UNIVERSITÀ DEGLI STUDI DI URBINO CARLO BO

Department of Pure and Applied Sciences

Ph.D. PROGRAMME IN: Research Methods in Science and Technology

CYCLE XXXVIII

Design of Experiments-Based Development of UHPLC Methods for Targeted Analysis of Phytoplankton Pigments and Phycocyanin in Aquatic Samples: Characterization for Chemotaxonomic and Remote Sensing Applications

ACADEMIC DISCIPLINES:
CHEM-01/A - ANALYTICAL CHEMISTRY
BIOS-05/A ECOLOGY

Coordinator: Prof. Luca Lanci

Supervisor: Prof. Antonella Penna

Co-Supervisor: Prof. Giorgio Famiglini

Ph.D. student: Elisabetta Canuti

ACADEMIC YEAR
2024/2025

Summary

Abstract

The accurate quantification of phytoplankton pigments is critical for understanding aquatic ecosystem dynamics, supporting satellite algorithm validation, and advancing chemotaxonomic and biodiversity studies. This doctoral thesis shows the development, optimization, and validation of two independent Ultra-High Performance Liquid Chromatography (UHPLC) methods for pigment analysis in marine and inland water samples, both guided by a Design of Experiments (DoE) strategy. The first method addresses the need for a more efficient and environmentally sustainable analysis of phytoplankton pigment diversity. A novel ethanol-based UHPLC method using a core-shell C8 chromatographic column was developed to resolve 33 ecologically relevant pigments, including challenging pairs such as divinyl/monovinyl chlorophylls and xanthophyll isomers. The optimization, based on a central composite DoE design, enabled full resolution in under 26 minutes, reducing both toxic solvent use and analysis time. The method was validated on standards, algal cultures, and natural samples from the northwestern Adriatic Sea. The second UHPLC method targets the quantification of phycocyanin (C-PC), a key pigment in cyanobacteria, extracted from concentrated filter-based samples—commonly collected in satellite validation campaigns. A fullfactorial DoE was applied to compare extraction conditions, optimize chromatographic separation on a reversed-phase C5 chromatographic column using an acetonitrile gradient, and validate the method on natural samples and algal cultures for sensitivity, accuracy, and reproducibility. The result is a fast and robust UHPLC protocol suitable for routine monitoring of cyanobacterial blooms and satellite product validation. Together, these analytical advancements contribute to the development of more reliable tools for aquatic pigment quantification, offering critical support for satellite remote sensing calibration, phytoplankton functional type discrimination, and environmental monitoring programs.

1. Introduction

Phytoplankton form the foundation of aquatic food webs and play a pivotal role in global biogeochemical cycles as primary producers. Understanding their composition, abundance, and functional traits is essential for assessing ecosystem health, modeling carbon and nutrient dynamics, and evaluating the impact of environmental change. Traditionally, phytoplankton were characterized by microscopy, a method that is labor-intensive and requires specialized expertise, limiting its scalability for broad ecological assessments.

High-performance liquid chromatography (HPLC) has become a widely used technique for pigment-based phytoplankton classification, offering detailed insights into community structure through pigment composition analysis (Mackey et al., 1996; Vidussi et al., 2001). These pigments act as chemotaxonomic markers, enabling researchers to identify the composition and relative abundance of phytoplankton functional types (PFTs), which vary in size, ecological function, and environmental response.

Compared to traditional microscopic methods, HPLC-based chemotaxonomy provides a more comprehensive and scalable approach, reducing the time and expertise required for species-level identification (Vidussi et al., 2001; Uitz et al., 2006; Brewin et al., 2010; Hirata et al., 2011). Pigment-based analyses are also central to the calibration and validation of ocean color remote sensing algorithms (Hooker et al., 2000). By linking in situ pigment data with satellite-derived reflectance, researchers can map PFTs and size classes across broad spatial and temporal scales (IOCCG, 2014).

This integration is particularly important under global change, as freshwater and coastal systems are increasingly affected by eutrophication, rising temperatures, and altered hydrological regimes. These changes contribute to the intensification of harmful algal blooms (HABs), particularly cyanobacteriadominated blooms, which degrade water quality, disrupt ecological balance, and pose risks to public health. In this context, remote sensing and pigment chemotaxonomy complement each other: remote sensing captures large-scale spatial dynamics, while pigment analysis provides taxonomic resolution through field validation.

However, both approaches face methodological challenges that limit their full potential. Chlorophyll a (Chl-a) is widely used as a proxy in remote sensing but cannot reliably distinguish cyanobacteria from other phytoplankton groups in optically complex waters (IOCCG, 2021). Additionally, detection of key cyanobacterial pigments, such as C-phycocyanin (C-PC), by satellite is limited by the absence of robust analytical protocols for field samples.

To address these limitations, this thesis introduces two methodological advances:

1. Optimization of phytoplankton pigment analysis through improved chromatographic conditions, enhancing the resolution of diagnostic pigments, including chlorophyll and carotenoid isomers, while employing less toxic and more environmentally friendly solvents.

2. Development of a robust UHPLC-based protocol for quantifying C-phycocyanin in seawater, supporting the validation of satellite algorithms for cyanobacterial detection.

Both methods were developed using a Design of Experiments (DoE) strategy to ensure analytical robustness and reproducibility (Myers et al., 2009; Hibbert, 2012; Ganorkar et al., 2017; Montgomery et al., 2019). HPLC remains the gold standard for pigment quantification, offering high sensitivity, specificity, and reproducibility. Intercomparison studies indicate uncertainties below 6% for Chl-a and within 25% for accessory pigments, thresholds suitable for satellite validation (Hooker & McClain, 2000).

Despite more than 26 published methods (Roy et al., 2011), including some UHPLC variants (Suzuki et al., 2015), most rely on conventional stationary phases (C8, C18, C30) and hazardous solvents such as methanol. Modern alternatives, such as core-shell or phenyl-based columns and environmentally friendly eluents (Plotka et al., 2013), are underutilized, especially in marine and limnological studies. Furthermore, pigments such as C-PC are underrepresented in routine protocols. Existing spectrophotometric and HPLCbased methods for C-PC (Kumar et al., 2014; Sobiechowska-Sasim et al., 2014; Lauceri et al., 2018; Yacobi et al., 2018) are rarely applied to field-collected material, limiting integration into satellite algorithm development.

Given that cyanobacterial blooms are most prevalent in inland and coastal waters (Mouw et al., 2015; Clark et al., 2017), developing analytically rigorous and environmentally conscious UHPLC methods fills a critical gap. The new protocols aim to enhance chemotaxonomic resolution, reduce environmental impact, and strengthen the link between in situ measurements and satellite observations. These contributions

support improved environmental surveillance, early warning systems for HABs, and more accurate assessments of phytoplankton community structure at regional to global scales. Ultimately, the methods will facilitate satellite data validation, improve PFT estimations, and advance understanding of phytoplankton ecology across diverse aquatic environments.

1.1. Methodology

The methodological approach adopted in this thesis for setting up the two complementary UHPLC-based methods—one for C-PC quantification and one for phytoplankton pigments—combined similar frameworks for development, optimization, and validation. Although targeting different analytes, both methods shared a common experimental design and were supported by systematic optimization strategies.

A central element of method development was DoE, which enabled simultaneous evaluation of multiple factors affecting chromatographic performance, reducing the number of runs compared to traditional one-factor-at-a-time approaches (Snyder et al., 2009; Kumar Sahu et al., 2018). DoE assessed column type, mobile phase composition, and operational parameters, facilitating the identification of conditions that ensured robust separation and reproducibility (Hibbert, 2012; Ganorkar et al., 2017; Montgomery et al., 2019).

Both UHPLC methods were tested on selected algal cultures for taxonomic and pigment diversity and on natural samples collected from the northwestern Adriatic Sea and lakes (e.g., Lake Varese) affected by cyanobacterial blooms. This dual strategy validated the methods across both controlled monocultures and environmentally complex communities, ensuring broad applicability.

Finally, the protocols underwent rigorous validation, including interlaboratory comparisons such as the HPLC Intercomparison on Phytoplankton Pigments (HIP) exercises (Canuti et al., 2025a) and cross-checks with established methods like spectrophotometry. Analytical uncertainties in pigments and C-PC quantification were explicitly evaluated, given their relevance for satellite bio-optical algorithm calibration and ecological modeling. Detailed statistical approaches for each method are presented in Chapters I and II.

1.2. Integration of Pigment-Based Datasets with Satellite Observations: Chemotaxonomy, Bio-Optical Algorithms, and Validation

This PhD research focused on developing two analytical methods: one for phytoplankton pigments and another for C-phycocyanin. While these methods are significant advances, the datasets generated during this study were relatively limited. To illustrate their potential, previously acquired, robust, and validated datasets were incorporated into the analyses. These datasets provide a strong empirical basis for demonstrating broader applications of pigment-based methods, particularly when combined with satellite observations.

Pigment datasets are of exceptional ecological and methodological relevance. They capture taxonomic and functional diversity and provide essential insights into aquatic ecosystems. When combined with satellite data, they enable calibration, validation, and refinement of bio-optical algorithms, extending their applicability from local to regional scales. Integration of chemotaxonomic pigment information with statistical and machine learning methods demonstrates how these datasets improve both ecological interpretation and reliability of satellite-derived products.

The datasets cover a wide range of environments, including the Adriatic, Mediterranean, Baltic, and Black Seas, as well as transitional and inland waters such as Lake Maggiore. This diversity spans oligotrophic open-ocean waters, turbid coastal zones, and inland waters with complex optical properties. Emphasis is placed on the northwestern Adriatic Sea, frequently impacted by HABs and recognized for its optical complexity (Melin, 2011).

In the northwestern Adriatic Sea, validation of satellite-derived chlorophyll a and accessory pigments was conducted using high-quality field datasets (Canuti et al., 2025 b, submitted). In the Baltic and Black Seas, HPLC-derived pigment concentrations combined with satellite observations enabled the reconstruction of PFTs, characterization of community structure, and development of regionally adapted bio-optical models. These datasets serve multiple purposes, supporting algorithm development, validation, and ecological interpretation of phytoplankton dynamics.

Building on this foundation, the datasets were applied in three principal directions:

1. Satellite data validation (Canuti et al., 2025b-submitted)
2. Bio-optical algorithm development (Canuti et al., 2024a; Canuti, 2025c, 2025d-submitted)
3. Chemotaxonomic characterization of phytoplankton communities (Canuti et al., 2024b)

The following subchapters provide a detailed presentation of each application.

1.2.1. Satellite Data Validation

Validation of satellite-derived chlorophyll a (Chl-a) and accessory pigment products is essential to ensure the accuracy of bio-optical algorithms across diverse water types. High-quality in situ datasets serve as an independent reference for assessing satellite performance, particularly in optically complex and ecologically dynamic coastal regions.

In the northwestern Adriatic Sea, a 16-year in situ time series collected monthly at two coastal stations by the University of Urbino was used to evaluate the performance of Copernicus Marine Service (CMEMS) Chl-a products (Canuti et al., 2025, submitted; ANNEX I). Three satellite products were assessed: the 100m high-resolution product (HR-L3), the 300-m multi-year product (MY-L3), and the 1-km gap-filled multiyear product (MY-L4). The analysis revealed that the 1-km MY-L4 product exhibited the most consistent performance across the entire dataset, making it particularly suitable for long-term monitoring of coastal waters, especially at intermediate Chl-a concentrations (1–10 mg m^{-3}). The 300-m MY-L3 products showed better correlation with in situ data at nearshore sites, while the 100-m HR-L3 products performed poorly overall. Performance varied with biomass, with the highest uncertainties observed at Chl-a concentrations exceeding 10 mg m^{-3}, typical of bloom conditions.

These results highlight the strengths and limitations of current CMEMS products for monitoring Chl-a in optically complex coastal waters and underscore the necessity of continuous in situ validation. By providing an independent benchmark, these datasets enable more reliable satellite-derived assessments of phytoplankton biomass and water quality, which are essential for ecological monitoring, operational management, and the early detection of harmful algal blooms.

1.2.2. Bio-Optical Algorithm Development for Remote Sensing

Accurate estimation of phytoplankton functional types (PFTs) from space remains challenging in optically complex waters, where global algorithms often fail to capture regional variability. In this thesis, regionalized bio-optical algorithms were developed and validated for two contrasting basins: the Baltic Sea and the Black Sea (Canuti, 2025 c and d; ANNEX II and III).

In the Baltic Sea, empirical diagnostic pigment (DP)-based approaches were refined using highperformance liquid chromatography (HPLC) datasets from multiple sub-basins. This allowed improved estimation of phytoplankton size classes (PSCs) and key functional groups, including cryptophytes, green algae, and dinoflagellates. Nanoplankton dominated basin-wide (~46% of chlorophyll a), picoplankton prevailed offshore, and microplankton peaked in nearshore regions. For the Black Sea, hierarchical clustering, principal component analysis, and network-based community detection were applied to HPLC measurements from 690 stations collected across 12 bio-optical campaigns. Region-specific PFT–pigment coefficients were derived and applied to multi-decadal satellite chlorophyll a datasets (1998–2024), reconstructing spatial and temporal patterns of PFTs. Microplankton dominated nutrient-rich coastal zones (70–80% of chlorophyll a), nanoplankton were broadly distributed (~30–40%), and picoplankton prevailed offshore (>60%). Both regionalized algorithms significantly reduced errors compared to global approaches, particularly for cryptophytes, haptophytes, and prochlorophytes, and results were consistent with microscopy-based observations. These findings demonstrate the potential of tailored, region-specific algorithms to enhance monitoring of phytoplankton community structure, supporting both ecological assessments and the development of satellite-based tools for operational monitoring of coastal and shelf seas.

1.2.3. Chemotaxonomic Characterization of Phytoplankton Communities

Chemotaxonomic approaches link pigment composition to phytoplankton taxonomy and functional traits, providing detailed insights into community composition beyond total chlorophyll a measurement. HPLC is

the gold standard for quantifying phytoplankton pigments and supports both ecological studies and satellite algorithm calibration.

In Lake Maggiore, cyanobacterial blooms have become increasingly frequent, impacting ecosystem function. Between May and September 2023, five measurement campaigns collected bio-geochemical and bio-optical data at 27 stations (Canuti et al., 2024 b; Annex IV). Chl a ranged from 1.13 to 6.9 mg m^{-3}. Phytoplankton pigments were analyzed by HPLC, and the CHEMTAX approach was applied for classification. Results were cross-validated using principal component analysis (PCA), hierarchical cluster analysis (HCA), and microscopic counts. Cyanobacteria were identified based on unique pigment markers such as carotenoids. HPLC-derived chemotaxonomic results aligned closely with PCA, HCA, and microscopy, confirming the reliability of pigment-based classification for cyanobacteria and other phytoplankton groups.

In the Western Mediterranean Sea, a comparative study between two laboratories evaluated uncertainties in HPLC pigment quantification for oligotrophic waters (Canuti et al., 2025 e). Significant differences were observed for chlorophyll c and peridinin, while total chlorophyll a agreed within 6.1%. This study highlighted both the limitations and robustness of HPLC methods for supporting satellite data validation and algorithm development, emphasizing the importance of consistent analytical procedures for chemotaxonomic characterization.

These studies toghether with other focused on Baltic Sea (Canuti et al., 2024 a, Canuti et al. 2025 f; ANNEX V and VI) demonstrate that chemotaxonomic approaches, when combined with statistical analyses and quality-controlled HPLC data, provide a reliable framework for characterizing phytoplankton communities, with direct implications for remote sensing applications and ecological monitoring.

2. Research Objectives

The overarching goal of this PhD research is to advance the analytical capabilities of phytoplankton pigment quantification by developing and validating efficient UHPLC methodologies. These methods are intended to overcome technical and ecological challenges in aquatic research, particularly supporting satellite remote sensing, bio-optical algorithm development, and environmental monitoring. The specific research objectives are:

(i) firstly, to develop chromatographic protocols that significantly reduce solvent consumption and total analysis time, without compromising the resolution and accuracy of pigment separation. This involves optimizing chromatographic conditions, such as column chemistry, gradient profiles, and flow parameters, to achieve faster, more environmentally friendly and cost-effective workflows suitable for high throughput analysis;

(ii) the second objective is to ensure that the developed UHPLC methods meet the precision and accuracy standards necessary for satellite algorithm calibration and validation. This includes generating high-quality pigment data to support bio-optical models and improve ocean color products, particularly in complex coastal and inland waters where interpreting the satellite signal is more challenging;

(iii) thirdly, the aim is to establish a robust and validated method for quantifying C-PC in field aquatic samples. Although C-PC is a key biomarker for cyanobacteria, it is underrepresented in pigment monitoring and remote sensing validation efforts due to analytical limitations. This research aims to fill a critical methodological gap and improve the detection and monitoring of cyanobacterial blooms by incorporating C-PC into routine UHPLC workflows;

(iv) fourth, the developed methodologies will be applied to a wide array of field-collected samples obtained from oceanographic campaigns and freshwater monitoring programs. These samples will be collected from marine, transitional, and freshwater environments with varying phytoplankton assemblages and bloom dynamics, with a particular focus on regions affected by cyanobacteria. This broad spatial and ecological coverage will enable a thorough evaluation of the applicability of the method and support comparative ecological studies across diverse aquatic systems.

3. Chapter I

Design of Experiment guided Development of an UHPLC Method for Phytoplankton Pigments Analysis

Elisabetta Canuti, Samuela Capellacci, Silvia Casabianca, Fabio Ricci, Antonella Penna

This chapter has been submitted and is currently under review in *Marine Chemistry*

Abstract

Phytoplankton pigment analysis using high-performance liquid chromatography (HPLC) is essential for characterizing community composition, supporting the satellite algorithm validation, and advancing aquatic biogeochemical research. However, traditional HPLC methods often rely on toxic solvents, take a long time to analyse, and find it difficult to resolve structurally similar pigments such as mono- and divinyl chlorophylls and xanthophyll isomers. In this study, we developed and validated a novel ultra-high-performance liquid chromatography (UHPLC) method optimized for separating of ecologically significant phytoplankton pigments. This method uses a core-shell C8 column (100 × 4.6 mm, 2.6 µm) at 60 °C, a linear ethanol–solvent A gradient over 22.5 minutes, and pH 7 buffering. This results in reduced solvent toxicity and lower solvent consumption. The method development was guided by a Design of Experiments (DoE) approach to optimize resolution for challenging pigment pairs such as divinyl/monovinyl chlorophylls and lutein/zeaxanthin. The final method achieved full resolution of 33 pigments, with excellent reproducibility (RT variation of less than 0.6% and an area variation of less than 12%), reducing the total analysis time reduced to 25.5 min. Validation was performed using pigment standards, laboratory cultures, and natural samples from the northern Adriatic Sea and by comparing results with those of another validated method. this demonstrated the robust performance across a range of concentrations and matrix complexities. This study provides a rapid, reproducible, and environmentally safer alternative for phytoplankton pigment analysis in marine research and operational monitoring programs.

1. Introduction

Marine phytoplankton, which contribute nearly half of the global primary productivity, play a central role in biogeochemical cycles and climate regulation (Behrenfeld et al., 2001; Demarcq et al., 2012; Alvain et al., 2013). Chlorophyll *a* (chl *a*), a key pigment present in all phytoplankton, is commonly used to assess biomass and primary production (Falkowski et al., 1998; Harding et al., 2016). However, chl *a* alone provides limited insight into community composition. Pigment-based chemotaxonomy is a robust approach that takes advantage of diagnostic pigments characteristic of specific taxonomic groups to infer phytoplankton functional composition (Mackey, M. et al., 1996, Jeffrey et al., 1997; Vidussi et al., 2001, Uitz et al., 2006, Brewin et al., 2010, Roy et al., 2011). High-performance liquid chromatography (HPLC), and more recently ultra- performance liquid chromatography (UHPLC), are widely recognized as reference techniques for the qualitative and quantitative analysis of phytoplankton pigments in aquatic environments (Roy et al., 2011, Suzuki et al., 2015). Their high specificity and sensitivity, coupled with their

ability to resolve complex pigment mixtures, make them fundamental tools for validating satellite-derived estimates of chl *a* (JGOFS 1994). Despite these strengths, pigment analysis via HPLC presents several analytical challenges. Phytoplankton pigments vary widely in polarity and include numerous structural isomers, which complicates their simultaneous separation and quantification. The mono- and divinyl forms of chlorophylls *a*, *b*, and *c* are particularly important due to their chemotaxonomic value, yet they are difficult to resolve chromatographically (Goericke et al., 1993; Barlow et al., 1997). Over than thirty methods for pigment separation have been reported for pigment analysis (Roy et al., 2011), with most of which employ C8 or C18 reversed-phase columns and diode array detectors (Gieskes et al., 1987; Wright et al., 1991; Goericke et al., 1993; Barlow et al., 1997; Van Heukelem et al., 2001; Zapata et al., 2005). Over the past few decades, the methods proposed by Wright et al. (1991), Zapata et al. (2000), and Van Heukelem et al. (2001) have been widely adopted in oceanographic research and in protocols for validating satellite data (JGOFS 1994; Jeffrey et al., 1997; Mueller et al., 2003; Roy et al., 2011). However, these methods traditionally involved relatively long analysis times, large sample and solvent volumes, and the use of toxic organic solvents such as methanol and acetonitrile. This raised concerns regarding throughput, environmental impact and operator safety. In response, to these concerns, recent advances have explored more efficient, environmentally friendly workflows. For instance, ethanol has been suggested as a safer alternative to methanol, achieving comparable extraction performance (Sanz et al., 2015), while UHPLC systems have enabled faster analyses without compromising pigment resolution (Suzuki et al., 2015).

As said, a key challenge in HPLC-based pigment analysis is achieving the simultaneous resolution of ecologically important pigment pairs. Examples include Chlorophyll *c1* (chl *c1*) and Chlorophyll *c2* (chl *c2*), and the monovinyl (MV) and divinyl (DV) forms of Chlorophyll *a* and *b* (chl *a* vs. DV chl *a*, and chl *b* vs. DV chl *b*). Divinyl Chlorophylls *a* and *b* are the primary photosynthetic pigments in the marine cyanobacterium *Prochlorococcus marinus*, which is a major contributor to phytoplankton biomass in subtropical and tropical ocean regions (Goericke et al., 1993). Moreover, to avoid overestimating chlorophyll *a* in satellite ocean color observations, it is essential to chromatographically separate DV chl *a* from its MV counterparts (Jeffrey et al., 1997, Muller et al, 2003). Comparative studies have assessed the performance of different stationary phases, including octylsilica (C8) and octadecylsilica (C18), in pigment separation. C8 phases often provide enhanced resolution of chlorophyll *c* variants and divinyl analogues (Mendez et al., 2007), while C18 columns are typically offer with shorter run times and lower flow rates. This is particularly advantageous for detecting pigments at trace concentrations. Advancements in column chemistry have further improved pigment selectivity. Pentafluorophenyl (PFP) phases, combined with octadecyl spacers,

have demonstrated enhanced selectivity toward structurally similar pigments. This configuration has enabled the simultaneous separation of nearly 70 pigments—including mono- and divinyl chlorophylls, as well as many taxonomically relevant carotenes and xanthophylls—in under 42 minutes (Sanz et al., 2015). Additionally, Jayaraman et al. (2011) used a C16-amide column method that successfully resolve several MV/DV pairs of chlorophylls *a* and *b*, as well as isomeric carotenoids such as lutein (lut) and zeaxanthin (zea).

In this context, the aim of the present study is to develop and validate a novel UHPLC method for phytoplankton pigment analysis that (i) reduces the toxicity of extraction and elution solvents, (ii) minimizes sample and solvent requirements, and (iii) shortens analytical run times, while maintaining or improving the resolution and reproducibility achieved by established HPLC methods used in satellite validation and chemotaxonomic studies. In a recent review, Kumar Sahu et al. (2018) highlighted the benefits of Design of Experiments (DoE) in HPLC method development and validation. In the present study, the UHPLC method development was guided by a DoE approach, optimizing key chromatographic parameters using pigment standards representative of typical phytoplankton pigment profiles, including the MV/DV chl *a* and zea/lut pairs. A variety of column chemistries and solvent systems were evaluated, including two core-shell C8 columns and one C18-phyenyl column. Following optimization, the method was validated using both laboratory cultures and natural samples collected during a year-long monthly monitoring campaign in the Northern Adriatic Sea. The focus was on ecologically significant groups such as diatoms and dinoflagellates of this geographical area. The results of the newly developed method were further validated against the widely used Van Heukelem protocol as implemented by Canuti (2023) to ensure ensuring compatibility with established reference procedures. This study contributes to the modernization of pigment analysis workflows, by offering a more sustainable, rapid, and scalable alternative for aquatic research, ecological monitoring, and satellite algorithm development.

2. **Materials and Methods**

2.1. Seawater samples

Surface seawater samples were collected monthly from March 2024 to February 2025 monthly at a depth of 0.5 m using Niskin bottles at 3000 m from the Foglia River (F3000: 43°56’.55N, 12°56’.18E, max depth 13 m) in the northwestern Adriatic Sea (Fig. 1). The water samples were filtered onto a Whatman GF/F filters (nominal pore size of 0.7 µm and a 25 mm diameter). The filters were kept frozen at –80°C prior to extraction.

