-specific neural network on a sample of just a few hundred pixels using the "regular" WASI module. However, this study found that the performance of WASI-AI depends heavily on proper convergence of the training pixels (see Appendix 1).

Other limitations of bio-optical model retrievals are related to SIOPs, which vary across space and time. Therefore, studies may use SIOP approximations measured close to satellite overflights or long-term lake averages for retrievals (see e.g. Giardino et al., 2014; Niroumand-Jadidi et al., 2022). While this provides a more accurate parametrization of the radiative transfer equations, it is not possible to take measurements for every overflight and long-term averages may change. This poses further challenges to the operational application of such retrievals.

Table 1: Literature review of studies that retrieved optically active constituents (OACs) using bio-optical models WASI and BOMBER. The literature review was conducted by a thorough keyword search of SCOPUS and the publication histories of major WASI and BOMBER contributors (i.e. Peter Gege, Milad Niroumand-Jadidi, Claudia Giardino, Mariano Bresciani). Listed studies do not include OAC retrievals in shallow water conditions. IOPs: inherent optical properties, AOPs: apparent optical properties, TSM: total suspended matter, CHL: chlorophyll a, CDOM: colored dissolved organic matter $R_{rs}(\lambda)$: remote sensing reflectance, $b_{b,TSM}(550)$: TSM backscattering at 550 nm, PRISMA: PRecursore IperSpettrale della Missione Applicativa, MODIS: Moderate Resolution Imaging Spectroradiometer (sensor on board of the Terra and Aqua satellites).

Study	Satellite & # images	OACs, IOPs, AOPs	Optically deep Case 2 waters
BOMBER			
Niroumand-Jadidi et al. (2022)	Sentinel-2, N=9 Landsat-9, N=9	CHL, TSM	Trasimeno, Maggiore, Varese, Mantova, San Francisco Bay
Niroumand-Jadidi et al. (2020)	Sentinel-2, N=5	CHL	Mantua Lakes
Giardino et al. (2014)	MODIS, N=1 Landsat-8, N=1	CHL, TSM, CDOM, $R_{rs}(\lambda)$	Garda
Fabbretto et al. (2023)	PRISMA, N=2	CHL, TSM	Mulargia, Peipsi
Pinardi et al. (2018)	Sentinel-2, N=3	CHL	Mantua Lakes
WASI			
Dörnhöfer et al. (2016)	Sentinel-2, N=1	TSM, $b_{b,TSM}(550)$, CDOM, $R_{rs}(\lambda)$	Starnberg
Hernando et al. (2024)	PRISMA, N=1 Sentinel-2, N=1	CHL	Taal
Niroumand-Jadidi et al. (2021)	Sentinel-2, N=80	CHL	Garda, Ledro, Idro, Trasimeno

Table 1 summarizes previous research on retrieving constituents using spaceborne sensors and the WASI and BOMBER bio-optical models. A review of the literature revealed several research gaps:

- Most studies considered a relatively small number of images (fewer than 10). These studies often served as proof of concept, focusing on accurate parametrization for

the day and location and sometimes employing sampling day SIOPs. The study by Niroumand-Jadidi et al. (2021) is an exception in that it validates CHL for 80 images.

- Past studies have validated an incomplete set of OACs, IOPs, and AOPs with a clear focus on CHL. Notably, $R_{rs}(\lambda)$ is rarely validated, despite being one of the major sources of error in such exercises (Giardino et al., 2019).
- Multispectral studies are currently limited to Sentinel-2, Landsat-8/9, and the Moderate Resolution Imaging Spectroradiometer and do not yet include S3. These missions were primarily designed for land applications (Giardino et al., 2019). S3 OLCI was specifically designed with water applications in mind and provides several advantages, such as higher spectral and temporal resolution (EUMETSAT, 2018).
- Past studies have assumed CHL as a single proxy for phytoplankton abundance. Retrievals of individual phytoplankton taxa, as demonstrated by Maire et al. (2025), have generally not been attempted from space using these models.
- Despite being the largest peri-alpine lake, no study has evaluated retrievals for Lake Geneva.

1.6 Objectives

Considering the current state of research into the retrieval of spaceborne bio-optical constituents, the aim of this thesis is to bridge the gap between custom, singular image retrievals and automated, operational processing. To address the identified research gaps, it validates WASI S3 OLCI retrievals from 365 valid matchup images against validation data, including OACs (CHL, TSM, CDOM, phytoplankton abundances and biovolumes), IOPs (backscattering and absorption), and AOPs ($R_{rs}(\lambda)$) in Lake Geneva.

Given computational limitations of WASI, such as long processing times and limited inversion accuracy under ill-posed conditions, the first objective of this thesis was to develop an identical Python model that addresses these limitations. Furthermore, this thesis seeks to answer the following research questions:

Using a default set of SIOPs, how does a WASI-based bio-optical inversion approach using S3 OLCI $R_{rs}(\lambda)$…

1) …perform in retrieving the three basic water constituents (i.e. TSM, CDOM, and CHL) in Lake Geneva?
2) …contribute to quantifying the abundance and biovolume of different phytoplankton taxa in Lake Geneva?

2. Data and Methods

2.1 Study area

Lake Geneva is a large, deep, clear, oligotrophic, and monomictic lake located on the northern outskirts of the Alps (Kiefer et al., 2015; Moisset, 2017; Nouchi, 2018; see Figure 3). Spanning 580 km^2, it straddles the border between France and Switzerland and provides drinking water for around one million people on both sides of the border (Moisset, 2017). Major settlements on its shores include Geneva and Lausanne, two of Switzerland's largest cities. Geographically, Lake Geneva is divided into the "small lake" (petit lac) and "large lake" (grand lac), which have different morphometric characteristics (Moisset, 2017).

Table 2: Morphometric and watershed characteristics of Lake Geneva, with subsets shown for the small and large lake where known. Note that this study is mainly concerned with the large lake (Moisset, 2017, Table 1, adapted).

Lake characteristic			
	Whole lake	Large lake	Small lake
Superficies [km^2]	580.1	498.9	81.2
Volume [km^3]	89.0	86	3
Average depth [m]	152.7	172	41
Main tributary [% of input]		Rhône river (86%)	La Versoix (1.6%)
Residence time [yr^{-1}]	11.3		
Watershed characteristic			
Superficies [km^2]	7419		
Main surface use	Agriculture 26%		

Due to increased agricultural use of fertilizers, urbanization, and insufficient wastewater treatment, Lake Geneva experienced a period of eutrophication from the 1960s onwards, which had a significant impact on the lake ecosystem. The average level of phosphorus increased from 12.4 µg L^{-1} in 1957 to 89.5 µg L^{-1} in 1979 (Moisset, 2017). This led to algal blooms becoming more frequent, especially in summer. To reduce eutrophication, the International Commission for the Protection of Lake Geneva (Commission Internationale pour la Protection des Eaux du Léman, CIPEL) implemented various phosphorus reduction policies. Despite the subsequent re-oligotrophication, the biomass of phytoplankton has recently remained stable in both the small and large lakes (Jacquet et al., 2026; Moisset, 2017).

2.2 *In situ* data: Sources and sampling stations

Lake Geneva is well adapted for the validation of spaceborne water quality and water constituent products because of its numerous *in situ* reference data. The reference data used in this study is exclusive to the large lake.

CIPEL regularly monitors various parameters, including CHL and taxa-specific phytoplankton biovolume, at the SHL2 station (46.453°N, 6.594°E) located at the deepest point of the lake (310 m). This data can be accessed via the Observatory on LAkes database (https://soere-ola.hub.inrae.fr; Rimet et al., 2020).

The LéXPLORE platform (46.500°N, 6.661°E) is a floating research laboratory moored at a depth of 110 m located 570 m from the northern shore (Gupana et al., 2025). It is a collaborative effort between five institutions, including the Swiss Federal Institute of Aquatic Sciences (Eawag), and provides data for a broad range of limnological research (Minaudo et al., 2021; Wüest et al., 2021).

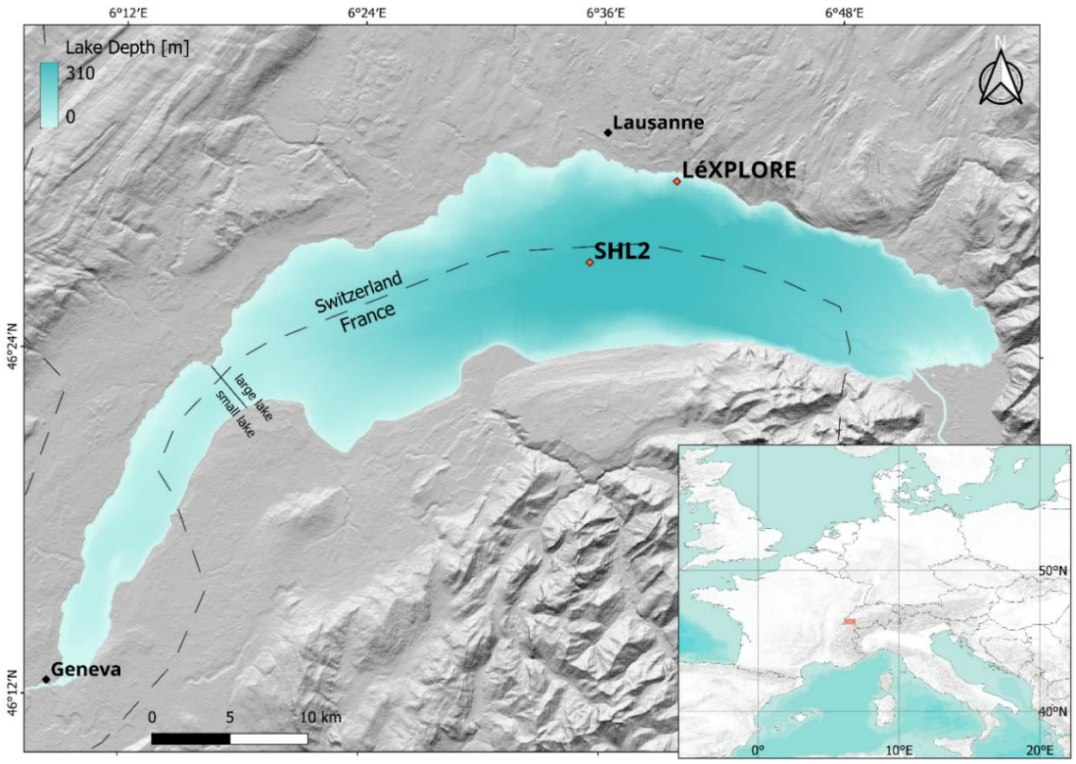

Figure 3: Overview map of Lake Geneva, including the location of the sampling station SHL2 and the LéXPLORE platform. The background hillshade and bathymetry maps are derived from the Shuttle Radar Topography Mission digital elevation model (©NASA) and swissBATHY3D datasets respectively (©swisstopo). The simplified country border is from Huwise Hub (©Huwise, OGL).

A variety of datasets are provided in near real-time and can be downloaded from the Datalakes website (www.datalakes-eawag.ch). The data used in this thesis originate from the autonomous Thetis profiler, which is located 30 m away from the platform and has been providing intraday measurements of various variables, such as IOPs, fluorometric CHL, and radiometric quantities since 2018 (Minaudo et al., 2021; Wüest et al., 2021). Measurements from the top 50 m of the water column are taken every three hours at high vertical resolution (Gupana et al., 2022). The Datalakes website provides Level 2 (L2, LéXPLORE Thetis Vertical Profiler Depth Time Grid) products that have been calibrated and quality controlled.

Additionally, three separate campaigns near LéXPLORE carried out by Eawag in May and June 2021 provide spatially distributed CHL and TSM data.

2.3 *In situ* data: Validation parameters

2.3.1 AOPs

The remote sensing reflectance $R_{rs}(\lambda)$ is the model input and not used for direct model evaluation. However, it is important to evaluate AC performance using *in situ* $R_{rs}(\lambda)$ to assess

the quality of the algorithm input and contextualize satellite retrievals (Giardino et al., 2019; Ogashawara et al., 2021). The *in situ* $R_{rs}(\lambda)$ measurements used in this study are derived from in-water upwelling radiance and downwelling irradiance (Table 3).

Table 3: Used measurements obtained from the Thetis profiler (adapted from Gupana et al., 2022, Table 1; Minaudo et al., 2021, supplementary material, Table S1).

Measurements	Instrument	# Instrum. Channels	Wavelength Region [nm]
Fluorometric chlorophyll *a*	WetLabs ECO Triplet BBFL2W	1	470 (excitation) and 695 (emission)
(Back)scattering	WetLabs ECO Triplet BBFL2W	1	700
(Back)scattering	WetLabs ECO Triplet BB3W	3	440, 532, 630
Absorption	WetLabs AC-S	81	400–730
Upwelling radiance	Satlantic HOCR R08W	180	300–1200
Downwelling irradiance	Satlantic HOCR ICSW	180	300–1200

2.3.2 CHL

A variety of CHL data was acquired for validation purposes (Table 5). This includes data from SHL2, the LéXPLORE platform Thetis profiler, and Eawag field campaigns.

At SHL2, water samples collected at nine discrete depths (0, 1, 2.5, 3.5, 7.5, 10, 15, 20, and 30 m) were analyzed using a state-of-the-art spectrophotometric approach (ISO10260:1992). Such measurements, denoted CHL_S, are taken approximately bi-weekly during summer while winter measurements are rarer.

The Thetis profiler's WetLabs ECO Triplet BBFL2W instrument (Table 3) provides fluorometric CHL measurements, denoted CHL_F. However, fluorometric measurements are highly susceptible to errors due to non-photochemical quenching (NPQ), which has been demonstrated for the Thetis profiler (Gupana et al., 2022).

Gupana et al. (2022) applied a procedure to minimize the effect of different quenching states. Accordingly, CHL_F was corrected using a linear relationship with nighttime absorption line height at 676 nm (aLH_{676}) measurements, yielding absorption-corrected CHL_F (hereafter referred to as CHL_A). The original code can be found in the Thetis GitHub repository (https://github.com/LeXPLORE-Platform/thetis-multi-instrument-profiler).

aLH_{676} is a reliable proxy for CHL (Boss et al., 2013; Bricaud et al., 1995; Minaudo et al., 2021) and was therefore also employed in the validation of CHL. According to Figure S3 in the supplementary material of Minaudo et al. (2021) (see Appendix 2), the local relationship between aLH_{676} and CHL_F which is not affected by NPQ can be approximately described as follows:

$$aLH_{676} \approx 0.0132 \times \mathrm{CHL}_F + 0.0033 \quad (1)$$

When aLH_{676} is converted to a CHL concentration for the comparison of absolute values, it is thereafter referred to as aLH_{676}→CHL to emphasize the potential uncertainties of such a conversion. Solving Equation 1 for CHL_F, the conversion is as follows:

$$aLH_{676} \rightarrow \mathrm{CHL} \approx \frac{aLH_{676} - 0.0033}{0.0132} \quad (2)$$

Phytoplankton concentration was also inverted from hyperspectral $R_{rs}(\lambda)$ (hereafter denoted $\mathrm{CPHY_R}$; see Section 2.9).

A limited number of high-performance liquid chromatography (HPLC) measurements are available from three Eawag field campaigns: $\mathrm{CPHY_{HPLC}}$ includes all pigments, while $\mathrm{CHL_{HPLC}}$ includes only CHL. These samples were collected around LéXPLORE at a depth of 0.5 m below the water surface, as well as at the measured Secchi disk depth.

2.3.3 Phytoplankton

Phytoplankton abundance and biovolume are sampled concurrently with $\mathrm{CHL_S}$ from a water mixture sample of the first 18 m (Rimet et al., 2020). The dataset includes 225 genera which were classified into seven different groups (chrysophyceae, cryptophyta, cyanobacteria, diatoms, dinoflagellates, green algae, others) using phytoplankton databases (www.algaebase.org, https:/diatoms.org, and www.ccap.ac.uk). "Others" includes six genera that are part of a rare group (choanoflagellates, dictyochophyceae, phaeothamniophyceae) and 13 genera that could not be classified. Note that these groups will thereafter be referred to as "taxa", even though they do not necessarily refer to a taxonomic group.

Although phytoplankton abundance and biovolume cannot be directly compared with phytoplankton pigment concentrations, they are expected to be correlated (Maire et al., 2025).

2.3.4 TSM and CDOM

Only a limited amount of TSM data was available from three Eawag field campaigns around LéXPLORE (see Table 5). TSM concentration was measured from water samples using a standard 1.5 µm glass fiber filter. Samples were taken 0.5 m below the water surface, as well as at the measured Secchi disk depth.

As with $\mathrm{CPHY_R}$, NAP and CDOM were inverted from hyperspectral *in situ* $R_{rs}(\lambda)$, denoted $\mathrm{NAP_R}$ and $\mathrm{CDOM_R}$ respectively.

2.3.5 IOPs

For Thetis L2 products, absorption is limited to few wavelengths. However, for our validation study, the full spectrum measured by the WetLabs AC-S (Table 3) is of interest. Therefore, the Thetis repository was cloned from GitHub and the full archive was reprocessed to include the hyperspectral constituent absorption $a_{wc}(\lambda)$. Note that Thetis processing removes pure water absorption.

For the same products, scattering at 440, 532, 630, and 700 nm measured at 117° is converted to backscattering, as described in the supplementary material of Minaudo et al., (2021), and corrected for the contribution of pure water ($b_{b,\mathrm{wc}}(440)$, $b_{b,\mathrm{wc}}(532)$, $b_{b,\mathrm{wc}}(630)$ and $b_{b,\mathrm{wc}}(700)$). Although these wavelengths do not match the S3 OLCI wavelengths, the values are expected to correlate strongly with the adjacent OLCI wavelengths, since backscattering in WASI is spectrally smooth.

2.3.6 Auxiliary data

The Secchi depth, measured alongside CHL_S, was used to analyze matchup representativeness of the CHL_S vertical distribution.

2.4 *In situ* data: Additional quality control

Standard Thetis processing includes basic quality control, such as range checks to ensure physical realism and despiking (for depth-resolved measurements, unreasonable outliers along the depth dimension are removed).

Hyperspectral measurements, i.e. $R_{rs}(\lambda)$ and $a_{wc}(\lambda)$, were filtered for spikes in the spectral domain. To detect spiky spectra, "spectral roughness" (own implementation) was calculated as the normalized mean absolute second derivative, taking $R_{rs}(\lambda)$ as an example:

$$\text{spectral roughness} = \frac{\sum_{i=1}^{N} \left| \frac{d^2 R_{rs}(\lambda_i)}{d\lambda^2} \right|}{\sum_{i=1}^{N} |R_{rs}(\lambda_i)|} \quad (3)$$

where N is the number of valid spectral bands after removing NaN values. Data with gaps (>10 nm) and spikes (spectral roughness > 0.001 for $a_{wc}(\lambda)$ and spectral roughness > 0.002 for $R_{rs}(\lambda)$) were consequently removed. $a_{wc}(\lambda)$ that passed previous checks were additionally smoothed using a Savitzky-Golay (Savitzky and Golay, 1964) filter (5-point window, 2nd-order polynomial).

Additional quality control measures included tightening the range checks to constrain backscattering to [0.0001, 0.1 m^{-1}], $a_{wc}(\lambda)$ to [0.0001, 1 m^{-1}], $R_{rs}(\lambda)$ to [0.001, 0.05 sr^{-1}], aLH_{676} to [0.001, 0.5 m^{-1}], and other CHL proxies to [0.001, 50 mg m^{-3}]. Note that range checks were applied to $a_{wc}(\lambda)$ and $R_{rs}(\lambda)$ at all wavelengths, filtering out the whole spectrum if the criteria were not met at a single wavelength. aLH_{676} was not recalculated based on the reprocessed $a_{wc}(\lambda)$ but retained as the original Thetis product to ensure consistency with earlier studies (e.g. Gupana et al., 2025; Werther et al., 2022).

The data at SHL2 were extensively reviewed by CIPEL (Rimet et al., 2020). Gaps in the vertical column were frequently observed for CHL_S (e.g. due to earlier quality control). These measurements were not removed to maintain a relatively large number of matchups. In-house campaign data was visually inspected and no removal of measurements was deemed necessary.

2.5 Satellite data: Acquisition and processing

Sentinel-3A (S3A) and Sentinel-3B (S3B), collectively referred to as S3, are two earth observation satellites operated by the European Space Agency (ESA), launched in February 2016 and April 2018 respectively (EUMETSAT, 2018). On board, the imaging multispectral radiometer OLCI captures 21 bands at 300 m spatial resolution. Since the launch of S3B, the combined revisit time for the study area is less than a day. Eleven of these bands are designed to support water constituent retrieval, while the others are most useful for atmospheric correction (EUMETSAT, 2018; see Table 4).

Level 1 images (L1 S3 OLCI full resolution) were atmospherically corrected using the POLYMER (POLYnomial-based algorithm applied to MERIS) radiative transfer model, which has been shown to produce the best results for Lake Geneva (Gupana et al., 2025). POLYMER performs AC by (1) applying corrections to the top-of-atmosphere reflectance ρ_{TOA} that can be reliably quantified (i.e. Rayleigh scattering, gaseous absorption, main sun-glint component) and (2) fitting the remaining top-of-atmosphere signal with a spectral optimization. It decomposes the remaining ρ_{TOA} into (a) an atmospheric and (residual) glint component modeled as a simple polynomial and (b) a bio-optical water reflectance model. POLYMER then iteratively adjusts both components so that their combined spectrum best matches the measured reflectance across all wavelengths (Steinmetz et al., 2011).

Table 4: Sentinel-3 OLCI Marine User Handbook band description. Wavelengths suited for phytoplankton or chlorophyll a (CHL) retrievals (excluding fluorescence) are marked blue. Note that Oa9 is not available for POLYMER outputs. AC: atmospheric correction, CDOM: colored dissolved organic matter, FWHM: full width at half maximum, TSM: total suspended matter. Table 1 adapted from EUMETSAT (2018).

Band number	Central wavelength [nm]	Spectral width (FWHM) [nm]	Main function
Oa1	400	15	AC, improved water constituent retrieval
Oa2	412.5	10	CDOM and detrital pigments
Oa3	442.5	10	CHL absorption maximum
Oa4	490	10	High CHL and other photosynthetic pigments
Oa5	510	10	CHL, TSM, turbidity, and red tides
Oa6	560	10	CHL reference (minimum)
Oa7	620	10	TSM
Oa8	665	10	CHL (2nd absorption maximum), TSM, and CDOM
Oa9 †	673.75	7.5	Improved fluorescence retrieval
Oa10	681.25	7.5	CHL fluorescence peak
Oa11	708.75	10	CHL fluorescence baseline
Oa12	753.75	7.5	O_2 (oxygen) absorption and clouds
Oa13	761.25	2.5	O_2 absorption band and AC aerosol estimation
Oa14	764.375	3.75	AC
Oa15	767.5	2.5	O_2 A-band absorption band used for cloud top pressure
Oa16	778.75	15	AC aerosol estimation
Oa17	865	20	AC aerosol estimation, clouds, and pixel co-registration
Oa18	885	10	Water vapour absorption reference band
Oa19	900	10	Water vapour absorption
Oa20	940	20	Water vapour absorption, linked to AC
Oa21	1020	40	AC aerosol estimation

Bad pixels (e.g. invalid pixels, clouds, and mountain shadows) were masked using the IdePix processor (Wevers et al., 2022), resulting in quality-controlled L2 POLYMER products. Images containing fewer than 2,000 pixels (i.e. containing fewer than roughly a third of valid pixels) were discarded. All image processing was performed using the Sencast toolbox, which was developed at Eawag (Odermatt et al., 2020).

2.6 Matchups

Table 5 shows the number of valid matchups per reference dataset. A matchup is valid when the paired image has more than 2,000 pixels and the sampling station's pixel itself is valid; a sample is a single (depth-resolved) measurement of one variable at one location on one day. For Thetis, the pixel just south of the LéXPLORE platform is used instead, to avoid contamination from the platform structure. When both S3A and S3B are valid for the same date, this counts as two matchups.

Table 5: In situ reference variables used for validation. $CPHY_{HPLC}$: total pigment concentration from HPLC (high-performance liquid chromatography) analysis, CHL_{HPLC}: chlorophyll a (CHL) from HPLC analysis, TSM: gravimetrically determined total suspended matter concentration, CHL_S: CHL from spectrophotometry, CHL_F: CHL from fluorometry, CHL_A: absorption-corrected CHL_F, aLH_{676}: absorption line height at 676 nm, aLH_{676}→CHL: aLH_{676} converted to CHL, $CPHY_R$: phytoplankton concentration inverted from in situ $R_{rs}(\lambda)$, NAP_R: non-algal particle concentration inverted from in situ $R_{rs}(\lambda)$, $CDOM_R$: colored dissolved organic matter absorption at 440 nm inverted from in situ $R_{rs}(\lambda)$. $a_{wc}(\lambda)$: hyperspectral constituent absorption, $b_{b,wc}(wvl)$: constituent backscattering at wavelength "wvl", $R_{rs}(\lambda)$: hyperspectral remote sensing reflectance.

In situ Reference Data	# *In situ* Samples	# Valid Matchups	Time Range
Eawag field campaigns			
$CPHY_{HPLC}$ & CHL_{HPLC}	18	29	2021
TSM	18	29	2021
SHL2			
CHL_S	159	42	2016 – 2023
Phytoplankton abundance	156	40	2016 – 2023
Phytoplankton biovolume	156	40	2016 – 2023
LéXPLORE Thetis profiler			
CHL_F	1125	326	2018 – 2025
aLH_{676} & aLH_{676}→CHL	1031	299	2018 – 2025
CHL_A	1031	309	2018 – 2025
$CPHY_R$	1035	215	2018 – 2025
NAP_R	1035	215	2018 – 2025
$CDOM_R$	1035	215	2018 – 2025
$a_{wc}(\lambda)$	1031	146	2018 – 2025
$b_{b,wc}(700)$	1042	326	2018 – 2025
$b_{b,wc}(630)$	1042	327	2018 – 2025
$b_{b,wc}(532)$	1042	325	2018 – 2025
$b_{b,wc}(440)$	1042	325	2018 – 2025
$R_{rs}(\lambda)$	1035	215	2018 – 2025

2.7 Methods: Overview

Atmospherically corrected and quality-controlled S3 images are inverted using the minimal WASI deep-water model (MiniWASI; Section 2.8). This process aims to retrieve seven unknown constituents using SIOPs of NAP and CDOM, as well as five phytoplankton-specific SIOPs (cryptophyta, cyanobacteria, diatoms, dinoflagellates, and green algae).

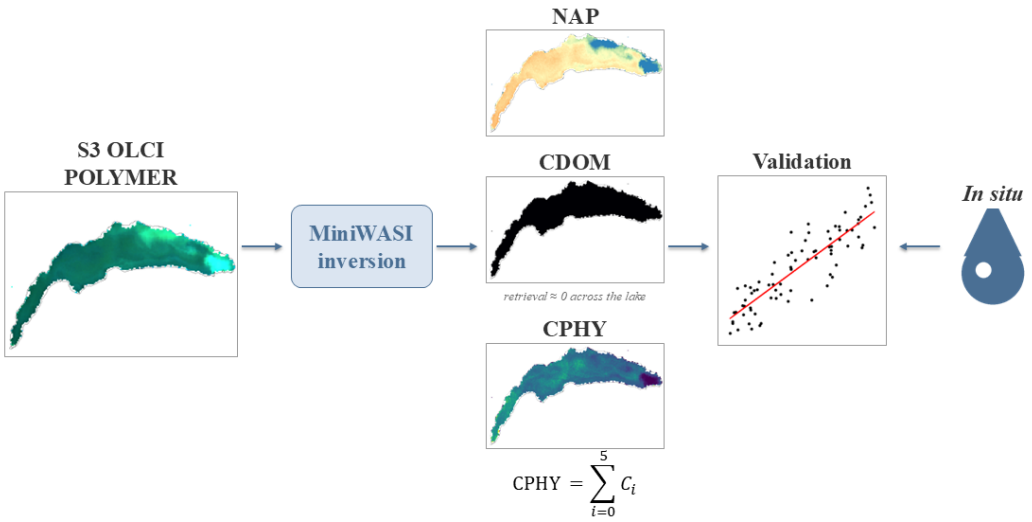

Figure 4: Schematic overview for the validation of the three basic constituents retrieved from Sentinel-3 (S3). NAP: non-algal particles, CDOM: colored dissolved organic matter, CPHY: phytoplankton concentration (sum of five classes).

NAP and CDOM retrievals (hereafter referred to as NAP_{S3} and $CDOM_{S3}$ respectively) are the product of a single SIOP each (Figure 4). To assess the retrieval of CHL, the concentrations of the five retrieved taxa are summed to give a total phytoplankton concentration (hereafter referred to as $CPHY_{S3}$). The resulting WASI-derived maps are evaluated by comparing them with *in situ* measurements taken at corresponding sampling stations (Table 5). Detailed descriptions of the bio-optical model configuration, the full processing chain, and the validation procedure are provided in Sections 2.8 – 2.10.

2.8 Bio-optical model MiniWASI

MiniWASI is a minimal Python clone of the WASI deepwater model developed by Peter Gege (Gege, 2004; Gege, 2014). MiniWASI was specifically designed for operational constituent retrieval from satellite images in optically deep Case 2 waters. It uses the lmfit (Non-Linear Least-Squares Minimization and Curve Fitting for Python) library (Newville et al., 2025) for inversion, which allows for faster and more accurate minimization (see Appendix 1).

Radiative transfer model inversion relies on a forward model that simulates the observed signal as a function of constituent concentrations. In WASI, $R_{rs}(\lambda)$ is modeled as a function of constituent concentrations and associated SIOPs that convert these concentrations to IOP (backscattering and absorption) values. IOPs are related to AOPs such as $R_{rs}(\lambda)$ using established equations commonly used in water optics (Albert and Mobley, 2003). An initial estimate yields a forward-modeled $R_{rs}(\lambda)$ that is compared to the observed (e.g. satellite) $R_{rs}(\lambda)$. The residuals between modeled and observed $R_{rs}(\lambda)$ are minimized by iteratively adapting constituent concentrations in the forward model. This so-called inverse problem is solved using a minimization algorithm (see Section 2.8.2).

2.8.1 Forward model

The MiniWASI forward model is based on the WASI (Version 7) manual (Gege, 2024). It follows the WASI deepwater model for Case 2 waters, barring some limitations outlined in this section. The reference equations in the manual are given in Table 6. For a comprehensive overview of WASI, this thesis refers to the most recent version of the WASI manual (e.g. Gege, 2024). This section provides the general model formulation; the exact model parametrization is presented in Section 2.8.3. The reference wavelengths of SIOPs match the WASI defaults and are fixed in MiniWASI.

To account for absorption by natural water, a pure water absorption coefficient spectrum at 20°C is read from file (original data from Kou et al., 1993; Lu, 2007; Pope & Fry, 1997; Wang, 2008; Warren, 1984).

$$a_{w,T}(\lambda) = a_w(\lambda) + (T - T_0)\frac{\partial a_w(\lambda)}{\partial T}, \qquad T_0 = 20°C \tag{4}$$

This spectrum $a_w(\lambda)$ [m^{-1}] is corrected using a gradient [m^{-1} °C^{-1}] (also read from file) scaled by the temperature difference to the reference (Röttgers et al., 2013).

The CDOM absorption coefficient can be modeled as an exponential function (Bricaud et al., 1981; Carder et al., 1989):

$$a_{\mathrm{CDOM}}(\lambda) = C_y\, a^n_{\mathrm{CDOM}}(\lambda), \qquad a^n_{\mathrm{CDOM}}(\lambda) = \exp[-S_{\mathrm{CDOM}}(\lambda - 440)] \tag{5}$$

where C_y denotes the CDOM absorption coefficient at 440 nm [m^{-1}], which is assumed to be a proxy for CDOM abundance. The absorption coefficient equals the product of C_y and a normalized absorption spectrum $a^n_{\mathrm{CDOM}}(\lambda)$ dependent on the slope parameter S_{CDOM} [nm^{-1}]. A Gaussian approximation of the normalized absorption spectrum representative of Lake Constance can also be read from file.

The phytoplankton absorption coefficient [m^{-1}] and concentration [mg m^3] are the sums of up to six classes (C_0 – C_5). The corresponding absorption coefficient is the product of a class's concentration C_i and its spectral absorption per unit pigment concentration [m^2 mg^{-1}]:

$$a_{\mathrm{phy}}(\lambda) = \sum_{i=0}^{5} C_i \cdot a^*_{\mathrm{phy},i}(\lambda), \qquad C_{\mathrm{phy}} = \sum_{i=0}^{5} C_i \tag{6}$$

C_0 is connected to a mixture spectrum representative of the phytoplankton taxa present in Lake Constance during the 1990s (Heege, 2000). The other five classes represent individual taxa, namely cryptophyta (original data from Soppa et al. 2020, unpublished), cyanobacteria (original data from Xi et al., 2017), diatoms, dinoflagellates, and green algae (Gege, 2012). The different taxa and their pigment-specific absorption spectra can be found in Figure 5.

The NAP absorption coefficient can be formulated as an exponential function (Babin et al., 2003b):

$$a_{\mathrm{NAP}}(\lambda) = C_{\mathrm{NAP}} \cdot a^*_{\mathrm{NAP}}(\lambda), \qquad a^*_{\mathrm{NAP}}(\lambda) = a^*_{\mathrm{NAP}}(440) \cdot \exp[-S_{\mathrm{NAP}} \cdot (\lambda - 440)] \tag{7}$$

$$with\colon\ C_{\mathrm{NAP}} = C_{\mathrm{X}} + C_{\mathrm{Mie}}$$

It is the product of the concentration C_{NAP} [g m^{-3}] and a specific NAP absorption coefficient spectrum. $a^*_{\mathrm{NAP}}(\lambda)$ is the product of the specific NAP absorption coefficient at 440 nm [m^2 g^{-1}] and a normalized absorption spectrum with slope S_{NAP} [nm^{-1}]. In WASI C_{NAP} is the sum of two sub-components with spectrally distinct backscattering (see Equation 11).

The total absorption coefficient is the sum of constituent absorption and the (temperature-corrected) absorption of natural water:

$$\mathrm{a}(\lambda) = a_{w,T}(\lambda) + a_{\mathrm{CDOM}}(\lambda) + a_{\mathrm{phy}}(\lambda) + a_{\mathrm{NAP}}(\lambda) \tag{8}$$

While the absorption coefficient of pure water is mainly temperature dependent, its backscattering coefficient is salinity dependent. The following empirical equation applies (after Morel, 1974):

$$b_{b,w}(\lambda) = b\left(\frac{\lambda}{500}\right)^{-4.32} \tag{9}$$

For freshwater the backscattering coefficient b is set to 0.00111 [m^{-1}]. For oceanic water backscattering it is higher; for a salinity between 35 and 38 ‰ b is approximately 0.00144 [m^{-1}]. By default, the freshwater configuration is used.

The phytoplankton backscattering coefficient is assumed to be similar for all taxa (Gege, 2024):

$$b_{b,\mathrm{phy}}(\lambda) = C_{\mathrm{phy}} \cdot b^*_{b,\mathrm{phy}}(\lambda), \qquad b^*_{b,\mathrm{phy}}(\lambda) = b^*_{b,\mathrm{phy}}(550) \cdot b^N_{\mathrm{phy}}(\lambda) \tag{10}$$

Therefore, it is the product of the total phytoplankton pigment concentration C_{phy} [mg m^3] and a single specific phytoplankton backscattering spectrum. $b^*_{b,\mathrm{phy}}(\lambda)$ is the product of the specific phytoplankton backscattering coefficient at 550 nm [m^2 mg^{-1}] and a normalized phytoplankton scattering spectrum that is read from file (original data from C. Giardino, personal communication with P. Gege).

The NAP backscattering coefficient is described by two spectrally distinct components with concentrations C_X and C_{Mie}. For Case 2 waters, the backscattering coefficient of NAP is generally considered spectrally flat (Babin et al., 2003a).

$$b_{b,\mathrm{NAP}}(\lambda) = C_x \cdot b^*_{b,\mathrm{X}} \cdot b^N_{\mathrm{X}}(\lambda) \; + \; C_{\mathrm{Mie}} \cdot b^*_{b,\mathrm{Mie}} \left(\frac{\lambda}{500}\right)^{-n} \tag{11}$$

$$with\colon \; b^*_{b,\mathrm{X}}(\lambda) = \; b^*_{b,\mathrm{X}} \cdot b^N_{\mathrm{X}}(\lambda)$$

The backscattering coefficient of C_X is the product of the concentration [g m^{-3}], the specific backscattering coefficient [m^2 g^{-1}], and a normalized scattering coefficient spectrum which equals 1 for all wavelengths. This results in a spectrally flat specific backscattering spectrum $b^*_{b,\mathrm{X}}(\lambda)$. C_{Mie} is formulated in a similar manner, but its scattering coefficient follows Angström's law instead, where n is the Angström exponent related to the particle size distribution (Gege, 2014).

CDOM backscattering is assumed to be negligible and thus not parametrized. The bulk backscattering coefficient is the sum of the constituent and natural water backscattering coefficients:

$$b_b(\lambda) = b_{b,w}(\lambda) + b_{b,\mathrm{phy}}(\lambda) + b_{b,\mathrm{NAP}}(\lambda) \tag{12}$$

The conversion from IOPs to AOPs implemented in WASI is largely based on the work of Albert & Mobley (2003) and earlier work by Mobley (1994). The unitless single backscattering albedo can be purely expressed as a function of the calculated IOPs from Equations 8 and 12:

$$\omega_b(\lambda) = \frac{b_b(\lambda)}{a(\lambda) + b_b(\lambda)}, \qquad 0 \le \omega_b \le 1 \tag{13}$$

It is related to the below-surface remote sensing reflectance $r^-_{rs}(\lambda)$ through the subsurface radiance reflection factor [sr^{-1}] (Albert and Mobley, 2003):

$$f_{rs}(\lambda) = 0.0512 \cdot (1 + 4.6659 \cdot \omega_b - 7.8387 \cdot \omega_b^2 + 5.4571 \cdot \omega_b^3)$$
$$\cdot \left(1 + \frac{0.1098}{cos(\theta'_s)}\right) \cdot \left(1 + \frac{0.4021}{cos(\theta'_v)}\right) \tag{14}$$

where θ'_s denotes the solar zenith angle in water, θ'_v the viewing zenith angle in water, and θ_v the corresponding viewing zenith angle in air [rad]. The above water viewing and solar angles, input in degrees, are converted to radians. To compute below-water angles, Snell's law is applied: $\theta_{below} = \arcsin\left(\frac{\sin(\theta_{above})}{1.33}\right)$.

Assuming no contribution by fluorescence, $r_{rs}^{-}(\lambda)$ [sr^{-1}] is the product $f_{rs}(\lambda)$ and $\omega_b(\lambda)$ (after Gege, 2014):

$$r_{rs}^{-}(\lambda) = f_{rs}(\lambda)\,\omega_b(\lambda) \tag{15}$$

The conversion from below- to above-water $r_{rs}(\lambda)$ requires the calculation of Fresnel reflectance at the air-water interface (Jerlov, 1976):

$$\rho_L = \frac{1}{2}\left[\frac{\sin^2(\theta_v - \theta_v')}{\sin^2(\theta_v + \theta_v')} + \frac{\tan^2(\theta_v - \theta_v')}{\tan^2(\theta_v + \theta_v')}\right] \tag{16}$$

as well as the water-air transmission factor (Gege, 2024):

$$\xi = \frac{(1-\sigma)(1-\rho_L)}{1.33^2}, \qquad \sigma = 0.03 \tag{17}$$

where ρ_L is the reflectance factor of upwelling radiance just below the water surface, which corresponds to the Fresnel reflectance and σ the reflection factor for downwelling irradiance with a typical value of 0.03 for cloud-free conditions and moderate solar zenith angles (<45°).

MiniWASI assumes no surface contribution (commonly referred to as sun-glint) to the remote sensing reflectance $r_{rs}(\lambda)$ since AC algorithms such as POLYMER attempt to correct for it (Steinmetz et al., 2011; Vanhellemont and Ruddick, 2018). This conveniently eliminates the need for parametrization of the atmosphere. Assuming no sun glint, $r_{rs}(\lambda)$ [sr^{-1}] equals the radiance reflectance $R_{rs}(\lambda)$ contributed by water. Subsequently, the remote sensing reflectance is denoted as $R_{rs}(\lambda)$ and is calculated as follows (after Gege, 2014):

$$r_{rs}(\lambda) = R_{rs}(\lambda) = \xi\frac{r_{rs}^{-}(\lambda)}{1 - \rho_u\, Q\, r_{rs}^{-}(\lambda)}, \qquad \rho_u = 0.54,\ Q = 5 \tag{18}$$

where ρ_u is the reflectance factor of upwelling irradiance just below the water surface with a typical value of 0.54. The ill-favored parameter Q [sr] is fixed at 5. It corresponds to the fraction of upwelling irradiance compared to the upwelling radiance (Albert and Mobley, 2003):

$$Q = \frac{E_u^{-}}{L_u^{-}} \tag{19}$$

Its value describes the angular distribution of the light field underwater. Unfortunately, no convenient parametrization of Q exists. Assuming a single constant, wavelength-independent value is not physically accurate (Albert and Mobley, 2003).

2.8.2 Inverse model

MiniWASI uses the lmfit library for inversion. Given a vector of constituent concentrations x lmfit minimizes the residual function:

$$\hat{x} = \arg\min_x \sum_i \left(f(x,\lambda_i) - R_{rs}^{\text{measured}}(\lambda_i)\right)^2 \tag{20}$$

where $f(x,\lambda_i)$ is the forward model evaluated at wavelength λ_i, i.e. the modeled remote sensing reflectance $R_{rs}^{\text{modeled}}(\lambda_i)$ and $R_{rs}^{\text{measured}}(\lambda_i)$ is the measured (e.g. satellite) remote sensing

reflectance. The inversion result $\hat{x}$ is the vector of parameters that best fits the observations in a least-squares sense. Note that x is constrained by minimum and maximum constituent values which are not considered in the formulation of Equation 20 (see Table 7).

Like WASI, MiniWASI allows for the individual weighting of wavelengths. Accordingly, Equation 20 can be modified by adding a weight term:

$$\hat{x} = \arg\min_{x} \sum_{i} w_i \left(f(x, \lambda_i) - R_{rs}^{\text{measured}}(\lambda_i) \right)^2 \tag{21}$$

where w_i is the weight for wavelength λ_i. Conveniently, this also provides an easy option to include or exclude certain wavelengths (binary weights).

MiniWASI uses the Levenberg-Marquardt algorithm (Levenberg, 1944; Marquardt, 1963) by default to solve the inverse problem. However, any minimization method implemented in lmfit can be used, including the Nelder-Mead algorithm implemented in WASI (Gege, 2004). lmfit also provides an estimate of the standard error for each inverted variable based on the covariance matrix. However, such an estimate may not be possible under ill-posed conditions (see https://lmfit.github.io/lmfit-py/parameters.html#lmfit.parameter.Parameter.stderr).

2.8.3 Model parametrization

As outlined in Section 1, this thesis relies on a standard set of SIOPs representative of Lake Constance. Accordingly, the default values in Table 6 are used for the SIOPs of CDOM, NAP, and bulk phytoplankton backscattering in Equations 4, 6, 9, and 10.

Equations 13 and 15 depend on the viewing zenith angle and the solar zenith angle. These values were input based on the average value of the image. The temperature T was fixed at the default value of 20°C since the temperature dependence of pure water absorption is rather weak. The backscattering coefficient b of pure water was set to 0.00111.

The initial, minimum, and maximum values used for inversion are given in Table 7. For bulk CHL retrieval, the concentrations of individual taxa ($C_1 - C_5$) were summed. This approach was favored over varying the mixture class C_0, which is representative of conditions in Lake Constance in the 1990s and should be used with caution (personal communication with Peter Gege). Conveniently, this also results in a single model configuration for both individual taxa and bulk CHL retrievals. Such an approach has also been employed in a recent study where phytoplankton SIOPs were unknown (see Detoni et al., 2026). Note that the chrysophyceae taxon present in Lake Geneva is not parametrized in WASI and could therefore not be retrieved.

For NAP retrievals, only C_X was varied while C_{Mie} was fixed at 0. To parametrize CDOM absorption, the exponential approximation was used. This is the common approach outlined in literature (Dörnhöfer et al., 2016; Maire et al., 2025; Niroumand-Jadidi and Gege, 2025).

The initial values and bounds were set based on the known data ranges from literature and the used *in situ* data, where available. It is important to set reasonable initial values: If they are too far off, the minimization algorithm may not converge. Furthermore, to prevent unrealistic values, the range of constituent concentrations should be limited (Gege, 2004). The effect of

initial values and the general minimization accuracy of WASI and MiniWASI are further discussed in Appendix 1.

Table 6: References for the MiniWASI equations in the WASI (version 7) manual with lake- and condition-specific parametrization of the equations. For lake-specific parameters, default values for Lake Constance are used.

Equation	Equation(s) in Manual	Lake-specific Parameters	Default Value(s)	Comments
5	2.3, 2.4, 2.5	S_{CDOM}	0.14	
6	2.7, 2.8	-	-	
7	2.9, 2.10	S_{NAP}, $a^*_{NAP}(440)$	0.011, 0.041	
8	2.1, 2.2	-	-	
10	2.15	$b^*_{b,phy}(550)$	0.001	
11	2.16	$b^*_{b,X}$, $b^*_{b,Mie}$, n	0.0086, 0.0042. -1	Case 1 parametrization is not implemented in MiniWASI
12	2.11	-	-	
13	2.20	-	-	
15	2.31	-	-	MiniWASI does not parametrize fluorescence
17	2.36	-	-	
18	2.36, 2.34	-	-	MiniWASI assumes sunglint-corrected data
19	*Not numerated, see manual Page 19*	-	-	
		Condition-dependent		
4	2.1, 2.2	T	20	Since T_0 also equals 20°C there is no temperature-dependency in the default configuration
9	2.14	b	0.00111	Standard parametrization assumes fresh water
14	2.32	θ_s, θ_v	40, 0.001	Angles [°] are input based on the viewing and sun geometry.
16	2.24	θ_s, θ_v	40, 0.001	*See comment above*

Table 7: MiniWASI initial, minimum, and maximum values. C_x, C_i, and C_y parametrization is based on knowledge of in situ data ranges (both literature and used data). Since the typical pigment concentrations of C_1 - C_5 are not well known, their initial value was set to 1 mg m^{-3} and their maximum value to that of maximum in situ chlorophyll a (15 mg m^{-3}).

Parameter	Model alias	Unit	Vary	Initial	Minimum	Maximum
NAP concentration (X)	C_x	g m^{-3}	True	1	0	10
NAP concentration (Mie)	C_{Mie}	g m^{-3}	False	0	0	0
CDOM absorption coefficient at 440 nm	C_y	m^{-1}	True	0.1	0	0.5
Pigment concentration of mixture class	C_0	mg m^{-3}	False	0	0	0
Pigment concentration of cryptophyta	C_1	mg m^{-3}	True	1	0	15
Pigment concentration of cyanobacteria	C_2	mg m^{-3}	True	1	0	15
Pigment concentration of diatoms	C_3	mg m^{-3}	True	1	0	15
Pigment concentration of dinoflagellates	C_4	mg m^{-3}	True	1	0	15
Pigment concentration of green algae	C_5	mg m^{-3}	True	1	0	15

Since the inversion of phytoplankton pigment concentrations proved to be unstable, only the bands affiliated with CHL or other phytoplankton pigments (Oa3 – Oa6 and Oa8; see Table 4) were used in the inversion process, barring fluorescence related ones.

Setting C_{Mie} to 0 and inserting the Lake Constance values for the lake-specific parameters into Equations 5, 6, 7, 10, and 11 yields a set of relevant SIOPs (Table 6). Using the condition-dependent pre-sets for Equations 4 and 10 determines the pure water IOPs. The resulting spectra are presented in Figure 5.

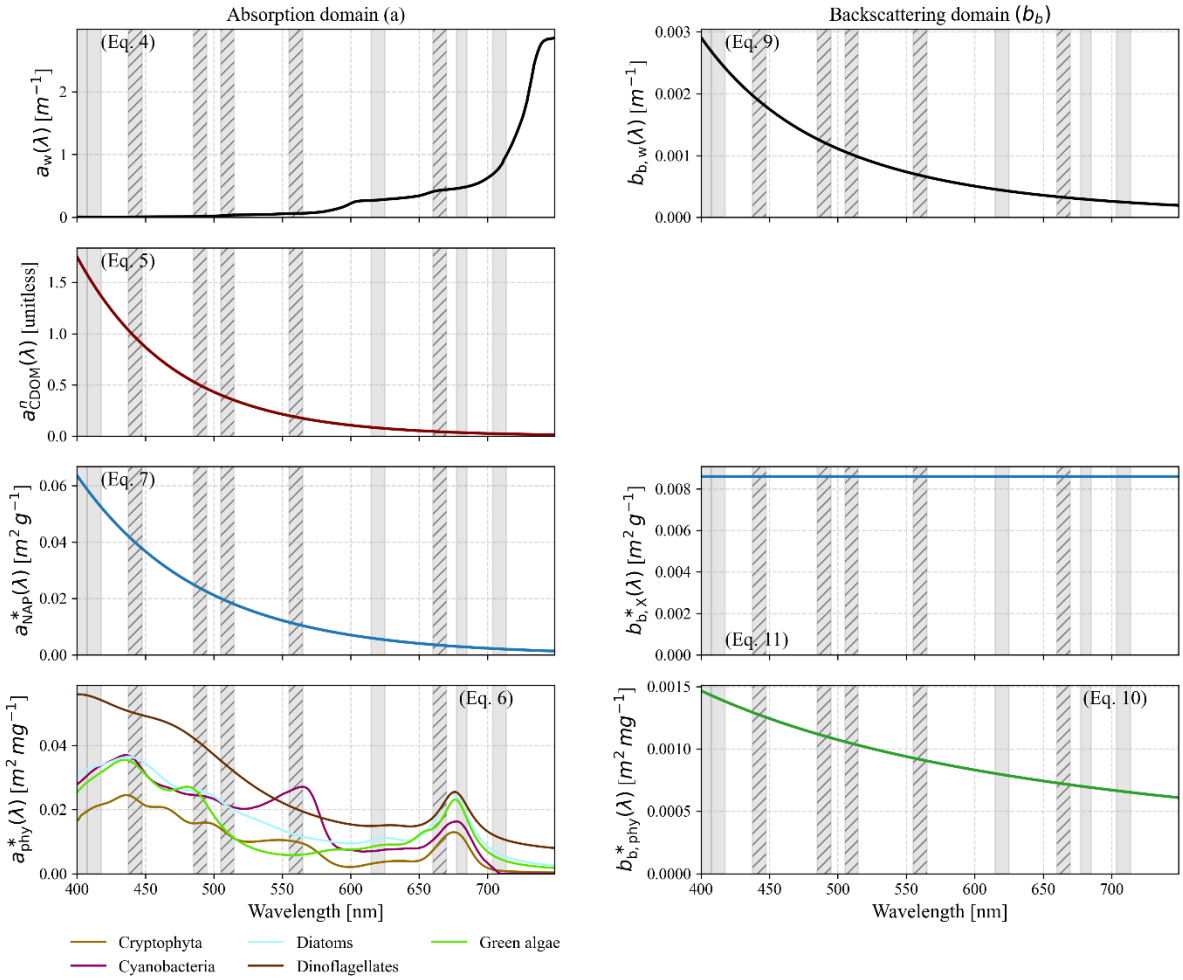

Figure 5: (Specific) inherent optical properties as configured with the pre-sets from Table 6. Reference to the Equations is given in brackets. Grey bars indicate Sentinel-3 OLCI wavelengths (full width at half maximum), hatched bars signify the used bands. From Equation 4: Absorption of pure water; note that $a_{w,T}(\lambda)$ equals $a_w(\lambda)$ given $T = 20°$. From Equation 9: Backscattering of pure water. From Equation 5: Normalized absorption coefficient of CDOM (exponential approximation). From Equation 7: Specific absorption coefficient of NAP. From Equation 11: Specific backscattering coefficient of NAP. From Equation 6: Normalized absorption coefficient of phytoplankton. From Equation 10: Normalized backscattering coefficient of phytoplankton.

Note that thereafter, WASI model variable names (e.g. C_X) and their respective OACs (e.g. NAP) are used interchangeably. In model-specific contexts, the variable names are used, whereas the OAC names are used when discussing results.

2.9 Inversion of *in situ* remote sensing reflectance

To obtain $CDOM_R$, NAP_R, and $CPHY_R$ the *in situ* $R_{rs}(\lambda)$ was inverted using the same model parametrization outlined in Section 2.8.3. Rather than using only five wavelengths, the entire available range between 400 and 730 nm was employed. θ_s was input based on the matchup image and θ_v was set to zero.

2.10 Matchup validation

To enable comparison with spaceborne retrievals, the vertically distributed *in situ* data was averaged over the first 7.5 m at SHL2 (which constitutes approximately the average auxiliary Secchi depth) and, in line with previous research (e.g. Gupana et al., 2025; Werther et al., 2022), over 5 m at the LéXPLORE platform. The median was used for the rather noisy absorption measurements and the mean for other variables. Eawag field campaign samples, taken 0.5 m below the surface and at the Secchi disk depth, were also averaged.

The metrics Pearson's r, Bias, root mean square error (RMSE), and median symmetric accuracy (MdSA) were used for validation. Pearson's r (Wilks, 2019) is a unitless statistic that measures the strength and direction of the correlation between two variables. In this case, it measures the correlation between the *in situ* and retrieved values of a constituent:

$$r = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^{n}(x_i - \bar{x})^2}\sqrt{\sum_{i=1}^{n}(y_i - \bar{y})^2}} \tag{22}$$

The relative Bias [%] (hereafter referred to as Bias) indicates whether retrievals systematically underestimate or overestimate compared to *in situ* data (Werther et al., 2022). A value close to 0% indicates little or no Bias:

$$\text{Bias} = \left(10^{|MR|} - 1\right)\text{sign}(MR) \times 100, \qquad MR = \text{median}\left(\log_{10}\left(\frac{x_i}{y_i}\right)\right) \tag{23}$$

The RMSE [constituent units] (Wilks, 2019) indicates the average difference between *in situ* measurements and retrievals:

$$\text{RMSE} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - x_i)^2} \tag{24}$$

The MdSA [%] (Werther et al., 2022) quantifies the typical relative difference between retrieved and *in situ* values while treating overestimation and underestimation symmetrically. Unlike the RMSE, the MdSA is less sensitive to outliers:

$$\text{MdSA} = 100\left(\exp\left(median\left(\left|\ln\left(\frac{y_i}{x_i}\right)\right|\right)\right) - 1\right) \tag{25}$$

In the above statistical metrics, n is the number of paired samples (i.e. valid matchups), x_i represents the i-th reference (*in situ*) measurement, y_i the i -th modeled (retrieved) value, $\bar{x}$ the *in situ* mean, and $\bar{y}$ the modeled average.

To assess the similarity between the reference and the retrieved spectra, variables with hyperspectral *in situ* reference (i.e. $R_{rs}(\lambda)$ and $a_{wc}(\lambda)$) were also evaluated in terms of the spectral angle (SA, Soppa et al., 2021):

$$\mathrm{SA} = \cos^{-1}\left(\frac{\sum_{i=1}^{n} a_i b_i}{\sqrt{\sum_{i=1}^{n} a_i^2}\sqrt{\sum_{i=1}^{n} b_i^2}}\right) \times \frac{180}{\pi} \quad (26)$$

where a_i and b_i are the spectral values at band i for the reference and retrieved spectra and n is the number of bands.

2.11 Operational processing chain

The processing steps described in Section 2 were combined into a single processing chain that links image acquisition, AC, quality control, image inversion, and validation (https://github.com/samlan2000/msc_thesis). In summary (Figure 6):

- Matchups (Section 2.6) were generated between *in situ* data (Section 2.3) and L1 S3 OLCI full resolution images (Section 2.5).
- These L1 images were downloaded, atmospherically corrected, and quality-filtered using the Sencast toolbox (Section 2.5), yielding an L2 POLYMER product.
- The L2 products were inverted using MiniWASI, applying a single model parametrization (Section 2.8; parametrization in Section 2.8.3)
- The resulting constituent maps were then compared to the *in situ* data using the validation metrics described in Section 2.10.

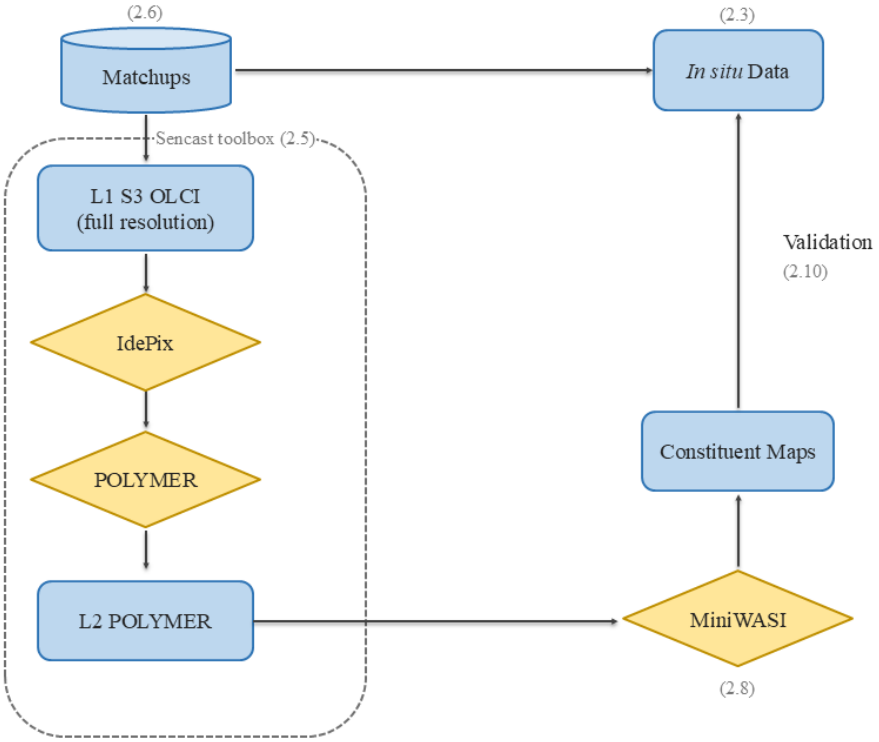

Figure 6: Schematic diagram of the processing chain presented in Section 2. References for the specific Section are given in brackets. L1: Level 1, L2: Level 2, S3: Sentinel-3, OLCI: Ocean and Land Color Instrument.

In situ and satellite data acquisition is documented in the repository. Re-running the processing chain may yield slightly different results due to the minimization algorithm.

3. Results

This section begins with AC validation, as this is a broad topic that is not directly related to model output validation. The results are then ordered by sampling location, as these offer distinct types of data sampled quasi-simultaneously, enabling valuable intercomparisons.

In accordance with the research questions, the aim of this section is not to rigorously assess the results of *in situ* hyperspectral inversions. Rather, the results are combined with those of satellite retrievals where appropriate. Additionally, although model validation is divided by S3A and S3B for completeness, differences in retrieval accuracy between the two are not discussed as this is outside the scope of this thesis.

3.1 AC performance

The performance of POLYMER AC varied across the bands (Figure 7). The median SA across all bands was 8.68°, with most SAs (168 out of 215) falling within the 5° – 15° range. AC performed best at longer blue and green wavelengths (490, 510, and 560 nm). Good agreement was observed with *in situ* $R_{rs}(\lambda)$, resulting in r values of 0.86 – 0.89, RMSE of around 0.002 sr^{-1}, MdSA of 21.4 – 25.3%, and a small positive Bias of 12.5 – 18.4%.

The shorter blue wavelengths (400, 412.5 and 442.5 nm) showed a Bias towards overestimating *in situ* $R_{rs}(\lambda)$ (18.1 – 41.1%). This was particularly evident for low to medium $R_{rs}(\lambda)$ values (Figure 7). The other quality metrics were very similar for the three bands: r-values varied around 0.85, RMSE was around 0.0025 sr^{-1}, and MdSA was around 40%.

In the red region of the spectrum, AC performance deteriorated towards longer wavelengths. While the performance was still quite good at 620 nm (r = 0.75, Bias = -2.1%, RMSE = 0.0007 sr^{-1}, MdSA = 30.8%), in the longer red wavelengths (665, 681.25, and 708.75 nm) errors were in the magnitude of the *in situ* $R_{rs}(\lambda)$ (MdSA = 78.2 – 140.3%). In the latter bands, satellite $R_{rs}(\lambda)$ was also clearly biased towards underestimating the *in situ* values (-71.6 to -125.2%).

Only five bands were used for inversion (see Section 2.8.3). Overall, the fitted $R_{rs}(\lambda)$ performed very similarly to the input satellite $R_{rs}(\lambda)$ (Table 8; Figure 8). However, for the 665 nm band, located at the second CHL absorption peak, the fitted $R_{rs}(\lambda)$ showed better performance when compared against *in situ* measurements. This resulted in a reduced MdSA (from 95.6% to 37.7%), less Bias (-95.6% to 2%), and a slightly higher correlation (from 0.59 to 0.64).

This improvement is also evident in the SA of the fitted bands. The input $R_{rs}(\lambda)$ showed an SA of 6.48°, whereas the SA decreased to 5.45° for the fitted $R_{rs}(\lambda)$.