2.2. Algal cultures

Algal strains representing variety of marine phytoplankton species including some peculiar of the Adriatic Sea were selected for the diversity in their pigment composition. The cultures, listed with their designation indicated in parentheses, used in the present study were: *Navicula sp*. (CBA 2 (D1')), *Chaetoceros sp.* (CBA 13), *Skeletonema marinoi* (CBA 3D), *Cylindrotheca fusiformis* (CBA 82), *Prorocentrum micans* (CBA 21), *Scripsiella trochoidea* (CBA 2- F500), *Gymnodium impudicum* (GY6V), *Alexandrium minutum* (CNR-AMI V1), *Pavlova gyrans* (CCMP 608) and *Heterosigma akashiwo* (HA2V). Dinoflagellate cultures were grown in F/4 -Si medium, while the classes Pavlovophyceae and Raphidophyceae were cultured in L1 medium at a temperature of 23 ± 1 °C. Bacillariophyceae cultures were maintained in F/2 medium at 18 ± 1 °C. Illumination was provided by cool-white fluorescent bulbs, with a photon flux density of 100 µE m^{-2} s^{-1}, under a standard 14:10 h light–dark photoperiod. In addition, the *Synechococcus* (PCC 6911) was grown under dim lighting conditions and at a temperature of 22 °C (as in Wojtasiewicz et al., 2016). The strains were cultured at the Laboratory of Marine Ecology, University "Carlo Bo" of Urbino, Italy. For each strain, 10 ml of culture aliquots was added and gently mixed with sea water samples (1 L). An amount of 250 mL of this mixture was filtered under low vacuum through 25 mm diameter glass fiber filters (0.7 µm pore size) and immediately flash frozen in liquid nitrogen and preserved at -80° C until analysis.

2.3. Pigment standards

Certified pigment standards used for method calibration, along with a pigment mixture (Mix-128, Lot No.) employed for method development and retention time verification, were obtained from DHI Laboratory Products (Hørsholm, Denmark). Each standard was accompanied by a certificate of analysis to ensure identity, purity, and concentration accuracy.

2.4. Pigment extraction

To evaluate pigment recovery throughout the extraction process, α-tocopherol was used as an internal standard due to its high stability and absence of degradation products in both methanol and acetone, regardless of extraction time (Van Heukelem et al., 2001, Esteban et al., 2009). Two extraction solvents were compared: acetone and ethanol. The filters were cut into pieces and transferred into 10-mL polypropylene tubes (Corning Inc., Arizona, USA) containing 2.5 mL of an internal standard solution—either 0.025 g/L α-tocopherol (Honeywell, North Carolina, USA) in acetone or 0.0125 g/L α-tocopherol in ethanol (HPLC gradient grade, Merck, Darmstadt, Germany)—along with 150 µL of Milli-Q water. The mixture was left to soak for 1 hour at −20 °C, after which it was sonicated for 90 seconds using a sonication

probe (BANDELIN electronic, Berlin, Germany) on ice. This was followed by an additional soaking step of 3.5–4 hours at −20 °C followed. The resulting extract was filtered through a 0.2-µm Teflon syringe filter into an amber vial, vortexed, and transferred into an UHPLC amber vial. The samples were then stored at 4 °C in a thermostated autosampler until analysis.

2.5. UHPLC Equipment

Pigment separation was carried out using an Agilent Infinity II 1260 UHPLC system, equipped with a quaternary pump (G7104A), diode array detector (G7115A DAD WR equipped with a 10 mm pathlength capillary cell), fluorescence detector (G1321A FLD), autosampler (G7129B), and a thermostated column compartment. Chromatograms were analyzed using OpenLab CDS software, with absorbance monitored at 436, 440, and 665 nm. Three different chromatographic columns were tested for performance and selectivity (specifications provided in Table 1). The mobile phase consisted of solvent A (ethanol, HPLC grade, Merck, Germany) and solvent B (28 mM tetrabutylammonium hydroxide [TBA-OH] in Milli-Q water, adjusted to pH 6 or 7 with acetic acid). The importance of flow rate, buffer pH, and column temperature in achieving efficient pigment separation is well documented (Snyder and Dolan, 2009). These parameters, along with gradient time and initial ethanol percentage, were systematically optimized using a DoE approach (see Section 2.6). A linear gradient elution was applied, and the experimental conditions were varied according to the DoE matrix. The linear gradient was followed by of 3-min at 100% of solvent A for all the analysis. A 3-minute of re-equilibration step at initial conditions was included between injections. Sample injection was performed by sequentially loading 40 µL of sample followed by the ion-pairing reagent (28 mM TBA-OH in Milli-Q water) to reduce peak distortion, as recommended in previous studies (Zapata et al., 1991; Latasa et al., 2001; Van Heukelem et al., 2001). The amount of ion-pairing added to the sample was also matter of optimization in the DoE approach. All samples were equilibrated at the temperature of 4 °C in the UHPLC autosampler compartment before the injection.

Table 1. Characteristics of the chromatographic columns used in this study

Columns	Bonded phase	Dimensions (mm)	Pore size (µm)	Particle size (Å)	Surface area (m^2/g)	Carbon load
Kinetex Core-shell	C8	100x4.6	2.6	100	200	8%
Kinetex Core-shell	C8	150x4.6	5	100	200	8%
ACE C18-AR	C18-Phenyl	150x4.6	3	100	300	15.5%

Chlorophylls and carotenoids were detected using diode-array spectroscopy across the 270–800 nm range. Pigment identification was performed by comparing the retention times and spectral features of chromatographic peaks with those of authenticated standards or reference extracts from phytoplankton cultures. Peak purity was assessed by evaluating spectral homogeneity. The resolution, R_s, between two adjacent peaks (a and b) was calculated using the equation [1]:

$$\mathrm{Rs,(a/b)} = \frac{2(\mathrm{Rtb} - \mathrm{Rta})}{(\mathrm{Wa} + \mathrm{Wb})} \qquad [1]$$

where Rt_a and Rt_b are the retention times of peaks a and b, respectively, and W_a and W_b are their corresponding baseline peak widths (Snyder et al., 2009). Resolution of 1 is considered acceptable, 1.5 is considered desirable (Van Heukelem et al., 2001, Snyder et al., 2009)

2.6. Method development and validation

A structured DoE approach was employed to develop and optimize a UHPLC method for separating chlorophylls and accessory pigments in phytoplankton samples, using ethanol as the organic component of the mobile phase. The experimental design included both continuous and categorical variables. The categorical factors comprised column type—two C8 core-shell and one C18 reversed-phase columns (see Table 1)—and mobile phase pH (levels 6 and 7). These were selected to ensure adequate separation performance while preserving pigment stability, particularly for monovinyl chlorophylls.

Continuous variables were tested across defined ranges: gradient time (5–25 min), initial ethanol concentration (20–60%), column temperature (25–60 °C), flow rate (0.8–1.3 mL/min), and ion-pair volume added to sample prior to injection (0–40 µL of 28 mM TBA-OH).

An I-optimal design was selected to minimize the average prediction variance across the factor space, allowing robust model building while limiting the number of experimental runs. This design enabled the evaluation of key chromatographic performance metrics, including resolution between critical peak pairs (e.g., DV/MV chl *a* and chl *b*, lut/zea), peak symmetry, and pigment detectability (i.e., number of compounds effectively separated). The approach also contributed to a more resource-efficient method development process, aligned with the study's sustainability objectives. The experimental matrix and run randomization were generated using Design-Expert® software (see Supplementary Table S1).

Following optimization, the UHPLC method selected via DoE was evaluated in terms of performance and robustness using a pigment mix-standard and pigment extracts from pure algal cultures. Moreover, the new method was compared with a validated an existing validated method (Canuti, 2023). Natural seawater

samples were analysed using both methods. Statistical comparisons were performed to assess the method linearity, limit of detection, precision, and reproducibility of the methods across pigment classes.

3. Results and Discussion

3.1. Criteria for conditions used in the experimental design

The tested columns differed in bonding chemistry (monomeric C18-phenyl vs. core-shell C8), particle size, pore size, and surface area—all factors that collectively influence retention strength and separation efficiency. Columns were selected based on demonstrated or expected performance in separating critical pigment pairs, focusing on two chemistries: C8 core-shell and C18-phenyl phases. Core-shell columns feature a solid core surrounded by a porous shell, where only the outer layer provides active sites for analyte interaction. This design reduces analyte diffusion paths and travel distances, producing sharper, narrower peaks with improved resolution and sensitivity. Their enhanced efficiency and lower backpressure enable faster analysis times while maintaining high separation performance—comparable to sub-2 µm fully porous columns, but with reduced solvent consumption. C18-phenyl columns, on the other hand, were expected to perform well particularly with ethanol-based mobile phases and in separating the critical co-eluting compounds (Sanz et al., 2015). In typical marine pigment analyses, injection volumes range from 100 to 150 µL (Wright et al., 1997, Zapata et al., 2000, van Heukelem et al., 2001, Mendes et al., 2008, Jarayama et al., 2011, Sanz et al., 2015). However, one of the objectives of this method was to minimize sample consumption and reduce injection time. To this end, the injection volume was reduced to 40 µL. This reduction not only conserves the sample but also contributes to better peak shape and faster injections, supporting higher-throughput workflows Ethanol was selected as the organic phase in a UHPLC system capable of withstanding higher pressures. Being a stronger elution solvent than methanol (Snyeder et al., 1978), with twice the viscosity of methanol, ethanol allows the use of an alternative and less harmful mobile phase composition (Sadek et al., 2002; Plotka et al., 2013). While the method followed a DoE approach, which precluded full optimization for each chromatographic column individually, the main goal was to systematically evaluate how key parameters affect pigment separation, while minimizing the number of runs and enabling subsequent fine-tuning. The primary aim was to optimize pigment separation using a TBA-OH-based buffer while replacing methanol with ethanol as solvent A. Column temperature is known to influence retention and selectivity (α) in reversed-phase systems (van Heukelem et al., 2001; Snyder et al., 2009; Sanz et al., 2015). Prior studies showed that changes of 20 °C or more can significantly impact α and resolution (R_s), as α does not vary linearly with temperature (van Heukelem et al., 2001; Snyder et al., 2009). In particular, it has been reported that lower

temperatures (30–40 °C) have been reported to improve the resolution of certain pigment classes on specific columns (Roy et al., 2011). Based on these findings, temperature ranging from 25 to 60 °C were tested. Gradient elution times varied from 5 to 25 minutes, with a linear gradient consistently applied across all runs to isolate the effect of column performance and factor interactions from effects related to mobile phase composition. Solvent B consisted of 28 mM TBA-OH, adjusted to pH 6 or 7, while ethanol was used exclusively as solvent A to provide sufficient elution strength for even strongly retained analytes. Initial organic solvent content (20–60%) was also varied to assess its impact. Additional variables included the volume of buffer (0–40 µL) added to the sample prior to injection, which can influenced peak shape and reproducibility in ion-pairing systems, and flow rate (0.8–1.3 mL/min), which affected resolution and run time. Chromatographic performance was evaluated by examining the peak symmetry, the number of compounds separated and resolution of critical pigment pairs such as MV/DV chl *a* and chl *b*, zea/lut and prasinoxanthi/violaxanthin/19'-Hexanoyloxyfucoxanthin (pras/viol/hex)). A total of 63 chromatographic runs were performed across the three columns (see Supplementary Table S1). These experiments revealed insight into how critical method parameters, such as column temperature, gradient time, buffer addition, and initial organic content, interact with column chemistry and pH to affect the efficiency of pigment separation.

Table 2. Gradient profile of the UHPLC used for the separation of chlorophylls and accessory pigments in microalgal cultures and seawater sample extracts (flow rate: 1.3 mL/min).

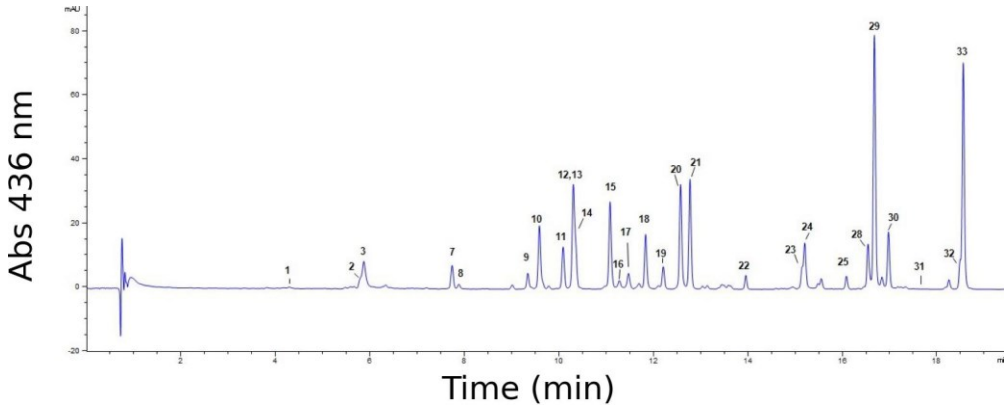
Time (min)	% Solvent A (Ethanol)	% Solvent B (Water Phase: 28 mM TBA-OH)
0	40	60
22.5	100	0
25.5	100	0
27	40	60

3.2. DoE results and method of choice

The results of the ANOVA from the DoE revealed that the most influential factor affecting the separation of MV/DV chl *a* and chl *b* was the type of column used. For the separation of zea and lut, the initial percentage of solvent B (28 mM TBA-OH) was identified as the primary influencing factor, while the column type also had a moderate but statistically significant effect ($p = 0.0035$ Three factors were found

to be significant regarding the total number of compounds separated: the elution gradient time, the initial percentage of solvent A and the flow rate. The latter showed a moderate yet statistically significant influence ($p = 0.0029$). In contrast, peak symmetry did not exhibit any significant dependence on the tested experimental factors. Among the tested columns, the C18-AR column provided the best separation for the α,β- and β,β-carotene isomers as well as for the pras-viol-hex group. However, priority was given to the separation of MV/DV chlorophylls, which are rarely resolved in C18-phase columns under the short chromatographic run times tested in this study. Additional methodological considerations were made to improve pigment separation and sample compatibility. A comparison of pigment extraction in ethanol versus acetone was conducted using microalgal strans. Previous extraction test in the identical condition and analysed using the Canuti (2023) method, showed that comparable results in terms of pigment quantification can be achieved (Kratzer et al., 2022). In the present case, , injecting the extracts in acetone leads to a low front tailing effect in the early elueters (i.e., chl *c*3 and chl *c*2) whereas this effect is not observed when the ethanol extracts were injected in a more compatible solvent mobile phase (i.e., ethanol and buffer). Overall, the ethanol extracts showed improved chromatographic behavior and higher reproducibility (see section 3.4.).

Figure 1. UHPLC chromatogram of the DHI mix-128 pigment mixture for DoE the selected method.



The method selected based on the DoE analysis employed a Kinetex C8 column (100 × 4.6 mm, 2.6 µm particle size) at 60 °C. The flow rate was set to 1.3 mL/min, and a linear ethanol–solvent B gradient was applied over 22.5 minutes, followed by a 3-minute hold at 100% ethanol and a 1.5-minute re-equilibration phase (Table 2). The initial proportion of solvent A was 40%. A 40 µL sample was injected and mixed with an additional 40 µL of buffer. Both solvent A and the buffer used for sample dilution were adjusted to pH 7. Compared to the Van Heukelem method, as implemented in Canuti (2023), this approach reduced

solvent consumption by 25% and shortened the total analysis time (including column equilibration and injection) from 40 minutes to 25.5 minutes. The method successfully resolved 33 pigments (Table 3 and Figure 1).

Table 3. Phytoplankton pigments and their identification sources, retention time (t_r), resolution (R_s) and maximum wavelength (λ_{max}) values as determined with the UHPLC. If not indicated, the resolutions >1.5

Peak no.	Pigment name	SCOR WG 78	Source*	t_R (min)	R_s	Absorption max (nm)			Quantified (nm)
1	Chlorophyll *c*3	Chl *c*3	DHI-mix, DHI-std, M	4.26		454	592	638	436
2	Chlorophyll *c*2	Chl *c*2	DHI-mix, DHI-std, A- I, M	5.82		446	582	634	436
3	Mg 2,4 divinyl pheoporphyrin a_5 monomethyl ester	Mg-DVP	DHI-mix, DHI-stdI,	5.86	0.41 (2/3)	442	632		436
4	Chlorophyll *c*1	Chl *c*1	DHI-std, B, E, I	6.02		442	581	634	436
5	Chlorophyllide *a*	Chlide *a*	DHI-mix, DHI std, I	6.32		434	468		665
6	Pheophorbide *a*	Pheo a	DHI-std, I	6.47		412	668		665
7	Peridinin	Peri	DHI-mix, DHI-std, A, C	7.61		480			436
8	Peridinin isomer	Peri iso	DHI-mix, A, , D, F, H	7.70	1.45 (7/8)	484			436
9	19'-Butanoyloxyfucoxanthi	But	DHI-mix, DHI-std	9.37		450			436
10	Fucoxanthin	Fuco	DHI-mix, DHI-std, B, C, D, E, G, H, I, M	9.53		456			436
11	Neoxanthin	Neo	DHI-mix, DHI-std, I	10.04		438	468		436
12	Prasinoxanthin	Pras	DHI-mix, DHI-std	10.25	co-elute (Pras,Viol)	460			436
13	Violaxanthin	Viol	DHI-mix, DHI-std, E	10.26	co-elute (Pras, Viol)	442	472	458	436
14	19'-Hexanoyloxyfucoxanthin	Hex	DHI-mix, DHI-std	10.32	0.47 (13/14)	450	470	466	436
15	Diadinoxanthin	Diad	DHI-mix, DHI-std, A, B, C , D, F, G, H, M	11.05		448	478	464	436
16	Dinoxanthin	Dino	DHI-mix, DHI-std, A, D, F	11.26		444	472	460	436
17	Antheraxanthin	Anth	DHI-mix, DHI-std	11.40		448	464	476	436
a	Unknown (myxo-xantholike spectra)		I	11.50					436
18	Alloxanthin	Allo	DHI-mix, DHI-std,I	11.81		454		482	436
19	Diatoxanthin	Diato	DHI-mix, DHI-std, C, F, G, I	12.18		454	480	472	436
20	Zeaxanthin	Zea	DHI-mix, DHI-std, E, L	12.53		458	480	470	436
21	Lutein	Lut	DHI-mix, DHI-std	12.74		446		474	436
22	Gyroxanthin	Gyro	DHI-mix, DHI-std	13.95		448		472	436
23	DV Chlorophyll *b*	DVhl *b*	DHI-mix, DHI-std	15.18		474	652		436
24	Chlorophyll *b*	Chl *b*	DHI-mix, DHI-std	15.20	0.58 (23/24)	464	650		436
25	Crocoxanthin	Croco	DHI-mix, DHI-std	16.06		448	474		436

26	Chlorophyll c2-monogalactosyldiacylglyceride	Chl-*c*2-MGDG	DHI-mix, DHI-std, G	16.10		460	590	638	436
27	Echinenone	Echi	DHI-mix, DHI-std	16.12		466			436
28	DV Chlorophyll *a*	DV chl *a*	DHI-mix, DHI-std	16.51	1.48 (27/28)	442	622	666	665
29	Chlorophyll *a*	Chl *a*	DHI-mix, DHI-std, A-M	16.65		416	432	666	665
30	Chlorophyll *a* epimer	Chl *a*'	DHI-mix, DHI-std, A, C, D, F, G, H, M	16.98	>1.5	416	432	666	665
31	Pheophytin *a*	Phy *a*	DHI-mix, DHI-std	17.27		410	664		665
32	β,ε-carotene	β,ε-car	DHI-mix, DHI-std	18.50		446	474	462	436
33	β,β-carotene	β,β-car	DHI-mix, DHI-std, A-M	18.53	1.02 (32/33)	452	478	470	436

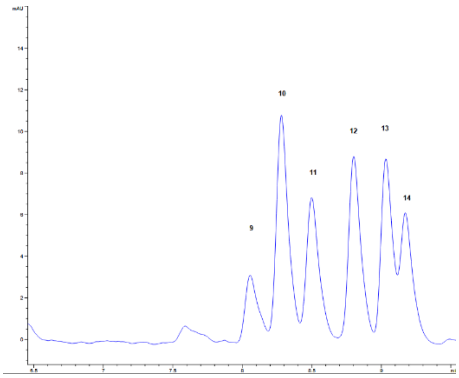
* DHI-Mix: DHI standard of mixed phytoplankton pigments, DHI-Std: DHI quantitative standard; A: *A. minutum* (CNR-AMI V1); B: *Chaetoceros* sp. (CBA 13); C: *C. fusiformis* (CBA 82); D: *G. impudicum* (GY6V); E: *H. akashiwo* (HA2V); F: *S. trochoidea* (CBA 2- F500); G: *P. gyrans* (CCMP 608); H: *P. micans* (CBA 21); I: *S. marinoi* (CBA 3D); N: *Synechococcus* (PCC 6911); M: *Navicula* sp. (CBA 2).

3.3. Pigment mixture and standards

The use of pigment standard mixtures was essential for benchmarking method performance and identifying resolution limitations before applying method to natural samples. These mixtures provided a controlled system, free from interference caused by unknown or minor peaks. This made them especially effective during the initial stages of method development. The pigments were selected based on their ecological importance, particularly with regards to association with major algal classes in diverse environments. Chromatograms obtained from these mixtures allowed to understand the separation performance for several key pigment pairs thus providing a solid foundation for the accurate quantification and identification of pigments in complex phytoplankton assemblages (Van Heukelem et al., 2001, Sanz et al., 2015, Suzuki et al., 2015). Mixtures of 28 pigments (i.e., mix-128 from DHI) were used to evaluate and refine the conditions for the optimal separation of taxonomically relevant pigments. Using this method, good resolution was achieved zea and lut, as well as MV and DV chl *a*. In addition, several other challenging and taxonomically important pigment pairs were successfully resolved, including phaeophorbide *a* and peridinin ($R_s > 1.5$), 19'-butanoyloxyfucoxanthin ($R_s > 1.5$), and fucoxanthin and 9'-cis-neoxanthin ($R_s > 1.5$). For chlorophyll *c*1 and *c*2, which were not both present in the mix-128 lot, a mixture of the two standards was used for checking the separation. However, viol and pras, MV/DV chl *b*, and α,β-and β,β-carotene only displayed only partial resolution under the tested gradient conditions, indicating some degree of co-elution. . Further optimization is required for these pairs, such as fine adjustments in gradient slope (i.e. longer analytical time) and adjustments to the ion-pairing solution to achieve full baseline

separation. Targeted tests were conducted using extended run times and three buffer molarity (56 mM, 112 mM, and 224 mM TBA-OH) to enhance the separation of overlapping pigments, particularly pras, viol, and hex which partially co-eluting in the method of choice. These trials aimed to optimize the resolution of these critical pigment groups, especially in complex phytoplankton matrices. It was observed that using 56 mM TBA-OH at pH 7 improved the resolution between viol, pras, and hex from co-elution (viol and pras) and a resolution of 0.47 (for hex) to resolution of 1.3 (viol/pras) and 0.88 (pras/hex), respectively (Figure 2). Further increases in the ion-pairing concentration did not result in improved separation of viol-pras-hex, and decreased the separation of the acidic chlorophylls (data not showed here). In addition, the increased molarity of the ion-pairing caused a broadening in the peak shape. Increasing the time gradient to 33.5 min only resolved partially resolved the hex co-elution. A longer gradient time was not an option as longer chromatographic times are outside of the scope of the present study.

Figure 2. Improved separation of co-elutes by increasing the molarity of TBA-OH to 56 mM.



The pigment concentrations obtained using the new UHPLC method for the pigment mix-128 were compared with the mean values reported by the quality reference group participating in the international inter-comparison exercise on phytoplankton pigments (HIP-8; Canuti et al., 2025). Relative percent differences (RPD%) between the two datasets were calculated. For chl *a*, the RPD was only 7.9%, while the average deviation across all other pigments was approximately 25%. The highest deviations were observed for chl *b*, hex, and peridinin (peri), each exceeding 35%. For the remaining pigments, the deviations remained below 25%. These results further support the suitability of the method as a candidate reference method for phytoplankton pigment analysis. (Table 4).