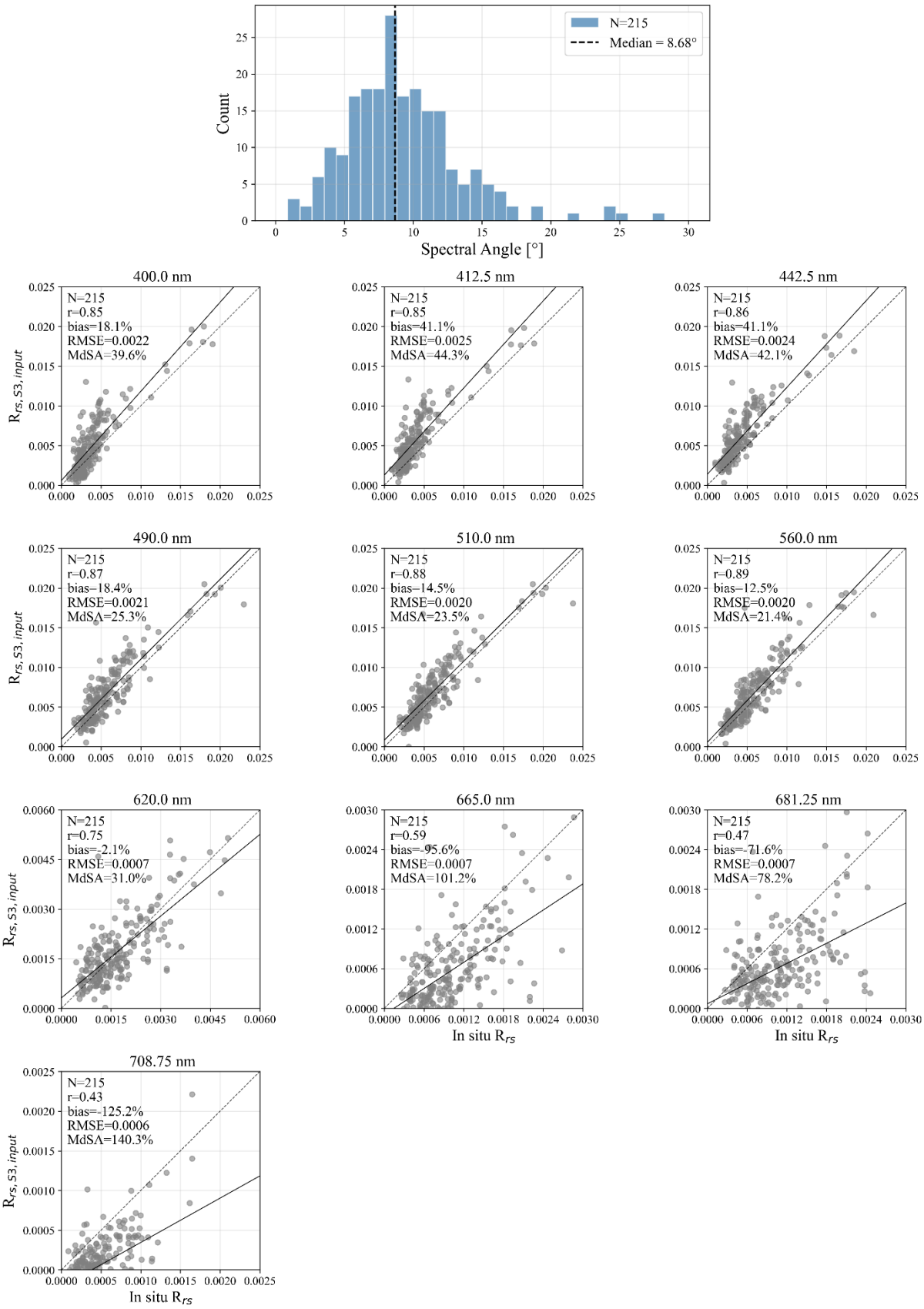

Figure 7: Validation of the atmospheric correction applied to the Sentinel-3 imagery, at all available bands. Top: histogram of the spectral angle between the atmospherically corrected and in situ reflectance spectra. Below: one scatter plot per band, comparing the atmospherically corrected Sentinel-3 remote sensing reflectance ($R_{rs,S3,input}$) against the in situ remote sensing reflectance (R_{rs}) at the same wavelength, with the 1:1 line (dashed) and best fit (solid). Axis bounds differ between panels.

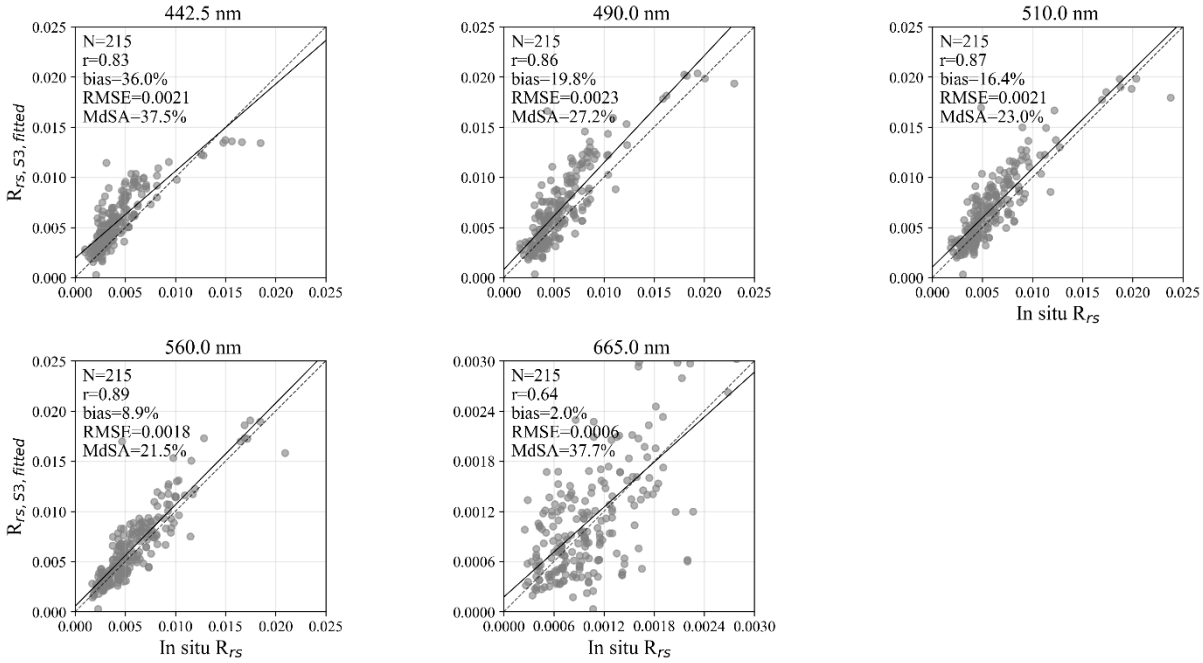

Figure 8: Validation of the Sentinel-3 remote sensing reflectance after it was re-fitted by the MiniWASI bio-optical inversion model ($R_{rs,S3,fitted}$), shown for the five bands (442.5, 490, 510, 560, and 665 nm) that are actively weighted during the MiniWASI inversion. Each panel plots $R_{rs,S3,fitted}$ against the in situ remote sensing reflectance (R_{rs}) at the same wavelength, with the 1:1 line (dashed) and best fit (solid).

Table 8: Band-by-band validation statistics for Sentinel-3 remote sensing reflectance ($R_{rs}(\lambda)$) at both processing stages shown in Figure 7 and Figure 8: the atmospherically corrected "input" R_{rs} from the POLYMER algorithm, and the "fitted" R_{rs} re-fitted by the MiniWASI bio-optical inversion model. Bands actively weighted during the MiniWASI inversion are shaded blue; statistics for unweighted bands are shown in parentheses and should be interpreted with care, as they do not influence the inversion output. Oa9 (673.75 nm) is not output by the POLYMER atmospheric correction algorithm. SA: spectral angle, given as the median across all bands ("SA all") and across weighted (re-fitted) bands only ("SA weighted").

Band number	Central wavelength [nm]	Pearson r	Bias [%]	RMSE [sr^{-1}]	MdSA [%]	N
Input satellite $R_{rs}(\lambda)$ (SA all: 8.68°, SA weighted: 6.48°)						
Oa1	400	(0.85)	(18.1)	(0.0022)	(39.6)	(215)
Oa2	412.5	(0.85)	(41.1)	(0.0025)	(44.3)	(215)
Oa3	442.5	0.86	41.1	0.0024	42.1	215
Oa4	490	0.87	18.4	0.0021	25.3	215
Oa5	510	0.88	14.5	0.0020	23.5	215
Oa6	560	0.89	12.5	0.0020	21.4	215
Oa7	620	(0.75)	(-2.1)	(0.00067)	(31)	(215)
Oa8	665	0.59	-95.6	0.00067	101.2	215
Oa9 †	673.75	-	-	-	-	-
Oa10	681.25	(0.47)	(-71.6)	(0.00072)	(78.2)	(215)
Oa11	708.75	(0.43)	(-125.2)	(0.00059)	(140.3)	(215)
Fitted $R_{rs}(\lambda)$ (SA all: 10.56°, SA weighted: 5.45°)						
Oa1	400	(0.65)	(62.5)	(0.0027)	(67.4)	(215)
Oa2	412.5	(0.73)	(56.6)	(0.0024)	(58.3)	(215)
Oa3	442.5	0.83	36.0	0.0021	37.5	215
Oa4	490	0.86	19.8	0.0023	27.2	215
Oa5	510	0.87	16.4	0.0021	23	215
Oa6	560	0.89	8.9	0.0018	21.5	215
Oa7	620	(0.73)	(0.1)	(0.00080)	(31.3)	(215)
Oa8	665	0.64	2.0	0.00056	37.7	215
Oa9 †	673.75	-	-	-	-	-
Oa10	681.25	(0.51)	(-11.8)	(0.00059)	(50.1)	(215)
Oa11	708.75	(0.5)	(10.7)	(0.00036)	(51.3)	(215)

3.2 SHL2

3.2.1 Spectrophotometric CHL

A comparison of $CPHY_{S3}$ with CHL_S revealed a high degree of agreement, resulting in a strong correlation (r = 0.8), relatively low Bias (12.2%), and moderate relative and absolute errors (see Figure 9; Table 9). $CPHY_{S3}$ reproduced the *in situ* CHL_S time series relatively well, capturing the general trends as well as two phytoplankton blooms in 2020 and 2021, respectively.

There were three instances of underestimation of CHL_S in 2019, mid-2021, and autumn 2023 (Figure 10, top panel). The depth of the maximum CHL_S was below the measured Secchi disk depth on 19 dates and within visibility on seven dates. On five dates the depth of the maximum CHL_S was within the top 7.5 m used as reference when it was simultaneously below the measured Secchi disk depth. These dates are part of the aforementioned instances in mid-2021 and autumn 2023 (Figure 10, middle panel).

Retrieved NAP_{S3} varied by up to 2 g m^{-3} and $CDOM_{S3}$ by up to 0.12 m^{-1}. On most dates $CDOM_{S3}$ was 0; it only exceeded very low absorption values once during the 2021 bloom (Figure 10, bottom panel).

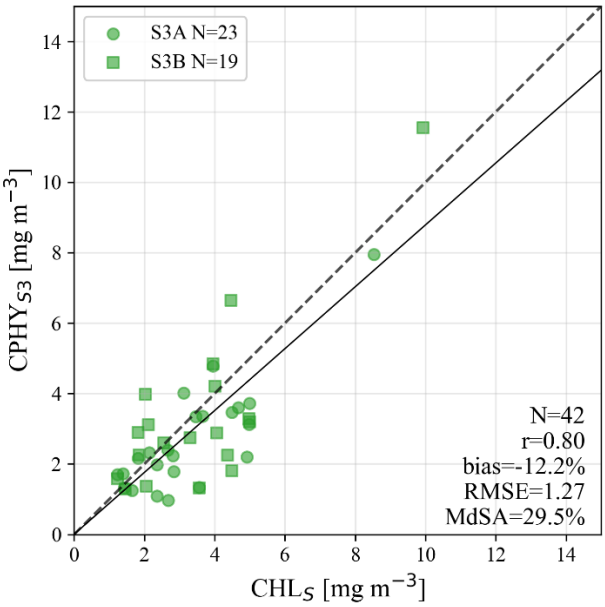

Figure 9: One-to-one validation of the satellite-modeled phytoplankton concentration from Sentinel-3 ($CPHY_{S3}$) against in situ spectrophotometric (laboratory-measured) chlorophyll a (CHL_S) at the SHL2 monitoring station. Sentinel-3A (S3A) and Sentinel-3B (S3B) matchups are shown as circles and squares respectively, with the 1:1 line (dashed) and best fit (solid).

Table 9: Validation statistics for the comparison in Figure 9: satellite-modeled phytoplankton concentration ($CPHY_{S3}$) from Sentinel-3 against in situ spectrophotometric chlorophyll a (CHL_S) at SHL2, split by satellite: Sentinel-3A (S3A), Sentinel-3B (S3B), and both combined ("all").

	Pearson r	Bias [%]	RMSE [mg m^{-3}]	MdSA [%]	N
S3A	0.82	-19.6	1.11	29.1	23
S3B	0.79	1.6	1.44	47.5	19
all	0.8	-12.2	1.27	29.5	42

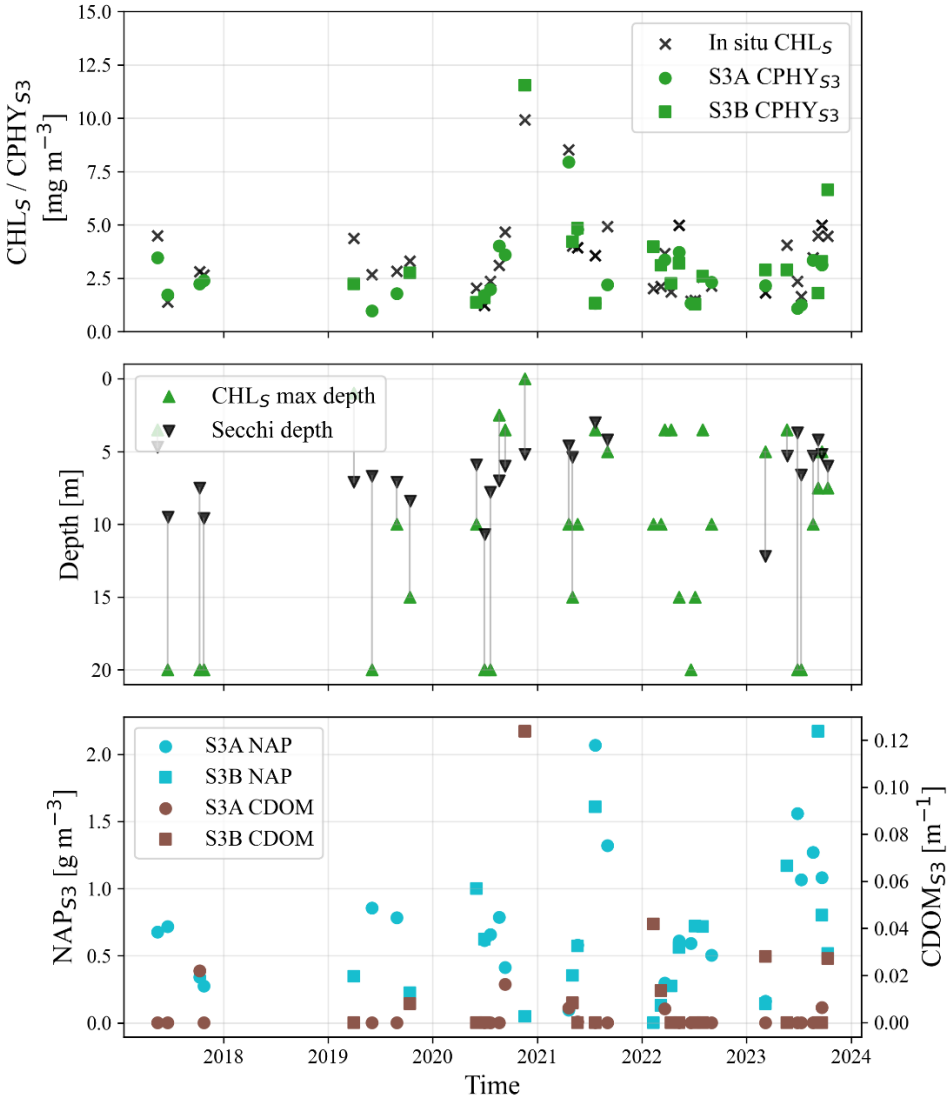

Figure 10: Time series at the SHL2 monitoring station, 2017 – 2024. Top: satellite-modeled phytoplankton concentration from Sentinel-3A and 3B (S3A and S3B; CPHY$_{S3}$) against in situ spectrophotometric chlorophyll a (CHL$_S$). Middle: depth of the CHL$_S$ maximum compared with the Secchi disk depth (the depth to which the water is visually transparent). Bottom: satellite-modeled non-algal particle concentration (NAP$_{S3}$) and colored dissolved organic matter absorption (CDOM$_{S3}$) over time.

3.2.2 Phytoplankton taxa

MiniWASI inversions mainly retrieved three taxa, namely diatoms, green algae, and dinoflagellates. Other taxa were rarely retrieved at all, with cyanobacteria and cryptophyta being present on only one and four dates respectively. Most taxa exhibited low or even negative correlation between modeled taxa pigment concentration and both biovolume and abundance. Only diatoms were significantly correlated with their *in situ* counterparts (r = 0.31 and r = 0.32 for biovolume and abundance, respectively; Table 10; Figure 12).

When the relative proportions of the measured abundances and biovolumes were compared to the proportions of the modeled concentrations (Figure 11), no clear relationships were apparent. The *in situ* data showed that significant quantities of chrysophyceae were present in Lake Geneva (Figure 11), which are not parametrized in MiniWASI.

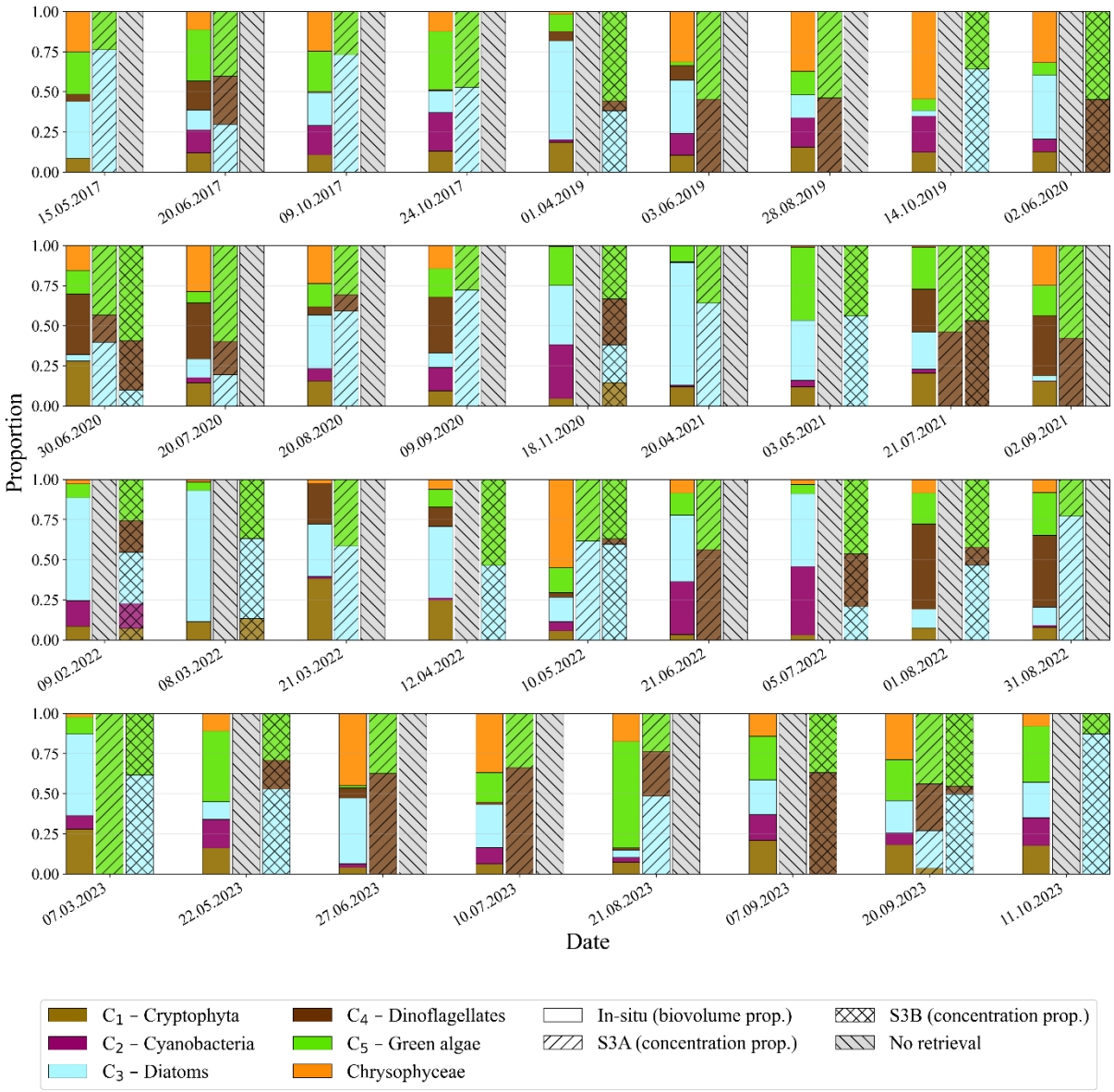

Figure 11: Relative proportions of the five phytoplankton taxa groups retrieved by the MiniWASI bio-optical inversion model from Sentinel-3A and 3B (S3A and S3B), i.e. cryptophyta (C_1), cyanobacteria (C_2), diatoms (C_3), dinoflagellates (C_4), and green algae (C_5), compared with the relative in situ biovolume proportions of the same taxa, for each sampling date. In situ chrysophyceae is shown for context but is not one of the taxa parametrized by MiniWASI. Bar shading distinguishes the in situ biovolume proportion from the S3A- and S3B-modeled concentration proportions; hatched grey bars mark dates without a valid retrieval.

Table 10: Pearson correlation (r) and significance (p) between the MiniWASI-modeled, per-taxon phytoplankton pigment concentrations retrieved from Sentinel-3 (taxon groups C_1 – C_5, see Figure 11) and the corresponding in situ biovolume and abundance measurements, for each of five phytoplankton taxa. Statistically significant correlations ($p < 0.05$) are shown in bold.

	Cryptophyta	Cyanobacteria	Diatoms	Dinoflagellates	Green algae
Biovolume (N=40)					
r	-0.15	-0.07	**0.31**	-0.14	0.06
p	0.370	0.673	**0.049**	0.3783	0.696
Abundance (N=40)					
r	-0.19	-0.07	**0.32**	-0.13	-0.10
p	0.235	0.652	**0.043**	0.420	0.556

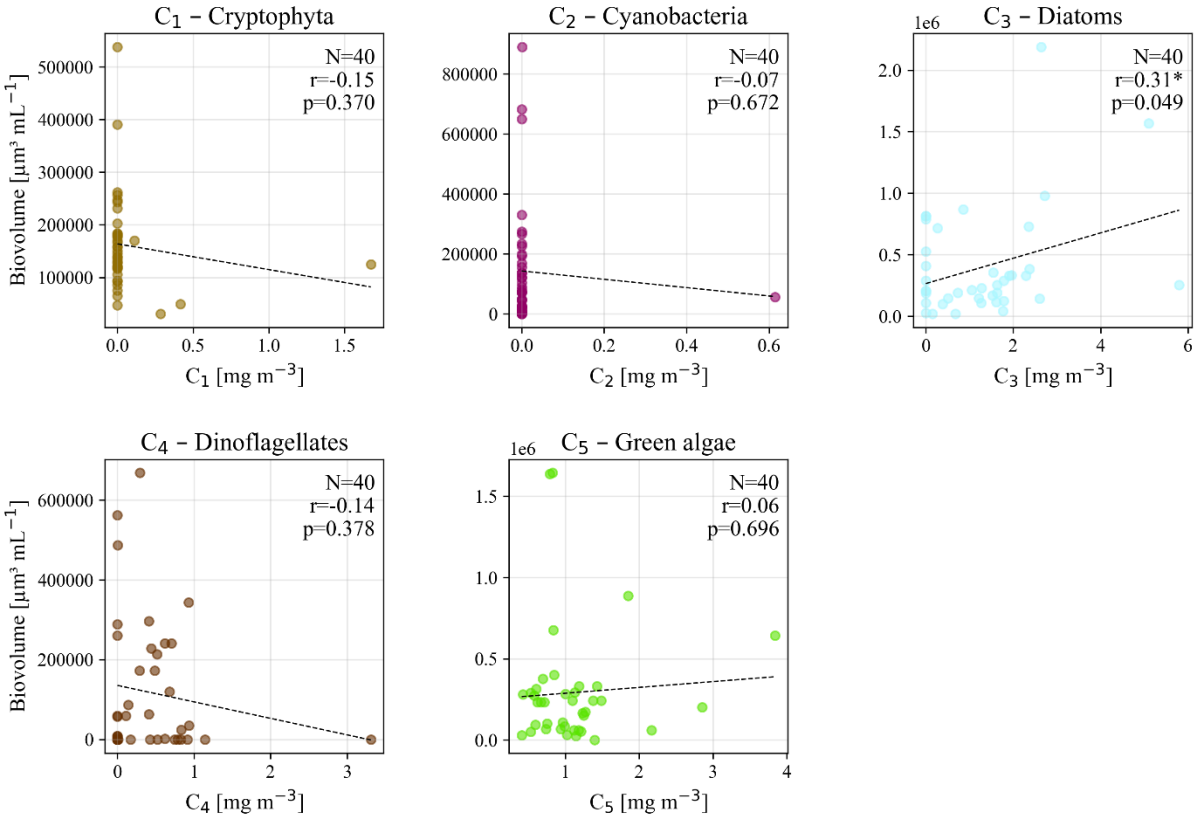

Figure 12: In situ phytoplankton biovolume plotted against the corresponding MiniWASI-modeled pigment concentration retrieved from Sentinel-3, for each of the five taxa groups: cryptophyta (C_1), cyanobacteria (C_2), diatoms (C_3), dinoflagellates (C_4), and green algae (C_5). N, Pearson r and p-value are given per panel; statistically significant correlations ($p < 0.05$) are marked with an asterisk. Note the differing axis scales.

3.3 LéXPLORE

3.3.1 Bulk phytoplankton proxies

Inter-comparisons of the *in situ* bulk phytoplankton data at LéXPLORE showed a high level of correlation (r = 0.76 – 0.87) across the board (Figure 13; Table 11). CHL_F showed a negative Bias compared to other methods (-29.7 to -74.2%), also resulting in relatively high MdSA (44.8 – 75.8%), and RMSE (1.86 – 2.36 mg m^{-3}). aLH_{676}→CHL and CHL_A agreed best, with r = 0.87 and MdSA = 25.7%. A visual comparison of the two methods revealed that CHL_A was lower at medium to high CHL but higher at low (<2 mg m^{-3}) CHL (Figure 13).

$CPHY_R$ retrievals from hyperspectral *in situ* $R_{rs}(\lambda)$ showed satisfactory agreement with CHL (Figure 13). The highest correlation was found with aLH_{676} (r = 0.78). $CPHY_R$ was generally higher than CHL_F (by 29.7%) but lower than aLH_{676}→CHL (by 28.9%) and CHL_A (by 20.6%). The correlation with CHL proxies was relatively high (r = 0.76 – 0.81), with an RMSE of between 1.77 and 1.88 mg m^{-3}, and an MdSA of between 32.3 and 44.8%. Visually, the agreement with $CPHY_R$ appears to be more strongly affected by outliers than in other inter-comparisons (Figure 13).

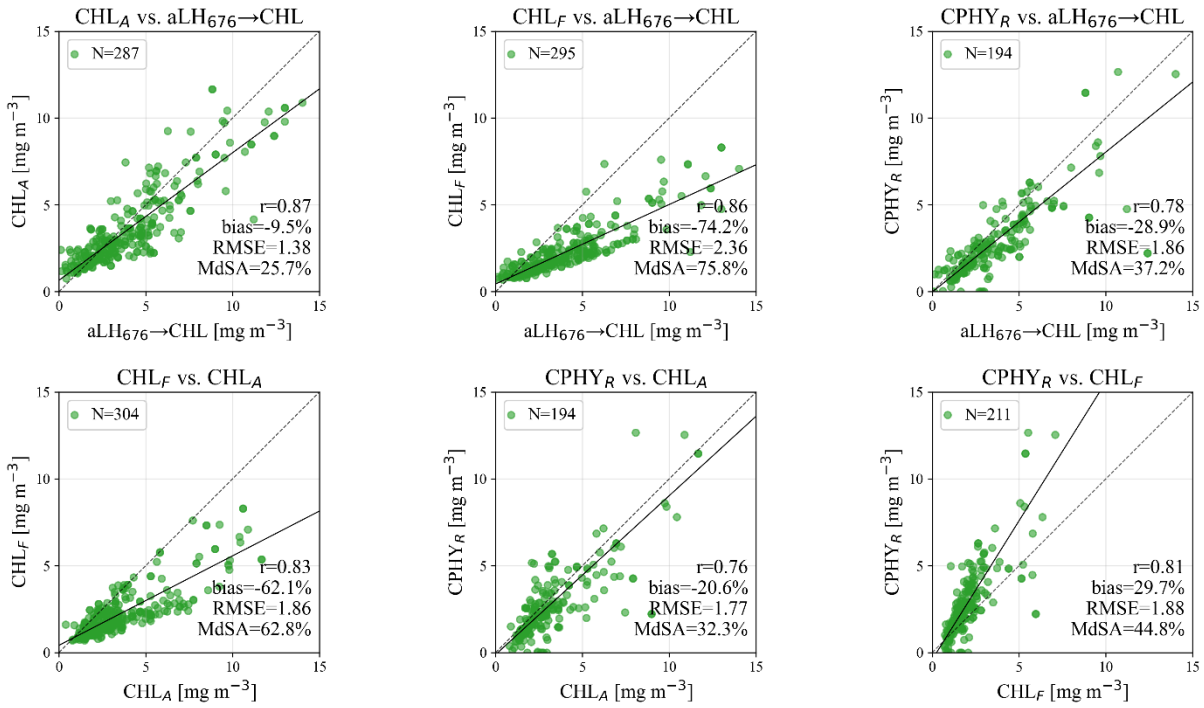

Figure 13: Pairwise inter-comparison of four in situ estimates of chlorophyll a / phytoplankton concentration at LéXPLORE: fluorometric chlorophyll a (CHL_F), absorption-corrected chlorophyll a (CHL_A, derived from CHL_F), the absorption line height at 676 nm converted to a chlorophyll-equivalent concentration (aLH_{676}→CHL), and the phytoplankton concentration modeled from hyperspectral in situ reflectance ($CPHY_R$). The dashed line corresponds to the 1:1 line and the solid line to the best fit. One $CPHY_R$ datapoint falls outside the plotted range.

Table 11: Validation statistics for the six pairwise in situ comparisons in Figure 13, between fluorometric chlorophyll a (CHL_F), absorption-corrected chlorophyll a (CHL_A), the absorption line height at 676 nm converted to chlorophyll a (aLH_{676}→CHL), and phytoplankton concentration modeled from hyperspectral in situ reflectance ($CPHY_R$).

	Pearson r	Bias [%]	RMSE [mg m^{-3}]	MdSA [%]	N
CHL_A vs. aLH_{676}→CHL	0.87	-9.5	1.38	25.7	287
CHL_F vs. aLH_{676}→CHL	0.86	-74.2	2.36	75.8	295
$CPHY_R$ vs. aLH_{676}→CHL	0.78	-28.9	1.86	37.2	194
CHL_F vs. CHL_A	0.83	-62.1	1.86	62.8	304
$CPHY_R$ vs. CHL_A	0.76	-20.6	1.77	32.3	194
$CPHY_R$ vs. CHL_F	0.81	29.7	1.88	44.8	211

Satellite retrievals of $CPHY_{S3}$ showed moderate to good agreement with *in situ* CHL and $CPHY_R$ (Figure 14; Table 12). The best overall agreement was found between $CPHY_{S3}$ and CHL_A, with r = 0.72, Bias = -18%, RMSE = 1.71 mg m^{-3}, and MdSA = 45%, being worse than the SHL2 agreement. The correlation was slightly higher with aLH_{676} (r = 0.73), accompanied by slightly higher Bias (-28.3%), RMSE (2.03 mg m^{-3}), and MdSA (51.8%) for aLH_{676}→CHL. At very low aLH_{676}→CHL values (<1 mg m^{-3}) satellite retrievals were higher in comparison. $CPHY_{S3}$ retrievals were positively biased when compared to CHL_F (30.1%), with r = 0.68, RMSE = 1.81 mg m^{-3}, and MdSA = 54.3%. For comparison with $CPHY_R$, lower $CPHY_{S3}$ values were consistently observed at high $CPHY_R$ (>8 mg m^{-3}). However, quality metrics were similar to those of the other comparisons (r = 0.69, Bias = -14.7%, RMSE = 1.88 mg m^{-3}, MdSA = 57%).

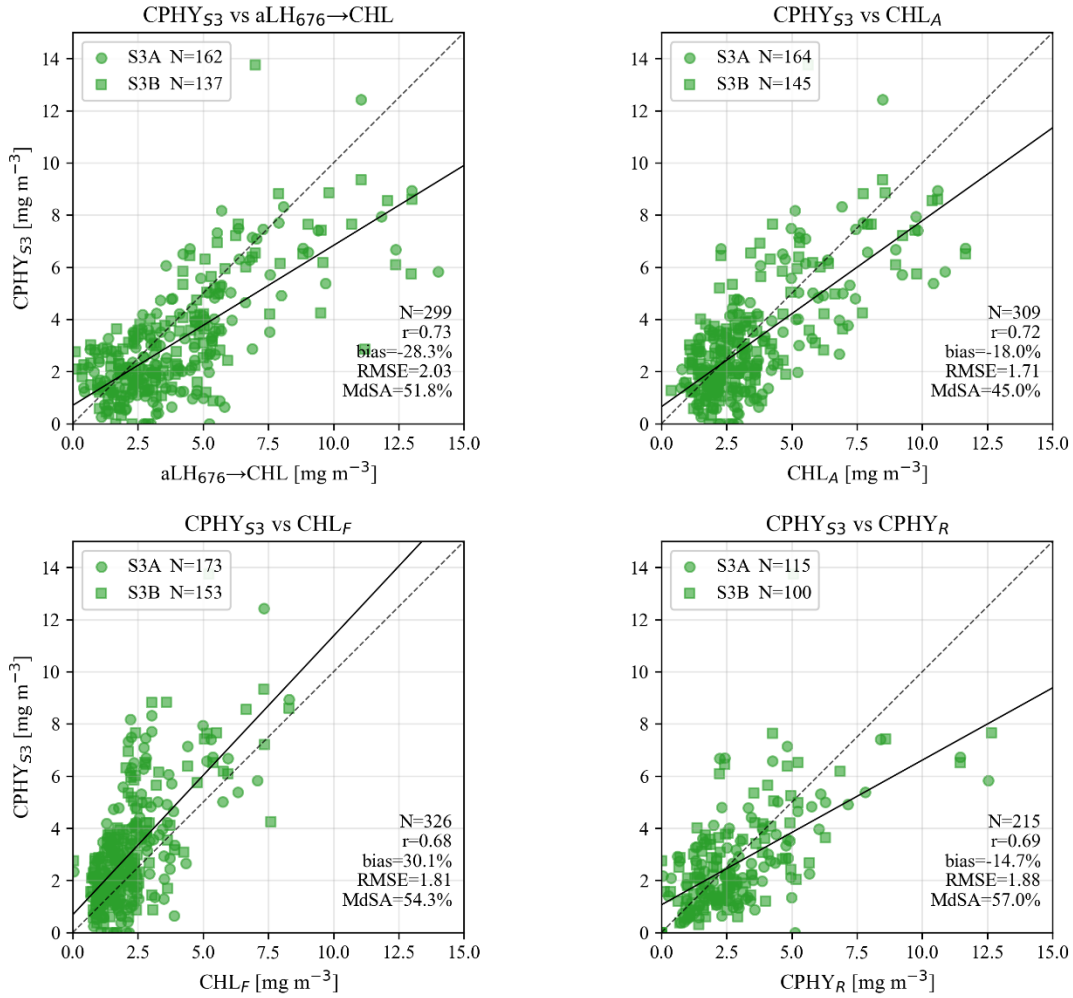

Figure 14: Validation of the satellite-modeled phytoplankton concentration from Sentinel-3A and 3B (S3A and S3B; CPHY$_{S3}$) against four in situ chlorophyll a / phytoplankton references at LéXPLORE: the absorption line height at 676 nm converted to a chlorophyll-equivalent concentration (aLH$_{676}$→CHL), absorption-corrected chlorophyll a (CHL$_A$), fluorometric chlorophyll a (CHL$_F$), and phytoplankton concentration modeled from hyperspectral in situ reflectance (CPHY$_R$). S3A and S3B matchups are shown as circles and squares respectively. The dashed line corresponds to the 1:1 line and the solid line to the best fit. One CPHY$_R$ matchup falls outside the plotted range.

Table 12: Validation statistics for the comparisons in Figure 14: satellite-modeled phytoplankton concentration from Sentinel-3 (CPHY$_{S3}$) against four in situ references (aLH$_{676}$→CHL, CHL$_A$, CHL$_F$, CPHY$_R$; see Figure 14 for full definitions. Split by satellite: Sentinel-3A (S3A), Sentinel-3B (S3B), and both combined ("all").

	Pearson r	Bias [%]	RMSE [mg m^{-3}]	MdSA [%]	N
CPHY$_{S3}$ vs. aLH676→CHL					
S3A	0.72	-31.6	2.01	52.2	162
S3B	0.74	-20.7	2.06	49.3	137
all	0.73	-28.3	2.03	51.8	299
CPHY$_{S3}$ vs. CHL$_A$					
S3A	0.73	-20.5	1.7	42.6	164
S3B	0.72	-5.8	1.71	46.1	145
all	0.72	-18	1.71	45	309
CPHY$_{S3}$ vs. CHL$_F$					
S3A	0.64	28.3	1.81	61.6	173
S3B	0.71	34.5	1.82	49.5	153
all	0.68	30.1	1.81	54.3	326
CPHY$_{S3}$ vs. CPHY$_R$					
S3A	0.72	-22.1	1.78	53.6	115
S3B	0.66	-4.7	2	64.3	100
all	0.69	-14.7	1.88	57	215

3.3.2 TSM and CDOM

Essentially, satellite CDOM retrievals failed, with $CDOM_{S3}$ values being extremely low, mostly well below the $CDOM_R$ values (Figure 15; Table 13). Consequently, agreement was poor, with a weak correlation (r = 0.2) and errors similar in magnitude to the $CDOM_R$ values (RMSE = 0.07 m^{-1}).

Satellite NAP_{S3}, however, showed satisfactory agreement with *in situ* NAP_R hyperspectral retrievals (r = 0.77, Bias = -5.3%, RMSE = 0.51 g m^{-3}, MdSA = 40.1%). NAP_{S3} retrievals were generally higher for high NAP_R values (>2 g m^{-3}) and generally lower for low NAP_R values (<0.5 g m^{-3}) (Figure 15).

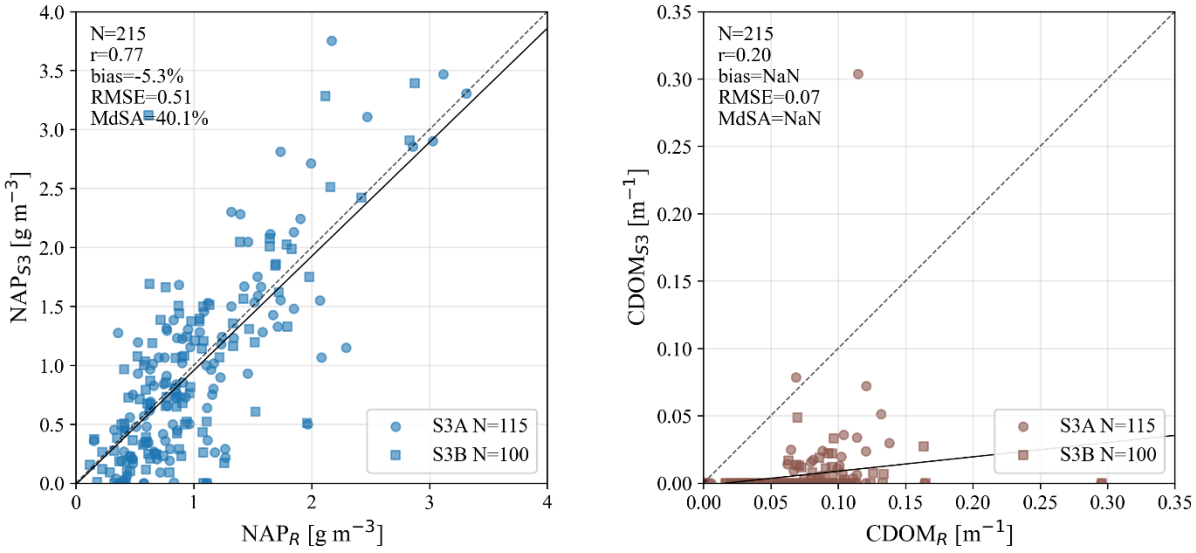

Figure 15: Validation of the satellite-modeled non-algal particle concentration (NAP_{S3}, left) and colored dissolved organic matter absorption ($CDOM_{S3}$, right) from Sentinel-3A and 3B (S3A and S3B) against the corresponding in situ estimates modeled from hyperspectral reflectance (NAP_R and $CDOM_R$). NaN values indicate numerical overflow in the bias/MdSA calculation, caused by near-zero $CDOM_{S3}$ values.

Table 13: Validation statistics for the comparisons in Figure 15: satellite-modeled non-algal particle concentration (NAP_{S3}) against the in situ estimate (NAP_R), and satellite-modeled colored dissolved organic matter absorption ($CDOM_{S3}$) against the in situ estimate ($CDOM_R$), split by satellite: Sentinel-3A (S3A), Sentinel-3B (S3B), and both combined ("all"). "Overflow" indicates numerical overflow in the statistic, caused by near-zero $CDOM_{S3}$ values.

	Pearson r	Bias [%]	RMSE	MdSA [%]	N
NAP_R vs. NAP_{S3}			[g m^{-3}]		
S3A	0.79	-8.5	0.51	41.9	115
S3B	0.74	0.9	0.52	38.3	100
all	0.77	-5.3	0.51	40.1	215
$CDOM_R$ vs. $CDOM_{S3}$			[m^{-1}]		
S3A	0.24	*overflow*	0.071	*overflow*	115
S3B	0.21	*overflow*	0.077	*overflow*	100
all	0.20	*overflow*	0.074	*overflow*	215

3.3.3 Constituent time series and interrelationships

This section presents time series and inter-comparisons for the three constituents, for both satellite retrievals and *in situ* data. It was decided that CHL_A would represent CHL, since it is

considered a reliable proxy and can be compared with earlier studies (Gupana et al., 2025; Werther et al., 2022).

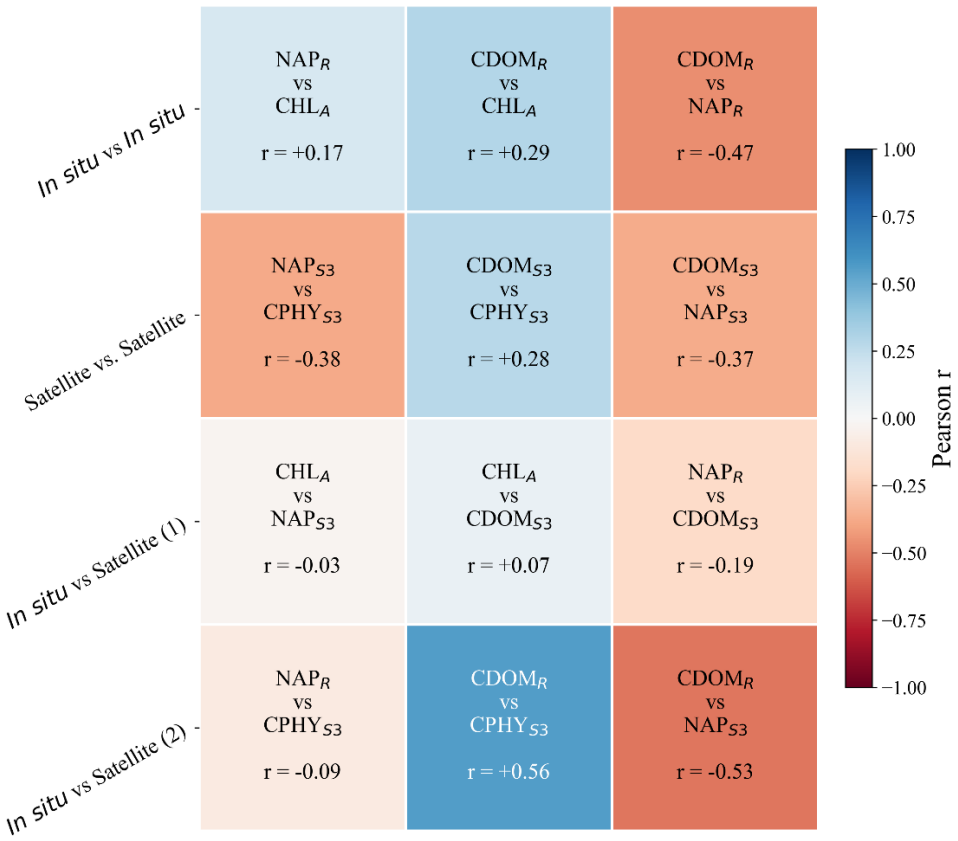

Figure 16: Pairwise Pearson correlation (r) between the three water constituents (chlorophyll a / phytoplankton, non-algal particles (NAP), and colored dissolved organic matter (CDOM). In situ quantities: CHL_A (absorption-corrected fluorometric chlorophyll a), NAP_R and $CDOM_R$ (NAP concentration and CDOM absorption at 440 nm, both modeled from hyperspectral in situ reflectance). Satellite quantities, from Sentinel-3: $CPHY_{S3}$ (phytoplankton concentration), NAP_{S3} (NAP concentration), $CDOM_{S3}$ (CDOM absorption at 440 nm).

Figure 16 shows the pairwise correlation of different constituents for all possible combinations of satellite and *in situ* data. Note that comparisons including $CDOM_{S3}$ should not be overinterpreted, as these retrievals are unreliable (see Section 3.3.2). It is notable that the relationship between NAP and CHL is rather random (r = -0.38 to 0.17), regardless of the mode of comparison. Furthermore, NAP and $CDOM_R$ exhibit moderate negative correlations (r = -0.47 to -0.53). $CDOM_R$ and CHL proxies are weakly to moderately positively correlated (r = 0.29 and 0.56 for CHL_A and $CPHY_{S3}$, respectively).

The equivalent time series (Figure 17) shows that $CPHY_{S3}$ and NAP_{S3} are mostly able to reproduce the temporal patterns of their *in situ* counterparts. $CPHY_{S3}$ captures most of the algal blooms visible in CHL_A, although there are some under- (e.g. spring 2020) and overestimations (e.g. fall 2021). In 2024, when no major blooms appeared in CHL_A, there was little visual agreement. The time series reveals prolonged periods when $CPHY_{S3}$ was significantly lower than CHL_A under high NAP conditions, most notably in summer 2024. The NAP_R time series consistently shows a summer peak between June and mid-September, when data is available, which NAP_{S3} mostly matches well. In 2020, NAP_{S3} retrievals were consistently below NAP_R.

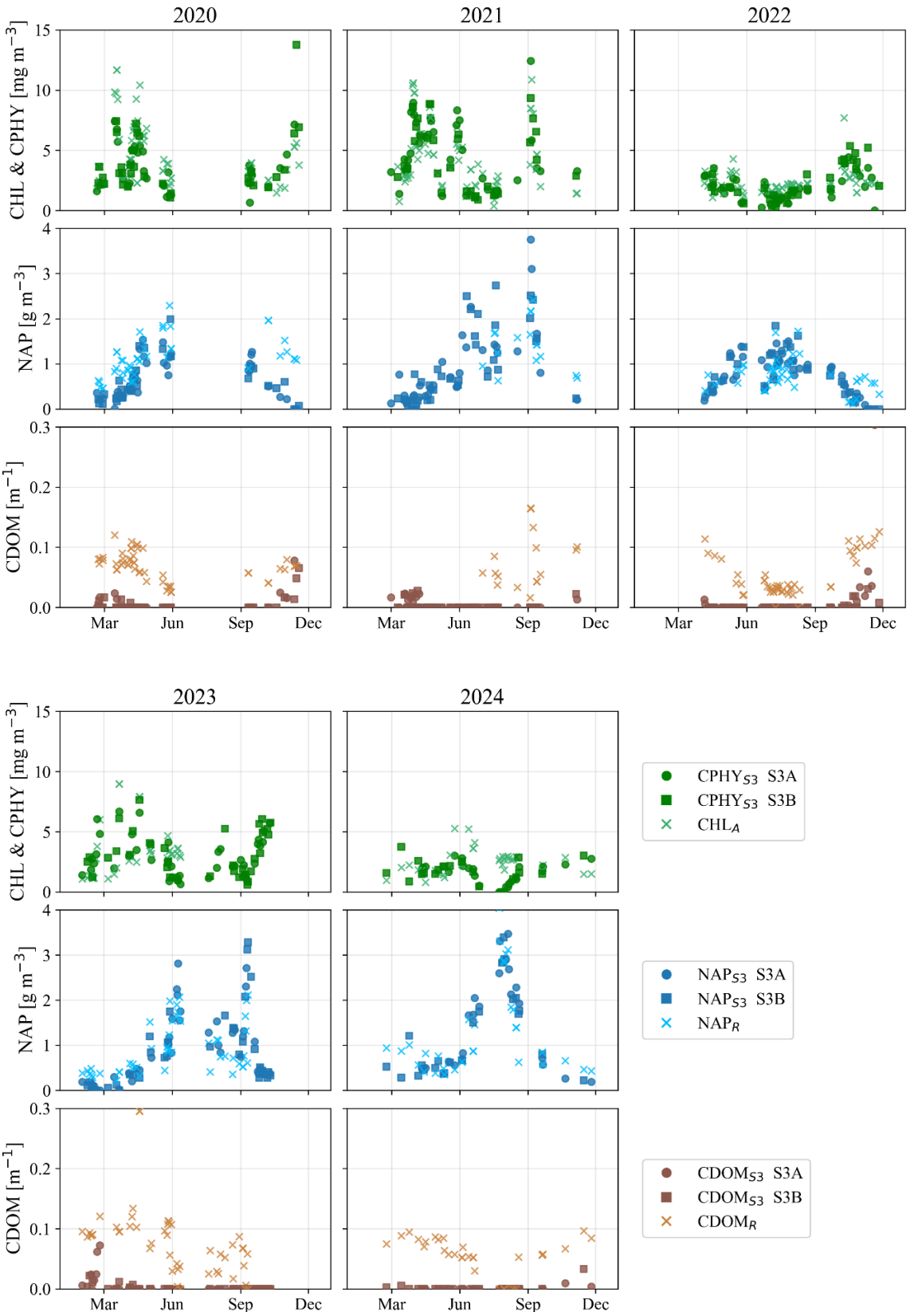

Figure 17: Time series (LéXPLORE, 2020 – 2024) of satellite and in situ water constituent estimates by year; only years with more than 10 matchups are shown. Top row: satellite-modeled phytoplankton concentration from Sentinel-3A and 3B (S3A and S3B; $CPHY_{S3}$) against absorption-corrected fluorometric chlorophyll a (CHL_A), used here as the in situ chlorophyll a reference. Middle row: satellite-modeled non-algal particle concentration (NAP_{S3}) against the in situ estimate modeled from hyperspectral reflectance (NAP_R). Bottom row: satellite-modeled colored dissolved organic matter absorption ($CDOM_{S3}$) against the in situ estimate ($CDOM_R$).

3.3.4 IOPs

Satellite $a_{wc,S3}(\lambda)$ underestimated *in situ* $a_{wc}(\lambda)$ across all bands used in the inversion (Table 14). In line with the failed CDOM retrieval reported in Section 3.3.2, Bias was strongest in the blue bands, decreasing towards the red (from -169.6% at 442.5 nm to -41.9% at 665 nm). Note that some biases were also found for hyperspectral *in situ* retrievals (see Appendix 3), which are discussed in Section 4.4. A similar pattern was observed for MdSA (a decrease from 169.6% to 52.6%) and RMSE (a decrease from 0.28 m^{-1} to 0.037 m^{-1}). The correlation was highest at 665 nm (r = 0.64) and lowest at 560 nm (r = 0.23). The median SA was 8.9° (Table 14), indicating moderate matching of the $a_{wc}(\lambda)$ spectral shape by $a_{wc,S3}(\lambda)$.

Visually, $a_{wc,S3}(\lambda)$ was highly correlated with $CPHY_{S3}$ (Figure 18). This was confirmed statistically, with the highest correlation being found for the second phytoplankton absorption peak (r = 1), with slightly lower correlations found at other wavelengths (r = 0.87 – 0.92). This pattern is also evident when comparing retrieved phytoplankton absorption (a_{phy}) with that of NAP absorption (a_{NAP}) and CDOM absorption (a_{CDOM}). a_{phy} generally dominates over a_{NAP} at high $a_{wc,S3}(\lambda)$, while at medium to low values the contributions are roughly equal at 442.5 nm. At 665 nm a_{phy} clearly dominates (Figure 18, second column), as does a_{phy} over a_{CDOM} in all bands (Figure 18, third column).

However, *in situ* $a_{wc}(\lambda)$ was not highly correlated with $CPHY_{S3}$, with r = 0.64 at 665 nm and consistently lower correlation coefficients for the other bands (r = 0.28 – 0.43).

Table 14: Validation statistics for satellite-modeled constituent absorption from Sentinel-3A and 3B (S3A and S3B) against the in situ hyperspectral estimate, at the five spectral bands used during the MiniWASI inversion (see Figure 18 for the underlying scatter plots at two of these bands). SA: spectral angle, reported here as the median value per satellite.

Band number	Central wavelength [nm]	Pearson r	Bias [%]	RMSE [m^{-1}]	MdSA [%]	N
S3A (median SA: 8.21°)						
Oa3	442.5	0.49	-159.8	0.27	159.8	81
Oa4	490	0.45	-124.9	0.15	124.9	81
Oa5	510	0.41	-111.2	0.11	111.2	81
Oa6	560	0.20	-84.2	0.062	86.7	81
Oa8	665	0.59	-42.0	0.039	54.9	81
S3B (median SA: 9.50°)						
Oa3	442.5	0.53	-177.7	0.29	177.7	65
Oa4	490	0.53	-128.0	0.15	128	65
Oa5	510	0.50	-121.6	0.11	121.6	65
Oa6	560	0.29	-100.9	0.058	101.8	65
Oa8	665	0.71	-41.7	0.035	50.7	65
All (median SA: 8.90°)						
Oa3	442.5	0.50	-169.6	0.28	169.6	146
Oa4	490	0.48	-127.8	0.15	127.8	146
Oa5	510	0.44	-120.1	0.11	120.1	146
Oa6	560	0.23	-87	0.06	93.9	146
Oa8	665	0.64	-41.9	0.037	52.6	146

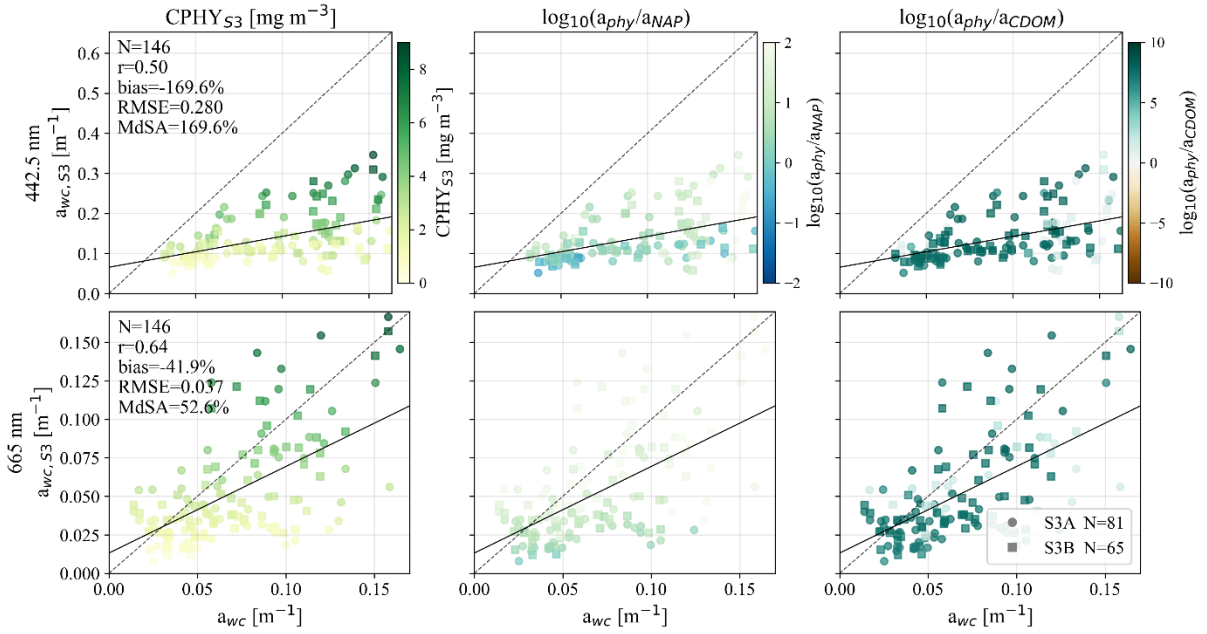

Figure 18: Validation of satellite-modeled constituent absorption from Sentinel-3A and 3B (S3A and S3B; $a_{wc,S3}$) against the in situ measurement (a_{wc}), at the wavelengths corresponding to the phytoplankton absorption peaks used in the MiniWASI inversion (442.5 and 665 nm). Left column: points colored by the satellite-modeled phytoplankton concentration ($CPHY_{S3}$). Middle column: points colored by the (log-scale) ratio of retrieved phytoplankton absorption (a_{phy}) to retrieved non-algal particle absorption (a_{NAP}). Right column: points colored by the (log-scale) ratio of a_{phy} to retrieved colored dissolved organic matter absorption (a_{CDOM}). The 1:1 line (dashed) and best fit (solid) are shown in each panel; note that rows use different axis bounds.

The correlation between satellite constituent backscattering and *in situ* measurements at adjacent wavelengths was high (r = 0.81 – 0.86; Table 15). Figure 19 shows the relationship between the retrieved backscattering at 620 nm ($b_{b,\mathrm{wc,S3}}(620)$) and *in situ* measurements at 630 nm ($b_{b,\mathrm{wc}}(630)$), exemplifying the general relationship between retrieved and measured backscattering. $b_{b,\mathrm{wc,S3}}(620)$ is largely dominated by NAP backscattering ($b_{b,\mathrm{NAP}}(\lambda)$). Accordingly, $b_{b,\mathrm{wc,S3}}(620)$ appears highly correlated with $\mathrm{NAP_{S3}}$ (Figure 19, bottom plot), which was confirmed statistically (r = 0.96).

Table 15: Validation statistics for satellite-modeled constituent backscattering from Sentinel-3A and 3B (S3A and S3B; $b_{b,wc,S3}(wvl)$, at wavelength "wvl") against the in situ estimate converted from scattering measurements ($b_{b,wc}(wvl)$), at four paired wavelengths (see Figure 19 for the scatter plot at 620/630 nm).