Table 4. Comparison of pigment concentrations (mg/L) in the standard mix-128 between the HIP-8 inter-comparison quality reference group (Canuti et al., 2025) and the new UHPLC method expressed as relative percent difference (RPD%).

	mix-128 [mg/L]		
	HIP-8 inter-comparison quality reference group	**new UHPLC method**	**RPD%**
Chl *a*	2.912	3.153	7.9
Chl *b*	1.113	0.779	-35.3
Chl *c*	0.217	0.175	-21.5
Caro	0.876	0.677	-25.6
But	0.083	0.070	-17.1
Hex	0.128	0.192	39.7
Allo	0.178	0.205	13.9
Diad	0.283	0.342	18.9
Diato	0.069	0.075	8.0
Fuco	0.292	0.349	17.7
Peri	0.273	0.404	38.6
Zea	0.436	0.395	-9.8

3.4. Method Evaluation for phytoplankton pigment analysis

To assess the analytical performance of the UHPLC method in resolving taxonomically relevant pigments, species-specific strain cultures of phytoplankton representing key functional groups were (Fig. 3). The selected strains were chosen to reflect the taxa typically found in the Adriatic Sea, where the seawater samples used for method validation were collected. These algal strains included diatoms and dinoflagellates, as well as representative strains of cyanobacteria (*Synechococcus*), haptophytes (*P. gyrans*), and raphidophytes (*H. akashiwo*). These strains are relevant for pigment-based chemotaxonomic and ecological applications (Fig. 3). All diatom strains (*Navicula* spp., *Chaetoceros* spp, *S. marinoi*, and *C. fusiformis*) exhibited prominent fucoxanthin peaks, confirming the method's ability to identify this primary carotenoid marker with high specificity. Additionally, Chls *c1* and *c2* were clearly resolved in both

Chaetoceros spp. and *S. marinoi*, demonstrating the method's efficiency in separating closely eluting chlorophyll derivatives. Chlorophyllide *a* and pheophorbide *a*, were detected in *S. marinoi* and in *Chaetoceros* spp*.,* both in ethanol and acetone extract. Among the dinoflagellates, *P. micans*, *S. trochoidea*, *G. impudicum*, and *A. minutum* presented peri, which was clearly resolved and exhibited excellent peak symmetry, even at lower concentrations. These findings provide further support for the suitability of the method for detecting group-specific pigments across a range of retention times. Both the haptophyte *P. gyrans* and the raphidophyte *H. akashiwo* both contained chls *c*1 and *c*2, with a high level of resolution between these components. *Navicula* alone presented chl *c*3. A minor peak, which was attributed to chlorophyll *c*2-mgdvp was uniquely observed in *Pavlova*, highlighting the method's sensitivity to lesser-known accessory pigments. Zeaxanthin was identified in both *Heterosigma* and *Synechococcus*, consistent with their known pigment composition. As expected, *Synechococcus* lacked chlorophyll *c* pigments, confirming its cyanobacterial lineage. These results were consistent with previous analyses on these algal cultures performed with different analytical methods (Wright et al., 1991, Van Heukelem et al., 2001, Zapata et al., 2004, Sanz et al, 2015, Nogeira et al., 2022, Wang et al., 2022, Bérard et al., 2024), demonstrating the robustness and taxonomic resolution offered by the UHPLC method. The clear separation of major and minor pigment components across phylogenetically diverse cultures supports the use of this method for identifying functional groups and species in complex natural phytoplankton assemblages. The choice of solvent for chlorophyll extraction—such as acetone, ethanol, or methanol—has long been debated (Jeffrey et al., 1997). In marine research, 90% acetone has traditionally been used (Jeffrey et al., 1997; Roy et al., 2011), while 96% ethanol is recommended by HELCOM (2001). For freshwater samples, 90% ethanol is the prescribed solvent according to DIN 38412 (1985) and ISO 10260 (1992). Wasmund et al. (2006) observed that generally, chl *a* concentrations tend to be lower in acetone extracts compared to ethanol, particularly in *Phaeodactylum* and *Microcystis*. Overall, the extraction efficiency of the various solvents seems to depend on factors, such as the taxonomic composition of the algal community (Wasmund et al., 2006).

To evaluate the reproducibility of pigment extraction from cultures, chl *a* concentration was performed on duplicate samples extracted using ethanol and acetone. For the *C fusiformis,* the relative standard deviation, RSD% (also known as coefficient of variation, CV%) for the chl *a* peak area was 3.5% for ethanol extracts and 8.7% for acetone extracts. Similar results were obtained for the chl *a* quantified in the *Pavlova* strain, having 3.6% for ethanol and 7.4% for the acetone. When considering all four replicates (two for each solvent), the overall inter-replicate variation was 14.6%, with values of 7.1% and 8.7% for the *C*

fusiformis and the *P. gyrans* respectively. This indicates that there are no systematic differences between the extraction methods. Ethanol yielded slightly more consistent results, and notably, did not lead to any degradation in reproducibility.

Table 5. Pigment quantification reproducibility in algal cultures using ethanol (Eth-OH) and acetone (Ac) as extraction solvents. Mean variation between replicates (expressed as RSD%) are reported for (A) chl *a* and (B) for selected taxon-specific marker pigments: fuco (diatoms), peri (dinoflagellates), and c2-mgdvp, (haptophytes). For each pigment, values obtained with ethanol, acetone, and across all extracts are shown. The difference in extraction efficiency between Eth-OH and Ac was expressed as Peak Area Relative Difference [%].

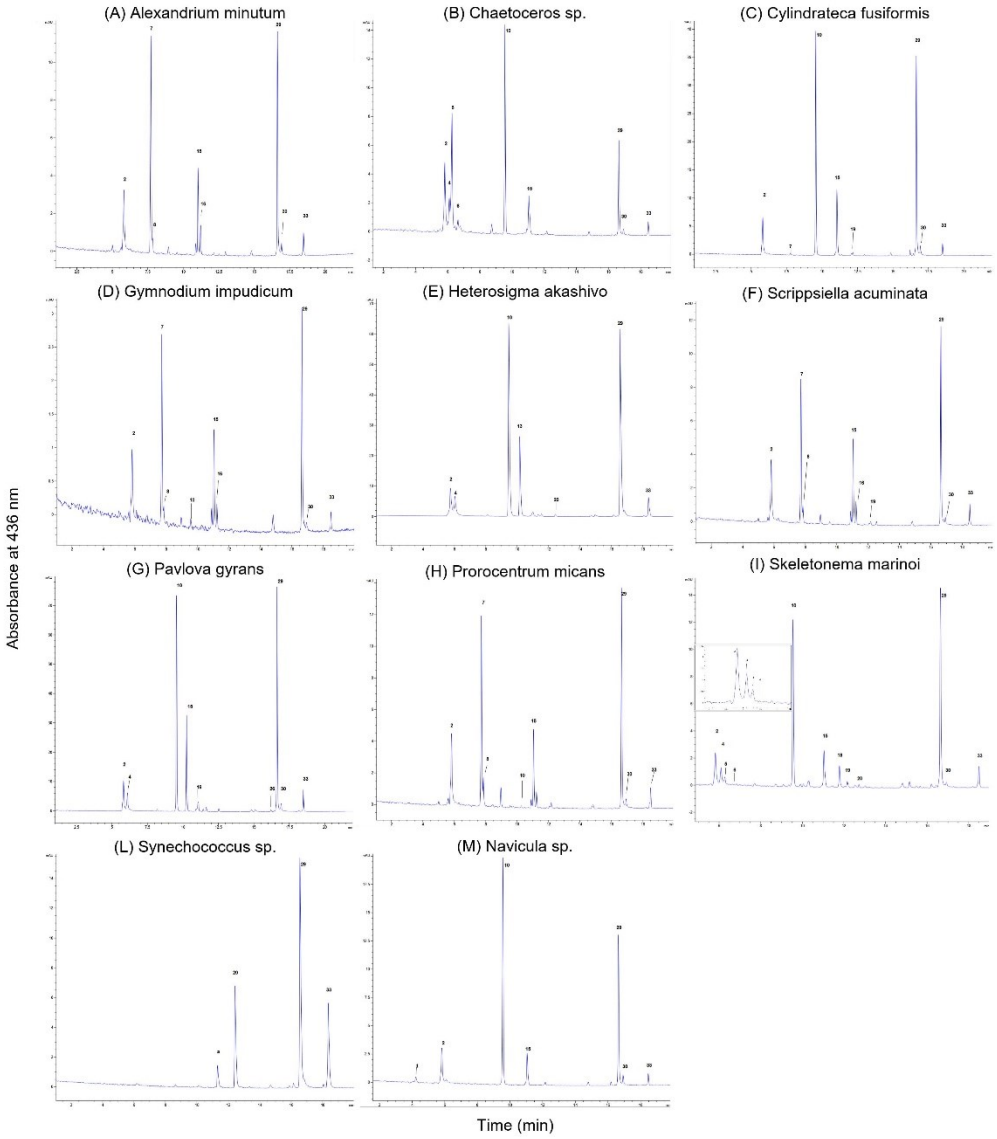
(A) Solvent / Metric	*Navicula*	*Charateroceros*	*Cylindrotheca fusiformis*	*Prorocentrum micans*	*Scrippsiella trochoidea*	*Gymnodinium impudicum*	*Alexandrium minutum*	*Pavlova gyrans*
RSD% Eth-OH	1.3	7.8	3.4	6.8	1.9	6.7	10.2	3.7
RSD% Ac	11.3	8.4	8.6	8.4	9.8	9.7	2.2	7.4
RSD% All	6.7	18.4	14	6.9	6.3	6.3	6.3	7.5
Peak Area RD [%]*	3	62.6	20.4	-4	4	12.6	12.4	-9.2

(B) Solvent / Metric	*Navicula*	*Charateroceros*	*Cylindrotheca fusiformis*	*Prorocentrum micans*	*Scrippsiella trochoidea*	*Gymnodinium impudicum*	*Alexandrium minutum*	*Pavlova gyrans*
Main Pigment	Fuco	Fuco	Fuco	Peri	Peri	Peri	Peri	c2-mgDVP
RSD% Eth-OH	5	4.8	4.3	3.6	3.9	4.9	10.1	4.6
RSD% Ac	11.2	9.3	9.1	4.3	8.6	8.8	2.9	7.8
RSD% All	10.3	26.8	18.8	8.2	14	14	14	6.8
Peak Area RD [%]*	12.6	37	26.9	12.2	20	30.3	21.2	-7.6

*Peak Area Relative Difference [%] = (ethanol − acetone) / max(ethanol, acetone) × 100

It is important to note that pigment extractions from algal cultures is inherently variable and technically challenging. Therefore, these values provide approximate indication of performance rather than an absolute benchmark (Canuti et al., 2025). Nevertheless, the observed variability remains within 10% for the chl *a* supporting the suitability of ethanol as an extraction solvent (Table 5).

Figure 3. Chromatograms of pigment extracts from selected microalgal cultures. UHPLC chromatograms recorded at 436 nm for 13 different microalgal strains, including diatoms (e.g., *Charatoceros* spp., *C. fusiformis*, *Navicula* spp., *S.marinoi*), dinoflagellates (e.g., *A. minutum*, *G. impudicum*, *P. micans*, *S. acuminata*), haptophytes (*P. gyrans*), raphidophytes (*H. akashiwo*), and cyanobacteria (*Synechococcus* sp.).



Furthermore, pigment-specific reproducibility was evaluated for key chemotaxonomic markers: fuco for diatoms, peri for dinoflagellates, and 19'-hexanoyloxy-neochlorophyll a (*c*2-mgdvp) for haptophytes. These pigments are essential for linking pigment profiles to taxonomic groups. As shown in Table 5, ethanol

extractions generally resulted in lower mean variation for these markers compared to acetone, supporting its suitability for accurate quantification across major phytoplankton groups. However, an exception was observed for *A. minutum*, where acetone yielded a lower RSD for chlorophyll *a* (2.2%) compared to ethanol (10.2%), suggesting that extraction efficiency may still vary by species and pigment composition. Moreover, for all the species considered, the extraction in ethanol yielded higher recovery of both chl *a* (from 3% in *Navicula* to 62% of *Chareoceros* spp.) and pigments, with the only exception of haptophytes, where the acetone was more efficient in extracting pigments (Table 5). These findings are consistent with the previous results reported (Wasmund et al., 2006, Kratzer et al. 2022), reinforcing which reinforce ethanol's applicability in the phytoplankton analysis for microalgal samples.

3.5. Method performance evaluation and calibration

Method performance was evaluated using two sets of replicate injections of a pigment standard (Mix-128): one undiluted set of six injections and one diluted set of five injections at ratio of 1:10. In the undiluted set, pigment concentrations ranged from 5.09 to 269.9 ng per injection, covering a range of retention times from 5.8 to 18.5 minutes. The overall variation in peak area quantification across all pigments was 12.03%, with lower variability observed for major pigments such as chl *a* (0.9%) and on the primary pigments of (7.8%). Retention time reproducibility across three separate sequences was also high, with a maximum variation of 0.6% observed for all pigments.

For the diluted standard (1:10), the analysis included 20 quantified pigments, covering a range of retention times from 5.8 to 18.5 minutes. The concentration in the injected samples spanned from 0.35 to 7.64 ng per injection. Retention time reproducibility remained high, with an average RSD% of 0.7% (range: 0.0–3.5%). Peak area reproducibility was more variable, with an average RSD% of 14.6% (range: 2.2–32.7%).

Calibration was performed using certified pigment standards (DHI Laboratory Products, Denmark) diluted at four concentration levels with the lowest level corresponding to approximately three times the signal-to-noise ratio (SNR). The limit of detection (LOD) for each pigment was determined based on the lowest quantifiable concentration from the standard dilutions. The average LOD across 22 pigments was 0.42 ± 0.26 ng/injection (mean ± standard deviation), corresponding to 0.012 ± 0.006 mg/L of standard. Calibration curves exhibited strong linearity, with coefficients of determination (R^2) greater than 0.99 for all 24 pigments, except for DVchl *b* (R^2 = 0.89). The response factors (RFs), expressed as slope values, varied between compounds, ranging from 1.22 for DVchl *b* to 9.37 for viol, reflecting intrinsic differences in

detector response and molar absorptivity. For chl *a*, a RF of 3.13 was observed, in agreement with values previously reported for the Van Heukelem method (Van Heukelem et al., 2001) (Supplementary Figure S1). Residual analysis revealed generally symmetrical distributions around zero. Root mean square error (RMSE) values were low for most compounds (e.g., MV chl *b*: 0.69; chl *c1*: 0.73), indicating high precision. However, pigments such as MV chl *a* (RMSE: 9.75) and alloxanthin (allo; RMSE: 8.63) exhibited higher dispersion, particularly at lower concentration levels, suggesting increased variability at trace levels. These patterns were visually confirmed in the residual plots, where larger deviations were most evident at the low end of the concentration range (Supplementary Figure S1). Overall, the method proved robust and reproducible for most of the pigments, with negligible bias and tight residuals, supporting its applicability for quantitative pigment profiling in environmental samples

Table 6. Statistical comparison of pigment concentrations measured by UHPLC and HPLC in natural phytoplankton samples from the Adriatic Sea.

compound	slope	intercept	R^2	Bias (µg/L)	RPD%	APD%	min	max	mean	n
chl *c3*	0.24	0.05	0.84	0.03	-48.4	48.4	0.0511	0.0641	0.0571	4
chl *c2*	1.08	0.01	0.95	0.03	-14.9	41.5	0.0403	1.0482	0.2936	9
Peri	1.4	0.06	0.98	0.08	-73.8	75.1	0.063	0.2077	0.1391	6
Fuco	0.89	0.02	1	-0.03	-1.2	9.9	0.0562	1.024	0.3667	10
Hex	0.77	0.03	0.84	0.01	26.3	59.4	0.0239	0.0527	0.0635	5
Diad	1.04	0.01	0.98	0.01	-19.6	19.6	0.0373	0.278	0.1036	10
Allo	0.78	0.01	1	0	3.2	9.6	0.0155	0.0583	0.0322	6
Diato	0.61	0.01	0.84	0	-11.6	12.7	0.0132	0.0152	0.0197	5
Zea	0.79	0	1	-0.03	25.4	25.4	0.0311	0.0311	0.0945	4
chl *b*	0.42	0.06	0.43	0.03	-24.3	32.2	0.0305	0.1478	0.0869	6
β,β-Caro	0.8	0.01	0.97	0	4.1	14.2	0.0191	0.0945	0.0457	9
chlide *a*	0.61	0.05	0.96	-0.04	28.4	28.4	0.0798	0.415	0.2026	4
chl *a*	0.87	-0.02	1	-0.17	15.5	15.5	0.2618	3.1168	0.9728	10

3.6. Intercomparison of UHPLC and HPLC Methods for Pigment Quantification

The agreement between UHPLC and HPLC methods for pigment quantification in 10 natural phytoplankton samples collected from the Adriatic Sea was assessed using both linear regression and Bland–Altman analyses. Only pigments for which at least four paired measurements were available from both laboratories were considered. The statistical results, including slope, intercept, R^2, mean bias, RPD%, and absolute percent differences (APD%), are summarized in Table 6.

These metrics were calculated for each compound, accounting for the range of concentrations observed, and further visualized through Bland–Altman plots (Supplementary Figures S2), to evaluate consistency and potential systematic differences across concentration levels. Most pigments exhibited strong linear relationships between methods, with R^2 values exceeding 0.95 for key chlorophylls and carotenoids such as chl *a* (R^2 = 1), chl *c2* (R^2 = 0.95), Peri (R^2 = 0.98), and Diadinoxanthin (diad, R^2 = 0.98). In these cases, the regression slopes were close to unity, and the bias values remained low (≤ ±0.08 µg/L), indicating a high degree of agreement. For example, Fucoxanthin (fuco) showed perfect correlation (R^2 = 1.00) with minimal bias (−0.03 µg/L), although its Bland–Altman plot revealed increasing dispersion at higher concentrations, indicating the presence of proportional error not captured by linear regression alone (Supplementary figure S1). Chl *a* showed an excellent correlation (R^2 = 1.00), but a consistent negative bias (−0.17 µg/L) and increasing variability at higher concentrations, as seen in its Bland–Altman plot. This suggests a systematic underestimation by UHPLC relative to HPLC, potentially due to matrix effects, pigment degradation, or differences in detector response. Nevertheless, the APD% for chlorophyll a was 15.5%, which is within the acceptable threshold for satellite validation applications, as suggested by Hooker et al. (2000). Therefore, despite the observed underestimation, UHPLC quantification of chl *a* remains suitable for satellite data product validation purposes. For other pigments, if APD% is less than 25%, the agreement can be considered acceptable for the development of bio-optical algorithms. This applies to several compounds such as allo (APD% = 9.6%), fuco (9.9%), diad (19.6%), and β,β-Carotene (β,β-caro, 14.2%), all of which showed negligible biases and high correlation values, confirming strong inter-method agreement. Peri displayed a slope of 1.40 and R^2 = 0.98, but with a positive bias (+0.08 µg/L) and high RPD% (75.1%), indicating a systematic overestimation by UHPLC, further corroborated by its Bland–Altman plot. In contrast, Chl *b* had the lowest correlation (R^2 = 0.43) and a shallow slope (0.42), indicating poor agreement between methods at low concentrations, even though the APD% remained within acceptable limits (32.2%). Several pigments such as allo, diad, and β,β-caro exhibited negligible bias, high R^2 values (1.00, 0.84, and 0.97, respectively), and tight limits of agreement in their Bland–Altman plots, confirming

excellent agreement between the two analytical methods. Notably, despite a slightly reduced slope (0.89), fuco maintained high inter-method agreement, as supported by its R^2 = 1.00, low bias, and low RPD% (9.9%). This highlights the importance of considering multiple statistical indicators—not just correlation—to evaluate method performance. Conversely, the other hand, high RPD% values observed for chl *c3* (48.4%), hex (59.4%), and peri (75.1%) indicate greater variability dispersion and reduced consistency in quantification, despite acceptable the R^2 values being acceptable. These cases warrant compound-specific scrutiny if the methods are to be used interchangeably in sensitive applications such as algorithm calibration or long-term trend analysis, , these cases warrant compound-specific scrutiny. The Bland–Altman analyses (Supplementary Figure S2) support these findings. Pigments such as fuco, allo, diato, hex, and β,β-caro showed narrow 95% limits of agreement and small mean differences, confirming robust inter-method reproducibility. Moderate agreement was found for chl *a*, chl *b*, diad, and chl *c2*, with chl *a* in particular exhibiting increasing variance at higher concentrations, a known feature in pigment analyses due to light saturation or detector sensitivity at high levels.

4. Conclusion

This study presented a novel UHPLC method for the rapid, reproducible, and environmentally conscious analysis of phytoplankton pigments. The method achieves high-resolution separation of 33 key pigments, including taxonomically important carotenoids and the challenging monovinyl/divinyl (MV/DV) chlorophyll pairs, by replacing methanol with ethanol — a less toxic solvent with lower disposal costs — and employing a DoE approach for chromatographic optimization. Unlike other pigment analysis protocols that relied on ternary gradients (e.g., Wright et al., 1991; Jayaraman et al., 2011) and thus required specialized quaternary pumping systems, this method utilized a binary elution gradient compatible with standard high-pressure binary UHPLC systems. Notably, it enabled the simultaneous separation of both polar (e.g., chlorophyll *c* and protochlorophyllides) and non-polar (e.g., chlorophylls *a* and *b*) MV/DV pigments. Compared to the widely adopted Van Heukelem and Thomas (2001) method, the proposed approach reduced solvent consumption by 25% and shortened the total analysis time to 25 min, without compromising pigment resolution or quantification accuracy. Validation of the method using standards, pigment mixtures, algal cultures, and natural samples from the northern Adriatic Sea confirmed its robustness and applicability across complex aquatic matrices. Pigment-specific uncertainty analysis, based on absolute percentage difference, supported the method's suitability for satellite validation (e.g., for

chlorophyll *a*) and bio-optical algorithm development, as the uncertainty remained below 15%. While the method performed reliably for most pigments, specific attention was warranted for compounds such as chlorophyll *c*3, for which a proportional bias was observed and prasinoxanthin, violaxanthin and 19'-hexanoyloxyfucoxanthin, which were not fully resolved. However, increasing the concentration of the ion-pairing reagent improved the resolution of 19'-hexanoyloxyfucoxanthin and successfully resolved the co-elution between prasinoxanthin and violaxanthin. Nevertheless, the proposed UHPLC method offered a robust, sustainable, and scalable solution for phytoplankton pigment analysis. Its balance of efficiency, accuracy, and environmental safety made it a strong candidate for routine application in ecological monitoring, chemotaxonomy, and ocean color validation workflows.

Acknowledgment

The authors would like to acknowledge Dr. Andrea Zaccagnini for the fruitful discussion on DoE analysis and critical aspect of the analytical separation.

Funding

The study has been supported by the European Commission Directorate General Joint Research Centre (JRC) and the Copernicus Program.

Conflict of Interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Author Contributions

First authorship: Elisabetta Canuti.

EC contributed to this paper by writing original draft, conceptualization and investigation, data curation, formal analysis, methodology and validation, reviewing and editing the content. SC, SC and FR contributed in the investigation, AP in reviewing and editing the contentment.

References

Alvain, S., Le Quéré, C., Bopp, L., Racault, M.F., Beaugrand, G., Dessailly, D., Buitenhuis, E., (2013) Rapid climatic driven shifts of diatoms at high latitudes. Remote Sens. Env. 132, 195-201

Barlow, R. G., Cummings, D.G., Gibb, S.W. (1997) Improved resolution of mono-and divinyl chlorophylls a and b and zeaxanthin and lutein in phytoplankton extracts using reverse phase C-8 HPLC. Mar. Ecol. Prog. Ser. 161: 303-307. doi: 10.3354/meps161303

Behrenfeld, M.J., Randerson, McClain, C.R., Feldman, G.C., Los, S.O., C J Tucker, C.J., Falkowski, P.G., C B Field, C.B., Frouin, R., Esaias, W.E., Kolber, D.D., Pollack, N.H. (2001), Biospheric primary production during an ENSO transition, Science, 291, 2594–2597.

Bérard, J.-B., Ouk, T.-S., Sol, V., Lefoulon, L., Landolt, C., Grenier, K., Saad, N., Robert, E., Sibat, M., Schreiber, N., Guenin, S., & Picot, L. (2024). Phototoxicity of the Ethanolic Extract of Skeletonema marinoi for the Dermocosmetic Improvement of Acne. Marine Drugs, 22(8), 343. https://doi.org/10.3390/md22080343

Brewin, R.J.W., Sathyendranath, S., Hirata, T. Lavender, S.J., Barciela, R.M., Hardman-Mountford, N.J., (2010). A three-component model of phytoplankton size class for the Atlantic. Oc. Ecol. Model., 221 pp. 1472-1483

Canuti E., (2023) Phytoplankton pigment in situ measurements uncertainty evaluation: an HPLC interlaboratory comparison with a European-scale dataset. Front. Mar. Sci. 10:1197311. doi: 10.3389/fmars.2023.1197311

Canuti, E., Allerup, M., Bracher, A., Brotas, V., Capellacci, S., Carabelli, C., Casabianca, S., Develoouse, I., Dias, A., Dimier, C., Flander-Putrle, V., Laloux, M., Lami, A., Kenemer, C., Kumar, S., Peeken, I., Penna, A., Raman, M., Ras, J., Ricci, F., Salaün, J., Schluter, L., Suzuki, K., Thomas, C., Tirkey, A., Tracana, A., Uitz, J., Vijaya Krishna, A., Vlaemnick, B. and Zaccagnini, A. (2025) The 5th Workshop on High-Performance Liquid Chromatography (HPLC) Inter-comparison on Phytoplankton Pigments (HIP), Publications Office of the European Union, Luxembourg, https://data.europa.eu/doi/10.2760/1257493, JRC140829.

Demarcq, H., Reygondeau, G., Alvain, S., Vantrepotte, V., (2012) Monitoring marine phytoplankton seasonality from space, Remote Sens. Env., 117, 211-222, https://doi.org/10.1016/j.rse.2011.09.019.