	Pearson r	Bias [%]	RMSE [m^{-1}]	MdSA [%]	N
$b_{b,wc,S3}$(442.5) vs. $b_{b,wc}$(440)					
S3A	0.79	-36.9	0.011	41.5	173
S3B	0.83	-38.7	0.010	39.6	152
all	0.81	-37.6	0.011	40.0	325
$b_{b,wc,S3}$(510) vs. $b_{b,wc}$(532)					
S3A	0.84	-70.8	0.010	70.8	173
S3B	0.85	-70.3	0.010	70.3	152
all	0.84	-70.4	0.010	70.4	325
$b_{b,wc,S3}$(620) vs. $b_{b,wc}$(630)					
S3A	0.86	-58.1	0.0083	58.3	174
S3B	0.87	-55.6	0.0088	55.6	153
all	0.86	-57.2	0.0085	57.9	327
$b_{b,wc,S3}$(708.75) vs. $b_{b,wc}$(700)					
S3A	0.86	-68.6	0.0088	70.4	173
S3B	0.87	-67.1	0.0086	68.4	153
all	0.86	-68.4	0.0087	68.6	326

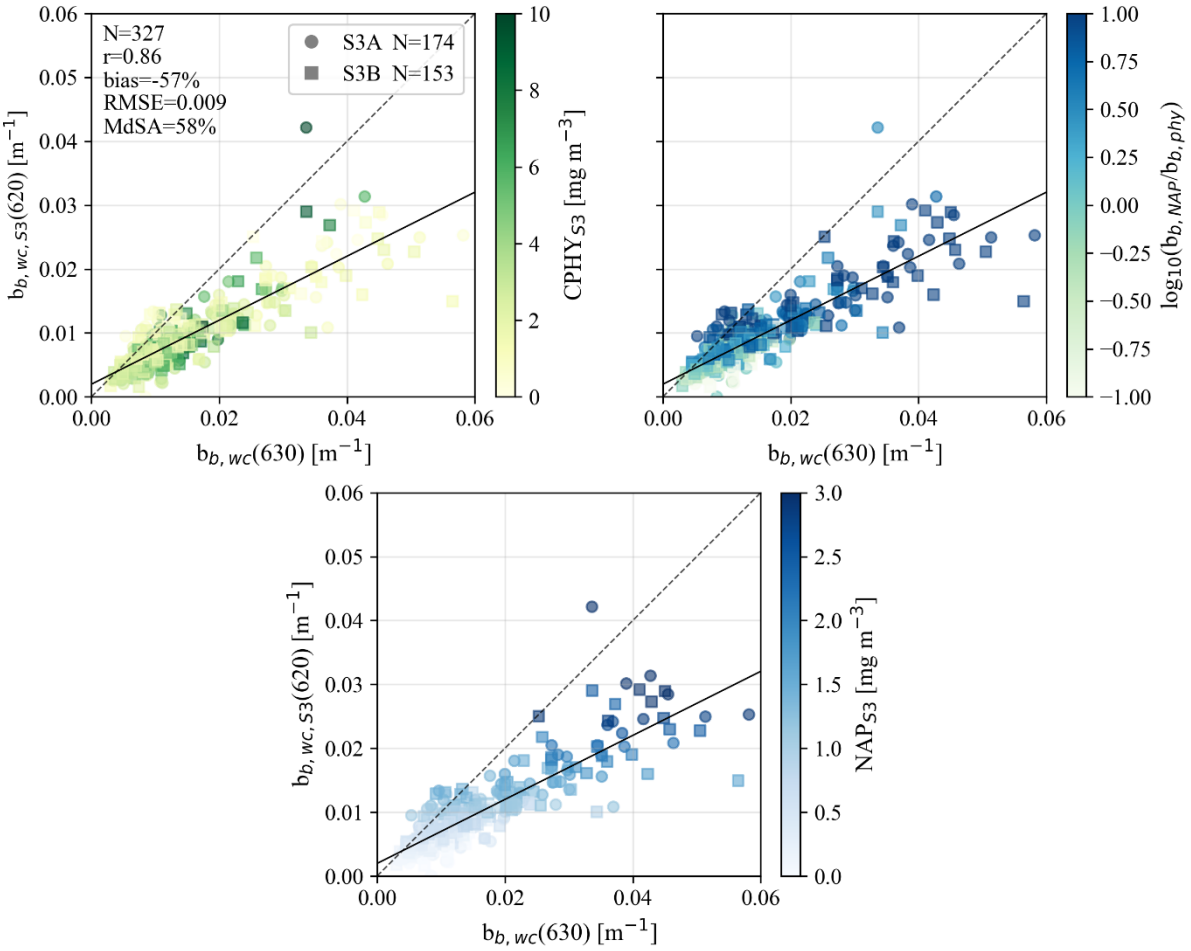

Figure 19: Validation of the satellite-modeled constituent backscattering at 620 nm from Sentinel-3 ($b_{b,wc,S3}(620)$) against the in situ estimate converted from scattering measurements at 630 nm ($b_{b,wc}(630)$); S3A and S3B matchups are shown as circles and squares. Top-left: points colored by satellite-modeled phytoplankton concentration ($CPHY_{S3}$). Top-right: points colored by the (log-scale) ratio of retrieved non-algal particulate backscattering ($b_{b,NAP}$) to retrieved phytoplankton backscattering ($b_{b,phy}$). Bottom: points colored by satellite-modeled non-algal particle concentration (NAP_{S3}). The 1:1 line (dashed) and best fit (solid) are shown in each panel.

Satellite retrievals were significantly lower than *in situ* measurements (-70.4% to -37.6%) with MdSA ranging from 40 to 70.4% and RMSE around 0.01 m^{-1} (Table 15). As for absorption, this pattern is not exclusive to satellite retrievals; a similar trend was observed in hyperspectral *in situ* inversions (see Appendix 3).

The $b_{b,\mathrm{wc,S3}}(442.5)$ vs. $b_{b,\mathrm{wc}}(440)$ comparison showed less Bias than the others (Table 15). In accordance with the spectral shape of $b^{*}_{b,\mathrm{phy}}(\lambda)$ (see Figure 5), $b_{b,\mathrm{wc,S3}}(442.5)$ naturally has a stronger $b_{b,\mathrm{phy}}(\lambda)$ component and higher total backscattering than the longer wavelengths.

Normalized *in situ* $b_{b,\mathrm{wc}}(\lambda)$ shows significant spectral variation but the spectral shapes are relatively well matched on average (Figure 20, A). Although NAP_{S3}-dominated spectra are on average more spectrally flat, they exhibit greater variability in shape than $CPHY_{S3}$-dominated spectra. Mismatch in the blue can be observed for $CPHY_{S3}$-dominated spectra (Figure 20, B) while NAP_{S3}-dominated spectra show a less prominent mismatch at 532 nm (Figure 20, C). Figure 20, D, which shows unnormalized spectra, further reinforces some of the observations

from Figure 19: a) Backscattering is generally lower for $CPHY_{S3}$-dominated spectra (for both *in situ* and satellite measurements). b) There is a mismatch in magnitude (regardless of the dominating constituent) between *in situ* measurements and satellite retrievals.

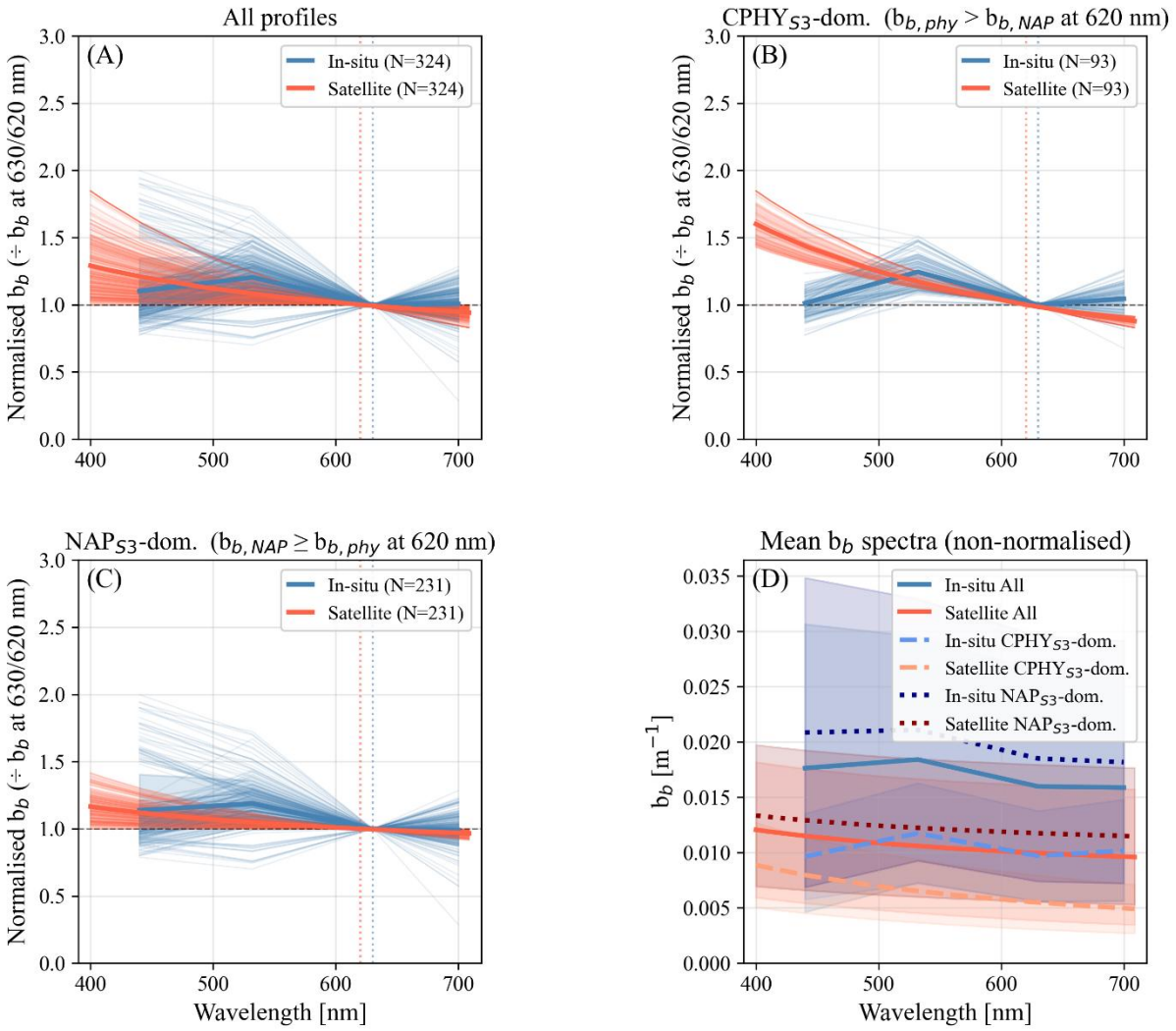

Figure 20: Comparison of satellite-modeled and in situ backscattering (b_b) spectral shapes. $b_{b,NAP}$: backscattering by non-algal particles, $b_{b,phy}$: backscattering by phytoplankton (both retrieved), $CPHY_{S3}$: satellite-modeled phytoplankton concentration, NAP_{S3}: satellite-modeled non-algal particle concentration. Spectra in panels A – C are normalized at 630 nm (in situ) or 620 nm (satellite); thick lines show the average spectrum per group. (A) All normalized spectra, in situ vs. satellite. (B) Spectra where $b_{b,phy}$ exceeds $b_{b,NAP}$ ("$CPHY_{S3}$-dominated"). (C) Spectra where $b_{b,NAP}$ exceeds $b_{b,phy}$ ("NAP_{S3}-dominated"). (D) The same six group-average spectra as A – C, but non-normalized, with shaded polygons showing the observed range within each group.

3.4 Eawag field campaigns

On 20 May 2021, both S3A and S3B covered Lake Geneva, coinciding with a field campaign on the same date. The true color RGB images reveal a gyre-like structure with a comparatively high $R_{rs}(\lambda)$ in the sampling zone (see Figure 21).

Four sampling points (P1, P2, P4, and P5) are located on the northern edge of the gyre. Accordingly, slightly higher NAP_{S3} values were retrieved (0.78, 0.99, 0.84, and 0.83 g m^{-3}) than for P3 and P6 (0.71 and 0.5 g m^{-3}) in the case of the S3A image. For $CPHY_{S3}$, values were generally lower in the main arm of the gyre; however, due to high noise this did not always translate to matchups.

The S3B image is very similar to the S3A image but is slightly darker; consequently, lower NAP_{S3} concentrations and higher $CPHY_{S3}$ concentrations were obtained (maps not shown; see Table 16). While the *in situ* samples do not necessarily follow the spatial distribution of the maps, the general range of values is in good agreement (Table 16).

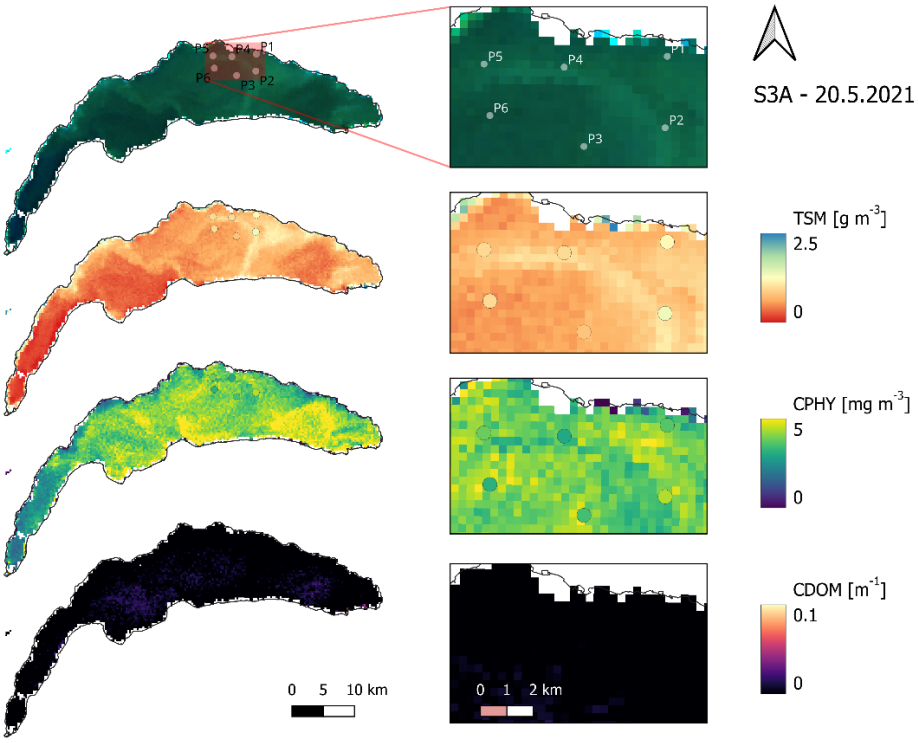

Figure 21: Map-based retrievals from a Sentinel-3A (S3A) image on 20 May 2021 in Lake Geneva, paired with in situ point samples (P1 – P6) from a spatially distributed field campaign. From top to bottom: true-color image (©Copernicus), total suspended matter (TSM, considered equivalent to the satellite-modeled particulate non-algal concentration), phytoplankton concentration (CPHY; in situ samples represent all pigments combined), and colored dissolved organic matter absorption at 440 nm (CDOM; no in situ samples available for comparison). The Lake Geneva shoreline was downloaded from OpenStreetMap (©OpenStreetMap contributors, openstreetmap.org/copyright).

Table 16: In situ vs. satellite-modeled matchup values at the six sampling points (P1 – P6) of the 20 May 2021 field campaign shown in Figure 21, for Sentinel-3A and 3B (S3A and S3B). Top block: in situ total suspended matter (TSM) against satellite-modeled non-algal particle concentration (NAP_{S3}). Bottom block: in situ total phytoplankton pigment concentration from laboratory HPLC analysis ($CPHY_{HPLC}$) against satellite-modeled phytoplankton concentration ($CPHY_{S3}$).

	P1	P2	P3	P4	P5	P6
TSM vs. NAP_{S3} [g m^{-3}]						
S3A	0.78	0.99	0.71	0.84	0.83	0.50
S3B	0.65	0.78	0.65	0.65	0.69	0.50
In situ	1.2	1.37	0.80	0.97	0.96	0.97
$CPHY_{HPLC}$ vs. $CPHY_{S3}$ [mg m^{-3}]						
S3A	3.66	3.28	4.31	3.61	4.15	4.44
S3B	4.31	4.57	4.23	4.58	4.32	4.55
In situ	3.67	4.43	3.50	3.00	3.86	3.19

On 11 June 2021, a campaign was covered only by S3B. This revealed very bright areas with high $R_{rs}(\lambda)$ in the sampling zone towards the north-eastern shore, with NAP_{S3} retrievals exceeding 2.5 g m^{-3}. $CPHY_{S3}$ retrievals were relatively low in the area of interest, particularly in regions with very high NAP_{S3} values (Figure 22).

NAP_{S3} retrievals matched the general spatial distribution of the *in situ* TSM samples. The two points within the bright area (P3 and P4) corresponded to the highest retrievals (2.15 and 2.56 g m^{-3}) and *in situ* samples (2.33 and 2.3 g m^{-3}). P2, located on the very edge of the bright area, exhibited similar NAP_{S3} (2 g m^{-3}) but somewhat lower *in situ* values (1.65 g m^{-3}). P5 and P6 had lower NAP_{S3} values (1.43 and 1.24 g m^{-3}), with *in situ* TSM values of 1.45 and 1.75 g m^{-3}.

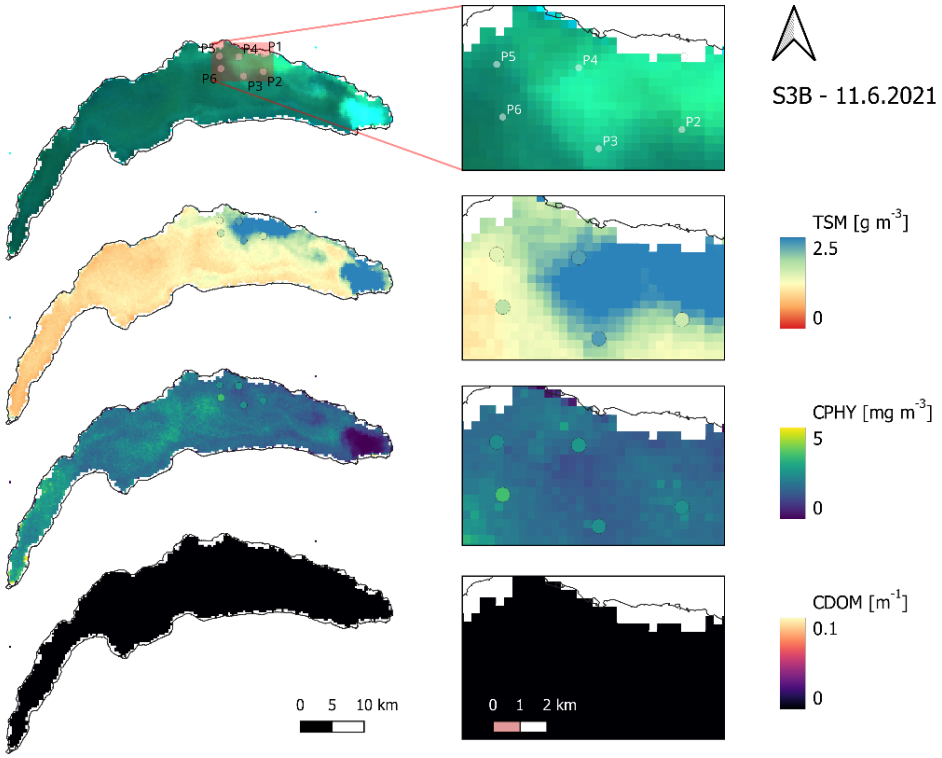

Figure 22: Map-based retrievals from a Sentinel-3B (S3B) image on 11 June 2021 in Lake Geneva, paired with in situ point samples (P2 – P6; P1 fell outside the valid image area) from a spatially distributed field campaign. From top to bottom: true-color image (©Copernicus), total suspended matter (TSM, considered equivalent to the satellite-modeled particulate non-algal concentration), phytoplankton concentration (CPHY; in situ samples represent all pigments combined), and colored dissolved organic matter absorption at 440 nm (CDOM; no in situ samples available for comparison). The Lake Geneva shoreline was downloaded from OpenStreetMap (©OpenStreetMap contributors, openstreetmap.org/copyright).

Both $CPHY_{S3}$ and $CPHY_{HPLC}$ indicate that P6 has the highest CPHY concentrations. While $CPHY_{S3}$ does not strictly adhere to the spatial distribution of $CPHY_{HPLC}$ and shows slightly lower concentrations, the absolute values are in a similar range (see Table 17).

Table 17: In situ vs. satellite-modeled matchup values at five of the six sampling points (P2 – P6; P1 fell outside the valid image area) of the 11 June 2021 field campaign shown in Figure 22, for Sentinel-3B (S3B). Top block: in situ total suspended matter (TSM) against satellite-modeled non-algal particle concentration (NAP_{S3}). Bottom block: in situ total phytoplankton pigment concentration from laboratory HPLC analysis ($CPHY_{HPLC}$) against satellite-modeled phytoplankton concentration ($CPHY_{S3}$).

	P1	P2	P3	P4	P5	P6
TSM vs. NAP_{S3} [g m^{-3}]						
S3B	-	2.0	2.15	2.56	1.43	1.24
In situ	-	1.65	2.33	2.30	1.45	1.75
$CPHY_{HPLC}$ vs. $CPHY_{S3}$ [mg m^{-3}]						
S3B	-	1.93	1.57	1.45	1.76	2.14
In situ	-	2.47	2.49	2.67	2.40	3.48

Images from S3A and S3B on 16 June 2021 show that the abundance of suspended matter recorded five days earlier had moved towards the western edge of the sampling zone. This is reflected in the NAP_{S3} retrievals (Figure 23). Sampling points P4, P5, and P6 were located in bright areas, with S3A retrievals of 2.27, 2.78, and 2.57 g m^{-3}, respectively. P1 was located right on the edge with a retrieval of 2.2 g m^{-3}, while P2 and P3 were located to the south-east with retrievals of 1.55 and 1.5 g m^{-3}, respectively. This pattern is only partly visible in the *in situ* data, with notably high concentrations at P2 and notably low concentrations at P6 (Table 18).

$CPHY_{S3}$ showed rather low concentrations (around 1.5 mg m^{-3}) across the entire area of interest. A horseshoe-like area of very low (often below 1 mg m^{-3}) $CPHY_{S3}$ encompassed P4 and P5, where the RGB image is distinctly turquoise. The $CPHY_{HPLC}$ data matched the map very well, making the *in situ* sampling points nearly invisible in the CPHY map in Figure 23. Retrievals for the S3B image (not shown) are almost identical to those of S3A, except for slightly lower concentrations of both NAP_{S3} and $CPHY_{S3}$ (see Table 18).

Table 18: In situ vs. satellite-modeled matchup values at the six sampling points (P1 – P6) of the 16 June 2021 field campaign shown in Figure 23, for Sentinel-3A and 3B (S3A and S3B). Top block: in situ total suspended matter (TSM) against satellite-modeled non-algal particle concentration (NAP_{S3}). Bottom block: in situ total phytoplankton pigment concentration from laboratory HPLC analysis ($CPHY_{HPLC}$) against satellite-modeled phytoplankton concentration ($CPHY_{S3}$).

	P1	P2	P3	P4	P5	P6
TSM vs. NAP_{S3} [g m^{-3}]						
S3A	2.20	1.55	1.50	2.27	2.78	2.57
S3B	2.14	1.48	1.37	1.99	2.45	2.26
In situ	1.70	2.10	1.60	2.12	1.9	1.45
$CPHY_{HPLC}$ vs. $CPHY_{S3}$ [mg m^{-3}]						
S3A	1.42	1.49	1.49	0.91	1.11	1.43
S3B	1.35	1.44	1.39	0.88	0.71	1.46
In situ	1.32	1.55	1.82	1.07	1.02	1.44

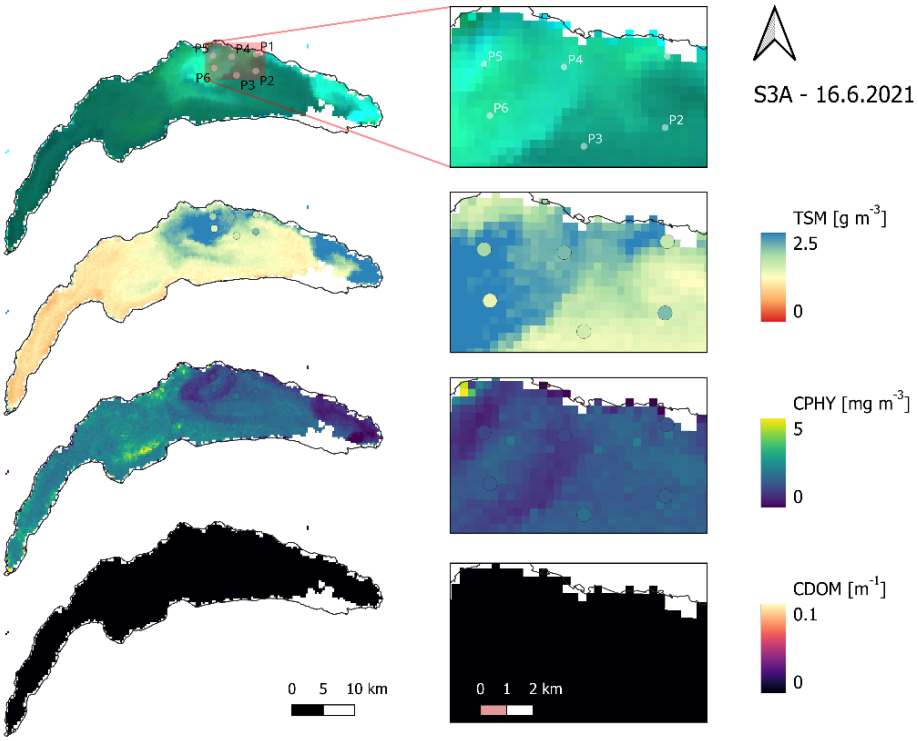

Figure 23: Map-based retrievals from a Sentinel-3A (S3A) image on 16 June 2021 in Lake Geneva, paired with in situ point samples (P1 – P6) from a spatially distributed field campaign. From top to bottom: true-color image (©Copernicus), total suspended matter (TSM, considered equivalent to the satellite-modeled particulate non-algal concentration), phytoplankton concentration (CPHY; in situ samples represent all pigments combined), and colored dissolved organic matter absorption at 440 nm (CDOM; no in situ samples available for comparison). The Lake Geneva shoreline was downloaded from OpenStreetMap (©OpenStreetMap contributors, openstreetmap.org/copyright).

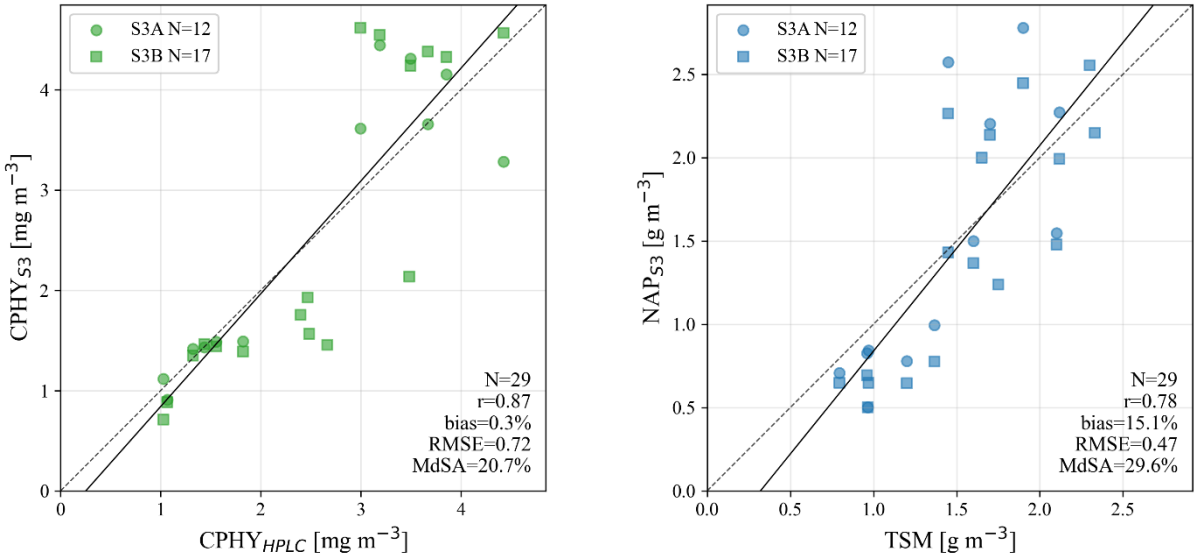

Figure 24: Combined validation of satellite-modeled phytoplankton concentration ($CPHY_{S3}$, left) and non-algal particle concentration (NAP_{S3}, right) against in situ references, pooling matchups from five Sentinel-3A and 3B (S3A and S3B) images across the three spatially distributed field campaigns shown in Figures 21 – 23. In situ references: total phytoplankton pigment concentration from laboratory HPLC analysis ($CPHY_{HPLC}$) and total suspended matter (TSM). The 1:1 line (dashed) and best fit (solid) are shown in each panel.

Combining the results from all three campaigns (Figure 24), very good agreement was achieved between $CPHY_{S3}$ and $CPHY_{HPLC}$ (r = 0.87, Bias = 0.3%, RMSE = 0.72 mg m^{-3}, MdSA = 20.7%). Similarly good results were also obtained when comparing NAP_{S3} and TSM (r = 0.78, Bias = 15.1%, RMSE = 0.47 g m^{-3}, MdSA = 29.6%).

However, when $CPHY_{S3}$ was compared to CHL_{HPLC} (also excluding chlorophyll *a* chlorophyllide and other chlorophylls) instead of $CPHY_{HPLC}$ (total pigment concentration), the correlation remained similar (r = 0.81), but a significant Bias was introduced (86.2%), resulting in apparent overestimates.

4. Discussion

4.1 Bulk phytoplankton retrieval: Strengths and limitations

4.1.1 SHL2 validation and matchup representativeness

Although retrievals of individual taxa were unsuccessful (as will be discussed in Section 4.5), bulk phytoplankton retrievals in the form of CHL largely achieved good results (Sections 3.2.1, 3.3.1, and 3.4).

Good results were achieved at SHL2 (r = 0.8) with good matching of the *in situ* time series. Underestimation of CHL_S coincided with the maximum being below the Secchi depth, yet still within the top 7.5 m of the water column, for five matchups in mid-2021 and autumn 2023. In such cases, the CHL maximum is not visible to the satellite, yet still relevant in the *in situ* reference.

Using the same dataset, Bonnier et al. (2024) weighted the *in situ* data according to optical visibility in order to improve the representativeness of the matchup. This resulted in better agreement between the S3-derived ESA climate change initiative (CCI) product and CHL_S than when a simple depth average was taken (r = 0.8 vs. r = 0.85). They also noted that agreement was generally highest when considering the first 2 or 5 m of the water column. However, we observed no improvements when decreasing the averaging depth, so it was kept at 7.5 m.

Although optically weighing the *in situ* data may increase the representativeness of the matchup, a simple depth-average CHL is a physically defined and directly measurable quantity. In contrast, optically weighted CHL is a derived quantity whose value depends on the chosen weighting scheme. Matchup representativeness of this study is further discussed in-depth in Section 4.7.

This study achieved the same correlation as the ESA CCI product (r = 0.8). However, the uncorrected ESA CCI product showed a clear Bias of -2.7 mg m^{-3}, whereas this was less of an issue for MiniWASI retrievals (see Section 3.2.1). This supports the general notion that physical models are more widely applicable (Giardino et al., 2019), whereas empirical models, such as the OC2 algorithm used by ESA CCI for Lake Geneva, require lake-specific calibration. Comparing different OCx algorithms, Bonnier et al. (2024) found best results for re-calibrated OC4 (r = 0.9); however, this comparison applied the aforementioned optical weighting strategy, making it difficult to directly compare the results with those of this study.

4.1.2 LéXPLORE: Intercomparison of *in situ* CHL and CPHY references

At LéXPLORE, there were substantial differences between the *in situ* products of CHL and CPHY. CHL_F was negatively biased compared to the other products, most likely due to systematic underestimations under NPQ conditions (see Section 2.3.2). $CPHY_R$ exhibited some outliers that negatively impacted the correlation with the other products. Nevertheless, its good overall agreement with CHL_A and $aLH_{676} \rightarrow CHL$ supports the chosen methodology as a means of phytoplankton retrieval.

While aLH_{676} showed a high degree of correlation with other products, cross-comparisons revealed that the linear conversion of aLH_{676} to CHL consistently produced overestimates at low concentrations and underestimates at high concentrations. Various equations have been suggested to calculate the conversion of aLH_{676} to CHL in Case 1 waters (see e.g. Boss et al., 2013; Cael et al., 2020). Conversion from phytoplankton absorption to CHL or CPHY depends on the packaging effect (Bricaud et al., 1995; Ferreira et al., 2017) and thus also cell size and the phytoplankton community composition (Boss et al., 2013; Cael et al., 2020). Furthermore, any Bias in the absorption measurements naturally propagates, so conversion was computed based on a local relationship for the Thetis profiler (Minaudo et al., 2021, supplementary material; see Appendix 2).