DIN 38412, 1985, Deutsche Einheitsverfahren zur Wasser-, Abwasser- und Schlammuntersuchung; Testverfahren mit Wasserorganismen (Gruppe L) –Bestimmung des Chlorophyll-a-Gehaltes von Oberfl¨achenwasser (L 16), 16. Liefer. 1986, Deutsch. Inst. Normung, Berlin, 7 pp

Esteban, R., Balaguer, L., Manrique, E., Rubio de Casas, R., Ochoa, R.P., Fleck, I., Pintó-Marijuan, M., Casals, I., Morales, D., Jiménez, M.S., Lorenzo, R., Artetxe, U., Becerril, J.M., García-Plazaola, J.I. (2009). Alternative methods for sampling and preservation of photosynthetic pigments and tocopherols in plant material from remote locations. Photosynthesis Research, 101, 77-88

Harding, J., Mallonee, M.E., Perry, E.S., Miller, W.D., Adolf, J.E., Gallegos, C.L., Paerl, H.W. (2016), Variable climatic conditions dominate recent phytoplankton dynamics in Chesapeake Bay Sci. Rep., 6 (1) p. 2377

HELCOM, 2001, Manual for marine monitoring in the COMBINE programme of HELCOM, Part C, http://sea.helcom.fi/Monas/CombineManual2/contents.html

Hooker, S.B., McClain, C.R. (2000). The calibration and validation of SeaWIFS data, Progress In Oceanography 45(3-4):427-465 doi:10.1016/S0079-6611(00)00012-4

Falkowski, P.G., Barber, R.T., Smetacek, V. V. (1998). Biogeochemical Controls and Feedbacks on Ocean Primary Production. Science. 281(5374):200-7. doi: 10.1126/science.281.5374.200. PMID: 9660741.

Gieskes, W.W., Kraay, G.W. (1986) Analysis of phytoplankton pigments by HPLC before, during and after mass occurrence of the microflagellate Corymbellus aureus during the spring bloom in the open northern North Sea in 1983. Marine Biology 92, 45–52

Goericke, R., Repeta, D.J. (1993) Chlorophylls a and b and divinyl chlorophylls a and b in the open subtropical North Atlantic Ocean. Mar. Ecol. Prog. Ser., 101, 307-313.

Kratzer, S., Harvey, E.T., Canuti, E. (2022) International Intercomparison of In Situ Chlorophyll-a Measurements for Data Quality Assurance of the Swedish Monitoring Program.Front. Remote Sens. 3:866712. doi: 10.3389/frsen.2022.866712

Kumar Sahu, P., Rao Ramisetti, N., Cecchi, T., Swain, S., Patro, C.S, Panda, J. (2018) An overview of experimental designs in HPLC method development and validation. J. Pharmaceutical and Biomedical Analysis 147 590–611

Jayaraman, S., Knuth, M.L., Cantwell, M., Santos, A. (2011) High performance liquid chromatographic analysis of phytoplankton pigments using a C16-Amide column. J. Chrom. A, 1218, 3432–3438.

Jeffrey, S.W, Mantoura R.F.C., Wright, S.W.(1997). Phytoplankton pigments in oceanography: guidelines to modern Methods. UNESCO publishing, Paris, 661 pp.
Joint Global Ocean Flux Study (1994): JGOFS Protocols for the Joint Global Ocean Flux Study Core Measurements. JGOFS Report No. 6, Scientific Committee on Oceanic Research, 170 pp.

Latasa, M., R. R. Bidigare, M. E. Ondrusek, and M. C. Kennicutt. (1996). HPLC analysis of algal pigments: A comparison exercise among laboratories and recommendations for improved analytical performance. Mar. Chem., 51, 315-324.

Mackey, M., Mackey, D. Higgins, H. W., Wright, S. W., 1996. CHEMTAX - a program for estimating class abundances from chemical markers: Application to HPLC measurements of phytoplankton. Mar. Ecol. Prog. Ser. 144: 265– 283

Mendes, C.R., Cartaxana, P, Brotas, V. (2007). HPLC determination of phytoplankton and microphytobenthos pigments: comparing resolution and sensitivity of a C 18 and a C 8 method Limnol. Oceanogr.: Methods 5, 2007, 363–370

Mueller, J. L.; Pietras, C.; Hooker, S. B.; Austin,R.W.; Miller, M.; Knobelspiesse, K. D.; Frouin, R. ; Holben, B. and Voss, K. (2003) Ocean Optics Protocols For Satellite Ocean Color Sensor Validation, Revision 5.

Volume V: Biogeochemical and Bio-Optical Measurements and Data Analysis Protocols. Greenbelt, MD. Goddard Space Flight Space Center. pp.1-43. (NASA/TM-2003-). DOI: http://dx.doi.org/10.25607/OBP-67

Nogueira, E., Bravo, I., Montero, P., Díaz-Tapia, P., Calvo, S., Ben-Gigirey, B., Figueroa, R. I., Garrido, J. L., Ramilo, I., Lluch, N., Rossignoli, A. E., Riobó, P., & Rodríguez, F. (2022). HABs in coastal upwelling systems: Insights from an exceptional red tide of the toxigenic dinoflagellate Alexandrium minutum. Ecological Indicators, 137, 108790. https://doi.org/10.1016/j.ecolind.2022.108790

Plotka, J., Tobiszewski, M., Sulej, A.M., Kupska, M., Gorecki, T. & Namiesnik, J. (2013) Green chromatography. J. Chrom. A,, 1307, 1–20.

Roy, S., Llewellyn, C., Egeland, E., Johnsen, G. (2011) (Eds.), Phytoplankton Pigments: Characterization, Chemotaxonomy and Applications in Oceanography (Cambridge Environmental Chemistry Series, pp. 314-342) Cambridge: Cambridge University Press. doi:10.1017/CBO9780511732263.01

Sadek, P.C. (2002) The HPLC Solvent Guide, 2nd edn. Wiley-Interscience, New York City, New York.

Sanz, N., Garcıa-Blanco, A., Gavalas-Olea, A., Loures, Garrido, P.J.l., (2015) Phytoplankton pigment biomarkers: HPLC separation using a pentafluorophenyloctadecyl silica column Methods in Ecology and Evolution, 6, 1199–1209

Snyder, L.R. (1978). Classification off the Solvent Properties of Common Liquids, Journal of Chromatographic Science, Vol, 16, Issue 6, 223–234, https://doi.org/10.1093/chromsci/16.6.223

Snyder, L. R. Dolan J. W. (2009). Introduction to Modern Liquid Chromatography, Third Edition, Wiley doi:10.1002/9780470508183

Suzuki, K., Kamimura, A., Hooker, S.B., (2015) Rapid and highly sensitive analysis of chlorophylls and carotenoids from marine phytoplankton using ultra-high performance liquid chromatography (UHPLC) with the first derivative spectrum chromatogram (FDSC) technique, Mar. Chem., 176, 96-109, https://doi.org/10.1016/j.marchem.2015.07.010.

Uitz, J., Claustre, H., Morel, A., Hooker, S.B., (2006). Vertical distribution of phytoplankton communities in open ocean: an assessment based on surface chlorophyll. J. Geophys. Res., 111 p. C08005 https://doi.org/10.1029/2005JC003207

Van Heukelem, L., C.S. Thomas, Computer-assisted high-performance liquid chromatography method development with applications to the isolation and analysis of phytoplankton pigments (2001). J. Chrom. A, 910, 31–49.

Vidussi, F., Claustre, H., Bustillos-Guzmàn, J., Cailliau, C. and J.-C. Marty (1996). Rapid HPLC method for determination of phytoplankton chemotaxinomic pigments: separation of chlorophyll a from divinylchlorophyll a and zeaxanthin from lutein. Journal of Plankton Research, 18, 2377-2382.

Zapata, M., Garrido, J.L. (1991). Influence of injection conditions in reversed-phase high-performance liquid chromatography of chlorophylls and carotenoids. Chromatographia 31:589-594.

Zapata, M., Rodriguez, F., Garrido, J.L. (2000) Separation of chlorophylls and carotenoids from marine phytoplankton: a new HPLC method using reverse phase C8 column and pyridine-containing mobile phase. Mar. Ecol. Prog. Ser., Vol. 195: 29-45.

Zapata, M. (2005). Recent advances in pigment analysis as applied to picophytoplankton. Vie et Milieu. 55. 233-248.

Zapata, M., Jeffrey, S.W., Wright, S., Rodríguez, F., Garrido, J., Clementson, L. (2004). Photosynthetic pigments in 37 species (65 strains) of Haptophyta. Marine Ecology-Progress Series. 270. 83-102. 10.3354/meps270083.

Wang, J., Kong, F., Wang, Y., Ji, N., Song, M., Hu, Z., Niu, Z., Liu, C., Wang, X., Sun, Y., Yu, R., & Yan, T. (2022). Newly recorded bloom-forming dinoflagellate Gymnodinium impudicum in Haizhou Bay, Yellow Sea, China. Journal of Oceanology and Limnology, 40(6), 2430–2445. https://doi.org/10.1007/s00343022-1402-0

Wasmund, N., Topp, I., Schories, D. (2006). Optimising the storage and extraction of chlorophyll samples. Oceanologia, 48(1), 125–144

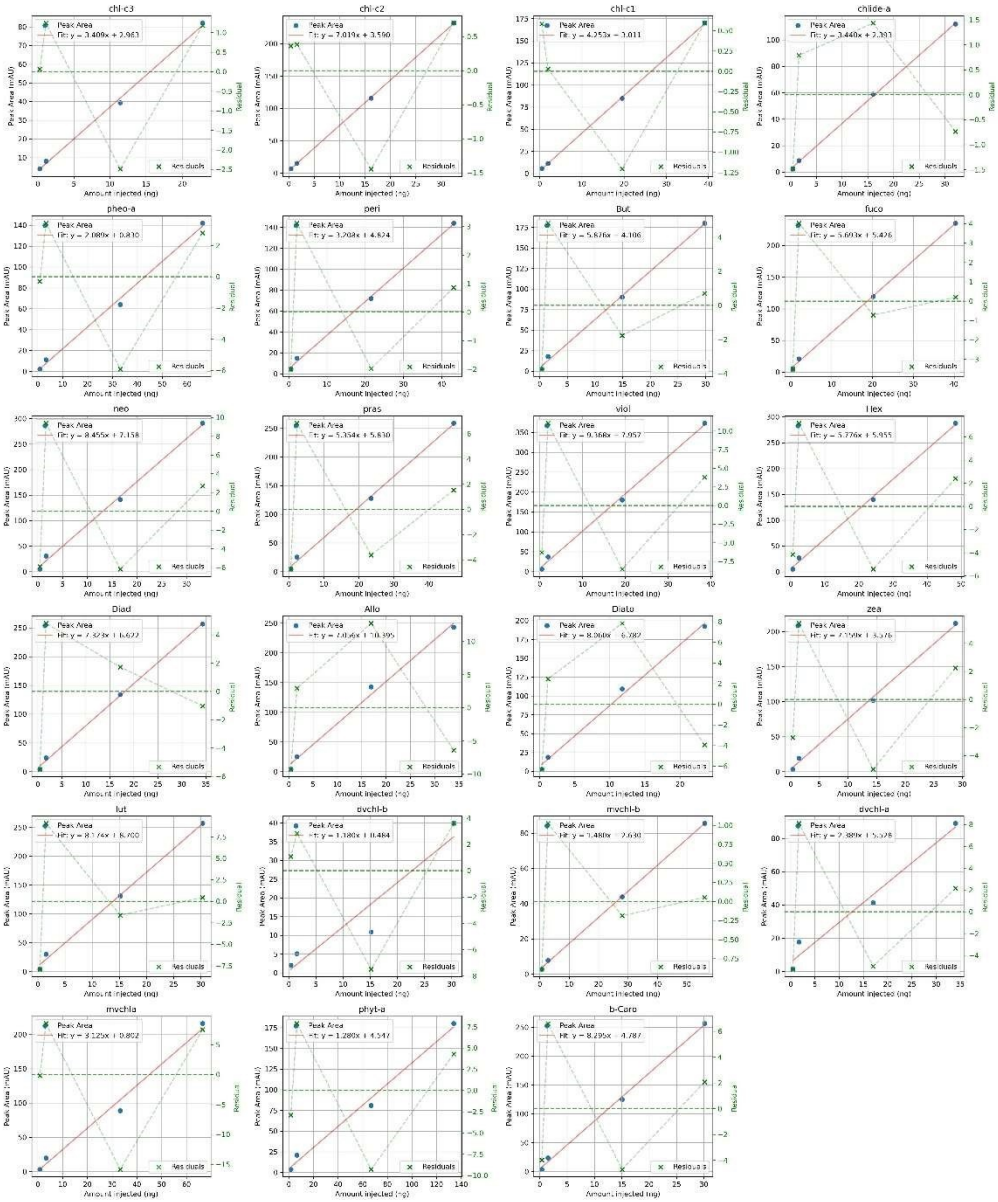
Wright, S. W., Jeffrey, S.W., Mantoura, R.F.C., Llewellyn, C.A., Bjornland, T., Repeta, D., Welschmeyer, N. (1991) Improved HPLC method for the analysis of chlorophylls and carotenoids from marine phytoplankton. Mar. Ecol. Prog. Ser., 77, 183-196.

Wright S.W., Jeffrey S.W. (2005). Pigment Markers for Phytoplankton Production. In: Volkman J.K. (eds) Marine Organic Matter: Biomarkers, Isotopes and DNA. The Handbook of Environmental Chemistry, vol 2N. Springer, Berlin, Heidelberg. https://doi.org/10.1007/698_2_003

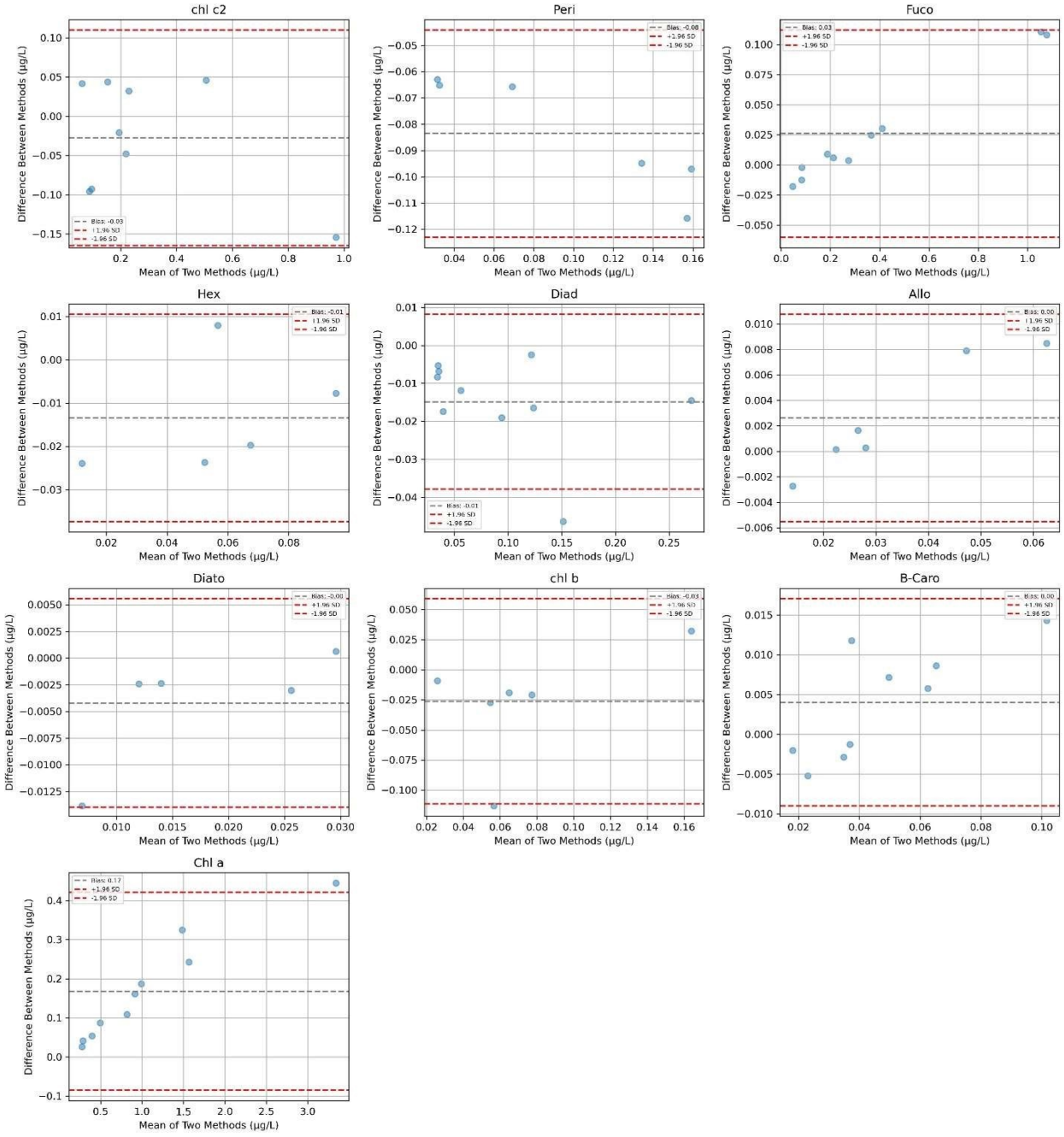
Supplementary table S1. DoE experiments and results

Run	definition	gradient_time	initial_A	T_column	flow	added_buffer	column	pH	DV/MV chla	DV/MV chlb	Zea/Lut	Pras/Viol/Hex	Compounds	Symmetry
1	Model	14.7	20	60.0	0.9	0.0	c8 core_150	7	0	0	0	0	12	0.92
9	Model	5	20	48.8	1.0	40.0	c8 core_150	6	1.2	0	0	0	21	1.05
12	Replicate	25	20	38.1	0.8	20.2	c8 core_150	6	0.8	0	0	0	18	1.32
13	Model	25	60	30.8	0.8	0.0	c8 core_150	7	0.82	0	1.61	0	41	1.26
18	Model	5	60	50.0	1.3	16.0	c8 core_150	7	0	0	0.79	0	22	1.03
19	Center	15	40	45.0	1.1	20.0	c8 core_150	7	0.9	0	1.38	0	29	0.96
22	Center	15	40	45.0	1.1	20.0	c8 core_150	6	1.41	0	1.07	0	30	1.20
24	Center	15	40	45.0	1.1	20.0	c8 core_150	7	1.1	0	1.4	0	30	1.02
26	Model	13	33	60.0	1.3	40.0	c8 core_150	6	1.22	0	1.43	0	24	1.01
36	Model	25	20	38.1	0.8	20.2	c8 core_150	6	0.89	0	0	0	19	1.53
45	Model	25	20	36.0	1.3	40.0	c8 core_150	7	0.94	0	0	0	21	0.99
48	Replicate	5	60	50.0	1.3	16.0	c8 core_150	7	0.79	0	0.77	0	22	1.00
49	Model	14.9	60	30.0	1.2	40.0	c8 core_150	6	0	0	1.47	0	37	1.11
53	Lack of Fit	7	44.2	30.2	0.8	0.6	c8 core_150	7	0	0	0.75	0	55	0.96
55	Model	5.5	31.2	30.0	0.8	35.8	c8 core_150	7	0	0	0.88	0	23	1.04
56	Model	5	60	54.8	0.8	0.0	c8 core_150	6	0.85	0	0	0	27	1.26
58	Model	25	54.8	60.0	0.8	39.0	c8 core_150	7	0.89	0	1.71	0	32	1.14
59	Model	25	56	60.0	1.3	0.0	c8 core_150	6	0	0	0	0	34	1.09
61	Center	15	40	45.0	1.1	20.0	c8 core_150	6	1.13	0	1.01	0	27	1.11
62	Model	5	27	30.0	1.3	0.0	c8 core_150	6	0	0	0	0	17	1.01
2	Lack of Fit	25	60	36.8	1.3	0.0	c8 core_100	7	1.28	0	2.49	0.8	33	1.25
3	Model	22	20	30.0	1.3	0.0	c8 core_100	7	0	0	0	0	8	0.62
4	Model	5	20	31.5	1.3	40.0	c8 core_100	7	0.78	0	0	0	21	1.02
5	Model	25	37.0697	53.1	0.8	0.0	c8 core_100	7	0.45	0	0	0	27	1.07
8	Model	5	20	60.0	0.9	28.0	c8 core_100	7	0.45	0	0	0	19	1.04
14	Model	25	38.6	37.2	1.3	40.0	c8 core_100	6	1.02	0	0	0	29	0.94
15	Model	5	20	32.1	0.8	0.0	c8 core_100	6	1.52	0.52	0	0	14	1.10
20	Replicate	25	38.6	37.2	1.3	40.0	c8 core_100	6	0.73	0	0	0	30	0.95
21	Model	25	20	30.0	0.9	40.0	c8 core_100	7	1.47	0.45	0	0	25	0.93
28	Model	6.6	60	41.1	0.8	40.0	c8 core_100	7	1.17	0	0	0	28	1.08
29	Model	24.9	60	30.0	0.8	11.0	c8 core_100	6	0	0	1.03	0	43	1.04
30	Model	5	60	30.0	1.1	0.0	c8 core_100	7	1.1	0	2.35	0	26	1.07
33	Model	5	29	60.0	1.3	0.0	c8 core_100	7	0	0	0.78	0	17	1.81
34	Model	16.5	35.4	60.0	0.8	40.0	c8 core_100	6	0	0	0	0	29	0.93
35	Model	5	60	60.0	1.3	24.0	c8 core_100	6	1.53	0.67	1.76	0	27	1.26
38	Model	22.5	60	60.0	1.3	40.0	c8 core_100	7	0	0	0	0	37	1.07
39	Center	15	40	45.0	1.1	20.0	c8 core_100	6	1.57	0.69	2.54	0.76	29	1.22
41	Center	15	40	45.0	1.1	20.0	c8 core_100	6	1.36	0.65	1.33	0	19	1.60
43	Center	15	40	45.0	1.1	20.0	c8 core_100	7	0	0	0	0	18	1.41
46	Replicate	25	37.0697	53.1	0.8	0.0	c8 core_100	7	0	0	0	0	27	1.07
50	Model	24.3	20	60.0	1.1	0.0	c8 core_100	6	0	0	0	0	24	1.00
63	Center	15	40	45.0	1.1	20.0	c8 core_100	7	1.37	0.65	1.58	0	27	0.96
6	Model	13.2	20	30.0	1.2	40.0	c18AR_150_3	6	0	0	0	0	10	0.81
7	Model	12	60	44.3	1.3	0.0	c18AR_150_3	6	0	0	0	0	26	1.09
10	Lack of Fit	16	20	30.0	0.8	14.4	c18AR_150_3	7	0	0	0	0	22	0.88
11	Center	15	40	45.0	1.1	20.0	c18AR_150_3	7	0	0	0.63	0	32	1.42
16	Model	19.9	20	47.0	0.8	40.0	c18AR_150_3	7	0	0	0.67	0	33	1.18
17	Model	25	35.4895	30.0	0.9	0.0	c18AR_150_3	6	0	0	0	0	30	1.06
23	Model	25	60	30.0	1.3	33.2	c18AR_150_3	7	0	0	0	0	39	1.13
25	Model	5	45.4	60.0	1.2	40.0	c18AR_150_3	7	0	0	0	0	20	0.84
27	Replicate	12	60	44.3	1.3	0.0	c18AR_150_3	6	0	0	0	0	29	1.09
31	Model	5	22	60.0	0.8	3.2	c18AR_150_3	6	0	0	0	0	15	1.27
32	Model	5	60	30.0	0.8	27.0	c18AR_150_3	6	0	0	0	0	22	1.12
37	Model	25	60	56.3	1.0	40.0	c18AR_150_3	6	0	0	0.94	0	39	1.12
40	Model	24.9	20	60.0	1.3	24.0	c18AR_150_3	6	0	0	0	0	24	1.00
42	Model	13.3	60	60.0	0.8	4.2	c18AR_150_3	7	0	0	0	0	34	1.23
44	Lack of Fit	23.2363	44.547	30.0	0.8	40.0	c18AR_150_3	6	0	0	0.67	1.4	31	1.19
47	Model	25	32.2	60.0	1.2	0.0	c18AR_150_3	7	0	0	0	0	28	1.56
51	Center	15	40	45.0	1.1	20.0	c18AR_150_3	6	0	0	0.65	0	29	1.39
52	Center	15	40	45.0	1.1	20.0	c18AR_150_3	6	0	0	0.65	0	23	0.96
54	Lack of Fit	5	20	50.7	1.3	15.6	c18AR_150_3	6	0	0	0	0	15	1.02
57	Model	5	20	32.0	1.1	0.0	c18AR_150_3	7	0	0	0	0	14	1.27
60	Center	15	40	45.0	1.1	20.0	c18AR_150_3	7	0	0	0.62	0	34	1.57

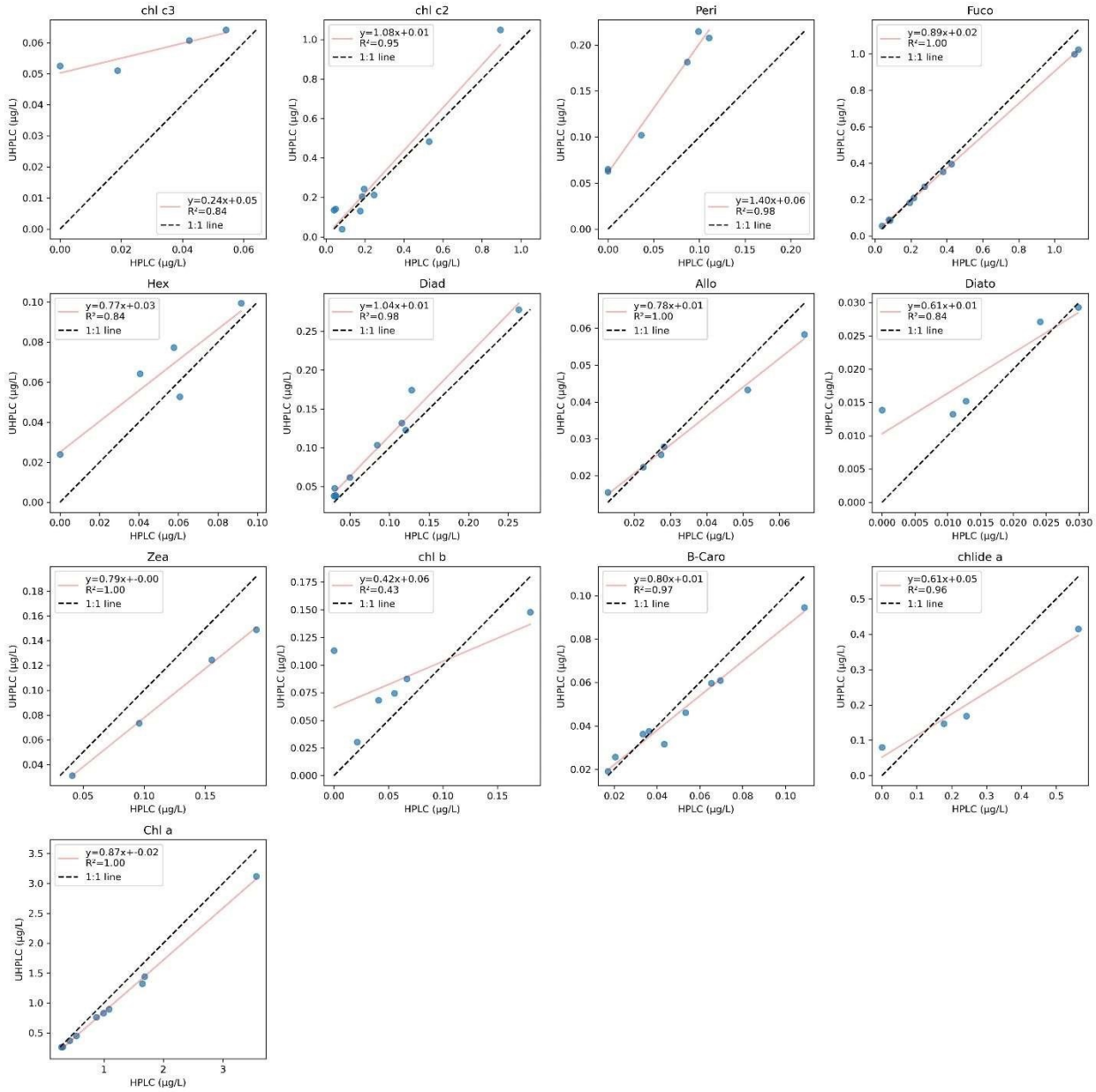
Supplementary Figure S1. Calibration curves and residual plots for 20 of the pigments quantified using UHPLC. Each subplot shows the linear regression fit of peak area versus injected pigment amount (blue circles), with red lines indicating the fitted model. The corresponding residual plots are in green crosses and dashed line. R^2 values and regression equations are reported for all calibration curves.