While such conversions are routinely used to assess CHL and CPHY in Case 1 waters, the used conversion method has not been evaluated for the Thetis profiler. Rather, past research at LéXPLORE used aLH_{676} to correct CHL_F for NPQ effects (Gupana et al., 2025; Werther et al., 2022). Although CHL_A was consistent with other products, uncertainties in the absorption measurements may propagate to aLH_{676} and, in turn, to CHL_A.

Quality control of the hyperspectral absorption profiles revealed that many of them had spikes and gaps. Such profiles were filtered out for absorption matchups (N = 146), but not for aLH_{676} matchups (N = 299). Using the standard L2 aLH_{676} provided by the Datalakes website ensures consistency with previous studies using the same dataset (Gupana et al., 2025; Werther et al., 2022), as well as general reproducibility. However, additional quality control may improve the accuracy of the products.

4.1.3 Factors limiting retrieval performance at LéXPLORE

The correlations between $CPHY_{S3}$ and the *in situ* data at LéXPLORE (ranging from 0.68 to 0.73) were lower than those at SHL2 (r = 0.8). Furthermore, $CPHY_{S3}$ tended to underestimate CHL and CPHY significantly, except for CHL_F, which was expected due to NPQ effects. Possible reasons for the degraded performance may include:

- Lower quality of *in situ* data: CHL_S samples are collected using a state-of-the-art spectrophotometric approach (ISO10260:1992). However, CHL_F, CHL_A, and aLH_{676}→CHL are not laboratory measurements of CHL, but rather convert either fluorescence (CHL_F), absorption (aLH_{676}→CHL), or a mix of both (CHL_A) to CHL. This data is further acquired autonomously by the Thetis profiler and is not routinely manually checked. $CPHY_R$ is, like $CPHY_{S3}$, a result of $R_{rs}(\lambda)$ inversion and should thus be interpreted with additional caution.
- Positional uncertainty: To avoid undesired effects of the LéXPLORE platform on $R_{rs}(\lambda)$, the pixel to the south of the platform was taken as reference. This resulted in a spatial mismatch: The distance between the Thetis profiler and the center of the southern pixel is approximately 250 m.
- Lower quality of satellite data due to adjacency effect (AE): Pixels located close to the shore are more strongly impacted, i.e. a significant part of the optical signal may originate from adjacent land pixels through atmospheric scattering rather than the lake (Castagna and Vanhellemont, 2025).

- Higher optical complexity: While NAP_{S3} was at most moderate (2 g m^{-3}) at SHL2 (Figure 10), high concentrations of up to 4 g m^{-3} were sometimes observed at LéXPLORE (Figure 17). This reflects large scale lake processes: At the northern shore, interflow dynamics may lead to the upwelling of suspended matter and the precipitation of calcite (Many et al., 2022; Many et al., 2024; Nouchi et al., 2019). The possible implications, which include the masking of the optical signal (Giardino et al., 2019) and degraded representativeness of the matchups, are discussed further.
- Differing timeframe: SHL2 data did not include 2024 data, where retrievals at LéXPLORE were in comparatively bad agreement with CHL_A (see Section 3.3.3).

Noticeable underestimations of CHL_A in summer 2024 may be caused by high NAP concentrations. On one hand, the accurate retrieval of low to intermediate CHL may be difficult given the large NAP contribution to the optical signal ("masking effect"). On the other hand, high NAP concentrations could affect the representativeness of the matchup, as the maximum CHL may not be visible to the satellite while still showing up in the depth-average CHL_A. This has also been demonstrated for the SHL2 station using auxiliary Secchi depth data. Although visibility is unknown, inspection of the corresponding *in situ* aLH_{676} depth profiles revealed non-surface CHL maxima, suggesting that this assumption is reasonable.

The issue of representativeness is illustrated by an outlier in Figure 14: On 5 September 2021, $CPHY_{S3}$ signaled a significant algal bloom (around 12 mg m^{-3}), whereas CHL_A was more conservative in its estimate (8 mg m^{-3}). This instance corresponds to the aforementioned *Uroglena* sp. bloom (see Figure 1). For the same event Irani Rahaghi et al. (2024) estimated surface concentrations of up to 50 mg m^{-3}. This discrepancy may be due to strong non-uniformity, with CHL being strongly concentrated at the surface (thus strongly influencing $R_{rs}(\lambda)$) but decreasing rapidly with depth, as can be seen in the corresponding aLH_{676} profile (see Appendix 4).

4.1.4 Benchmarking against empirical algorithms and literature

A comparison with CHL_A (MdSA = 45%) revealed that the MiniWASI retrievals outperformed various empirical algorithms that were not specifically calibrated for Lake Geneva, achieving an MdSA ranging from 49.79% to 175.18% (Werther et al., 2022). However, re-calibrated OC3 (MdSA = 26.59%) outperformed retrievals presented in this study (Gupana et al., 2025). It should be noted that the study by Gupana et al. (2025) did only include data until August 2022 and that MiniWASI performance was somewhat weaker thereafter, especially in 2024 (see Figure 17).

4.1.5 Pigment composition and SIOP considerations

The best results were obtained for matchups with $CPHY_{HPLC}$ (see Section 3.4), resulting in a high correlation coefficient (r = 0.87) and low absolute and relative errors (RMSE = 0.72 mg m^{-3}, MdSA = 20.7%). The poorer results obtained when considering only CHL_{HPLC} demonstrate that careful consideration of which pigments are to be retrieved on the model side and which are measured *in situ* is required for such a validation exercise.

The SIOPs used in this study encompass a variety of absorption features that do not originate from CHL. For example, green algae exhibit features of chlorophyll *b*, while the spectrum of the used cyanobacteria has a strong phycoerythrin feature at 560 nm (Maire et al., 2025). Accessory pigments such as chlorophyll *b* and *c*, as well as various photosynthetic and photoprotective carotenoids can account for a significant proportion of the total phytoplankton pigment concentration (Babin et al., 2003a; Ferreira et al., 2017). For the three field campaigns, HPLC analysis showed accessory pigment concentrations of about 50%, mainly consisting of chlorophyll *b*, chlorophyll *a* chlorophyllide, and carotenoids such as Fucoxanthin and Alloxanthin (pigments ordered by total concentration).

However, in WASI, phytoplankton absorption was normalized by either CHL (e.g. for cyanobacteria; Xi et al. (2017)), or, in some cases, the sum of CHL and phaeophytin (e.g. for diatoms; Gege (2012)). Therefore, although accessory pigments affect the spectral shape of SIOPs, the "phytoplankton concentration" retrieved in WASI is primarily the phytoplankton CHL concentration. Hence, $CPHY_{S3}$ should, in theory, be in better agreement with CHL_{HPLC} than $CPHY_{HPLC}$, but this was not the case. No satisfactory explanation could be found for this surprising result.

While HPLC is the gold standard for the analysis of phytoplankton pigment composition (A. Aminot; F. Rey, 2001), the other methods this study used as reference are aimed at only retrieving CHL. However, the quantity retrieved may include contributions from other pigments. For example, spectrophotometric procedures cannot distinguish between chlorophyllides and CHL, which naturally leads to slight overestimation (A. Aminot; F. Rey, 2001).

Similar validation studies of constituent retrievals often overlook the optical complexity of phytoplankton pigments. Since CHL and the concentration of accessory pigments are highly correlated (Babin et al., 2003b), empirical models for CHL retrievals may implicitly correct for the accessory pigment signal during calibration. For bio-optical models, careful consideration of both the used SIOPs and the *in situ* measurements is advised.

4.2 NAP: Stable retrieval

NAP retrievals were in good overall agreement with TSM field samples from three spatially distributed campaigns (see Section 3.4). Notable outliers existed for the 16 June 2021 campaign: For P2, high TSM was sampled, whereas a dark water color resulted in low NAP_{S3} retrievals. The opposite occurred for P6, where whiting of the water and high NAP_{S3} were observed, but low TSM was sampled *in situ*.

From a visual interpretation of the water color, it seems unlikely that P2 experienced higher TSM concentrations than P6. Possible explanations unrelated to the quality of the *in situ* data could include small-scale variations that cannot be resolved adequately by the spatial resolution of the S3 OLCI (300 m), or issues arising from a temporal mismatch (four to five hours).

A relatively high correlation has been found between NAP_{S3} and NAP_R ($r = 0.77$). This result is supported by the high correlation observed for the predominantly NAP-dominated backscattering IOPs ($r > 0.8$ for all comparisons; see Section 3.3.4). However, it should be noted

that NAP_R is not a direct measurement, but rather the result of *in situ* hyperspectral inversion using essentially the same model configuration as NAP_{S3}. The general range of values is consistent with past inter-annual averages (2002 – 2015) reported for Lake Geneva. These show low TSM concentrations of less than 0.5 g m^{-3} during winter and higher concentrations of more than 1 g m^{-3} during summer, which coincides with the hydrological cycle of the Rhône (Nouchi, 2018).

Although some studies have investigated whiting events in Lake Geneva using spaceborne RS (see e.g. Li et al., 2022; Many et al., 2022; Nouchi et al., 2019), to the authors knowledge no previous studies have directly validated retrievals of suspended matter concentrations. For the application of particle tracking Li et al. (2022) retrieved TSM using the S3 OLCI 665 nm band. Indirect validation using turbidity measurements at SHL2 revealed a higher RMSE (0.76 g m^{-3}, supplementary material) than achieved for this study (0.47 – 0.51 g m^{-3}).

4.3 CDOM: A retrieval failure and its causes

CDOM retrievals were unsuccessful, with $CDOM_{S3}$ retrievals being well below the $CDOM_R$ references. Average CDOM values for Lake Geneva are 0.1 m^{-1} (Nouchi, 2018), hence this discrepancy is due to a retrieval failure and not related to $CDOM_R$. The failure is most likely due to the interplay of several factors:

- AC errors in the blue bands: For matchups with (presumably) significant CDOM, i.e. with low to intermediate $R_{rs}(\lambda)$ in the 400 – 490 nm range, there is a clear overestimation of $R_{rs}(\lambda)$. Consequently, the absence of CDOM, which strongly absorbs at these wavelengths, may be compensating for these AC errors.
- Limited weighting of the blue wavelengths: Since CDOM strongly absorbs in the blue wavelengths, CDOM retrieval algorithms rely on them. This study did not use the first two OLCI bands (400 and 412.5 nm) in order to stabilize $CPHY_{S3}$ retrievals. One intended use of these two bands is to help improve CDOM retrievals (Table 4).
- The CDOM absorption in Lake Geneva is far below that exhibited by most waters in comparable studies. Odermatt et al. (2012) showed that previous studies had considered CDOM absorption ranges at 400 nm to be between 0.1 and 10 m^{-1}. Griffin et al. (2018) define visibly colored CDOM waters as having a CDOM value greater than 3 m^{-1}.
- Spectral ambiguity: Retrieving individual phytoplankton taxa is strongly ill-posed, as is retrieving CDOM. It "competes" with six other unknown absorbers (C_1 – C_5 and C_X), whose absorption peaks are largely superimposed (see Figure 5). This study observed through experimentation that the amount of retrieved CDOM decreased with an increasing number of unknown C_i.

The CDOM absorption configuration with S_{CDOM} = 0.14 should accurately represent Lake Geneva, given that reported values of S_{CDOM} range between 0.1 and 0.18 (Nouchi, 2018). Therefore, failure to retrieve data due to the configuration of CDOM absorption is unlikely.

This retrieval failure propagates through the whole range of relevant AOP, IOP, and OAC comparisons: There is too high an input of satellite $R_{rs}(\lambda)$ in the blue wavelengths (Table 8), a

very strong underestimation of $a_{wc}(\lambda)$ in the blue wavelengths (Figure 18) and consequently a low level of CDOM absorption (Figure 15).

4.4 IOPs: Implications for model configuration

The retrieved $a_{\mathrm{wc,S3}}(\lambda)$ strongly underestimated absorption, resulting in negative biases and significant errors (see Section 3.3.4). This is to a significant part due to missing CDOM absorption which was close to 0 at LéXPLORE for most of the time. Accordingly, Bias decreased with increasing wavelength as the CDOM absorption became smaller. The erroneous domination of the absorption spectrum by $\mathrm{CPHY_{S3}}$ is indicated by high correlations with $a_{\mathrm{wc,S3}}(\lambda)$ but low to moderate correlations with *in situ* $a_{wc}(\lambda)$. This phenomenon is also reflected in the stronger correlation between $\mathrm{CDOM_R}$ and $\mathrm{CPHY_{S3}}$ (r = 0.56) than between $\mathrm{CDOM_R}$ and $\mathrm{CHL_A}$ (r = 0.29).

Significant negative biases (>100% in the blue-green region) show that, even though CDOM absorption is small for Lake Geneva, CDOM is still a major absorber comparatively, even at low concentrations. For typical Lake Geneva concentrations (C_X = 1 g m^{-3}, C_y = 0.1 m^{-1}, C_0 = 3 mg m^{-3}) WASI forward modelling shows that C_y (i.e. CDOM) makes up approximately 50% of total $a_{wc}(\lambda)$ in the short blue wavelengths and significantly impacts $R_{rs}(\lambda)$ (Figure 25).

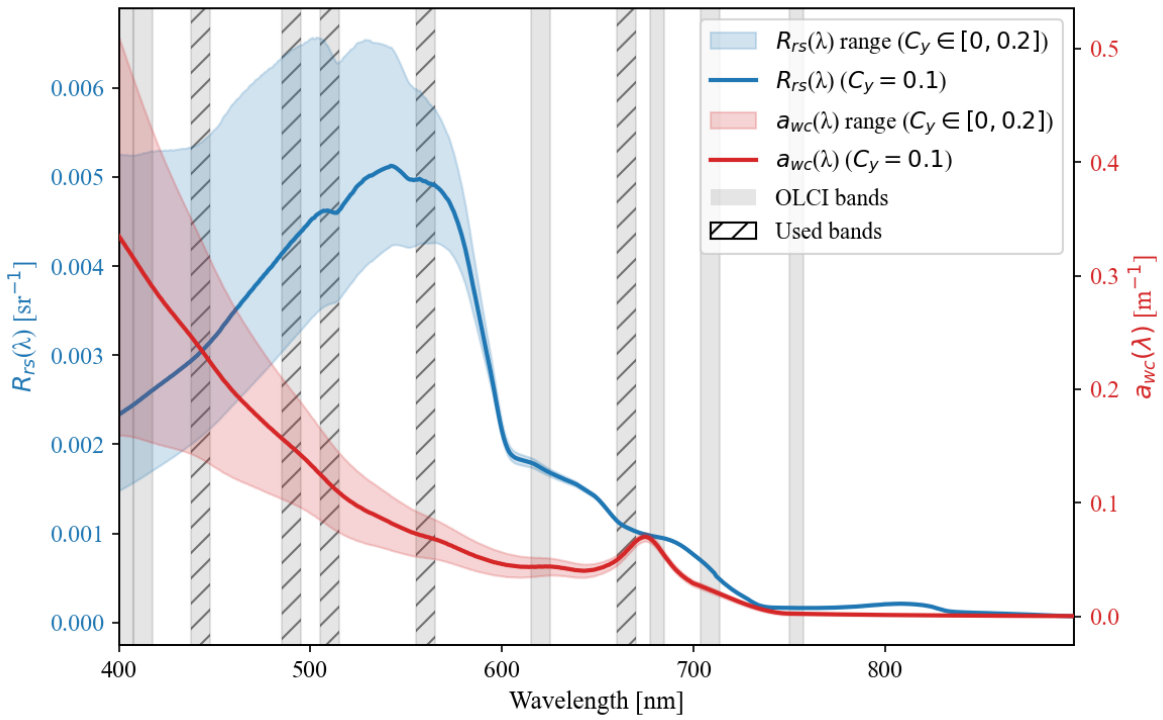

Figure 25: WASI forward-modeled remote sensing reflectance ($R_{rs}(\lambda)$) and constituent absorption ($a_{wc}(\lambda)$) for "standard" Lake Geneva concentrations (C_0 = 3 mg m^{-3}, C_x = 1 g m^{-3}) with C_y varying between 0 and 0.2 m^{-1}. Grey bars indicate Sentinel-3 OLCI wavelengths (full width at half maximum); hatched bars signify the bands used for MiniWASI image inversion.

The absence of CDOM and the accompanying Bias make it difficult to interpret the absorption bands used below 600 nm. At 665 nm, the contribution of CDOM absorption should be minimal.

However, Bias was still present at this wavelength, albeit to a much lesser extent than at smaller wavelengths (-41.9%).

To rule out effects of any biases in the satellite images themselves, hyperspectral *in situ* retrievals of $a_{wc}(\lambda)$ were cross-checked, revealing high negative biases of -63.8% to -100.5% (see Appendix 3).

This Bias prompted a thorough investigation into the processing and calibration of absorption data, which revealed AC-S instrument drift (see Appendix 5). This drift can be up to one order of magnitude higher than the constituent absorption. Although processing the L2 Thetis dataset attempts to account for this drift by linearly interpolating between calibration measurements, the magnitude of the drift and the infrequent nature of these measurements suggest that the quality of $a_{wc}(\lambda)$ is limited. As drift does not appear to be strongly wavelength-dependent, its impact on aLH_{676} is assumed to be comparatively small.

$b_{b,\mathrm{wc,S3}}(\lambda)$ underestimated *in situ* $b_{b,\mathrm{wc}}(\lambda)$, especially at high values. As the same pattern was also observed for hyperspectral *in situ* retrievals (see Appendix 3), the observed Bias cannot be attributed to errors in OLCI $R_{rs}(\lambda)$ or reduced spectral resolution.

Observed biases in absorption and backscattering could indicate a SIOP mismatch. The underestimation of $a_{wc}(\lambda)$ would enable an underestimation of $b_{b,\mathrm{wc}}(\lambda)$ while maintaining similar $R_{rs}(\lambda)$. However, issues observed in the absorption profiles (see Appendix 5) call into question the accuracy of the unnormalized $a_{wc}(\lambda)$.

Retrieved backscattering was found to be largely dominated by NAP_{S3}. As no significant Bias was observed in the retrieved NAP (see Figure 24) it is assumed that the magnitude of $b^*_{b,\mathrm{X}}(\lambda)$ is generally appropriate. This notion is supported by similar $b^*_{b,\mathrm{NAP}}(\lambda)$ magnitudes for other peri-alpine lakes such as Lake Garda (Giardino et al., 2014), Lake Maggiore, and Lake Varese (Niroumand-Jadidi et al., 2022) based on exponential approximations of *in situ* measurements.

Note that these studies generally assume $b^*_{b,\mathrm{NAP}}(\lambda)$ to be higher than in this study. To increase $b_{b,\mathrm{wc.S3}}(\lambda)$ even lower $b^*_{b,\mathrm{X}}(\lambda)$ would be needed. Even when $b^*_{b,\mathrm{X}}(\lambda)$ was set unreasonably low in experiments (e.g. to 0.003 $m^2\ g^{-1}$ at all wavelengths), which sometimes resulted in very unlikely $NAP_{S3} > 10\ g\ m^{-3}$, the negative Bias of $b_{b,\mathrm{wc,S3}}(\lambda)$ was still not fully removed.

Of course, $b_{b,\mathrm{wc}}(\lambda)$ is influenced not only by backscattering SIOPs such as $b^*_{b,\mathrm{X}}(\lambda)$ but also indirectly by absorption SIOPs (especially $a^*_{\mathrm{NAP}}(\lambda)$) and anisotropy modelling (e.g. the ill-favored Q factor). Removing the observed Bias would require parameter tuning of $b^*_{b,\mathrm{NAP}}(\lambda)$ and $a^*_{\mathrm{NAP}}(\lambda)$ leading to unreasonable constituent outputs. Therefore, while uncertainties in SIOPs may contribute to the Bias, it is deemed unlikely that they fully explain it.

Rather than measuring *in situ* backscattering directly, scattering at a single angle is converted to backscattering using the equations of Boss & Pegau (2001). This conversion may introduce uncertainties (see, for example, Vaillancourt (2004)), although not in the observed magnitude. Issues due to calibration (e.g. extrapolation to large $b_{b,\mathrm{wc}}(\lambda)$) or air bubbles might also be present but cannot be assessed. As no direct measurements of $b_{b,\mathrm{wc}}(\lambda)$ and backscattering

SIOPs have been published for Lake Geneva the exact nature of the observed Bias remains unknown.

As with the magnitude of backscattering, the spectral shapes do not necessarily match the *in situ* measurements (Figure 20). The observed spectra usually do not follow a spectrally flat or exponential function, as assumed by $b^*_{b,\mathrm{X}}(\lambda)$ and $b^*_{b,\mathrm{phy}}(\lambda)$ respectively.

Therefore, these commonly used SIOP approximations for backscattering may not accurately capture wavelength dependency. However, the effect of these mismatches should not be overstated, as the normalization of rather spectrally flat spectra strongly highlights spectral features.

Studies examining real world backscattering (see, for example, Figure 26) have also observed spectral shapes that do not strictly adhere to common exponential or spectrally flat approximations (Gallie and Murtha, 1992; Gordon et al., 2009; Vaillancourt, 2004; Whitmire et al., 2010). Nevertheless, especially the exponential approximation is widely used and generally aligns with *in situ* measurements to a satisfactory degree (see e.g. Reynolds et al., 2016).

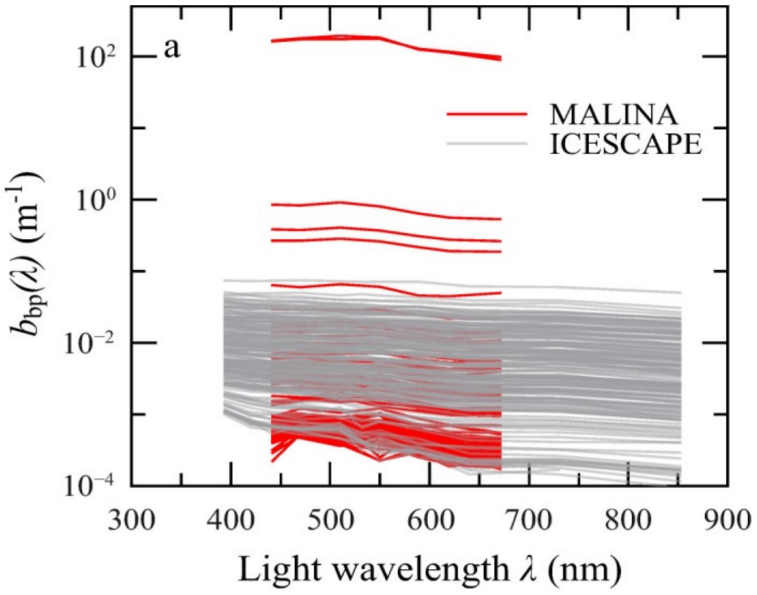

Figure 26: Measured particle backscattering coefficient spectra ($b_{bp}(\lambda)$, generally referred to as $b_{b,wc}(\lambda)$ in this thesis) from the MALINA and ICESCAPE cruises (Reynolds et al., 2016, Figure 8a, adapted, CC BY-NC).

Based on the comparison between phytoplankton-dominated modeled and *in situ* measurements, we suspect that the modeled phytoplankton backscattering is assumed too steep for this study. Note that this may not be necessarily towards the blue wavelengths (as Figure 20, B might suggest) as such an interpretation would implicitly assume correct modelling at the normalization reference wavelength.

The overall variability of *in situ* backscattering shapes indicates that backscattering processes in Lake Geneva are more complex than modeled. The spectral shape and magnitude of (concentration-specific) backscattering vary depending on a variety of factors, such as cell size, chemical composition, and taxonomy (Brewin et al., 2012; Gege, 2004; Reynolds et al., 2016; Vaillancourt, 2004), making it challenging to model.

4.5 Phytoplankton taxa retrieval: Limitations

Retrieval of individual phytoplankton taxa was largely unsuccessful; significant correlation could only be achieved with diatoms. While it has been demonstrated that such retrievals are feasible for Case 2 waters using *in situ* hyperspectral data (Maire et al., 2025), applying this to spaceborne multispectral data remains challenging. A variety of limitations exist on the sensor side, such as low spectral and spatial resolutions, AC errors, and instrument noise (i.e. low signal-to-noise ratio) (Giardino et al., 2019).

In terms of the model, including different phytoplankton absorption SIOP spectra can result in the inversion becoming highly ill-posed, even for hyperspectral measurements (Cael et al., 2020). Conversely, introducing errors can degrade hyperspectral spaceborne retrievals (Maire et al., in preparation). For multispectral sensors, the smaller number of bands naturally increases the ill-posedness of the inverse problem.

Another significant modelling uncertainty is that the SIOPs used in this study were mostly obtained from Lake Constance. There is no guarantee that they are representative of Lake Geneva, so this study is effectively working with an uncalibrated model. In the case of phytoplankton, one taxon with significant abundance and biovolume (chrysophyceae) was not parametrized.

The SIOPs of phytoplankton species can vary significantly even within a single taxon (Maire et al., 2025). Nevertheless, the absorption of bulk phytoplankton can generally be described by just a few spectral shapes (Cael et al., 2020). This may explain why individual phytoplankton taxa could not be identified, whereas the retrieved bulk concentration mostly aligned with *in situ* measurements across various methods and locations.

The exact concentrations of individual taxa are unknown, but their biovolumes suggest that they are generally low (Figure 12). Retrieval failure for Lake Geneva conditions is consistent with Maire et al. (in preparation), who report a detection limit of 5 mg m^{-3} for retrieving individual phytoplankton taxa (mono-bloom case). Diatoms are the only taxon for which retrievals correlate with *in situ* data, and they also show the highest biovolumes, which further supports this explanation.

For this study, we suspect that the main reasons for the poor performance of taxa retrievals are the combination of significant errors in the input $R_{rs}(\lambda)$ (21.4 – 99% depending on the band), low spectral resolution (five used bands) and relatively low concentrations. Section 4.8.3 and Section 4.8.4 further discuss how spaceborne retrievals of phytoplankton taxa could be improved.

4.6 AC: A limiting factor

Water constituent retrievals should also include an assessment of AC accuracy (Ogashawara et al., 2021). AC accuracy is a limiting factor, as these errors strongly propagate downstream. This notion is not new: According to early estimates by Gordon and Clark (1981), errors of less than 5% in the blue wavelengths are needed to achieve ocean products with errors of less than 30% in CHL. While this early benchmark does not directly apply to this study and other more

sophisticated retrieval algorithms, it clearly shows that even 45 years later AC errors (here 21.4 – 99% in the relevant bands) are still far beyond what would be required for highly accurate retrievals.

For the retrieval of CHL and phytoplankton, information about the second absorption peak around 665 nm is crucial (Cael et al., 2020). However, evaluation of the 665 nm band revealed limited AC accuracy (MdSA = 99%). Previous studies (Soppa et al., 2021; Werther et al., 2022) have also observed the underperformance of POLYMER at longer wavelengths. As past studies using the Thetis dataset have shown POLYMER to be the best performing AC algorithm for Lake Geneva and that errors in the 665 nm band are high for all algorithms (Gupana et al., 2025; Werther et al., 2022), switching to a different AC algorithm is unlikely to improve the results. MiniWASI was somewhat resistant to AC errors at 665 nm. Fitted $R_{rs}(665)$ showed significantly better results, with an MdSA improvement from 99 to 37.7%.

The impact of the AE remains a big unknown. The validation of POLYMER $R_{rs}(\lambda)$ implicitly includes errors due to the AE, which are technically not part of AC itself. AE is strongest close to the shore and in the near-infrared bands (Castagna and Vanhellemont, 2025). Although POLYMER can by design implicitly correct for (some of) the adjacency signal, it does so less effectively than explicit physical modelling (Castagna and Vanhellemont, 2025).

While adjacency effects generally add to the total $R_{rs}(\lambda)$, AE in the near-infrared may propagate to different errors in the visible wavelengths for POLYMER (Wu et al., 2024). Wu et al. (2024) found that AE correction significantly improved near-shore performance of POLYMER across the whole spectrum. Visual interpretation suggests that AE has a significant impact on the visual spectrum for about 3 pixels from shore (~1 km). However, visual interpretation alone cannot confirm a strong impact on the pixel to the south of the LéXPLORE platform, which served as reference.

Errors in the blue and red bands propagated directly into the CDOM (Section 4.3) and absorption IOP (Section 4.4) results discussed above.

4.7 Matchup representativeness

Sections 4.1 – 4.6 repeatedly attributed validation discrepancies to matchup representativeness, e.g. the non-visible CHL_S maximum at SHL2 (Section 4.1.1). This section synthesizes these recurring issues into four general categories:

- Vertical mismatch: For phytoplankton, samples include water along the first 18 m of the column, whereas the visibility at SHL2 is, on average, only 7 m. Consequently, a significant portion of the depth column contributing to the data is not visible to the satellite. Examples of the CHL maximum not being visible to the satellite, yet being present in the *in situ* reference, are shown and discussed in Sections 3.2.1 and 4.1.1, respectively. A comparable effect was observed at LéXPLORE, where high NAP concentrations were suspected to obscure a non-surface CHL maximum from the satellite (Section 4.1.3).
- Non-uniformities: Bio-optical models such as WASI assume that constituents are distributed uniformly throughout the vertical column. This is rarely the case, especially

for phytoplankton. Consequently, the impact of near-surface constituent concentrations is most significant for the $R_{rs}(\lambda)$ spectrum. Furthermore, such non-uniformities result in wavelength-dependent effects (different penetration depths) which are difficult to account for. Radiative transfer simulations have demonstrated that this is a significant source of error (Nouchi, 2018). Appendix 4 presents an example of a strongly non-uniform aLH_{676} profile, which has been discussed in Section 4.1.3.

- Spatial mismatch: The pixel area may not be representative of the sampling location (e.g. small scale patterns at low spatial resolution). In the case of Thetis matchups, using the pixel south of the platform further adds a spatial offset of approximately 250 m (see Section 4.1.3). Similarly, small-scale variations unresolved by the S3 OLCI's 300 m spatial resolution were suggested as one possible explanation for a field-campaign matchup outlier (Section 4.2).
- Temporal mismatch: Significant changes in constituent concentrations may have occurred between image acquisition and the time of sampling. Temporal mismatch at LéXPLORE is usually less than an hour. It is unknown at SHL2 (as there are no timestamps) and is four to five hours for the field campaigns, which was suggested as an alternative explanation for the same outlier (Section 4.2).

4.8 Future study directions and general limitations

4.8.1 Uncertainty

Although crucial for end users (Giardino et al., 2019; IOCCG, 2019) the assessment of uncertainties is rarely performed for constituent retrievals. While this need has been recognized (IOCCG, 2019) and some algorithms provide an uncertainty estimate (see e.g. Werther et al., 2022), bio-optical models such as WASI and BOMBER do not directly report uncertainty. For WASI, uncertainty is typically assessed by comparing it to WASI-AI outputs (Niroumand-Jadidi & Gege, 2025, see also Appendix 1). lmfit tries to estimate the standard error (see Section 2.8.2), but the extreme ill-posedness resulting from inverting seven unknown variables makes this estimate impossible (showing as “None”).

Retrieval errors are relatively complex due to a chain of major uncertainties (e.g. sensor noise, AC errors, AE, and different model uncertainties). For bio-optical models, separating these errors using an uncertainty budget would be optimal, but this is a challenging task given the complexity and diversity of OACs (IOCCG, 2019). It should also be noted that *in situ* data are not error-free, as evidenced by the disagreement between different CHL proxies. Hence, validation studies such as this one should not be interpreted as providing a strict quantification of uncertainty in end products (IOCCG, 2019). Future studies should explore how uncertainties can be quantified most effectively for bio-optical modelling.

4.8.2 AC and AE advancements

Although the exact effect of AC errors on the uncertainty budget is difficult to pinpoint (IOCCG, 2019), it is clear that the errors are substantial and that more accurate AC is required to improve constituent retrievals in Case 2 waters (Giardino et al., 2019). This study concludes

that higher accuracy in the 665 nm as well as the blue bands could help improve OLCI bio-optical CHL and CDOM retrievals.

Near-shore $R_{rs}(\lambda)$ validation inherently includes errors due to AE. In order to potentially improve constituent retrievals, especially in the near-shore region, future research should explore newly developed AE correction algorithms, such as RAdCor (Castagna and Vanhellemont, 2025). The impact of the prior application of such algorithms on the performance of AC algorithms is the subject of ongoing research (Pühringer et al., in preparation).

4.8.3 Bio-optical modelling: Limitations and possible improvements

Any study-specific issues encountered with IOPs and SIOPs have already been discussed in detail, including observed mismatches for absorption and backscattering, the absence of the chrysophyceae taxon, and the transfer of SIOPs from Lake Constance. This section will focus on some general limitations of bio-optical modelling, which are difficult to quantify yet still relevant to the bigger picture and future advancements.

A well-known issue with absorption-based estimates of phytoplankton concentrations, first raised by Bricaud et al. (1995) in the 1990s, is the so-called "packaging effect": The shape and magnitude of phytoplankton absorption SIOPs depend not only on accessory pigments, but also on the packaging effect, whereby bundling pigments inside cells reduces their light-absorbing efficiency (Bricaud et al., 1995). Using a variety of specific absorption spectra may partly account for the packaging effect (i.e. taxa-dependent cell sizes), but its impact on the results of this study remains unclear.

It should also be noted that, as with phytoplankton, NAP and CDOM are broad categories with variable optical properties, even within a single location (Gege, 2024). SIOP measurements could help to identify the optical properties of different OACs and their variability in Lake Geneva. Such measurements could also help to determine whether the identified IOP mismatches (see Section 4.4) are due to the SIOPs used or to other factors, such as the limited quality of the Thetis IOP data.

A potential approach for future research would be to model the absorption of different phytoplankton pigments rather than phytoplankton taxa. According to Cael et al. (2020) this would provide a less ill-posed configuration while accounting for phytoplankton optical variability. Furthermore, regularisation of the inverse problem (e.g. through the use of priors) could enhance the accuracy of the results, particularly in the case of strongly ill-posed conditions (Cael et al., 2020).

4.8.4 Hyperspectral missions: PACE

This study supports the idea that accurately inverting phytoplankton taxa from spaceborne multispectral measurements is extremely challenging (or even mathematically impossible) in the presence of significant errors in the input data (Cael et al., 2020; Maire et al. 2026, in preparation). Hyperspectral sensors offer new possibilities for identifying constituents, particularly for detecting different phytoplankton taxa (Giardino et al., 2019).

While past studies have relied on on-demand missions such as PRISMA (PRecursore IperSpettrale della Missione Applicativa) (see Table 1), these may not be suitable for operational monitoring. However, NASA's newly launched Plankton, Aerosol, Cloud, ocean Ecosystem (PACE) mission's Ocean Color Instrument provides 240 continuous bands between 340 and 890 nm at 2.5 nm spectral resolution, with a revisit time of approximately two days (Xu et al., 2026). PACE's spectral range (including parts of the ultraviolet spectrum) may also enable it to distinguish between NAP, CDOM, and phytoplankton signals more effectively (Cael et al., 2020). The use of the PACE mission to assess different phytoplankton taxa in Case 2 waters is the subject of ongoing research (Maire et al., in preparation).

4.9 Conclusion

This thesis developed a bio-optical model MiniWASI tailored to operational retrievals of OACs and IOPs in optically deep Case 2 waters. By validating retrievals from 365 images using a default set of SIOPs and automated processing, the thesis took a significant step towards a more operational approach in WASI inversions.

A WASI-based bio-optical inversion approach using a default set of SIOPs and S3 OLCI $R_{rs}(\lambda)$ performed well in retrieving CHL (r = 0.72 – 0.8; for datasets used in previous studies) and TSM (r = 0.77 – 0.78) in Lake Geneva but was unable to retrieve CDOM or differentiate between different phytoplankton taxa. Validation showed similar accuracy to that of empirical, locally calibrated CHL retrieval algorithms.

Retrieval errors for phytoplankton taxa and CDOM were attributed to the strong ill-posedness of the inverse problem, driven by low concentrations, limited spectral resolution, and a large number (seven) of unknown variables, as well as to errors in the input remote sensing reflectance. These results support the notion that retrieving individual phytoplankton taxa with multispectral sensors such as Sentinel-3 OLCI is challenging, if not mathematically impossible.

Conversely, NAP showed robust agreement with a limited number of field samples and estimates from hyperspectral *in situ* data. Consistency in NAP retrievals is further supported by the high correlation between modeled and measured $b_{b,\mathrm{wc}}(\lambda)$.

Future research directions include quantifying retrieval uncertainties, applying and improving AC and AE algorithms, and performing explicit model calibration for Lake Geneva (e.g. through SIOP *in situ* measurements). There is also scope for application to hyperspectral missions such as PACE.

Bibliography

A. Aminot; F. Rey, 2001. Chlorophyll a: Determination by spectroscopic methods. ICES. https://doi.org/10.17895/ICES.PUB.5080

Albert, A., Mobley, C., 2003. An analytical model for subsurface irradiance and remote sensing reflectance in deep and shallow case-2 waters. Opt. Express 11, 2873. https://doi.org/10.1364/OE.11.002873

Ansari, M., Knudby, A., Amani, M., Sawada, M., 2025. Retrieving Inland Water Quality Parameters via Satellite Remote Sensing: Sensor Evaluation, Atmospheric Correction, and Machine Learning Approaches. Remote Sens. 17, 1734. https://doi.org/10.3390/rs17101734

Babin, M., Morel, A., Fournier-Sicre, V., Fell, F., Stramski, D., 2003a. Light scattering properties of marine particles in coastal and open ocean waters asrelated to the particle mass concentration. Limnol. Oceanogr. 48, 843–859. https://doi.org/10.4319/lo.2003.48.2.0843

Babin, M., Stramski, D., Ferrari, G.M., Claustre, H., Bricaud, A., Obolensky, G., Hoepffner, N., 2003b. Variations in the light absorption coefficients of phytoplankton, nonalgal particles, and dissolved organic matter in coastal waters around Europe. J. Geophys. Res. Oceans 108, 2001JC000882. https://doi.org/10.1029/2001JC000882

Bi, S., Hieronymi, M., Röttgers, R., 2023. Bio-geo-optical modelling of natural waters. Front. Mar. Sci. 10, 1196352. https://doi.org/10.3389/fmars.2023.1196352

Bonnier, M., Anneville, O., Iestyn Woolway, R., Thackeray, S.J., Morin, G.P., Reynaud, N., Soulignac, F., Tormos, T., Harmel, T., 2024. Assessing ESA Climate Change Initiative data for the monitoring of phytoplankton abundance and phenology in deep lakes: Investigation on Lake Geneva. J. Gt. Lakes Res. 50, 102372. https://doi.org/10.1016/j.jglr.2024.102372

Boss, E., Pegau, W.S., 2001. Relationship of light scattering at an angle in the backward direction to the backscattering coefficient. Appl. Opt. 40, 5503. https://doi.org/10.1364/AO.40.005503

Boss, E., Picheral, M., Leeuw, T., Chase, A., Karsenti, E., Gorsky, G., Taylor, L., Slade, W., Ras, J., Claustre, H., 2013. The characteristics of particulate absorption, scattering and attenuation coefficients in the surface ocean; Contribution of the Tara Oceans expedition. Methods Oceanogr. 7, 52–62. https://doi.org/10.1016/j.mio.2013.11.002

Brandão, L.P.M., Brighenti, L.S., Staehr, P.A., Asmala, E., Massicotte, P., Tonetta, D., Barbosa, F.A.R., Pujoni, D., Bezerra-Neto, J.F., 2018. Distinctive effects of allochthonous and autochthonous organic matter on CDOM spectra in a tropical lake. Biogeosciences 15, 2931–2943. https://doi.org/10.5194/bg-15-2931-2018

Brewin, R.J.W., Dall'Olmo, G., Sathyendranath, S., Hardman-Mountford, N.J., 2012. Particle backscattering as a function of chlorophyll and phytoplankton size structure in the open-ocean. Opt. Express 20, 17632. https://doi.org/10.1364/OE.20.017632

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

Bricaud, A., Morel, A., Prieur, L., 1981. Absorption by dissolved organic matter of the sea (yellow substance) in the UV and visible domains1. Limnol. Oceanogr. 26, 43–53. https://doi.org/10.4319/lo.1981.26.1.0043

Cael, B.B., Chase, A., Boss, E., 2020. Information content of absorption spectra and implications for ocean color inversion. Appl. Opt. 59, 3971. https://doi.org/10.1364/AO.389189

Carder, K.L., Steward, R.G., Harvey, G.R., Ortner, P.B., 1989. Marine humic and fulvic acids: Their effects on remote sensing of ocean chlorophyll. Limnol. Oceanogr. 34, 68–81. https://doi.org/10.4319/lo.1989.34.1.0068

Castagna, A., Vanhellemont, Q., 2025. A generalized physics-based correction for adjacency effects. Appl. Opt. 64, 2719. https://doi.org/10.1364/AO.546766

Detoni, A.M.S., Oliveira, N.R., Zoffoli, M.L., Gege, P., Caballero, I., Navarro, G., Fontela, M., Álvarez Fernandéz, M.J., Serrano, M.L., Moreno, L., Teixeira, I.G., Velo, A., Padín, X.A., 2026. Unveiling optical and phytoplankton variability in the northwestern Iberian Peninsula using in-situ radiometry and the Water Colour Simulator – WASI. Estuar. Coast. Shelf Sci. 340, 109976. https://doi.org/10.1016/j.ecss.2026.109976

Dörnhöfer, K., Göritz, A., Gege, P., Pflug, B., Oppelt, N., 2016. Water Constituents and Water Depth Retrieval from Sentinel-2A—A First Evaluation in an Oligotrophic Lake. Remote Sens. 8, 941. https://doi.org/10.3390/rs8110941

EUMETSAT, 2018. Sentinel-3 OLCI Marine User Handbook (No. EUM/OPS-SEN3/MAN/17/907205). EUMETSAT, Darmstadt.

Fabbretto, A., Pellegrino, A., Giardino, C., Bresciani, M., Alikas, K., Braga, F., Vaičiūtė, D., Lima, T.M.A.D., Mangano, S., Ghirardi, N., Daraio, M.G., Brando, V.E., 2023. Hyperspectral Prisma Data Processing For Water Quality Research And Applications, in: IGARSS 2023 - 2023 IEEE International Geoscience and Remote Sensing Symposium. Presented at the IGARSS 2023 - 2023 IEEE International Geoscience and Remote Sensing Symposium, IEEE, Pasadena, CA, USA, pp. 1744–1747. https://doi.org/10.1109/IGARSS52108.2023.10283366

Ferreira, A., Ciotti, Á.M., Mendes, C.R.B., Uitz, J., Bricaud, A., 2017. Phytoplankton light absorption and the package effect in relation to photosynthetic and photoprotective pigments in the northern tip of A ntarctic P eninsula. J. Geophys. Res. Oceans 122, 7344–7363. https://doi.org/10.1002/2017JC012964

Gallie, E.A., Murtha, P.A., 1992. Specific absorption and backscattering spectra for suspended minerals and chlorophyll-a in chilko lake, British Columbia. Remote Sens. Environ. 39, 103–118. https://doi.org/10.1016/0034-4257(92)90130-C

Gege, P., 2024. The Water Colour Simulator WASI — User manual for WASI version 7 (No. WASI version 7). German Aerospace Center (DLR), Institut für Methodik der Fernerkundung.

Gege, P., 2014. WASI-2D: A software tool for regionally optimized analysis of imaging spectrometer data from deep and shallow waters. Comput. Geosci. 62, 208–215. https://doi.org/10.1016/j.cageo.2013.07.022

Gege, P., 2012. Estimation of phytoplankton concentration from downwelling irradiance measurements in water. Isr. J. Plant Sci. 60, 193–207. https://doi.org/10.1560/IJPS.60.1-2.193

Gege, P., 2004. The water color simulator WASI: an integrating software tool for analysis and simulation of optical in situ spectra. Comput. Geosci. 30, 523–532. https://doi.org/10.1016/j.cageo.2004.03.005

Giardino, C., Brando, V.E., Gege, P., Pinnel, N., Hochberg, E., Knaeps, E., Reusen, I., Doerffer, R., Bresciani, M., Braga, F., Foerster, S., Champollion, N., Dekker, A., 2019. Imaging Spectrometry of Inland and Coastal Waters: State of the Art, Achievements and Perspectives. Surv. Geophys. 40, 401–429. https://doi.org/10.1007/s10712-018-9476-0

Giardino, C., Bresciani, M., Cazzaniga, I., Schenk, K., Rieger, P., Braga, F., Matta, E., Brando, V., 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., 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

Gordon, H.R., Clark, D.K., 1981. Clear water radiances for atmospheric correction of coastal zone color scanner imagery. Appl. Opt. 20, 4175. https://doi.org/10.1364/AO.20.004175

Gordon, H.R., Lewis, M.R., McLean, S.D., Twardowski, M.S., Freeman, S.A., Voss, K.J., Boynton, G.C., 2009. Spectra of particulate backscattering in natural waters. Opt. Express 17, 16192. https://doi.org/10.1364/OE.17.016192

Griffin, C.G., Finlay, J.C., Brezonik, P.L., Olmanson, L., Hozalski, R.M., 2018. Limitations on using CDOM as a proxy for DOC in temperate lakes. Water Res. 144, 719–727. https://doi.org/10.1016/j.watres.2018.08.007

Gupana, R.S., Damm, A., Rahaghi, A.I., Minaudo, C., Odermatt, D., 2022. Non-photochemical quenching estimates from in situ spectroradiometer measurements: implications on remote sensing of sun-induced chlorophyll fluorescence in lakes. Opt. Express 30, 46762. https://doi.org/10.1364/OE.469402

Gupana, R.S., Odermatt, D., Rahaghi, A.I., Minaudo, C., Werther, M., Giardino, C., Damm, A., 2025. Remote Sensing of Sun-Induced Fluorescence in a Deep Lake: Disentangling Quenching Mechanisms Improves Relationship With Chlorophyll-a Concentration Estimates. IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens. 18, 4410–4426. https://doi.org/10.1109/JSTARS.2025.3528911

Heege, T., 2000. Flugzeuggestützte Fernerkundung von Wasserinhaltsstoffen im Bodensee. DLR, Köln.

Hernando, A.M., Piano, A.M., Blanco, A.C., Medina, J.M., Manuel, A.Y., 2024. Comparative Analysis of Prisma Hyperspectral and SENTINEL-2 Multispectral Images for Chlorophyll-A and Turbidity Mapping of Taal Lake. Int. Arch. Photogramm. Remote Sens. Spat. Inf. Sci. XLVIII-4/W8-2023, 293–299. https://doi.org/10.5194/isprs-archives-XLVIII-4-W8-2023-293-2024

Hou, X., Feng, L., Dai, Y., Hu, C., Gibson, L., Tang, J., Lee, Z., Wang, Y., Cai, X., Liu, J., Zheng, Y., Zheng, C., 2022. Global mapping reveals increase in lacustrine algal blooms over the past decade. Nat. Geosci. 15, 130–134. https://doi.org/10.1038/s41561-021-00887-x

IOCCG, 2019. Uncertainties in Ocean Colour Remote Sensing. https://doi.org/10.25607/OBP-696

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

Irani Rahaghi, A., Odermatt, D., Anneville, O., Sepúlveda Steiner, O., Reiss, R.S., Amadori, M., Toffolon, M., Jacquet, S., Harmel, T., Werther, M., Soulignac, F., Dambrine, E., Jézéquel, D., Hatté, C., Tran-Khac, V., Rasconi, S., Rimet, F., Bouffard, D., 2024. Combined Earth observations reveal the sequence of conditions leading to a large algal bloom in Lake Geneva. Commun. Earth Environ. 5, 229. https://doi.org/10.1038/s43247-024-01351-5

Jacquet, S., Soulignac, F., Anneville, O., 2026. A complete overview of algal blooms in Lake Geneva: shall the past shed light on the future? Aquat. Sci. 88, 3. https://doi.org/10.1007/s00027-025-01234-7

Jerlov, N.G., 1976. Marine optics, Elsevier oceanography series. Elsevier Scientific Pub. Co, Amsterdam New York.

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. Sci. Total Environ. 527–528, 493–506. https://doi.org/10.1016/j.scitotenv.2015.05.011
Kou, L., Labrie, D., Chylek, P., 1993. Refractive indices of water and ice in the 065- to 25-μm spectral range. Appl. Opt. 32, 3531. https://doi.org/10.1364/AO.32.003531
Levenberg, K., 1944. A method for the solution of certain non-linear problems in least squares. Q. Appl. Math. 2, 164–168. https://doi.org/10.1090/qam/10666
Li, C., Odermatt, D., Bouffard, D., Wüest, A., Kohn, T., 2022. Coupling remote sensing and particle tracking to estimate trajectories in large water bodies. Int. J. Appl. Earth Obs. Geoinformation 110, 102809. https://doi.org/10.1016/j.jag.2022.102809
Lu, Z., 2007. Optical absorption of pure water in the blue and ultraviolet. Texas A&M University.
Maire, L., Gege, P., Damm, A., Odermatt, D., 2025. Differentiating phytoplankton taxa in lakes using hyperspectral in situ reflectance and imaging microscopy. Sci. Total Environ. 1003, 180718. https://doi.org/10.1016/j.scitotenv.2025.180718
Many, G., Escoffier, N., Ferrari, M., Jacquet, P., Odermatt, D., Mariethoz, G., Perolo, P., Perga, M.-E., 2022. Long-Term Spatiotemporal Variability of Whitings in Lake Geneva from Multispectral Remote Sensing and Machine Learning. Remote Sens. 14, 6175. https://doi.org/10.3390/rs14236175
Many, G., Escoffier, N., Perolo, P., Bärenbold, F., Bouffard, D., Perga, M.-E., 2024. Calcite precipitation: The forgotten piece of lakes' carbon cycle. Sci. Adv. 10, eado5924. https://doi.org/10.1126/sciadv.ado5924
Marquardt, D.W., 1963. An Algorithm for Least-Squares Estimation of Nonlinear Parameters. J. Soc. Ind. Appl. Math. 11, 431–441. https://doi.org/10.1137/0111030
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
Mobley, C., 1994. Light and Water: Radiative Transfer in Natural Waters, Academic Press.
Moisset, S., 2017. Investigation of the link between phytoplankton and nutrients dynamic in Lake Geneva. Université de Genève. https://doi.org/10.13097/ARCHIVE-OUVERTE/UNIGE:96830
Morel, A., 1974. Optical Properties of Pure Water and Pure Sea Water. Opt. Asp. Oceanogr. 1–24.
Nechad, B., Ruddick, K.G., Park, Y., 2010. Calibration and validation of a generic multisensor algorithm for mapping of total suspended matter in turbid waters. Remote Sens. Environ. 114, 854–866. https://doi.org/10.1016/j.rse.2009.11.022
Newville, M., Otten, R., Nelson, A., Stensitzki, T., Ingargiola, A., Allan, D., Fox, A., Carter, F., Rawlik, M., 2025. LMFIT: Non-Linear Least-Squares Minimization and Curve-Fitting for Python. https://doi.org/10.5281/ZENODO.598352
Niroumand-Jadidi, M., Bovolo, F., Bresciani, M., Gege, P., Giardino, C., 2022. Water Quality Retrieval from Landsat-9 (OLI-2) Imagery and Comparison to Sentinel-2. Remote Sens. 14, 4596. https://doi.org/10.3390/rs14184596
Niroumand-Jadidi, M., Bovolo, F., Bruzzone, L., 2020. Water Quality Retrieval from PRISMA Hyperspectral Images: First Experience in a Turbid Lake and Comparison with Sentinel-2. Remote Sens. 12, 3984. https://doi.org/10.3390/rs12233984
Niroumand-Jadidi, M., Bovolo, F., Bruzzone, L., Gege, P., 2021. Inter-Comparison of Methods for Chlorophyll-a Retrieval: Sentinel-2 Time-Series Analysis in Italian Lakes. Remote Sens. 13, 2381. https://doi.org/10.3390/rs13122381
Niroumand-Jadidi, M., Gege, P., 2025. WASI-AI: Synergistic Integration of AI and Physics for Retrieving Water Quality and Benthic Parameters from Multi- and Hyperspectral Images. IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens. 18, 22832–22846. https://doi.org/10.1109/JSTARS.2025.3605061

Nouchi, V., 2018. Toward a systematic integration of optical remote sensing for inland waters studies. EPFL.
Nouchi, V., Kutser, T., Wüest, A., Müller, B., Odermatt, D., Baracchini, T., Bouffard, D., 2019. Resolving biogeochemical processes in lakes using remote sensing. Aquat. Sci. 81, 27. https://doi.org/10.1007/s00027-019-0626-3
Odermatt, D., Gege, P., 2022. Lake Colors: Interpreting Apparent Optical Properties, in: Encyclopedia of Inland Waters. Elsevier, pp. 474–489. https://doi.org/10.1016/B978-0-12-819166-8.00041-4
Odermatt, D., Gitelson, A., Brando, V.E., Schaepman, M., 2012. Review of constituent retrieval in optically deep and complex waters from satellite imagery. Remote Sens. Environ. 118, 116–126. https://doi.org/10.1016/j.rse.2011.11.013
Odermatt, D., Runnals, J., Damm, A., 2020. SenCast: Copernicus Satellite Data on Demand.
Ogashawara, I., Kiel, C., Jechow, A., Kohnert, K., Ruhtz, T., Grossart, H.-P., Hölker, F., Nejstgaard, J.C., Berger, S.A., Wollrab, S., 2021. The Use of Sentinel-2 for Chlorophyll-a Spatial Dynamics Assessment: A Comparative Study on Different Lakes in Northern Germany. Remote Sens. 13, 1542. https://doi.org/10.3390/rs13081542
O'Reilly, J.E., 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. J. Geophys. Res. Oceans 103, 24937–24953. https://doi.org/10.1029/98JC02160
Pinardi, M., Bresciani, M., Villa, P., Cazzaniga, I., Laini, A., Tóth, V., Fadel, A., Austoni, M., Lami, A., Giardino, C., 2018. Spatial and temporal dynamics of primary producers in shallow lakes as seen from space: Intra-annual observations from Sentinel-2A. Limnologica 72, 32–43. https://doi.org/10.1016/j.limno.2018.08.002
Pope, R.M., Fry, E.S., 1997. Absorption spectrum (380–700 nm) of pure water II Integrating cavity measurements. Appl. Opt. 36, 8710. https://doi.org/10.1364/AO.36.008710
Reynolds, R.A., Stramski, D., Neukermans, G., 2016. Optical backscattering by particles in Arctic seawater and relationships to particle mass concentration, size distribution, and bulk composition. Limnol. Oceanogr. 61, 1869–1890. https://doi.org/10.1002/lno.10341
Rimet, F., Anneville, O., Barbet, D., Chardon, C., Crépin, L., Domaizon, I., Dorioz, J.-M., Espinat, L., Frossard, V., Guillard, J., Goulon, C., Hamelet, V., Hustache, J.-C., Jacquet, S., Lainé, L., Montuelle, B., Perney, P., Quetin, P., Rasconi, S., Schellenberger, A., Tran-Khac, V., Monet, G., 2020. The Observatory on LAkes (OLA) database: Sixty years of environmental data accessible to the public: The Observatory on LAkes (OLA) database. J. Limnol. 79. https://doi.org/10.4081/jlimnol.2020.1944
Roesler, C., 2021. Non-Algal Particles (NAP). Ocean Opt. Web Book.
Roth, 1990. Eaux du Léman malodorantes. La Suisse.
Röttgers, R., Doerffer, R., McKee, D., Schönfeld, W., 2013. Algorithm Theoretical Basis Document: The Water Optical Properties Processor (WOPP): Pure Water Spectral Absorption, Scattering, and Real Part of Refractive Index Model (No. Revision 7). ESA/ESRIN, Frascati, Italy.
Savitzky, Abraham., Golay, M.J.E., 1964. Smoothing and Differentiation of Data by Simplified Least Squares Procedures. Anal. Chem. 36, 1627–1639. https://doi.org/10.1021/ac60214a047
Soppa, M.A., Silva, B., Steinmetz, F., Keith, D., Scheffler, D., Bohn, N., Bracher, A., 2021. Assessment of Polymer Atmospheric Correction Algorithm for Hyperspectral Remote Sensing Imagery over Coastal Waters. Sensors 21, 4125. https://doi.org/10.3390/s21124125

Steinmetz, F., Deschamps, P.-Y., Ramon, D., 2011. Atmospheric correction in presence of sun glint: application to MERIS. Opt. Express 19, 9783. https://doi.org/10.1364/OE.19.009783
Vaillancourt, R.D., 2004. Light backscattering properties of marine phytoplankton: relationships to cell size, chemical composition and taxonomy. J. Plankton Res. 26, 191–212. https://doi.org/10.1093/plankt/fbh012
Vanhellemont, Q., Ruddick, K., 2018. Atmospheric correction of metre-scale optical satellite data for inland and coastal water applications. Remote Sens. Environ. 216, 586–597. https://doi.org/10.1016/j.rse.2018.07.015
Wang, L., 2008. Measuring optical absorption coefficient of pure water in UV using the inte grating cavity absorption meter. Texas A&M University.
Warren, S.G., 1984. Optical constants of ice from the ultraviolet to the microwave. Appl. Opt. 23, 1206. https://doi.org/10.1364/AO.23.001206
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 Sens. Environ. 283, 113295. https://doi.org/10.1016/j.rse.2022.113295
Wevers, J., Müller, D., Kirches, G., Quast, R., Brockmann, C., 2022. IdePix for Sentinel-3 OLCI Algorithm Theoretical Basis Document. https://doi.org/10.5281/ZENODO.6517333
Whitmire, A.L., Pegau, W.S., Karp-Boss, L., Boss, E., Cowles, T.J., 2010. Spectral backscattering properties of marine phytoplankton cultures. Opt. Express 18, 15073. https://doi.org/10.1364/OE.18.015073
Wilks, D.S., 2019. Statistical methods in the atmospheric sciences, Fourth edition. ed. Elsevier, Amsterdam, Netherlands ; Cambridge, MA.
Wu, D., Liu, W., Makowski, D., Tang, T., Greenwood, E.E., Huang, Y., Ciais, P., Zhang, H., Du, T., Xia, X., Odermatt, D., 2026. COVID-19 containment and control reduced lake turbidity around the world. Commun. Earth Environ. 7, 201. https://doi.org/10.1038/s43247-026-03311-7
Wu, Y., Knudby, A., Pahlevan, N., Lapen, D., Zeng, C., 2024. Sensor-generic adjacency-effect correction for remote sensing of coastal and inland waters. Remote Sens. Environ. 315, 114433. https://doi.org/10.1016/j.rse.2024.114433
Wüest, A., Bouffard, D., Guillard, J., Ibelings, B.W., Lavanchy, S., Perga, M., Pasche, N., 2021. LéXPLORE : A floating laboratory on Lake Geneva offering unique lake research opportunities. WIREs Water 8, e1544. https://doi.org/10.1002/wat2.1544
Xi, H., Hieronymi, M., Krasemann, H., Röttgers, R., 2017. Phytoplankton Group Identification Using Simulated and In situ Hyperspectral Remote Sensing Reflectance. Front. Mar. Sci. 4, 272. https://doi.org/10.3389/fmars.2017.00272
Xu, J., Xue, K., Fang, C., Shen, M., Cao, Z., Duan, H., 2026. Capabilities of PACE Ocean color Instrument (OCI) for phytoplankton pigments of inland lakes. Int. J. Appl. Earth Obs. Geoinformation 149, 105265. https://doi.org/10.1016/j.jag.2026.105265

Appendices

Appendix 1: MiniWASI and WASI7 inversion performance

When using WASI, it was found that when varying three basic constituent classes (C_0, C_X, and C_y), the phytoplankton proxy (C_0) tended to remain at its initial value under ill-posed conditions. Figure A1. 1 shows an example in which the WASI version 7 (WASI7) and WASI-AI retrievals remained at an initial value of 3 mg m^{-3}; meanwhile, MiniWASI achieved lower residuals and was less dependent on the initial value.

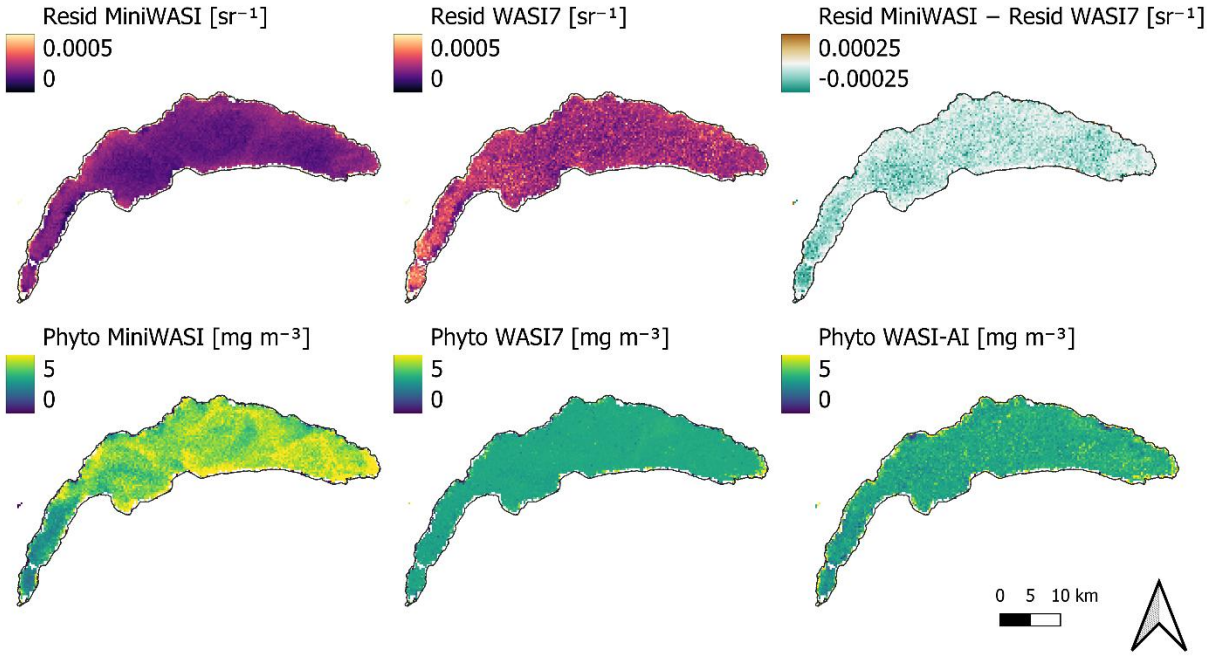

Figure A1. 1: Example of inversion performance of MiniWASI and WASI7 under equal conditions. WASI7 shows higher residual values than MiniWASI. WASI7 and WASI-AI phytoplankton retrievals are "stuck" at the initial value of 3 mg m^{-3}. Variables and initial values used for this test: C_0: 3, C_x: 1, C_y: 0.1. The number of iterations was doubled from the default 200 to 400 for WASI7.