Supplementary Figure S2. Bland–Altman analysis of pigment concentration differences between UHPLC and HPLC methods.



Supplementary Figure S3. Regression plot (UHPLC vs HPLC concentrations) including line of best fit, 1:1 reference line, R^2, and regression statistics.



5. Chapter II

UHPLC-Based C-Phycocyanin Quantification Method to Support the Validation of Remote Sensing Models for Harmful Cyanobacterial Blooms

Elisabetta Canuti

This chapter has been published as E. Canuti (2025) UHPLC-based C-phycocyanin quantification method to support the validation of remote sensing models for harmful cyanobacterial blooms, *Journal of Chromatography B*, 1265, https://doi.org/10.1016/j.jchromb.2025.124759 .

Journal of Chromatography B 1265 (2025) 124759

Contents lists available at ScienceDirect

Journal of Chromatography B

journal homepage: www.elsevier.com/locate/jchromb

UHPLC-based C-phycocyanin quantification method to support the validation of remote sensing models for harmful cyanobacterial blooms

Elisabetta Canuti

European Commission, Joint Research Centre (JRC), Ispra, Italy

ARTICLEINFO

Keywords:
UHPLC analysis
Remote sensing
C-Phycocyanin
Cyanobacterial blooms Design of experiment

ABSTRACT

The deployment of hyperspectral satellite missions has opened new opportunities for integrated approaches to address the escalating issue of cyanobacterial algal blooms in marine and inland waters. Despite these advancements, the validation of satellite data concerning C-phycocyanin (C-PC) content requires robust analytical methods. Currently, the available techniques, predominantly spectrophotometric or fluorometric, exhibit low reproducibility due to the interference of chlorophyll *a* and perform optimally with water samples. The samples collected during ship-based campaigns and commonly used for satellite data validation are concentrated samples on filters, which limiting the application of these methods. In this study, we aim to establish an Ultra High Pressure Liquid Chromatography (UHPLC) -based method for C-PC quantification in water samples concentrated on filters to validate satellite derived algorithm for monitoring cyanobacteria blooms. This involves the extraction of C-PC from pure algal cultures (*Synechococcus* spp. and *Anabaena* spp.) and natural samples fixed on filters and the quantification employing UHPLC analysis—a recognized gold standard for phytoplankton pigment analysis. Based on a Design of Experiment (DoE) full factorial design approach, we compared different extraction techniques. Subsequently, we compared different methods and developed and validated a rapid, simple, and sensitive reverse phase UHPLC method. The optimized chromatographic parameters for C5 phase, were acetonitrile with trifluoroacetic acid (0.1 % *v*/v) in a 5′ linear gradient (from 20 % to 100 %), flowrate 0.8 mL min^{-1} with an accuracy of 10 %, a reproducibility of 9.8 % and a limit detection of 0.025 mg L^{-1}.

1. Introduction

In response to todays environmental challenges,has become urgent to address the increasing occurrence of cyanobacterial harmful algal blooms (HABs) in both marine and inland waters has become urgent. Traditional monitoring methods, which rely on discrete in situ observations are inadequate due to their lack the spatial and temporal resolution needed for effective, large-scale surveillance of phytoplankton communities [1]. To overcome these limitations, alternative approaches have been explored, including the development of advanced water quality models [2–4] and remote-sensing-based monitoring and forecasting systems [5–7]. While water quality models offer valuable insights into bloom dynamics, their predictive accuracy is often limited by the challenge of calibrating the complex interactions between physical transport and biological processes, which vary across ecosystems and timescales [3]. Recent advancements in hyperspectral satellite missions—such as Germany's EnMAP, Italy's PRISMA, and NASA's PACE—have opened up new opportunities for large-scale HAB monitoring from space [8]. While remote sensing holds promise for water quality assessment, its application in optically complex and high-biomass waters remains technically challenging [9–11]. HAB detection from satellite data is still hindered by inadequate atmospheric correction in optically complex and coastal waters, and by a limited understanding of how optical signals respond to changes in phytoplankton functional types (PFTs) [12]. Chlorophyll *a* (Chl *a*) has traditionally served as the primary satellite proxy for phytoplankton biomass. However, in waters dominated by cyanobacteria, Chl *a* fails to distinguish between cyanobacteria and eukaryotic algae containing chlorophyll. In contrast, C- phycocyanin (C-PC)—a water-soluble pigment unique to cyanobacteria—has distinctive absorption features in the 600–650 nm range. Integrating these features into satellite retrieval algorithms can enhance the accuracy of C-PC concentration estimates and improve HAB detection, especially in ecosystems with mixed phytoplankton communities or seasonal variability. However, the lack of concurrent datasets with both optical properties and pigment concentrations during bloom events continues to hinder the development of accurate pigment-based satellite algorithms [5]. C-PC quantification in support to remote sensing modelling is often based on in vivo optical properties, such as

E-mail address: Elisabetta.Canuti@ec.europa.eu.

https://doi.org/10.1016/j.jchromb.2025.124759 Received 20 December 2024; Received in revised form 25 July 2025; Accepted 8 August 2025 Available online 14 August 2025

fluorescence [5]. On the other hand, benchtop spectrophotometric and fluorometric methods are limited by interference from Chl *a*, low reproducibility, and optimization for liquid samples rather than the concentrated biomass typically collected on filters during oceanographic campaigns [13–15]. Chromatographic techniques, particularly High-Performance Liquid Chromatography (HPLC), offer a powerful solution due to their ability to separate overlapping spectral signals. HPLC is widely considered the gold standard for Chl *a* validation in satellite applications [16], with a generally accepted uncertainty margin of 15 % for refining bio-optical algorithms [17]. However, HPLC methods conventional used in support of remote sensing do not allow C-PC quantification from filter pad samples, creating a critical gap in pigment-based satellite validation workflows. Nonetheless, published HPLC methods for C-PC quantification in algae have mainly focused on macroalgae [18], and those applied to phytoplankton have only been tested on pure cultures, not natural samples fixed on filters [19–21]. The primary objective of this study was to address this methodological shortfall by developing a robust and reliable analytical protocol specifically suited for quantifying C-PC from cyanobacterial biomass collected on filter pads—a sample format widely used in field campaigns. We hypothesized that adequate recovery, precision, and accuracy of C-PC could be achieved from filtered biomass through a tailored combination of extraction and chromatographic separation procedures. To this end, we compared several cell disruption techniques and optimized an Ultra-High Performance Liquid Chromatography (UHPLC) method for C-PC quantification using a Design of Experiments (DoE) framework. A key contribution of this work is the development of a UHPLC-based protocol suitable for natural, field-collected samples, thereby bridging a critical gap between laboratory pigment analysis and the in situ data needs of remote sensing algorithm development. Additionally, UHPLC results were compared with conventional UV–Vis spectrophotometry to assess differences in quantification accuracy. Overall, this study provides methodological advancements and practical tools to support the validation of satellite- derived C-PC products, thereby enhancing the monitoring and management of cyanobacterial blooms in diverse aquatic ecosystems.

2. Materials and methods

2.1. Monocultures and natural samples

In this study, two pure algal strains representing different aquatic environments—marine, brackish, and freshwater—were used: the *Synechococcus PCC 6911* and *Anabaena PCC 7120*. The *Synechococcus PCC 6911* is a marine, unicellular cyanobacterium adapted to saline environments, whereas *Anabaena* PCC 7120 is a filamentous, nitrogen-fixing freshwater strain commonly found in nutrient-rich inland waters. The strains were cultured at the Laboratory of Marine Ecology, University "Carlo Bo" of Urbino, Italy, under dim light conditions and at a temperature of 22 °C (as in [22] Wojtasiewicz et al., 2016). The pure cultures were harvested during their exponential growth phase, and cell densities were measured at optical densities (OD) of 600 nm and 750 nm (see Section 2.2). Two series of samples replicates, A1 and S1, were prepared simulating concentrations relevant to natural sample conditions:

- S1 (*Synechococcus PCC 6911*): A dilution was prepared by adding 10 mL of culture into 1 L of Milli-Q (MQ) water. Subsequently, 25 mL of this diluted culture. A total of 24 replicates were prepared.
- A1 (*Anabaena PCC 7120*): A dilution was prepared by adding 5 mL of culture into 1 L of MQ water. From this mixture, 50 mL was filtered for each replicate, resulting in 18 samples.

The replicates were filtered under gentle vacuum pressure (≤ 0.5 bar) using a Sartorius filtration ramp system (Sartorius, Germany) onto GF/F filter (0.7 µm pore size, 25 mm diameter, Whatman). These sample sets were specifically used to evaluate cell disruption and extraction methods for subsequent analyses. The choice to dilute the cultures in MQ water aimed to ensure a consistent medium and induce an initial osmotic shock in the cells, thus facilitating the cells rupture and releasing the intracellular components. After filtration, samples were preserved in petri dishes, covered with aluminum foil, flash-frozen in liquid nitrogen and stored at − 80 °C (Ultrafreeze, Thermo Fisher, USA) until further analysis. Additionally, natural samples were collected in replicates under algal bloom conditions, during the occurrence of *Woronichina* bloom in Lake Varese (Italy). The samples were collected on 11th December 2023 at coordinates 45.828055° N, 8.722186° E at the surface layer (0–0.5 m depth) using pre-cleaned polyethylene 10 L bottles. To preserve sample integrity and minimize biological and chemical alteration, the polyethylene bottles were stored in a dark and transported immediately to the nearby laboratory facility. Within one hour of collection, each replicate was filtered under gentle vacuum pressure (≤ 0.5 bar) using a Sartorius filtration ramp system (Sartorius, Germany) equipped with glass fiber filters (Whatman GF/F, 25 mm, nominal pore size 0.7 µm). A total of 18 replicate water samples (300 mL each) were prepared. Following filtration, filters were rapidly folded (biomass side inward), wrapped in pre-labeled aluminum foil, and flash-frozen in liquid nitrogen before being transferred to a − 80 °C freezer for long- term storage until analysis.

2.2. Determination of cell density in the culture

Biological conversion factors between optical density (OD) at 600 nm and cell density (gDW/L) and cell count (CFU/mL) are commonly used to estimate the growth of bacteria and unicellular organisms such as algae [23]. When the attenuation coefficient and path length remain constant, the Beer-Lambert law allows for a proportional correlation between biomass concentration ([X], g/L) and optical density (OD) at a given wavelength and a 1 cm path length:

$$[X] = m*OD \quad (1)$$

However, 600 nm is not always the optimal wavelength, particularly when cellular pigments interfere with absorbance measurements [23]. Myers (2013), [23] evaluated several bacterial and algal strains, including *Synechococcus*, to determine the most suitable wavelength for estimating cell density and cell count (Supplementary Table S1). For *Synechococcus*, 750 nm was identified as the most reliable wavelength. No optimized values, however, are currently available in the literature for *Anabaena*.

Therefore, the initial concentrations in pure cultures and the dilutions used to prepare filter pads were determined using OD measurements at 600 nm for *Anabaena* and 750 nm for *Synechococcus*. Cell density was verified using a Perkin Elmer Lambda 35 double-beam spectrophotometer (Perkin Elmer, USA), equipped with 1 cm path length quartz cuvettes and measurements taken at both wavelengths.

2.3. Spectrophotometric characterization of standard

The C-PC partially purified from *Spirulina* sp. (*Spirulina* sp. P2172- 10MG, Sigma–Aldrich, Germany) was used in this study because more representative of natural sample extracts and it was used for spectrophotometric characterization and UHPLC method development (see Section 2.6). The C-PC standard was diluted to a concentration of 1 mg mL^{-1} concentration in the 0.1 M, pH 6.8 phosphate buffer and stored at − 20 °C until analysis. Previous studies [24] indicated that the introduction of the non-aqueous solvents can alter the tertiary structure of the C-PC protein. Specifically, organic solvents can cause the release of the C-PC chromophore, allowing it to adopt a cyclic conformation. Therefore, before proceeding with HPLC method development, the behavior of the C-PC standard was characterized in the acetonitrile organic solvent (HPLC gradient, Merck, Germany) with 0.1 % TFA *v*/v (HPLC gradient, Sigma-Aldrich, Germany). To this end, a series of dilutions of the C-PC standard (*Spirulina* sp. P2172-10MG, Sigma–Aldrich, Germany) were prepared using

three different concentrations of acetonitrile with 0.1 % TFA *v*/v. These were analyzed using a double- beam spectrophotometer (Perkin Elmer Lambda 35, USA) with 1 cm path length quartz cuvettes. As a reference, corresponding dilutions of 0.1 M phosphate buffer at pH 6.8 with equivalent amounts of acetonitrile were used.

2.4. Spectrophotometric C-PC measurement on extract

C-PC extracts were analyzed with a double beam spectrophotometer (Perkin Elmer Lambda 35, USA), with 1 cm path length quartz cuvettes, scanning in the 250–800 nm range (scan speed: 240 nm/min; data acquisition interval: 1 nm). An additional test was conducted by mixing the C-PC standard (diluted in 0.1 M phosphate buffer at pH 6.8) and chlorophyll *a* standards (prepared in HPLC-grade acetone, Merck, Germany) in a 50:50 *v*/v ratio, and measuring the spectrum against a corresponding blank, to evaluate potential spectral interference between the two pigments.

C-PC concentration was determined using Eq. [2], originally proposed by Bennett and Bogorad (1973) [25], which was developed for extracts in 0.01 M sodium phosphate buffer at pH 7:

$$Conc_{C\text{-}PC} = (OD_{615} - 0.47 * OD_{652})/0.534 \quad (2)$$

where $Conc_{C\text{-}PC}$ represents the C-PC concentration in mg L^{-1}, OD_{615} is the optical density of the sample at 615 nm, OD_{652} is the optical density at 652 nm.

To account for sample and buffer volumes and cuvette path length, the equation was modified as follows:

$$Conc_{C\text{-}PC} = (OD_{615} - 0.47 * OD_{652}) * V_e / (0.534 * V_s * l) \quad (3)$$

Where V_e is the volume of buffer extract (mL), V_s is the volume of the original water sample (L), and l is the cuvette path length (cm).

The purity of the C-PC fraction was assessed spectrophotometrically by calculating the ratio of absorbance at 620 nm to that at 280 nm, following the method developed by Abalde [26], with minor modifications. Purity (EP) was calculated using Eq. [4]:

$$EP = OD_{620}/OD_{280} \quad (4)$$

where EP represents the extract purity, OD_{620} is the optical density of the sample at 620 nm, and OD_{280} is the optical density at 280 nm. Absorbance at 280 nm in these preparations is mainly due to aromatic amino acids, roughly proportional to the total protein concentration in the solution, including C-PC. The respective absorbance at 620 nm reflect only the concentrations of C-PC,

2.5. UHPLC method development

The separation and characterization of C-PC—a water-soluble phycobiliprotein—can be achieved through various chromatographic techniques, including ion-exchange, size-exclusion, hydrophilic interaction, hydrophobic interaction, and reversed-phase chromatography (RPC). Among these, RPC stands out for its high efficiency and rapid separation capability, making it a preferred method for protein analysis. In RPC, proteins are typically separated using C4 or C5 stationary phases with large pore sizes (300–400 Å) and a binary mobile-phase gradient composed of water with trifluoroacetic acid (TFA) and acetonitrile (ACN). TFA acts as an effective ion-pairing agent, enhancing protein retention and improving peak shape. Several studies [18–21] have successfully applied this configuration to C-PC analysis using both C4 and C5 columns with 0.1 % TFA in both solvents. In our study, we developed and optimized a UHPLC-DAD method using a C-PC standard (*Spirulina* sp., Sigma–Aldrich, P2172-10MG), prepared at 0.1 mg mL^{-1} in 0.1 M phosphate buffer (pH 6.8) and stored at − 20 °C until analysis. This method was then applied to extracts from pure algal cultures (*Synechococcus PCC 6911* and *Anabaena PCC 7120*) as well as from natural water samples collected on GF/F filter pads. The UHPLC-DAD results were compared with conventional techniques used for satellite product validation, including spectrophotometric analysis and HPLC- based pigment profiling.

2.5.1. Chromatographic conditions and optimization

To identify the best conditions for C-PC selectivity and resolution, the following parameters were tested:

- Mobile Phase Composition:
- Solvent A: 0.1 % TFA in water
- Solvent B: 0.1 % TFA in acetonitrile (ACN)
- Flow Rate: Constant at 0.8 mL min^{-1}
- Gradient: Initial compositions at 20 %, 32.5 %, or 45 % B (pre- equilibration for 3 min), followed by a linear gradient to 100 % B over 5, 30, or 55 min. This was followed by a 5 min isocratic elution at 100 % B.
- Column Temperatures Tested: 25 °C, 32.5 °C, and 40 °C

Additional tests (data not shown) were conducted to evaluate the effects of different solvents and ion-pairing agents. Acetonitrile was either diluted (25 % or 50 %) or replaced with methanol to assess the impact on separation performance. Overall, when comparing ACN–water mixtures with alcohol–water systems, acetonitrile consistently provided better resolution. Trifluoroacetic acid produced sharper and better-defined peaks than an alternative ion-pairing agent, tetrabutylammonium acetate (TBAA, 28 mM, pH 5), despite the latter being chosen to maintain a pH closer to the optimal 5–6 range for C-PC stability.

2.5.2. Columns and equipment

C4 and C5 columns were chosen for their known performance in protein separations: BIOshell Protein C4, 3.4 µm, 400 Å, 4.6 × 150 mm (SUPELCO, Germany) and Discovery BIO Widepore C5, 5 µm, 300 Å, 4.6 × 250 mm (SUPELCO, Germany). A Zorbax Eclipse XDB-C8 column (4.6 × 150 mm, 3.5 µm, Agilent) was also tested, as C8 columns are often used for simultaneous analysis of chlorophylls and carotenoids. However, significant peak tailing was observed for C-PC in all C8 configurations, confirming the poor suitability of C8 phases for C-PC analysis [27]. The UHPLC system used was an Agilent Infinity 1260 II, equipped with: quaternary pump (G7104A 1290 flexible pump); thermostated autosampler (G7129B); diode array detector (G7115A 1260 DAD WR); fluorescence detector (G1321A FLD); Thermostated column oven. Detection was performed using OpenLab CDS software, monitoring at 580, 620, and 640 nm via DAD, and via fluorescence (excitation at 565 nm, emission at 640 nm), following Homoelle and Beck [28].

2.5.3. Sample handling and temperature effects

All samples were maintained at 14 °C prior to injection. In triplicate tests, maintaining samples at room temperature showed no significant impact on retention time, peak shape, or area. However, storing samples at 4 °C led to visible precipitation in some cases (data not shown), highlighting the importance of consistent temperature control during analysis. For natural samples, lower concentrations of C-PC were expected; therefore, a larger injection volume of 100 µL was used to enhance detectability. To ensure consistency with the analysis of natural samples, the method calibration was also performed using 100 µL injections of low-concentration C-PC standards (0.2, 0.1, and 0.01 mg L^{-1}). The raw pigment concentration obtained directly from the UHPLC system—prior to any dilution correction—is referred to as $\tilde{P}_{ip}$. To determine the final C-PC concentration in the original water sample, was applied the following equation:

$$C - PC_{UHPLC} = \frac{V_e \, \tilde{P}_P^i}{V_f \, V_e} \quad (5)$$

where, $\tilde{P}_{ip}$ is the concentration of C-PC as measured by UHPLC before dilution correction, V_e is the extraction volume, V_f is the volume of water filtered for each sample, V_c is the volume of extract injected into the UHPLC column.

This correction accounts for dilution introduced during extraction and injection, allowing the reported $C- PC_{UHPLC}$ to reflect the true concentration in the original water sample.

2.5.4. Design of experiment: UHPLC method optimization

To optimize the UHPLC analysis of C-PC, we employed a Design of Experiments (DoE) strategy. DoE is a structured and systematic approach used to investigate the effects of multiple experimental factors on one or more response variables. Unlike traditional one-factor-at-a- time methods, DoE enables the simultaneous assessment of several variables, facilitating the identification of significant factors, interactions, and optimal conditions with a reduced number of experimental runs [28,29]. This methodology is widely used in chemical and analytical method development, including HPLC optimization [30,31]. In our study, DoE was chosen to screen and determine the optimal conditions for enhancing UHPLC performance. The goal was to develop a robust and reproducible method for extracting and quantifying C-PC from water samples collected on filter pads. To this end, we applied a Full Factorial Design (FFD) that included three continuous variables (each tested at two levels) and one categorical variable (at two levels), along with center points to evaluate curvature in the response.

The three continuous factors (at 2 levels, coded as − 1 and + 1) were X_1, column oven temperature, X_2, initial % of Solvent B, and X_3, gradient time (min) to reach 100 % Solvent B. The categorical variable (C) was the column type (i.e., C4 or C5, coded as 0 and 1). Center points (e. g., $X_1 = 0$, X_2, =0, $X_3 = 0$) were added to assess non-linearity and estimate pure error. The general form of the full factorial model with center points is given in Eq. [6]:

$$Y = \beta_0 + \beta_1X_1 + \beta_2X_2 + \beta_3X_3 + \beta_4C + \beta_{12}(X_1X_2) + \beta_{13}(X_1X_3) + \beta_{14}(X_1C) + \beta_{23}(X_2X_3) + \beta_{24}(X_2C) + \beta_{34}(X_3C) + \beta_{123}(X_1X_2X_3) + \beta_{124}(X_1X_2C) + \beta_{134}(X_1X_3C) + \beta_{234}(X_2X_3C) + \beta_{1234}(X_1X_2X_3C) + \epsilon \quad (6)$$

Where β_0 is the intercept, β_1, β_2, β_3, β_4 are main effect coefficients, β_{12}, β_{13}, β_{14}, β_{23}, β_{24} are coefficients for two-, β_{123}, β_{124}, β_{134}, β_{234}, for three-, and β_{1234} four-way interactions and ε is the residual error. The inclusion of center points improves the model by allowing detection of non-linear trends and refining the estimation of coefficients and error variance. Although the full model includes all possible interactions, it can later be simplified by excluding non-significant higher-order terms. The statistical analysis of the FFD was carried out using Design-Expert® software. The measured responses in this study were the peak area (mAU) (Response 1) and peak symmetry at 640 nm (Response 2) of the main C-PC peak. These responses were used to evaluate the chromatographic performance and extract quality under different experimental conditions (Table 1).