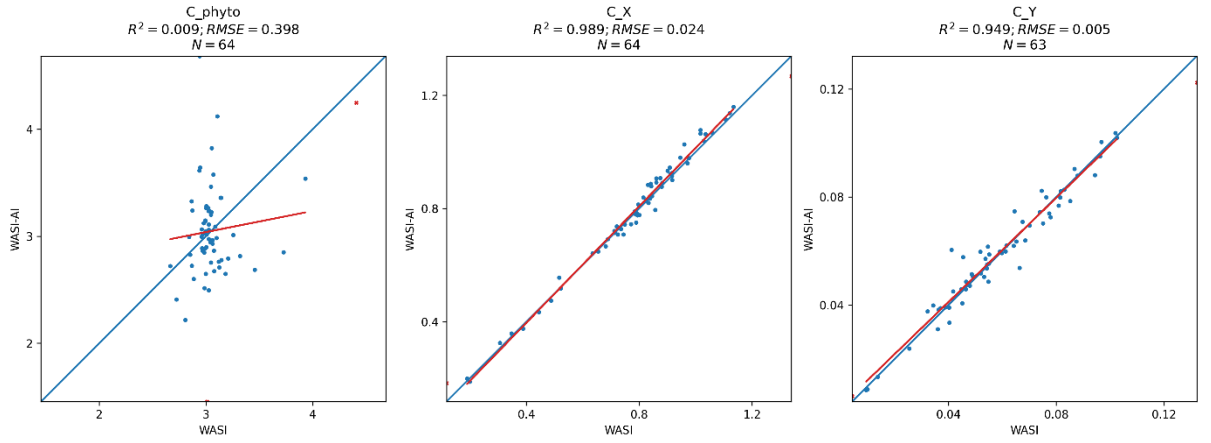

Figure A1. 2: For the same image as in Figure A1. 1: WASI vs. WASI-AI retrievals at 64 select pixels, as automatically generated by the WASI7 graphical user interface. The strong disagreement of phytoplankton retrievals of WASI-AI and WASI7 indicates strong ill-posedness for this parameter.

Figure A1. 2 shows that the phytoplankton retrievals for WASI7 and WASI-AI were around 3 mg m^{-3} with low agreement among the validation pixels (WASI7 vs. WASI-AI, r-squared = 0.009). Poor correlation between WASI7 and WASI-AI indicates high ill-posedness of the parameter (Niroumand-Jadidi and Gege, 2025).

This is not limited to the image presented in Figure A1. 1. The campaign images from Section 3.4 were inverted using the exact same model configuration with WASI7, WASI-AI, and MiniWASI. Figure A1. 3 shows that for WASI7 retrievals largely remained around the initial value of 3 mg m^{-3} while MiniWASI retrieved a broader range of values. As WASI-AI retrievals are image-based and learn from WASI7 retrievals, they are impacted by any issue in WASI7 inversion accuracy (see Figure A1. 1).

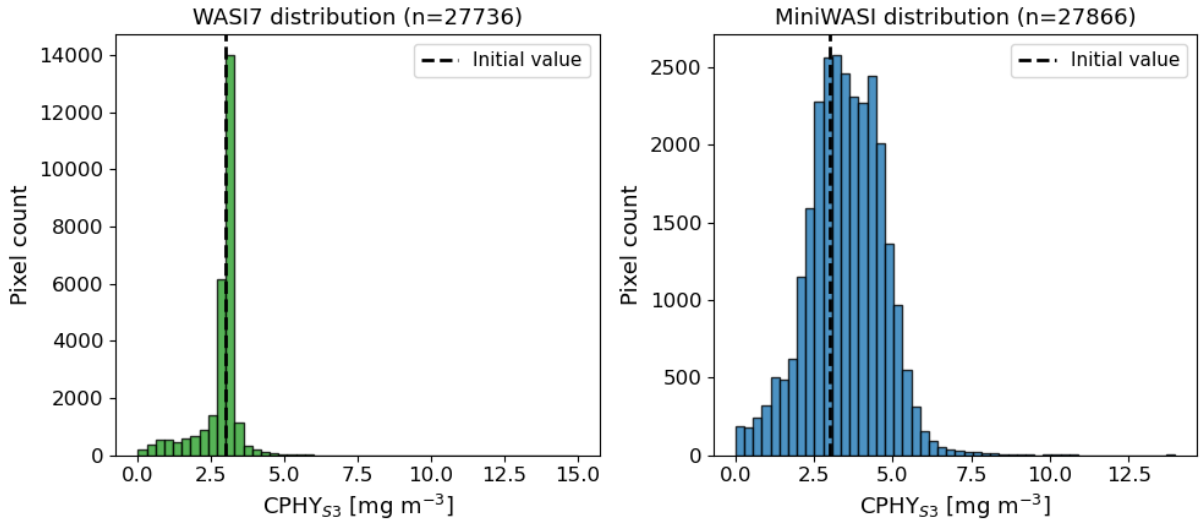

Figure A1. 3: Histograms of phytoplankton concentration values of the five campaign images (see Section 3.4; Figure A1. 1). WASI7 retrievals of phytoplankton are strongly biased towards the initial value (3 mg m^{-3}) while MiniWASI retrievals show greater variability at higher overall inversion accuracy. The number of pixels varies slightly due to the water masking procedure applied in WASI7.

Table A1. 1 shows that MiniWASI was approximately two orders of magnitude faster than WASI7, producing lower residuals for all campaign images.

Table A1. 1: The mean residual and overall runtime for each campaign image of Section 3.4 when inverted with MiniWASI and WASI7 respectively. MiniWASI shows lower average residuals while being two orders of magnitude faster using parallel processing. Variables and initial values used for this test: C_0: 3, C_x: 1, C_y: 0.1. Since it was not possible to adjust the viewing angle in WASI7 in image inverse mode it was set to 0 for both WASI7 and MiniWASI.

Image	Mean Resid MiniWASI [sr^{-1}]	Runtime MiniWASI [sec]	Mean Resid WASI7 [sr^{-1}]	Runtime WASI7 [sec]
S3A_20210520T101659	**0.0002655**	**8**	0.0003578	857
S3A_20210616T101659	**0.0004936**	**2**	0.0006733	451
S3B_20210520T093729	**0.0005574**	**1**	0.0008048	450
S3B_20210611T100729	**0.0004078**	**2**	0.0005581	494
S3B_20210616T093735	**0.0007878**	**2**	0.001095	424

The hyperspectral RAMSES measurements provided as part of the WASI demo data were inverted using all the available taxa-specific phytoplankton classes ($C_1 - C_5$), using both WASI7 and MiniWASI (Table A1. 2). The aim was to establish whether residuals are calculated in the

same way in MiniWASI and WASI7 and how they perform when inverting hyperspectral data with many phytoplankton classes (C_i).

Contrary to what is reported in the manual (Gege, 2024), the WASI7 residuals displayed in the user interface appear to be square-rooted. The following equation uses the same variable formulation as in the WASI7 user manual to facilitate comparison:

$$\text{Residuum} = \sqrt{\frac{1}{N}\sum_{i=1}^{N} g_i\,(m_i - f_i)^2}$$

where m = $R_{rs}^{\text{measured}}(\lambda_i)$, f = $R_{rs}^{\text{modeled}}(\lambda_i)$ (see Equation 21), g_i = per-band weight, N = number of nonzero-weight bands. MiniWASI residual reporting was adjusted for this exercise accordingly. While this formula does not exactly follow Equation 21, simple scaling does not impact the inversion result.

The spectra were inverted using both WASI7 (Table A1. 2, column "Resid WASI7") and MiniWASI (Table A1. 2, column "Resid MiniWASI"), yielding inverted parameters and residuals for each method.

To check whether a given WASI7 parameter set yields the same residuals in MiniWASI, the WASI7 parameter set was retrieved and used to forward calculate a spectrum in MiniWASI. The residuals resulting from this calculation were then manually recalculated against the original RAMSES spectrum (Table A1. 2, column "Resid WASI7 in MiniWASI").

The same procedure cannot be applied in reverse, since it is not possible to manually calculate a residual in WASI7. Therefore, to check whether a given MiniWASI parameter set yields the same residual in WASI7, the MiniWASI parameter set was retrieved and used to invert the original RAMSES spectrum with a single iteration (Table A1. 2, column "Resid MiniWASI in WASI7"). This single iteration had hardly any impact on the parameter set: Therefore, the residual could still be calculated without changing the parameters.

Table A1. 2: Inversion of hyperspectral RAMSES measurements distributed as part of the WASI7 demo data using MiniWASI and WASI7 respectively. Residual calculations in WASI7 and MiniWASI, while nearly identical, show small differences. Hence MiniWASI retrievals were also assessed in WASI7 and vice versa. MiniWASI retrieves a lower or the same residual, depending on the MiniWASI residual calculation. For WASI-type residuals MiniWASI outperforms WASI7 in 6 out of 8 cases. Variables and initial values used for this test: C_1: 1, C_2: 1, C_3: 1, C_4: 1, C_5: 1, C_x: 1, C_y: 0.1. The number of iterations was set to 10000 for WASI7.

File	Resid MiniWASI [sr^{-1}]	Resid WASI7 in MiniWASI [sr^{-1}]	Resid WASI7 [sr^{-1}]	Resid MiniWASI in WASI7 [sr^{-1}]
Ramses_STA10_09.rrs	**3.41E-03**	3.75E-03	3.75E-03	**3.41E-03**
Ramses_STA11_09.rrs	**1.06E-02**	**1.06E-02**	**1.05E-02**	1.06E-02
Ramses_STA12_09.rrs	**7.44E-03**	7.90E-03	7.82E-03	**7.43E-03**
Ramses_STA13_09.rrs	**6.05E-03**	6.06E-03	**5.9E-03**	6.04E-03
Ramses_STA14_09.rrs	**4.84E-03**	4.90E-03	4.89E-03	**4.84E-03**
Ramses_STA15_09.rrs	**1.26E-03**	1.31E-03	1.28E-03	**1.27E-03**
Ramses_STA16_09.rrs	**1.18E-03**	1.21E-03	1.21E-03	**1.18E-03**
Ramses_STA18_09.rrs	**5.00E-04**	5.54E-04	5.55E-04	**5.02E-04**

The residuals were largely the same in MiniWASI and WASI7 for the same parameter set (i.e. “Resid MiniWASI” vs. “Resid MiniWASI in WASI7” and “Resid WASI7 in MiniWASI” vs. “Resid WASI7” closely aligned). Slight differences may occur due to unknown differences in the residual or forward calculation. It should be noted that the forward calculations are visually indistinguishable when plotted on top of each other (not shown).

When judged by the MiniWASI replica of the WASI7 residual reporting, MiniWASI performed as well as or better than WASI7 for all RAMSES measurements. In only two cases was WASI7 as good as or better than MiniWASI when judged by the WASI7 user interface residual reporting (see Table A1. 2).

Appendix 2: Local relationship between aLH_{676} and CHL_F

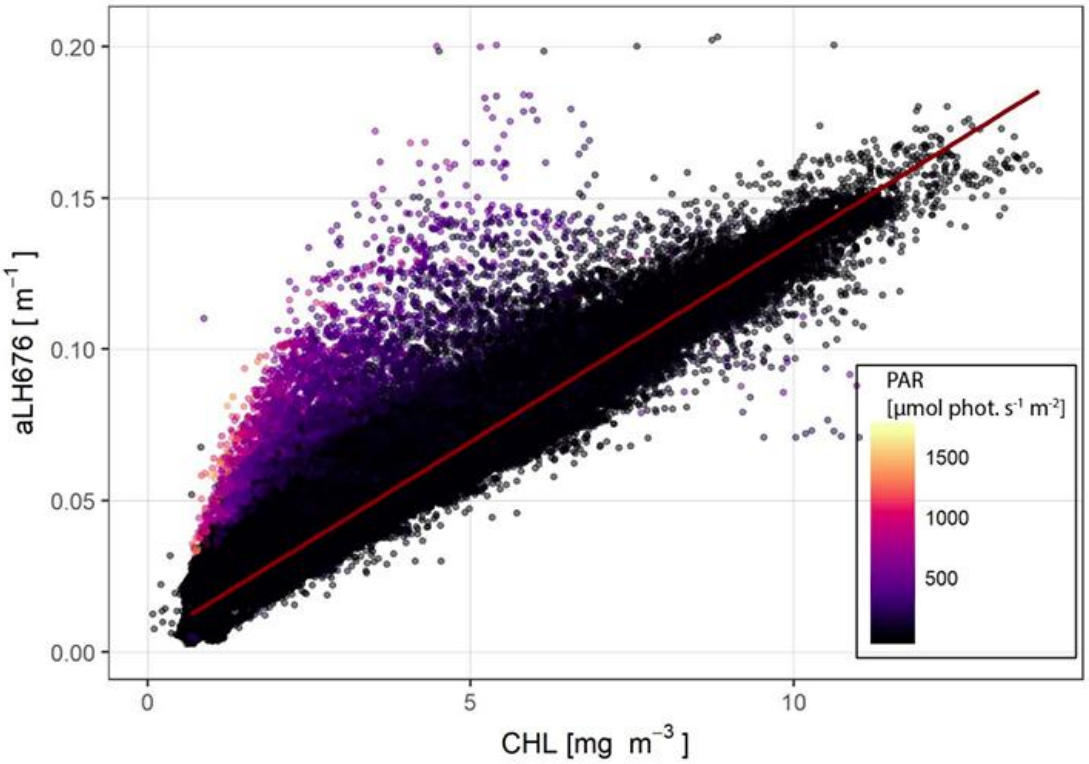

Figure A2. 1: Fluorescence-based CHL (chlorophyll a) plotted against aLH_{676} (absorption line height at 676 nm) for the full Thetis dataset, colored by downwelling PAR (Photosynthetically Active Radiation) irradiance. The color gradient illustrates non-photochemical quenching (NPQ): as PAR rises, the CHL-to-aLH_{676} ratio declines by up to 40%, introducing bias into fluorescence-based CHL estimates under bright conditions (Minaudo et al., 2021). The red line, fit to low-PAR (nighttime) data where quenching is negligible, gives the reference relationship used in Equations 1 and 2. Reprinted from Minaudo et al., 2021, Figure S3, CC BY-NC-ND 4.0.

Appendix 3: Relationship between modeled and measured *in situ* IOPs

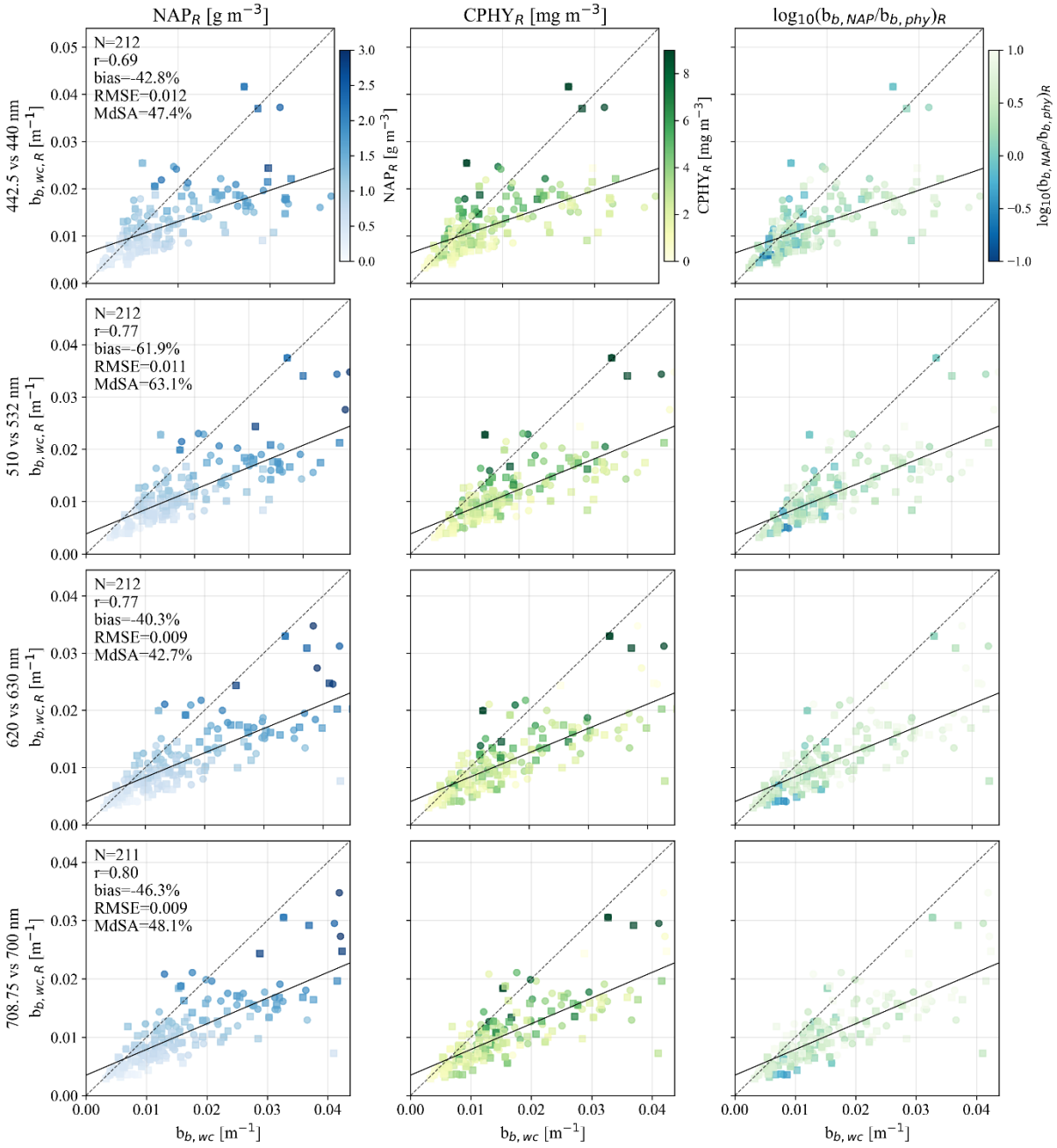

Figure A3. 1: Validation of constituent backscattering modeled from hyperspectral in situ $R_{rs}(\lambda)$ at the four wavelength pairs used for satellite validation (442.5 vs. 440, 510 vs. 532, 620 vs. 630, and 708.75 vs. 700 nm). $b_{b,wc,R}(wvl)$: constituent backscattering as inverted from hyperspectral in situ $R_{rs}(\lambda)$ at wavelength "wvl". $b_{b,wc}(wvl)$: constituent backscattering from in situ scattering conversion at wavelength "wvl". The first column is colored by non-algal particle concentration inverted from hyperspectral in situ $R_{rs}(\lambda)$ (NAP_R), the second by phytoplankton concentration inverted from hyperspectral in situ $R_{rs}(\lambda)$ ($CPHY_R$), the third by the (log-scale) proportion of retrieved NAP backscattering ($b_{b,NAP}$) to retrieved phytoplankton backscattering ($b_{b,phy}$). The 1:1 line (dashed) and best fit (solid) are shown in each panel; note that panels use different axis bounds.

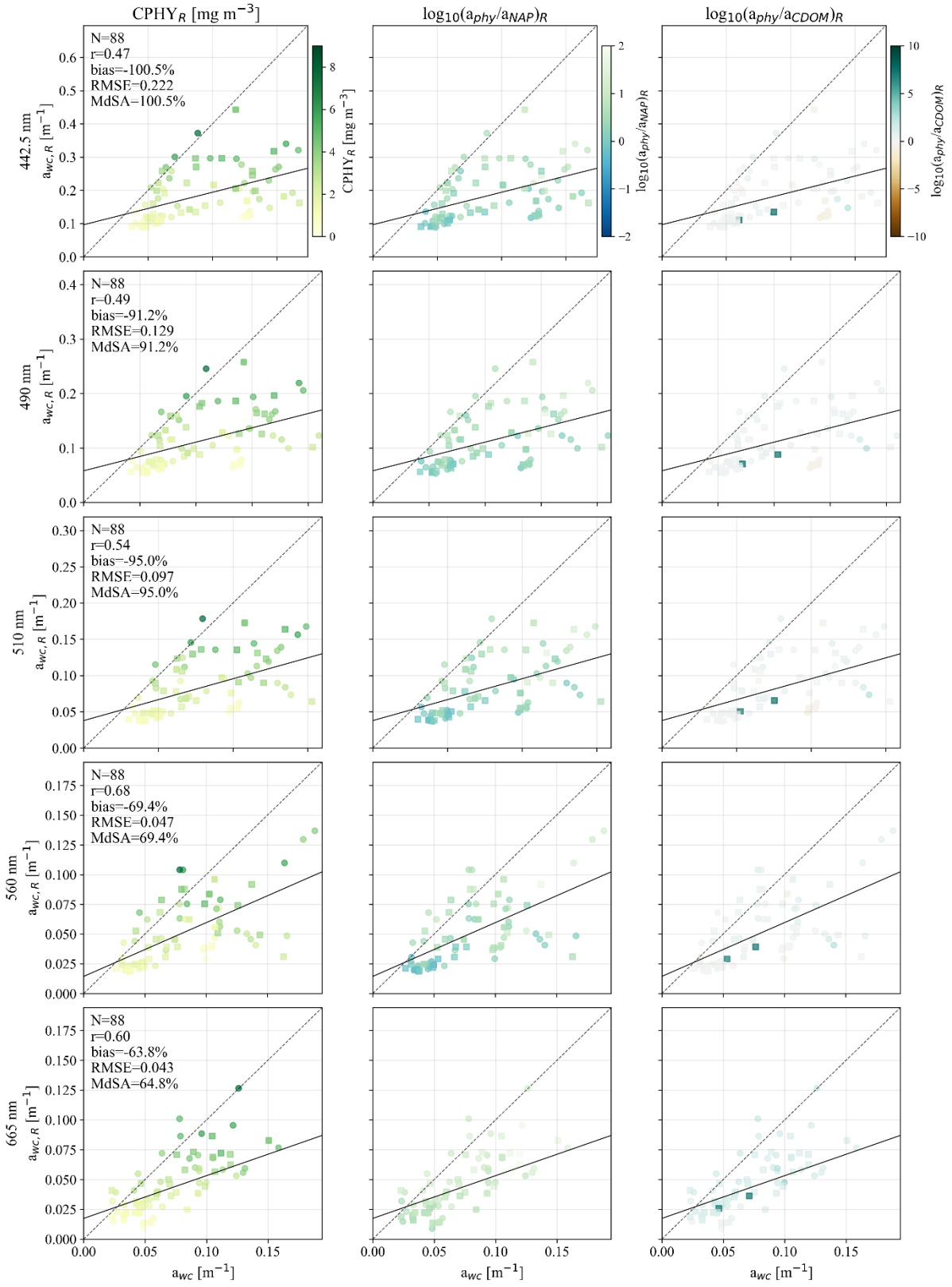

Figure A3. 2: Validation of water constituent absorption modeled from hyperspectral in situ $R_{rs}(\lambda)$ at the five wavelengths used during inversion (442.5, 490, 510, 560, and 665 nm). a_{wc}: absorption by water constituents measured in situ, $a_{wc,R}$: absorption by water constituents as inverted from hyperspectral in situ $R_{rs}(\lambda)$. The first column is colored by phytoplankton concentration inverted from hyperspectral in situ $R_{rs}(\lambda)$ ($CPHY_R$), the second by the (log-scale) proportion of retrieved phytoplankton absorption (a_{phy}) to retrieved non-algal particle absorption (a_{NAP}), the third by the (log-scale) proportion of a_{phy} to retrieved colored dissolved organic matter absorption (a_{CDOM}). The 1:1 line (dashed) and best fit (solid) are shown in each panel; note that panels use different axis bounds.

Appendix 4: aLH_{676} profile on 5.9.2021

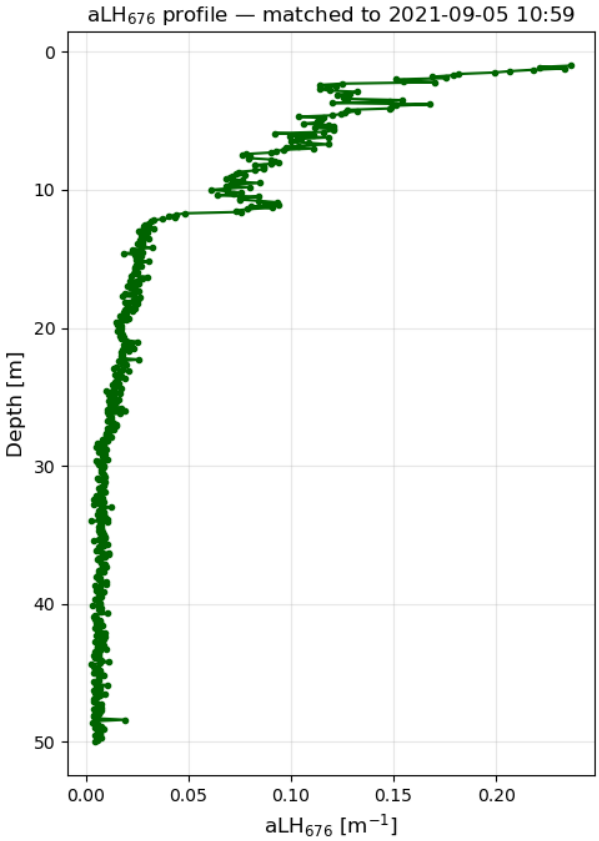

Figure A4. 1: aLH_{676} profile on 5.9.2021, corresponding to a bloom of Uroglena sp. Note the strong non-uniform vertical distribution, including in the range (0-5 m) used as average for in situ references. While near surface aLH_{676} is approximately 0.2 m^{-1} (corresponding to CHL ~ 15 mg m^{-3} when converted using Equation 2), it decreases to approximately 0.1 m^{-1} at a depth of 5 m (CHL ~ 7.5 mg m^{-3}).

Appendix 5: Drift of the AC-S instrument

Figure A5. 1 shows the observed AC-S instrument drift at Thetis using a pure water MiliQ reference where absorption should equal 0. Note that drift is sometimes an order of magnitude higher than usual constituent absorption in Lake Geneva (e.g. when >1 m^{-1} at 676 nm). Cleaning seems to only partly correct the drift (e.g. by removing the effect of biofouling). Absorption measurements are corrected by linearly interpolating for drift between the MiliQ references.

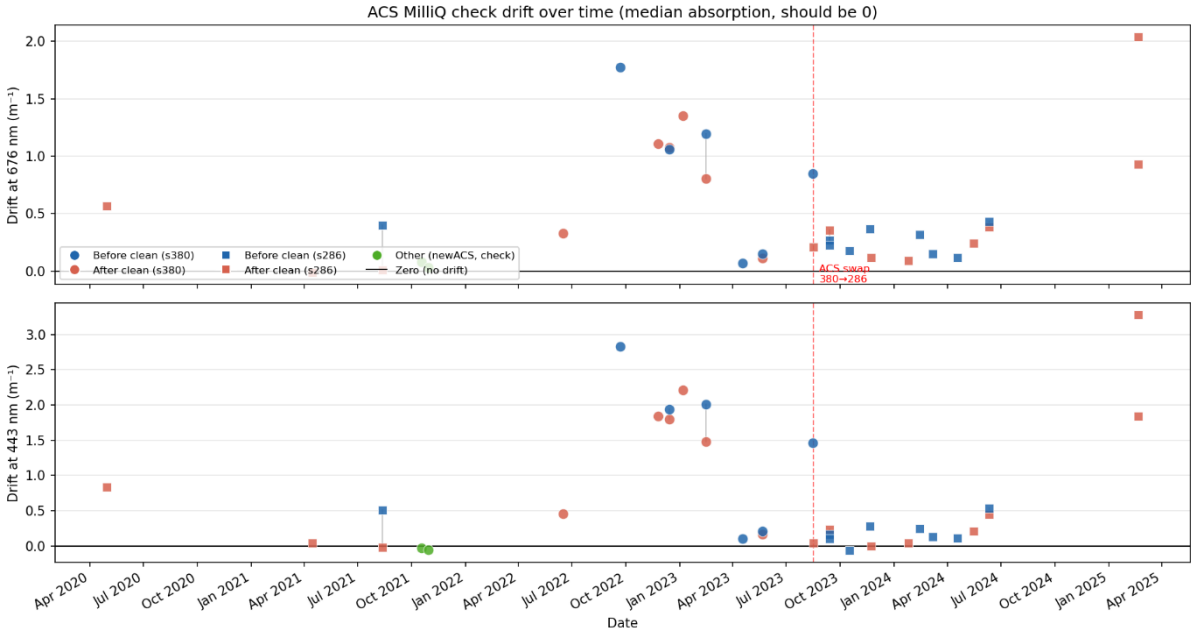

Figure A5. 1: Drift of the AC-S instrument absorption measurements at Thetis. An AC-S instrument switch is indicated in 2023. Courtesy of Jonas Wydler.

AI statement

When writing this thesis, I used the ChatGPT (Version 5, OpenAI) and Claude (Sonnet 5, Anthropic) language models to support my writing process. Throughout the thesis, I provided them with text I had written myself and asked them to correct grammar and syntax and to improve the wording of individual sentences and figure captions. Claude was further used to brainstorm section restructuring, help improve figures, and perform consistency checks (e.g. whether abbreviations are introduced at first occurrence). In Figure 4 the waterdrop and correlation schematics were generated by Claude using Python. DeepL was used to improve grammar, wording, and clarity.

Additionally, Claude (Sonnet 5) was used as an assistant for coding.

All ideas, arguments, and interpretations presented in this thesis are my own.

Statement of authorship

Personal declaration: I hereby declare that the submitted thesis is the result of my own, independent work. All external sources are explicitly acknowledged in the thesis.

Zurich, August 2026
Samuel Nils Landolt