2.6. C-PC pigment extraction from samples

The choice of cell disruption method plays a critical role in ensuring reproducibility and analytical reliability. Traditional approaches—such as detergent lysis, freeze-thaw cycles, sonication, enzymatic digestion, and mechanical disruption—each offer distinct advantages and limitations. For marine and freshwater cyanobacteria, commonly applied methods include detergent lysis [32], enzymatic digestion [33,34], freeze-thawing, sonication, mechanical disruption, or combinations thereof [22,26,35–37]. Although filter-concentrated samples generally yield lower extractable C-PC concentrations [15], immediate processing is often impractical during fieldwork, particularly aboard research vessels. As a result, long-term storage is required. The standard procedure for sample preparation during oceanographic field campaigns for subsequent HPLC pigment analysis involves filtering known volumes of water onto GF/F filters at 0.5 atm pressure, followed by flash-freezing in liquid nitrogen. Sample preservation until analysis is carried out in liquid nitrogen or at − 80 °C [16] until the analysis. In the present study the filters for the analysis are prepared following a similar procedure (see Section 2.1.) While full C-PC recovery remains challenging [34], this study focuses on improving and standardizing extraction protocols to enhance reproducibility and analytical reliability for satellite validation purposes.

Table 1

The factors considered were: (Factor 1) the column oven temperature, (Factor 2) the initial % of the Solvent B, and (Factor 3) the gradient time (min) to reach 100 % of Solvent B and the (Factor 4) the column type. The targets were the Area (mAU) of the main peak identified (response 1) and the peak symmetry (response 2) at 620 nm.

Std	Run	Factor 1 A:Temp. °C	Factor 2 B:Solv. B initial %	Factor 3 C:time min	Factor 4 D:Column	Response 1 Area (620 nm) mAU (maximize)	Response 2 Symmetry (620 nm) (target = 1)
1	22	25	20	5	c4	139.4	0.63
2	10	40	20	5	c4	210.7	1.5
3	11	25	45	5	c4	176.8	0.8
4	24	40	45	5	c4	219.7	1.47
5	9	25	20	55	c4	11.36	2.83
6	15	40	20	55	c4	27.43	1.96
7	18	25	45	55	c4	26.4	2.76
8	21	40	45	55	c4	26.22	2.71
9	17	25	20	5	c5	318.48	0.92
10	12	40	20	5	c5	324.91	0.96
11	6	25	45	5	c5	315.60	1.1
12	23	40	45	5	c5	315.7	0.99
13	7	25	20	55	c5	285.4	0.79
14	14	40	20	55	c5	309.0	0.47
15	20	25	45	55	c5	317	0.71
16	3	40	45	55	c5	158.7	0.95
17	2	32.5	32.5	30	c4	41.03	2.27
18	19	32.5	32.5	30	c5	222.9	1.13
19	8	32.5	32.5	30	c4	41.6	2.33
20	1	32.5	32.5	30	c5	201.1	1.11
21	13	32.5	32.5	30	c4	19.1	2.59
22	16	32.5	32.5	30	c5	212.1	1.13
23	4	32.5	32.5	30	c4	35.1	3.15
24	5	32.5	32.5	30	c5	205.3	1.05
25	25	32.5	32.5	30	c4	41.0	2.27
26	26	32.5	32.5	30	c5	212.2	1.13

2.6.1. Extraction procedures

Two extraction methods were tested: freeze-thaw (FT) with sonication, and enzymatic digestion with lysozyme (Ly). The FT method employed a 0.1 M phosphate buffer (pH 6.8) and included three to five freeze-thaw cycles, followed by probe sonication. Literature suggests that three to five cycles are effective for C-PC extraction [15,34], with no significant gain beyond five cycles. Each sample was defrosted for 5 min, cut into small pieces, and mixed with either 2.5 or 5 mL of buffer. Probe sonication (Bandeline, USA) was performed in an ice bath for 90 s (1 cycle, 25 W), following the protocol used for organic-soluble pigment extraction (Canuti, 2023 [38]) and supported by Horvath (2013) [´ 15], who reported optimal C-PC recovery with 90-s sonication. The freeze- thaw cycles involved freezing at − 20 °C for 1.5 h and defrosting in a 37 °C water bath for 10 min. After the final cycle, samples were centrifuged at 4000 rpm for 4 min (EBA 270, Hettich, Germany), and the supernatant was filtered using a 0.22 µm PFTE syringe filter (Millipore, Germany) to remove debris [25]. The Ly method combined gentle mechanical grinding with enzymatic lysis. The extraction buffer consisted of 0.25 M Trizma base, 10 mM disodium EDTA (both from Sigma- Aldrich, Italy), and 1.25 mg mL^{-1} lysozyme (Merck, Germany), each prepared separately in Milli-Q water (Millipore, Germany). The initial pH (~9) was adjusted to 5.5 using 1 M hydrochloric acid, and the solution was equilibrated for ~3 h. The buffer was stored in the dark at 4 °C and used within 48 h. Under dim light, 2.5 or 5 mL of buffer was added to each natural sample. Cells were disrupted through gentle grinding followed by enzymatic digestion. For half of the samples, an additional 90-s sonication step was included. Samples were vortexed (Heidolph, Germany), incubated at 37 °C for 2 h (AU 32, Argo Lab, Italy), and stored in the dark at 4 °C overnight (18 h). Centrifugation was performed at 4000 rpm for 10 min, and the supernatant was filtered using a 0.22 µm PFTE syringe filter (Millipore, Germany). Total extraction time did not exceed 24 h [12,13].

2.6.2. Design of Experiment: Extraction procedure

To determine the optimal extraction conditions, that maximizing C- PC extraction efficiency two sets of experiments were performed (Supplementary Table S2):

- For the Lysozyme (Ly) method, the tested variables were (A) Extraction buffer volume (2.5 or 5 mL), (B) Disruption type (mechanical vs. sonication).
- For the Freeze-Thaw (FT) method, the tested variables were (C) Number of freeze-thaw cycles (3 or 5), (D) Number of sonication steps (1 or 2).

The first sonication (90 s with probe) was applied to all FT samples after the first freeze-thaw cycle. A second sonication step was applied to half the samples using a sonication bath during the defrosting phase (10 min). Due to sample limitations for pure cultures and natural samples, a 2^2 factorial design was adopted. Each experiment was run in duplicate for reproducibility. The objective was to maximize C-PC recovery (peak area) while ensuring high reproducibility.

The statistical model for the two-factor, two-level factorial design (with duplicates) is shown in Eq. [7]:

$$Y_{ijk} = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_{12}(X_1 X_2) + \beta_4 C + \beta_5 C X_1 + \beta_6 C X_2 + \varepsilon \quad (7)$$

Where Y is the response variable (e.g., C-PC recovery/peak area). β_0 is the intercept, β_1 and β_2 are the main effects of factors X_1 and X_2, which represent the two experimental factors (e.g., sonication and another factor like buffer type), β_3 represents the interaction effect between X_1 and X_2, β_4 represents the effect of the categorical variable C (species, coded as 0 and 1), β_5 and β_6 represent the interaction between the categorical variable and the two factors and ε represents the random error term. As a finale remark, it has to be added that the extraction tests were conducted only after the identification of the UHPLC method.

3. Results

3.1. Determination of cell density in the culture

The analysis of the cell density of the cultures used for the extraction experiments indicated that the *Synechococcus PCC 6911* was in a phase of active growth, likely the exponential or log phase, where cell division is rapid, leading to a high number of viable cells per unit of optical density. Conversely, *Anabaena PCC 7120* was in a slower phase of growth, in the late exponential phase or transitioning into the stationary phase, where the rate of cell division is slower, resulting in fewer viable cells per OD unit (Supplementary Table S3).

3.2. Spectrophotometric characterization of standard

The standard characterization was done under varying condition. In the absorption spectrum of a *Spirulina* sp. standard diluted in phosphate buffer (Fig. 1) the absorption profile exhibits characteristic peaks corresponding to C-PC at 280 nm and 620 nm. The purity and concentration of the standard preparation was obtained accordingly the eq. [1,2] (Supplementary Table S4). Considering that in our chromatographic method the C-PC is in an ACN/water gradient, the effect of the acetonitrile (0.1 % TFA) solvent on C-PC spectral signature was evaluated for different dilution percent using as reference a solution of 0.1 M sodium phosphate buffer at pH 6.8 and acetonitrile (0.1 % TFA) solvent at corresponding percent dilution. We analyzed the absorption spectra of a series of dilutions of a 1 mg mL^{-1}C-PC standard (*Spirulina* sp. P2172- 10MG, Sigma–Aldrich) at three different percentage of ACN 0.1 % TFA *v*/v (20 %, 40 % and 60 %) against a corresponding phosphate- buffer/ACN dilution percentage (Fig. 1). The addition of acetonitrile resulted in slight shifts and changes in peak intensities, indicating the solvent's effect on the stability and optical properties of C-PC. The introduction of ACN 0.1 % TFA v/v at concentrations of 20 %, 40 %, and 60 % (*v*/v) induced a reduction in the intensity of the 620 nm absorption band. This decrease occurs in the specified concentration sequence compared to the maximum intensity observed in the absence of acetonitrile. Simultaneously, the 350 nm absorption band experiences an increase in intensity as acetonitrile was progressively incorporated into the solution. The reference samples, phosphate buffer with equivalent acetonitrile concentrations, showed no absorbance changes, underscoring that the observed variations in the C-PC standards were attributable to interactions between acetonitrile and the pigment. These results demonstrate the influence of solvent composition on the spectroscopic behavior of C-PC resulting in a shift of peak intensity and position. [24] Homoelle (1997) reported a shift in the maximum peak absorption when the C-PC was diluted in Methanol and a decreasing in the peak maximum in the blue band in favor to the absorption in the red.

3.3. Spectrophotometric measurement of standards and extracts

In agreement with previous studies [8,10], the quantification of C-PC through spectrophotometry presented several limitation due to the overlapping in natural samples of chlorophyll *a* and C-PC and for low concentration recovered from natural samples. In [19] was assumed that the absorbance ration A_{620}/A_{280} should be higher than four to be relevant, however in case of the lysate, the lysozyme and the cellular debris absorbed strongly in the UV region. By using the Eq. [3], we calculate the concentration in mg L^{-1} for some of the lysate extracts of *Anabaena* and *Synechococcus* Supplementary Table S5) for comparing with the results obtained by UHPLC analysis.

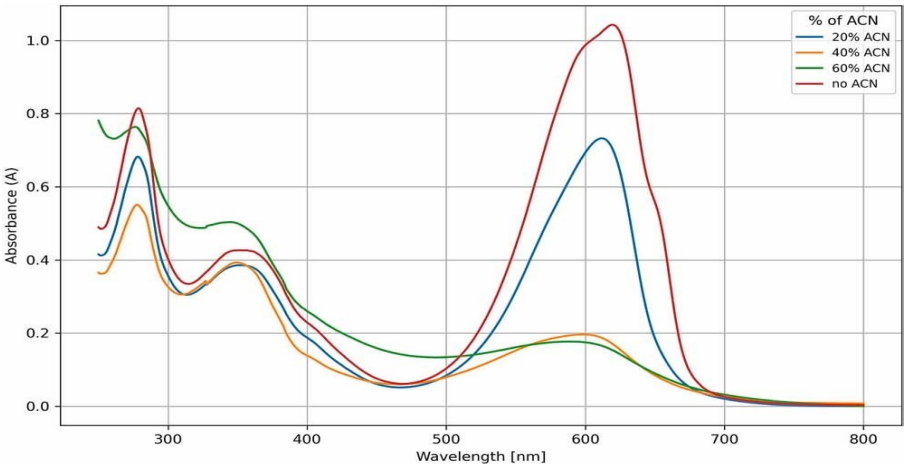


Fig. 1. C-PC *Spirulina* sp. standard diluted to 1 mg mL^{-1} in the 0.1 M, pH 6.8 phosphate buffer, measured in double beam Perkin Elmer Lambda 35 spectrophotometer, 1 cm quartz cuvette, reference 0.1 M, pH 6.8 phosphate buffer (in red), room temperature. The C-PC standard at concentration of 1 mg mL^{-1}, in three different solutions of ACN 0.1 % TFA *v*/v (20 % ACN in blue, 40 % ACN in orange and ACN in 60 % green), room temperature. In the reference cuvettes, a series of dilution of 0.1 M, pH 6.8 phosphate buffer solution with acetonitrile at equivalent than the standard were used. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

3.4. *UHPLC method development*

The experiments followed the Design of Experiments (DoE) scheme outlined in Table 1. For the Area response (Response 1), the normal plot (Fig. 2a) indicated that the analysis time (C), column type (D), and their interaction (CD) significantly affected the model. Two models were evaluated for the Area response based on the included terms and their statistical metrics. The first model included the interaction terms C, D, and CD, while the second, more complex model incorporated additional terms A (column oven temperature), B (initial % of the Solvent B), C, D and all possible interactions. The simpler 2FI model (C, D, CD) yielded better predictive accuracy, evidenced by a lower Predictive Residual Sum of Squares (PRESS = 35,57) compared to the extended model (PRESS = 178,87), indicating reduced prediction error. Although the extended model had a lower − 2 Log Likelihood (166.26) than the simpler model (248.41), suggesting a better fit to the training data, its Bayesian Information Criterion (BIC = 218.39) and the corrected Akaike Information Criterion (AICc = 258.71) were not substantially better than those of the simpler model (BIC = 261.44, AICc = 258.32). The slight improvement in model fit does not justify the added complexity, especially considering the significant increase in PRESS, which reflects poor generalization. Based on these statistics, the simpler 2FI model (C, D, CD) was preferred due to its better predictive performance and parsimony, aligning with the principle of model simplicity without sacrificing accuracy. The corresponding Analysis of Variance (ANOVA) for Area (Table 2) revealed that the model had an F-value of 90.37, demonstrating statistical significance. The main effects of C, D, and CD were all significant, with *p*-values less than 0.05. The significant Lack of Fit value ($p = 0.0001$) suggests that the model fit the data. (See Fig. 3.)

For the Symmetry response (Response 2), the normal plot (Fig. 2b) highlighted analysis time (C), column type (D), and their interaction (CD) as significant factors. The selected 2FI model, which includes these terms, demonstrated solid predictive performance with PRESS = 3.28, AICc = 18.36, BIC = 21.49, and − 2 Log Likelihood = 8.46. Although a more complex hierarchical model (including terms A (column temperature), C, D, and their quadratic and cubic interactions) yielded slightly improved statistics (PRESS = 3.01, AICc = 20.37), the marginal benefit did not justify the added complexity. Thus, the 2FI model was selected as the optimal model for Symmetry, balancing simplicity with predictive reliability. The ANOVA results (Table 3) further supported this choice, with a model F-value of 42.81, indicating high statistical significance. All key terms (C, D, CD) had *p*-values <0.05, confirming their importance. The Lack of Fit F-value of 1.67 ($p = 0.24$) was not significant relative to the pure error, suggesting the model adequately fits the data, with a 23.68 % probability that such a Lack of Fit could arise from random variation.

The best conditions (i.e., maximize the area and obtaining a symmetry close to value of 1, constrain Table 4) were reached for column C5 and the combination of T of 25 °C, 5 min of gradient and a 20 % of initial Solvent B. It was observed that longer gradient time causes a degradation of the protein, driving to lower Area.

Once chosen the best column, C5, the interaction of the different parameter was study on the single column based on a full factorial (FI3) with 4 central points, keeping Area and Symmetry as responses (data selected from Table 1 and summarized in Supplementary Table S6). The 3D surface plots (Fig. 4) indicated the curvatures and the influences of the different parameters (i.e., analysis time, column temperature and initial % of solvent B). The best condition were reached out with a 5′ linear gradient (from 20 % to 100 %) and a temperature of 25 °C.

A three points calibration curve was performed using the method but with an injection volume of 100 µL, as a low concentration is expected in the

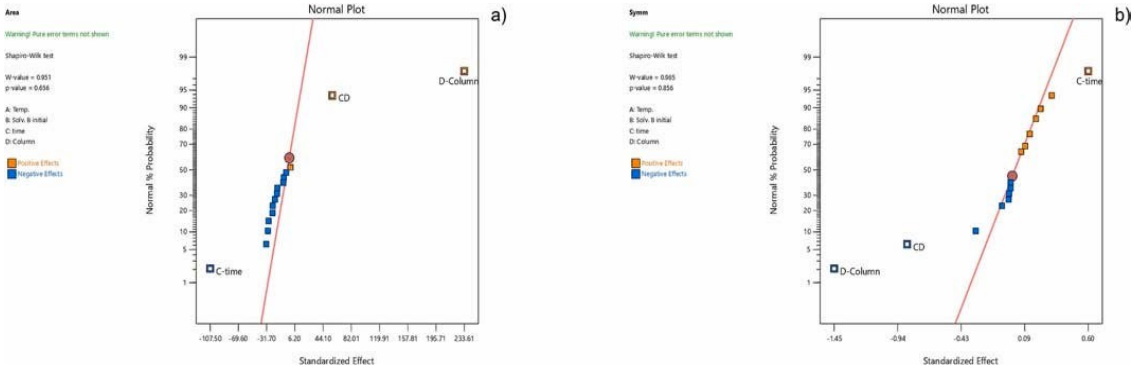


Fig. 2. The Normal plot for the different parameters considered in the factorial approach for the (**2a**) Area and (**2b**) Symmetry response parameters.

natural samples. The 1 mg mL^{-1} mother solution of C-PC from *Spirulina*, was diluted in 0.1 M, 6.8 pH phosphate buffer at 0.2, 0.1 and 0.01 mg L^{-1} (Fig. 5). Five injections at each of the three concentration were done. The regression curve fitting was of R^2 = 0.99, the reproducibility (defined as Standard Deviation of Residuals) was of 9.8 %. The low limit of quantitation is of 0.01 mg L^{-1}. The accuracy of the UHPLC method was evaluated using calibration standards at three concentration levels. Accuracy was assessed by comparing the measured concentrations—derived from the calibration regression—to the known nominal values, a standard approach in analytical method validation to reflect practical performance across the working range. The mean relative bias ranged from +62.0 % at the lowest level to +2.0 % at the highest, with the best agreement observed at 0.20 mg L^{-1}. At the intermediate level (0.10 mg L^{-1}), a slight underestimation was observed, though the bias remained within an acceptable margin of approximately 10 %. The method's sensitivity, estimated from the standard deviation of the residuals (σ = 0.0076 mg L^{-1}), yielded a limit of detection (LOD =3.3 × σ) of 0.025 mg L^{-1} and a limit of quantification (LOQ 10 × σ) of 0.077 mg L^{-1}. These values indicate that the method is suitable for quantifying C-phycocyanin in natural samples with moderate to high pigment concentrations.

3.5. C-PC pigment extraction

The extraction procedure was tested for both *Anabaena* and *Synechococcus* in duplicates to evaluate the reproducibility for the series S1 and A1 of replicates. The FT procedure did not recover any of the C-PC present in the algal replicates S1 and A1 while the LY procedure allowed the extraction of pigment both in the *Anabaena* and *Synechococcus*. Tests were done also with highest concentration of cyanobacteria on filter pad: in this case, the FT procedure gave positive results (data not shown here). However, as the aim of the method is to detect C-PC in concentration close to those encountered in natural samples, the A1 and S1 series were to be preferred for the method evaluation. The results for the LY procedure (Supplementary Table S7) were then analyzed through the FF model approach. For the extraction procedure, which evaluated the effects of lysis buffer volume (A), lysis disruption method (B), and strain species (C), the analysis focused on identifying the most relevant factors influencing peak area. The ANOVA results (Table 5) showed that the overall model was statistically significant (F-value = 4.33, p = 0.0276) when the strain species was excluded. This indicates that the extraction method was comparably effective across both cyanobacterial strains, and that buffer volume (A) and disruption method (B), along with their interaction (AB), were the main contributors to variability in pigment extraction. Among these, only the disruption method showed a statistically significant effect (p = 0.0264), while the buffer volume (p = 0.0860) and AB interaction (p = 0.1044) did not reach significance.

Importantly, the Lack of Fit test was not significant (F = 0.14, p = 0.9620), indicating a good fit of the model to the experimental data. Model comparison further supported the selection of a two-factor interaction (2FI) model, which included terms A, B, and AB. This model provided better predictive power than more complex hierarchical models that also included factor C and additional interactions. Specifically, the selected 2FI model yielded PRESS = 46.31, AICc = 64.84, BIC = 64.30, and − 2 Log Likelihood = 53.20, compared to higher PRESS (79.26) and AICc (80.60) values for the expanded model. This confirmed that increasing model complexity did not yield substantial performance gains and risked overfitting. Further validation of the selected model (Supplementary Table S8) demonstrated acceptable Adequate Precision = 4.40, indicating a strong signal-to-noise ratio (values above 4 are desirable). These results underscore the robustness of the 2FI model and highlight the mechanical disruption method combined with 2.5 mL of lysis buffer as the most effective extraction condition, maximizing peak area. Overall, the findings emphasize that optimization of the disruption technique has a more significant impact on pigment extraction efficiency than buffer volume.

The concentration obtained from each extraction procedure was calculated based on UHPLC results (Supplementary Table S9) and compared with concentrations measured spectrophotometrically from

Table 2
UHPLC DoE method development: ANOVA for selected factorial model Response 1: Area.

Source	Sum of Squares	df	Mean Square	F-value	p-value	
Model	2.772E+05	3	92,407.15	90.37	< 0.0001	significant
C-time	46,223.36	1	46,223.36	45.21	< 0.0001	
D-Column	2.183E+05	1	2.183E+05	213.49	< 0.0001	
CD	12,707.89	1	12,707.89	12.43	0.0020	
Curvature	35,344.80	1	35,344.80	34.57	< 0.0001	
Residual	21,472.47	21	1022.50			
Lack of Fit	20,828.23	13	1602.17	19.90	0.0001	significant
Pure Error	644.23	8	80.53			
Cor Total	3.340E+05	25				

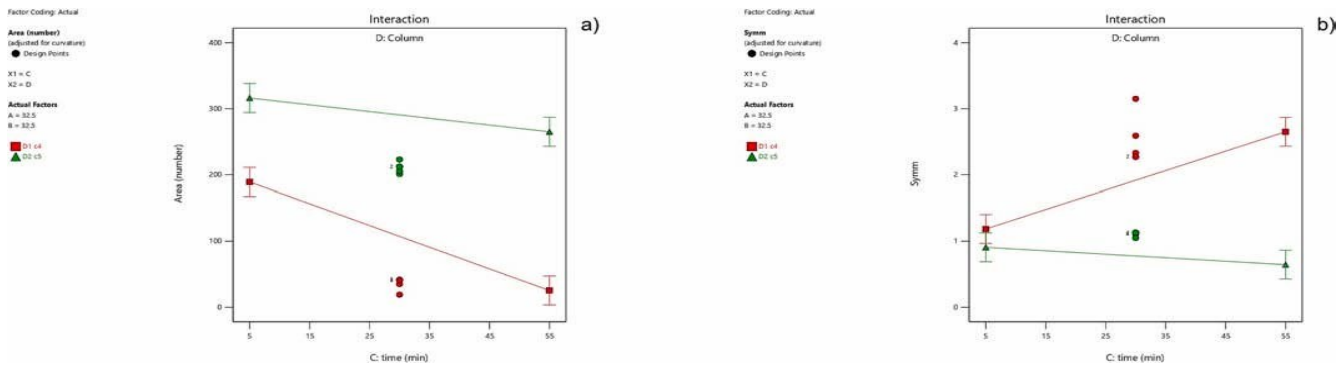


Fig. 3. Influence of analysis time (Time) and column type (Column) on Peak Area (Area) and Peak Symmetry (Symm).

Table 3

UHPLC DoE method development: ANOVA for selected factorial model Response 2: Symmetry (Symm).

Source	Sum of Squares	df	Mean Square	F-value	p-value	
Model	12.89	3	4.30	42.81	< 0.0001	significant
C-time	1.45	1	1.45	14.41	0.0011	
D-Column	8.46	1	8.46	84.29	< 0.0001	
CD	2.98	1	2.98	29.74	< 0.0001	
Curvature	1.35	1	1.35	13.49	0.0014	
Residual	2.11	21	0.10			
Lack of Fit	1.54	13	0.12	1.67	0.2368	not significant
Pure Error	0.56	8	0.07			
Cor Total	16.35	25				

the same extracts. The comparison between spectrophotometric and UHPLC measurements revealed a good correlation, as indicated by the high Pearson correlation coefficient (r = 0.79). This suggests that both methods yield consistent results, thereby supporting the validity of the UHPLC approach for estimating pigment concentrations relative to the established spectrophotometric technique (Supplementary Fig. S1). However, precision analysis highlighted high variability among the triplicates for both methods. This discrepancy may reflect differences between replicates (Supplementary Table S10), which could be attributed to inconsistencies in the extraction procedure or to the challenges associated with preparing homogeneous triplicate samples from algal cultures [39].

The best extraction method from the test on algal cultures emerged to be the addiction of 2.5 mL lysate combined with mechanical disruption, while the addiction of 2.5 mL lysate and the sonication disruption, was indicated as the second best method for C-PC recovery. We use these methods for extracting and analyzing Lake Varese samples in triplicates (Fig. 6). The results (Table 6) were analyzed by both basic statistical measures and inferential tests to assess the consistency, reliability, and potential differences between the two methods used to process samples from Lake Varese. Basic statistics, including standard deviation, and coefficient of variation (CV), were employed to evaluate the central tendency and variability within each method. Additionally, inferential tests such as the *t*-test and Mann-Whitney *U* test were conducted to determine whether the observed differences between the lysed and mechanically disrupted methods were statistically significant [40]. The t-test [41] compare the means between lysed and mechanically disrupted samples and help you understand if the differences observed are statistically significant. A non-parametric test like the Mann-Whitney *U* test [42] were of help in case of small sample size or non-normality. The results of the analysis revealed that the lysed samples exhibited better repeatability and consistency compared to the mechanically disrupted samples. The sonication-based method showed slightly higher C-PC concentrations, with a mean of 30.82 μg L^{-1} and a CV of 9.2 %, indicating good internal consistency. In contrast, the mechanical method yielded a lower mean concentration of 24.20 μg L^{-1} and a higher CV of 19.9 %, suggesting increased variability and lower repeatability. Despite these observed differences, the inferential tests did not reveal statistical significance. The Welch's *t*-test returned a test statistic of 2.06 and a *p*- value of 0.085, while the Mann–Whitney U test yielded a U statistic of 2.0 and a p-value of 0.10, both above the conventional threshold of 0.05. These results suggest that the differences in mean concentrations could be attributed to random variation rather than to a true effect of the disruption method. Therefore, while the sonication method may offer some practical benefits, further investigation with larger sample sizes is recommended to conclusively determine the superiority of one method over the other.

4. Discussion

In this study, the UHPLC method development and C-PC pigment extraction were critically analyzed to optimize and validate the chromatographic process and extraction efficiency. The UHPLC method,

Table 4
Constrains' table for searching the optimal UHPLC method (maximize the area and symmetry close to value of 1).

Name	Goal	Lower Limit	Upper Limit	Lower Weight	Upper Weight
A:Temp.	is in range	25	40	1	1
B:Solv. B initial	is in range	20	45	1	1
C:time	is in range	5	55	1	1
D:Column	is in range	C4	C5	1	1
Area	maximize	11.3	324.9	1	1
Symm	is target = 1	0.47	3.15	1	1

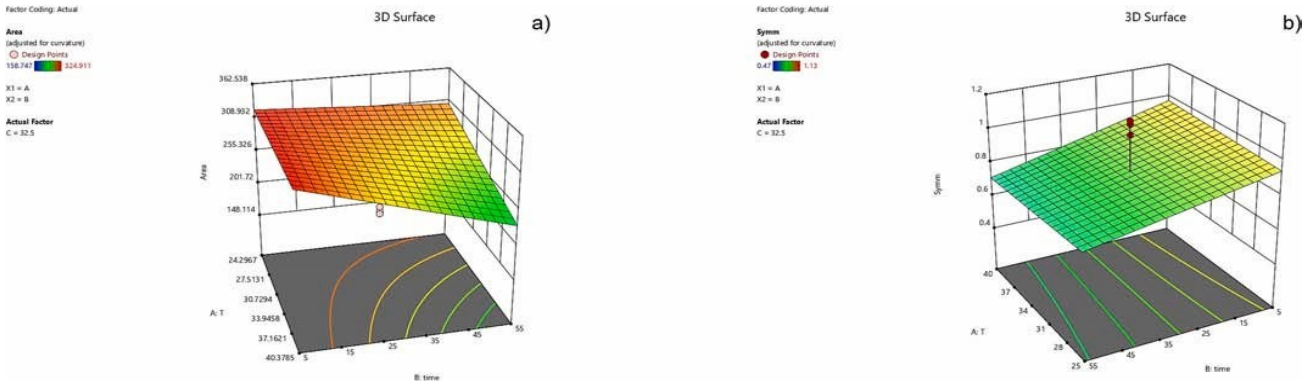


Fig. 4. The surface plot of Area and Symmetry response of the Optimization DoE of C5 column.

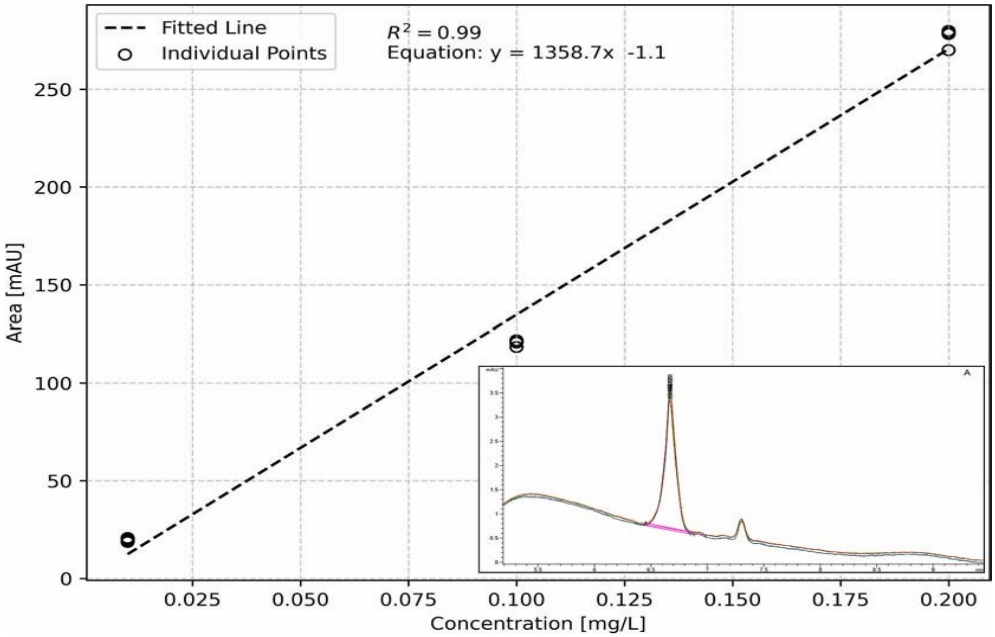


Fig. 5. The 3 points C-PC calibration curve for the chosen method (top figure) and the repeatability of the lower concentration point (inset bottom figures).

Table 5
Full Factorial model results for the lysate experiments: ANOVA results for the model considering the buffer of Lysate added (A-Ly-buffer) and the disruption method (B-Ly-disruption).

Source	Sum of Squares	df	Mean Square	F-value	*p*-value	
Model	28.19	3	9.40	4.33	0.0276	significant
A-Ly-buffer	7.59	1	7.59	3.50	0.0860	
B-Ly- disruption	13.90	1	13.90	6.40	0.0264	
AB	6.70	1	6.70	3.09	0.1044	
Residual	26.05	12	2.17			
Lack of Fit	1.72	4	0.43	0.14	0.9620	not significant
Pure Error	24.33	8	3.04			
Cor Total	54.24	15				

guided by a DoE approach, focused on identifying significant factors affecting chromatographic performance, specifically area and symmetry responses. The analysis showed that both time and column type, along with their quadratic effects, played significant roles in the chromatographic area and symmetry. The statistical significance of these factors was supported by the F-values obtained from ANOVA, indicating that the model was well-suited for the optimization of the chromatographic conditions. The optimized conditions, particularly the use of the C5 column with a specific gradient and temperature, were found to maximize the chromatographic area and maintain an appropriate symmetry, critical for accurate pigment quantification. The chosen parameters, particularly the C5 column, a temperature of 25 °C, a gradient time of 5 min, and an initial solvent B concentration of 20 %, were found to be optimal for maximizing the area while maintaining a symmetry value close to 1. The method was further validated by a three-point calibration curve, which demonstrated good linearity and reproducibility, essential for detecting C-PC at low concentrations in natural samples. The separation of the C-PC occurs for a percent of acetonitrile close to 45 %. The observed degradation of the protein with longer gradient times highlighted the importance of careful parameter selection to avoid compromising the chromatographic performance. Following the method development, the extraction of C-PC pigment from *Anabaena* and *Synechococcus* was evaluated, with particular attention to the reproducibility of the extraction process across different sample series. The factorial analysis of the lysate buffer volume, species type, and disruption method demonstrated that only the disruption method influenced the chromatogram area, as indicated by the significant *p*-values in the ANOVA results. This suggests that variations in lysate buffer volume, whether 2.5 mL or 5 mL, do not lead to substantial differences in extraction efficiency, nor do the species type. The determination of cell density, spectrophotometric characterization, and pigment quantification, integrates well with the findings from the UHPLC method development and C-PC pigment extraction. The determination of cell density revealed that *Synechococcus* was in an active growth phase, likely the exponential phase, with a high number of viable cells per OD unit. In contrast, *Anabaena* was in a slower growth phase, transitioning into the stationary phase, which contributed to fewer viable cells per OD unit. However, interpreting these results solely based on cell density may overlook significant differences in cell biovolume between cyanobacterial taxa. For instance, the much larger cell size of Anabaena compared to *Synechococcus* means that biovolume-based measurements would provide a more accurate reflection of total biomass or pigment content. This difference in growth phases likely influenced the overall pigment concentration and the extraction efficiency observed in subsequent experiments. The spectrophotometric characterization of C-PC standards in various solvent conditions, including phosphate buffer and ACN with 0.1 % TFA, showed significant effects of solvent composition on the absorption spectra. The introduction of ACN caused a decrease in the intensity of the 620 nm absorption band and an increase in the 350 nm band, indicating that solvent composition affects the stability and optical properties of C-PC. These findings underscore the importance of considering solvent effects when interpreting chromatographic results, as variations in solvent composition can lead to shifts in peak intensities and positions, which must be accounted for in the UHPLC method development. The spectrophotometric measurements of pure cultures and natural samples further highlighted the complexity of pigment composition in different cyanobacterial strains and environmental samples. The absorption spectra of *Synechococcus* showed higher overall intensity compared to *Anabaena*, reflecting its greater pigment concentration or efficiency in light absorption. High variability was observed in the triplicates analysis of pure strains both in spectrophotometric traditional approach and the UHPLC

analysis: these results can be attributed also to the difficulties of having homogeneous replicate from pure strains. The results on natural samples (LAV2), obtained from the specific application of the validated UHPLC method to a replicates of lake water, further test the applicability of the developed methods to the natural samples. The LAV2 analysis showed consistent chromatographic performance across different replicates, confirming the method's suitability for real-world applications. The photosynthetic pigment composition obtained from LAV2 samples exhibited a variability consistent with the differences observed between *Anabaena* and *Synechococcus* in the initial method validation. The DoE-guided approach effectively identified the critical parameters influencing chromatographic performance, leading to an optimized UHPLC method that is both accurate and reproducible. Similarly, the extraction method proved to be consistent across different experimental conditions, with no significant variability introduced by changes in lysate buffer volume, species type, or disruption method. These results provide a foundation for the application of these methods in the detection and quantification of C-PC in natural samples, ensuring that the analytical process is both efficient and reliable for future studies. The findings of this study have important implications for the validation of satellite-derived C-phycocyanin data. The diversity and complexity of pigment compositions in natural samples underline the challenge of using spectrophotometric methods alone for accurate pigment quantification. The UHPLC method, developed and optimized through DoE, offers a reliable alternative that can be used to validate satellite observations with greater precision, particularly for C-PC, which is a key indicator of cyanobacterial presence.

5. Conclusion

HPLC is widely used for quantifying photosynthetic pigments, such as Chl *a*, to support satellite data validation. In this study, we proposed and evaluated a method for extracting C-PC from filter pads and quantifying it using an UHPLC/DAD system. The UHPLC/DAD technique was selected for its recognized performance in Chl *a* quantification and growing relevance in remote sensing applications. Our results demonstrate that this method provides good reproducibility, operates effectively on natural samples collected on filter pads—commonly used in oceanographic and limnological fieldwork—and is compatible with widely available laboratory instrumentation. These advantages support its practical adoption by research groups focused on satellite validation and pigment monitoring. The proposed UHPLC method successfully separates C-PC from Chl *a*, offering a significant improvement over traditional spectrophotometric techniques by reducing overlap and increasing quantification accuracy. We recommend the use of larger- diameter filters (e.g., 47 mm) during sampling to increase biomass collection and improve the reliability of C-PC measurements. However, several limitations were identified. The extraction efficiency may vary due to the heterogeneous distribution of cyanobacteria on filters and inconsistencies in cell density, even in controlled pure culture conditions. Additionally, while the method shows strong potential, further validation using observational datasets—including comparisons with in situ pigment concentrations from diverse aquatic environments—is necessary to fully assess its performance under varying field conditions. Future work should aim to optimize extraction

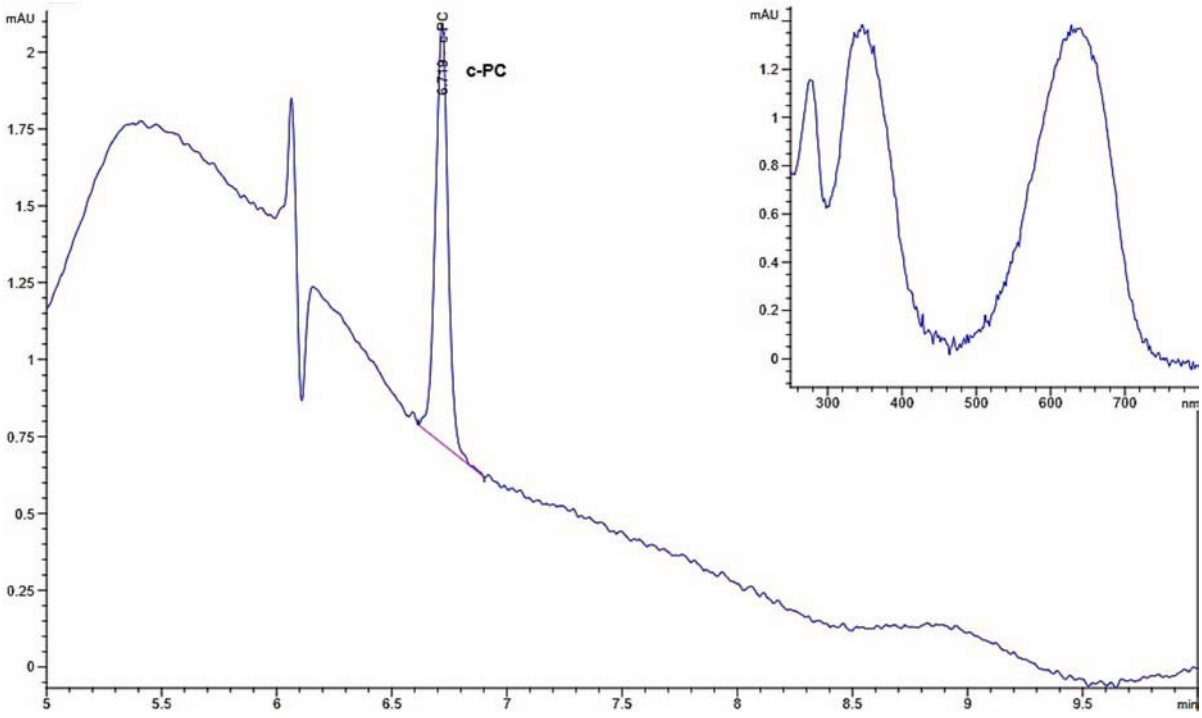


Fig. 6. Chromatographic acquisition of Lake Varese, LAV2 sample and C-PC signature.

Table 6
Results from the C5-UPLC analysis of the LAV2 natural samples.

	Extraction Volume	Volume of water filtered per sample	Peak Amount	Concentration (μg L- 1)
Sample info	V_e (mL)	V_f (mL)	C-PC	
LAV2_ly _2.5_s_a	2.5	300	0.375	31.25
LAV2_ly _2.5_s_a	2.5	300	0.358	29.83
LAV2_ly _2.5_s_a	2.5	300	0.367	30.58
LAV2_ly _2.5_s_b	2.5	300	0.410	34.17
LAV2_ly _2.5_s_c	2.5	300	0.315	26.25
LAV2_ly _2.5_m_a	2.5	300	0.329	27.42
LAV2_ly _2.5_m_b	2.5	300	0.317	26.42
LAV2_ly _2.5_m_c	2.5	300	0.225	18.75

protocols and implement standardized sampling strategies to enhance comparability and reproducibility.

In conclusion, this study presents a reproducible and accurate UHPLC method for C-PC analysis that holds promise for advancing satellite-based monitoring of cyanobacterial blooms. While further validation with observational datasets is needed, this approach contributes meaningfully to the development of robust bio-optical tools for aquatic ecosystem monitoring and water quality management.

CRediT authorship contribution statement

Elisabetta Canuti: Writing – review & editing, Writing – original draft, Visualization, Validation, Methodology, Investigation, Formal analysis, Data curation, Conceptualization.

Funding

The study has been supported by the European Commission, Joint Research Centre (JRC) and the Copernicus Program.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

We gratefully acknowledge Dr. Andrea Zaccagnini for the fruitful discussions on DoE analysis and critical aspects of analytical separation; Dr. Andrea Beghi and Dr. Franca Pandolfi from ARPA Lombardy, Italy, for their support in water sample collection during the cyanobacterial bloom in December 2023; Dr. Antonella Penna, Dr. Silvia Casabianca, Dr. Samuela Capellacci, Dr. Fabio Ricci and the Ecology Laboratory of the University 'Carlo Bo' of Urbino, Italy, for providing the *Anabaena* and *Synechococcus* cultures; Chiara Migliozzi for the English language review; and the anonymous reviewers for their constructive feedbacks.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.jchromb.2025.124759.

Data availability

Data will be made available on request.

References

[1] M. Zahir, Y. Su, M.I. Shahzad, G. Ayub, S.U. Rahman, J. Ijaz, A review on monitoring, forecasting, and early warning of harmful algal bloom, Aquaculture 593 (2024), https://doi.org/10.1016/j.aquaculture.2024.741351.

[2] H. Shen, W. Perrie, Q. Liu, Y. He, Detection of macroalgae blooms by complex SAR imagery, Mar. Pollut. Bull. 78 (2014) 190–195.

[3] J. Shen, Q. Qin, Y. Wang, M. Sisson, A data-driven modeling approach for simulating algal blooms in the tidal freshwater of James River in response to riverine nutrient loading, Ecol. Model. 398 (2019) 44–54.

[4] A.G. Sheik, A. Kumar, R. Patnaik, S. Kumari, F. Bux, Machine learning-based design and monitoring of algae blooms: recent trends and future perspectives–A short review, Crit. Rev. Environ. Sci. Technol. 54 (2024) 509–532.

[5] IOCCG (2021). Observation of harmful algal blooms with ocean colour radiometry. Bernard, S., Kudela, R., Robertson Lain, L., and Pitcher, G.C. (eds.), IOCCG Report Series, vol. No. 20, IOCCG, Dartmouth, Canada.

[6] N. Janatian, U. Raudsepp, P. Broomandi, K. Fickas, K. Olli, T. Heimovaara, A. Mannik, R. Uiboupin, N. Pahlevan, A review on remote-sensing-based harmful cyanobacterial bloom monitoring services, Remote Sens. Appl.: Soc. Environ. 37 (2025), https://doi.org/10.1016/j.rsase.2025.101488.

[7] M. Ananias, P. H. M., Negri, R. G., Bressane, A., Dias, M. A., Silva, E. a., Casaca, W., ABF: A data-driven approach for algal bloom forecasting using machine intelligence and remotely sensed data series, Software Impacts 17 (2023), https://doi.org/10.1016/j.simpa.2023.100518.

[8] I. Cetini´c, C.S. Rousseaux, I.T. Carroll, A.P. Chase, S.J. Kramer, P.J. Werdell, D. A. Siegel, H.M. Dierssen, D. Catlett, A. Neeley, I.M. Soto Ramos, J.L. Wolny, N. Sadoff, E. Urquhart, T.K. Westberry, D. Stramski, N. Pahlevan, B.N. Seegers, E. Sirk, P.K. Lange, R.A. Vandermeulen, J.R. Graff, J.G. Allen, P. Gaube, L.I. W. McKinna, S.M. McKibben, C.E. Binding, V. Sanjuan Calzado, M. Sayers, Phytoplankton composition from sPACE: requirements, opportunities, and challenges, Remote Sens. Environ. 302 (2024) 113964, https://doi.org/10.1016/j. rse.2023.113964.

[9] S. Maalouf, C. Adams, Monitoring inland water quality using remote sensing: potential and limitations of spectral indices, bio-optical simulations, machine learning, and cloud computing, Earth-Sci. Rev. 205 (2020) 103187.

[9] S. Maalouf, C. Adams, Monitoring inland water quality using remote sensing: potential and limitations of spectral indices, bio-optical simulations, machine learning, and cloud computing, Earth-Sci. Rev. 205 (2020) 103187.

[10] IOCCG (2000). Remote sensing of ocean colour in coastal, and other optically- complex, waters. Ed. by S. Sathyendranath. Vol. vol. No. 3. Reports of the International Ocean Colour Coordinating Group. Dartmouth, Canada: IOCCG.

[11] IOCCG (2010). Atmospheric correction for remotely-Sensed Ocean-colour products. Ed. by M. Wang. Vol. vol. No. 10. Reports of the International Ocean Colour Coordinating Group. Dartmouth, Canada: IOCCG.

[12] IOCCG (2014). Phytoplankton functional types from space. Ed. by S. Sathyendranath. Vol. vol. No. 15. Reports of the International Ocean Colour Coordinating Group. Dartmouth, Canada: IOCCG.

[13] R. Lauceri, M. Bresciani, A. Lami, G. Morabito, Chlorophyll a interference in phycocyanin and allophycocyanin spectrophotometric quantication, J. Limnol. 77 (2018), https://doi.org/10.4081/jlimnol.2017.1691.

[14] Y.Z. Yacobi, J. Kohler, F. Leunert, A. Gitelson, Phycocyanin-specific absorption¨ coefficient: eliminating the effect of chlorophylls absorption, Limnol. Oceanogr. Methods 13 (2015) 157–168.

[15] H. Horvath, A.W. Kov´ acs, C. Riddick, M. Pr´ esing, Extraction methods for ´ phycocyanin determination in freshwater filamentous cyanobacteria and their application in a shallow lake, Eur. J. Phycol. 48 (3) (2013) 278–286, https://doi.org/10.1080/09670262.2013.821525.

[16] JGOFS, Protocols for the joint Global Ocean flux study Core measurements. Intergovernmental oceanographic commission, scientific committee on oceanic research. Manual and Guides, UNESCO, 29, 91–96, 1994.

[17] S.B. Hooker, C.R. McClain, The calibration and validation of SeaWIFS data, Prog. Oceanogr. 45 (3–4) (2000) 427–465, https://doi.org/10.1016/S0079-6611(00) 00012-4.

[18] M. Kissoudi, I. Sarakatsianos, V. Samanidou, Isolation and purification of food- grade C-phycocyanin from Arthrospira platensis and its determination in confectionery by HPLC with diode array detection, J. Sep. Sci. 41 (4) (2018) 975–981, https://doi.org/10.1002/jssc.201701151.

[19] L. Zolla, M. Bianchetti, High-performance liquid chromatography coupled on-line with electrospray ionization mass spectrometry for the simultaneous separation and identification of the *Synechocystis* PCC 6803 phycobilisome proteins, J. Chromatogr. A 912 (2) (2001) 269–279, https://doi.org/10.1016/s0021-9673 (01)00532-5.

[20] S. Benedetti, S. Rinalducci, F. Benvenuti, S. Francogli, S. Pagliarani, L. Giorgi, M. S. Micheloni, G. D'Amici, L. Zolla, F. Canestrari, Purification and characterization of phycocyanin from blue-green alga *Aphanizomenon flos-aquae*, J. Chrom. B, Analytical technologies in the biomedical and life sciences. 833 (2006) 12–18, https://doi.org/10.1016/j.jchromb.2005.10.010.

[21] D. Kumar, P. Nehra, S. Saxena, D. Dhar, Quantification and purification of C- phycocyanin from cyanobacterial strains *Anabaena* and *Phormidium*, Ind. J. Exp. Biol. 57 (2019) 625–629.

[22] B. Wojtasiewicz, J. Ston-Egiert, 2016, bio-optical characterization of selected ´ cyanobacteria strains present in marine and freshwater ecosystems, J. Appl. Phycol. 28 (2016) 2299–2314, https://doi.org/10.1007/s10811-015-0774-3.

[23] J.A. Myers, B.S. Curtis, W.R. Curtis, Improving accuracy of cell and chromophore concentration measurements using optical density, BMC Biophys. 6 (2013) 4, https://doi.org/10.1186/2046-1682-6-4.

[24] B. Homoelle, W. Beck, Solvent accessibility of the Phycocyanobilin chromophore in the R subunit of C-Phycocyanin: implications for a molecular mechanism for inertial protein-matrix solvation dynamics, Biochemistry 36 (1997), 12970-12975 27.

[25] A. Bennett, L. Bogorad, Complementary chromatic adaptation in a filamentous blue green alga, J. Cell Biol. 58 (1973) 419–435.

[26] J. Abalde, L. Betancourt, E. Torres, A. Cid, C. Barwell, Purification and characterization of phycocyanin from the marine cyanobacterium *Synechococcus sp*. Plant Sci. 136 (1998) 1109–1120.

[27] Snyder, L. R, Kirkland, J. J., Dolan, J. W. 2009 Introduction to Modern Liquid Chromatography, 3rd Edition, John Wiley & Sons.

[28] Myers, R. H., Montgomery, D. C., Anderson-Cook, C. M. 2009 Response Surface Methodology: Process and Product Optimization Using Designed Experiments, 3rd Edition, John Wiley & Sons.

[29] Montgomery, D. C. 2019. Design and Analysis of Experiments, 10th Edition, John Wiley & Sons, Inc., New York, 2019.

[30] S. Ganorkar, A. Shirkhedkar, Design of experiments in liquid chromatography (HPLC) analysis of pharmaceuticals: analytics, applications, implications and future prospects, Rev. Anal. Chem. 36 (2017), https://doi.org/10.1515/revac- 2016-0025.

[31] D.B. Hibbert, Experimental design in chromatography: A tutorial review, J. Chromatogr. B 910 (2012) 2–13, https://doi.org/10.1016/j.

jchromb.2012.01.020.
[32] P.V. Zimba, An improved phycobilin extraction method, Harmful Algae 17 (2012) 35–39.
[33] D.E. Stewart, F.H. Farmer, Extraction, identification, and quantification of phycobiliprotein pigments from phototrophic plankton, Limnol. Oceanogr. 29 (1984) 392–397.
[34] M. Sobiechowska-Sasim, J. Ston-Egiert, A. Kosakowska, Quantitative analysis of ´ extracted phycobilin pigments in cyanobacteria - an assessment of spectrophotometric and spectrofluorometric methods, J. Appl. Phycol. 26 (5) (2014) 2065–2074, https://doi.org/10.1007/s10811-014-0244-3.
[35] E. Lawrenz, E.J. Fedewa, T.L. Richardson, Extraction protocols for the quantification of phycobilins in aqueous phytoplankton extracts, J. Appl. Phycol. 23 (2011) 865–871, https://doi.org/10.1007/s10811-010-9600-0.
[36] D.P. Jaeschke, I.R. Teixeira, L. Damasceno Ferreira Marczak, G. Domeneghini Mercali, Phycocyanin from *Spirulina*: a review of extraction methods and stability, Food Res. Int. 143 (2021), https://doi.org/10.1016/j.foodres.2021.110314.
[37] Lee, H.Y., Ryu, G.H., Choi, W.Y., Yang, W.S., Lee, H.W., Ma, C.J. Protective Effect of Water Extracted *Spirulina* maxima on Glutamate-induced Neuronal Cell Death in Mouse Hippocampal HT22 *Cell. Pharmacog*., (2018) n 14(54):242–247. doi:https://doi.org/10.4103/pm.pm_191_17.
[38] E. Canuti, Phytoplankton pigment in situ measurements uncertainty evaluation: an HPLC interlaboratory comparison with a European-scale dataset, Front. Mar. Sci. 10 (2023), https://doi.org/10.3389/fmars.2023.1197311.
[39] E. Canuti, M. Allerup, A. Bracher, V. Brotas, S. Capellacci, The 5th Workshop on High-Performance Liquid Chromatography (HPLC) Inter-Comparison on Phytoplankton Pigments (HIP), Publications Office of the European Union, 2025, https://doi.org/10.2760/1257493.
[40] Hollander, M., Wolfe, D. A., Chicken, E. 2013. Nonparametric Statistical Methods. 3rd edition. John Wiley & Sons.
[41] Student, The probable error of a mean, Biometrika 6 (1) (1908) 1–25, https://doi.org/10.2307/2331554.
[42] H.B. Mann, D.R. Whitney, On a test of whether one of two random variables is stochastically larger than the other, Ann. Math. Stat. 18 (1) (1947) 50–60, https://doi.org/10.1214/aoms/1177730491.

6. Conclusions

This work presents two ultra-high-performance liquid chromatography (UHPLC) protocols tailored for analyzing pigments in aquatic environments. These protocols have been demonstrated to be accurate, reproducible and environmental sustainable. Together, the methods together effectively quantify a broad range of diagnostic pigments, including C-phycocyanin, enabling more reliable assessments of phytoplankton functional types (PFTs) in both marine and freshwater systems.

The analytical performance meets the precision required for remote sensing validation, offering a valuable tool for improving satellite-derived products, particularly in optically complex waters where existing algorithms are limited. By ensuring compatibility with current satellite missions, the protocols facilitate the better detection and monitoring of phytoplankton dynamics, including harmful algal blooms. Looking to the future, the method will be applied to expand pigment time series in the Adriatic Sea and other coastal regions, providing essential information for ecological assessments and long-term ecosystem modeling. Integrating it into bio-optical workflows will strengthen the link between field observations and satellite-based indicators, supporting applications ranging from water quality monitoring to climate impact studies.

In summary, this study establishes scalable, field-ready pigment analysis protocols that advance both routine monitoring and scientific research, with clear pathways for ecological application and regional expansion.

References

Benedetti, S., Rinalducci, S., Benvenuti, F., Francogli, S., Pagliarani, S., Giorgi, L., Micheloni, M. S., D'Amici, G., Zolla, L., Canestrari, F. (2006) Purification and characterization of phycocyanin from blue-green alga Aphanizomenon flos-aquae. J. *Chrom. B*, Analytical technologies in the biomedical and life sciences. 833. 12-8. https://doi.org/10.1016/j.jchromb.2005.10.010

Brewin, R.J.W., Sathyendranath, S., Hirata, T. Lavender, S.J., Barciela, R.M., Hardman-Mountford, N.J., (2010) A three-component model of phytoplankton size class for the Atlantic. *Oc. Ecol. Model.,* 221 pp. 1472-1483

Canuti E., (2023) Phytoplankton pigment in situ measurements uncertainty evaluation: an HPLC interlaboratory comparison with a European-scale dataset. *Front. Mar. Sci.* 10:1197311. https://doi.org/0.3389/fmars.2023.1197311

Canuti, E., Penna, A. (2024 a) Dynamics of phytoplankton communities in the Baltic Sea: Insights from a multi-dimensional analysis of pigment and spectral data, Part I: Pigment dataset. *Front. Mar. Sci*, 11, 1425347. https://doi.org/10.3389/fmars.2024.1425347

Canuti, E., Austoni, M. (2024 b) Characterization of Phytoplankton Composition in Lake Maggiore: Integrated Chemotaxonomy for Enhanced Cyanobacteria Detection. *Microorganisms*, 12(11), 2211. https://doi.org/10.3390/microorganisms12112211

Canuti, E., Allerup, M., Bracher, A., Brotas, V., Capellacci, S., Carabelli, C., Casabianca, S., Develoouse, I., Dias, A., Dimier, C., Flander-Putrle, V., Laloux, M., Lami, A., Kenemer, C., Kumar, S., Peeken, I., Penna, A., Raman, M., Ras, J., Ricci, F., Salaün, J., Schluter, L., Suzuki, K., Thomas, C., Tirkey, A., Tracana, A., Uitz, J., Vijaya Krishna, A., Vlaemnick, B. and Zaccagnini, A. (2025 a) The 5th Workshop on High-Performance Liquid Chromatography (HPLC) Inter-comparison on Phytoplankton Pigments (HIP), Publications Office of the European Union, Luxembourg, https://data.europa.eu/doi/10.2760/1257493 , JRC140829.

Canuti, E., Cazzaniga, I., Ricci, F., Capellacci, S., Casabianca, S., Mélin, F., & Penna, A. (2025 b, submitted) Validation of Copernicus-derived chlorophyll a concentrations using an in situ time series on the northwestern Adriatic coast. [*IEEE Geoscience and Remote Sensing Letters*].

Canuti, E. (2025 c) Phytoplankton functional types in the Black Sea: A regionalized algorithm for improving ecosystem monitoring by integrating in-situ HPLC pigment analysis and satellite observations. *Mar. Pol. Bulletin*, 220, 118360. https://doi.org/10.1016/j.marpolbul.2025.118360

Canuti, E. (2025 d, under review) Satellite-Based Estimation of Phytoplankton Functional Types in the Baltic Sea Using a Regionalized Algorithm [*Ocean Modelling*]

Canuti, E., Artuso, F., Di Cicco, A. (2025 e) Comparative analysis of HPLC methods for measuring phytoplankton pigments in the Western Mediterranean Sea: A contribution to satellite Cal/Val activities. *Mar. Chem*., 270, 104516. https://doi.org/10.1016/j.marchem.2025.104516

Canuti E. and Penna A. (2025 f) Dynamics of phytoplankton communities in the Baltic Sea: insights from a multidimensional analysis of pigment and spectral data: part II, spectral dataset. *Front. Mar. Sci*. 12:1518057. https://doi.org/10.3389/fmars.2025.1518057

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

Demarcq, H., Reygondeau, G., Alvain, S., Vantrepotte, V., (2012) Monitoring marine phytoplankton seasonality from space, Remote Sens. Env., 117, 211-222, https://doi.org/10.1016/j.rse.2011.09.019.

Hooker, S.B., McClain, C.R. (2000) The calibration and validation of SeaWIFS data, *Progress In Oceanography* 45(3-4):427-465 https://doi.org/10.1016/S0079-6611(00)00012-4

Kumar Sahu, P., Rao Ramisetti, N., Cecchi, T., Swain, S., Patro, C.S, Panda, J. (2018) An overview of experimental designs in HPLC method development and validation. *J. Pharmaceutical and Biomedical Analysi*s 147 590–611

Kumar, D., Nehra, P., Saxena, S., Dhar, D. (2019). Quantification and purification of C-phycocyanin from cyanobacterial strains Anabaena and Phormidium. *Ind. J. Exp. Biol*., Vol. 57, 625-629

IOCCG (2014). Phytoplankton Functional Types from Space. Ed. by S. Sathyendranath. Vol. No. 15. Reports of the International Ocean Colour Coordinating Group. Dartmouth, Canada: IOCCG.

IOCCG (2021). Observation of Harmful Algal Blooms with Ocean Colour Radiometry. Bernard, S., Kudela, R., Robertson Lain, L. and Pitcher, G.C. (eds.), IOCCG Report Series, No. 20, IOCCG, Dartmouth, Canada.

Lauceri, R., Bresciani, M., Lami, A., Morabito, G. Chlorophyll a interference in phycocyanin and allophycocyanin spectrophotometric quantication. *J. Limn*., 77 (2018)
https://doi.org/10.4081/jlimnol.2017.1691

Yacobi, Y. Z. Köhler, J. Leunert, F. Gitelson, (2015) A. Phycocyanin-specific absorption coefficient: Eliminating the effect of chlorophylls absorption Limnol. *Oceanogr. Methods* 13, 157–168

Mackey, M., Mackey, D. Higgins, H. W., Wright, S. W., (1996) CHEMTAX - a program for estimating class abundances from chemical markers: Application to HPLC measurements of phytoplankton. *Mar. Ecol. Prog. Ser.* 144: 265– 283

Mélin, F. (2011) Comparison of SeaWiFS and MODIS time series of inherent optical properties for the Adriatic Sea, *Ocean Sci*., 7, 351–361, https://doi.org/10.5194/os-7-351-2011

Myers, R. H., Montgomery, D. C., Anderson-Cook, C. M., (2009) Response Surface Methodology: Process and Product Optimization Using Designed Experiments, 3rd Edition, John Wiley & Sons

Montgomery, D. C. (2019) Design and Analysis of Experiments, 10th Edition, John Wiley & Sons, Inc., New York, 2019

Mouw, C. B., Greb, S., Aurin, D., DiGiacomo, P. M., Lee, Z., Twardowski, M., Binding, C., Hu, C., Ma, R., Moore, T., Moses, W., & Craig, S. E. (2015). Aquatic color radiometry remote sensing of coastal and inland waters: Challenges and recommendations for future satellite missions. *Remote Sensing of Environment,* 160, 15–30. https://doi.org/10.1016/j.rse.2015.02.001

Ganorkar, S., Shirkhedkar, (2017). A. Design of experiments in liquid chromatography (HPLC) analysis of pharmaceuticals: Analytics, applications, implications and future prospects. *Reviews in Analytical Chemistry.* 36. https://doi.org/10.1515/revac-2016-0025

Hibbert, D. B.: (2012) Experimental design in chromatography: A tutorial review. *J. Chromatogr. B*, 910, 2-13. DOI https://doi.org/10.1016/j.jchromb.2012.01.020

Nogueira, E., Bravo, I., Montero, P., Díaz-Tapia, P., Calvo, S., Ben-Gigirey, B., Figueroa, R. I., Garrido, J. L., Ramilo, I., Lluch, N., Rossignoli, A. E., Riobó, P., & Rodríguez, F. (2022). HABs in coastal upwelling systems: Insights from an exceptional red tide of the toxigenic dinoflagellate Alexandrium minutum. *Ecological Indicators*, 137, 108790. https://doi.org/10.1016/j.ecolind.2022.108790

Plotka, J., Tobiszewski, M., Sulej, A.M., Kupska, M., Gorecki, T. & Namiesnik, J. (2013) Green chromatography. *J. Chrom. A*, 1307, 1–20.

Roy, S., Llewellyn, C., Egeland, E., Johnsen, G. (2011) (Eds.), Phytoplankton Pigments: Characterization, Chemotaxonomy and Applications in Oceanography (Cambridge Environmental Chemistry Series, pp. 314-342) Cambridge: Cambridge University Press. https://doi.org/10.1017/CBO9780511732263.01

Snyder, L. R. Dolan J. W. (2009). Introduction to Modern Liquid Chromatography, Third Edition, Wiley https://doi.org/10.1002/9780470508183

Sobiechowska-Sasim, M., Stoń-Egiert J, Kosakowska A. (2014) Quantitative analysis of extracted phycobilin pigments in cyanobacteria - an assessment of spectrophotometric and spectrofluorometric methods. *J. Appl. Phycol.*, 26(5):2065-2074 https://doi.org/10.1007/s10811-014-0244-3

Suzuki, K., Kamimura, A., Hooker, S.B., (2015) Rapid and highly sensitive analysis of chlorophylls and carotenoids from marine phytoplankton using ultra-high performance liquid chromatography (UHPLC) with the first derivative spectrum chromatogram (FDSC) technique, *Mar. Chem.,* 176, 96-109, https://doi.org/10.1016/j.marchem.2015.07.010.

Uitz, J., Claustre, H., Morel, A., Hooker, S.B., (2006). Vertical distribution of phytoplankton communities in open ocean: an assessment based on surface chlorophyll*. J. Geophys. Res.,* 111 p. C08005 https://doi.org/10.1029/2005JC003207

Vidussi, F., Claustre, H., Bustillos-Guzmàn, J., Cailliau, C. and J.-C. Marty (1996). Rapid HPLC method for determination of phytoplankton chemotaxinomic pigments: separation of chlorophyll a from divinylchlorophyll a and zeaxanthin from lutein. *Journal of Plankton Research*, 18, 2377-2382.

Annexes

ANNEX I Canuti, E., Cazzaniga, I., Ricci, F., Capellacci, S., Casabianca, S., Mélin, F., Penna, A. (2025, submitted) Validation of Copernicus-derived chlorophyll a concentrations using an in situ time series on the northwestern Adriatic coast. [*IEEE Geoscience and Remote Sensing Letters*].

ANNEX II Canuti, E. (2025 c) Phytoplankton functional types in the Black Sea: A regionalized algorithm for improving ecosystem monitoring by integrating in-situ HPLC pigment analysis and satellite observations. *Marine Pollution Bulletin*, 220, 118360. https://doi.org/10.1016/j.marpolbul.2025.118360

ANNEX III Canuti, E. (2025, under review) Satellite-Based Estimation of Phytoplankton Functional Types in the Baltic Sea Using a Regionalized Algorithm [*Ocean Modelling*]

ANNEX IV Canuti, E., Austoni, M. (2024) Characterization of Phytoplankton Composition in Lake Maggiore: Integrated Chemotaxonomy for Enhanced Cyanobacteria Detection. *Microorganisms*, 12(11), 2211. https://doi.org/10.3390/microorganisms12112211

ANNEX V Canuti, E., Penna, A. (2024 a) Dynamics of phytoplankton communities in the Baltic Sea: Insights from a multi-dimensional analysis of pigment and spectral data, Part I: Pigment dataset. *Front. Mar. Sci*, 11, 1425347. https://doi.org/10.3389/fmars.2024.1425347

ANNEX VI Canuti E. and Penna A. (2025 f) Dynamics of phytoplankton communities in the Baltic Sea: insights from a multidimensional analysis of pigment and spectral data: part II, spectral dataset. *Front. Mar. Sci*. 12:1518057. https://doi.org/10.3389/fmars.2025.1518057

ANNEX I

This annex has been submitted and is currently under review in IEEE Geoscience and Remote Sensing Letters

Validation of Copernicus-Derived Chlorophyll a Concentrations Using an In Situ Time Series on the Northwestern Adriatic Coast

Elisabetta Canuti, Ilaria Cazzaniga, Fabio Ricci, Samuela Capellacci, Silvia Casabianca, Frédéric Mélin and Antonella Penna

Validation of Copernicus-Derived Chlorophyll a Concentrations Using an In Situ Time Series on the Northwestern Adriatic Coast

Elisabetta Canuti, Ilaria Cazzaniga, Fabio Ricci, Samuela Capellacci, Silvia Casabianca, Frédéric Mélin and Antonella Penna

***Abstract*—Coastal waters are optically complex and ecologically dynamic environments, where chlorophyll a (Chl-a) concentration serves as a key proxy for phytoplankton biomass and water quality. In the semi-enclosed, river-influenced Adriatic Sea, routine monitoring of Chl-a is essential due to marked seasonal and interannual variability in phytoplankton productivity. This study evaluates the performance of satellite-derived Chl-a products from the Copernicus Marine Service (CMEMS) against a 16-year in situ time series collected monthly at two coastal stations off the central-western Adriatic coast. Three products were validated against in situ measurements: 100-m highresolution (HR-L3), 300-m multi-year (MY-L3), and 1 km gapfilled multi-year (MY-L4). Our results showed that the 1 km MYL4 product performed consistently across the entire dataset, making it suitable for long-term monitoring, especially at intermediate Chl-a concentrations (1–10 mg m^{-3}). The 300 m MYL3 products exhibited a higher level of correlation at nearshore sites. Conversely, the 100 m HR L3 products showed poor agreement with in situ data. Performance also varied significantly with biomass levels, with the highest uncertainties found at Chl-a concentrations exceeding 10 mg m^{-3}, which is typical of bloom conditions. Our findings provide insights into the suitability of current CMEMS products for monitoring coastal water quality and highlight the need for continuous validation in regions with complex optical characterization.**

Index Terms— Adriatic Sea; Chlorophyll-a Estimation; Copernicus Marine Service; Phytoplankton; Coastal water quality; Optical remote sensing; Ocean Color.

I. INTRODUCTION

Coastal waters are among the most dynamic and productive environments in the world's oceans. Phytoplankton blooms in these waters respond rapidly to changes in nutrient inputs, light availability, and hydrodynamic conditions. Chlorophyll-a (Chl-a) concentration is a widely used proxy for phytoplankton biomass and an indicator of ecosystem health [1], [2], as well as being used in regular monitoring programs for water-related policies, such as the European Water and Marine Strategy Framework Directives [3], [4]. Long-term monitoring of Chl-a along the Adriatic coast is especially important. This semi-enclosed basin experiences strong riverine inputs, seasonal stratification and variable anthropogenic pressures that together drive pronounced spatial and temporal variability in primary productivity [5].

This study was supported by the European Commission, Joint Research Centre (JRC) through the Wat€rs project. (Corresponding author: Elisabetta Canuti; e-mail: elisabetta.canuti@ec.europa.eu)

Elisabetta Canuti, Ilaria Cazzaniga, and Frédéric Mélin are with the European Commission, Joint Research Centre (JRC), 21027 Ispra, Italy Fabio Ricci, Samuela Capellacci, Sivia Casabianca and Antonella Penna are with the University of Urbino "Carlo Bo", Urbino, Italy.

Conventional methods of measuring Chl-a, such as spectrophotometry and fluorimetry, yield in situ data, but are limited to specific locations and infrequent sampling. In contrast, satellite Earth Observation (EO) techniques provide information on a synoptic scale at a higher frequency and at relative low cost (e.g. [6]). This enables the regular, frequent monitoring of extensive water bodies, including inaccessible areas. The potential of EO in monitoring coastal aquatic ecosystems has been amply demonstrated in recent decades ([7] and references therein). However, the uncertainties affecting satellite-derived products need always to be quantified [8], [9] to assess the suitability of these data for various applications. This fitness may in fact vary according to water type and boundary conditions.

The European Commission Copernicus Marine Service (CMEMS) provides daily observations of biogeochemical variables estimated from various optical satellite sensors at different spatial scales ([10] see Section II.B). CMEMS highresolution Chl-a products have never been validated in this area. Conversely, this area was included in the quality assessment of the coarser-resolution products in the Mediterranean Sea. According to the CMEMS Quality information Document (QUID, [11] and [12]), Chl-a products show a consistent distribution around the line of best fit, with increasing dispersion for higher values (above 0.3 mg m^{-3}). However, distinct ecological dynamics are seen in our study area due to the influence of the of Po River plume and the contribution of the Foglia River, with Chl-a values generally exceeding 0.3 mg m^{-3} and, during blooming seasons, surpassing the entire range covered by the QUID validation analysis. This suggested the need for a further quality assessment to determine whether these products are fit for purpose in monitoring water quality in the study area.

In this study, we use a unique time series of Chl-a measurements, collected at two stations off the north-central Adriatic coast at monthly intervals from January 2008 to April 2024. By pairing these in situ data with the corresponding CMEMS Chl-a products, we aim to quantify the accuracy of satellite-derived Chl-a values in this optically complex coastal environment. Our objectives are twofold: firstly, to assess the overall agreement between in situ and satellite Chl-a measurements; and secondly, to examine how uncertainties in satellite data vary with Chl-a concentration and across seasonal cycles. More generally, the applicability of CMEMS products for operational monitoring in this area is evaluated, providing

guidance for the future development of regionalized algorithms and applications.

II. Materials and Methods

A. Study area and in situ data

The study focuses on a coastal area located off the Pesaro shoreline in the Marche Region (Italy), in the central-western Adriatic Sea. This area is influenced by riverine inputs and seasonal stratification, and exhibits pronounced variability in phytoplankton biomass [13] and in optical properties [14]. In situ time-series for Chl-a are available for two stations located at 500 m (F500: 43°56′.50N, 12°54′.50E, max depth 6.5 m) and 3km (F3000: 43°56′.55N, 12°56′.18E, max depth 13 m) away from the coast, respectively.

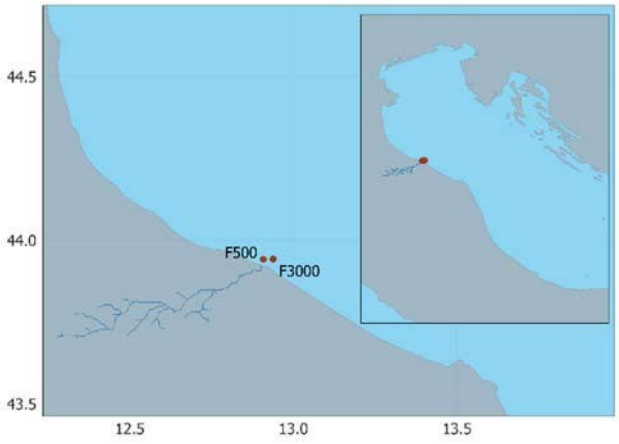


Figure 1. Study area in the central western Adriatic coast. The stations are located along a transect in front of the river Foglia at 500 m (F500) and 3000 m (F3000) from the coast.

Surface seawater samples were collected monthly from March 2008 to October 2024 at a depth of 0.5 m using Niskin bottles. The water was filtered at a rate of 1 L through Nitrocellulose filters. The samples were extracted in 90% acetone and analyzed for Chl-a content using the spectrophotometric method following the protocol described by APHA-AWWA- [15], on a double-ray UV spectrophotometer (mod. UV-1700, Shimadzu, Japan) as described in [16].

Starting from October 2021, fluorometric quantification was available for matching samples (unpublished dataset, using the [17] method). From November 2023 onwards, HPLC-based Chl-a quantification was also available, following the protocol described in [18].

In addition, at each station, phytoplankton analysis were collected [13]. The identification of taxa major groups (i.e, diatoms, dinoflagellates, and other phytoflagellates) were carried out using an inverted microscope (Zeiss Axiovert 40 CFL) equipped with phase contrast according to the [19].

B. Satellite data and match-up criteria

As mentioned, the CMEMS provides daily products at various spatial and temporal scales, combining data from different satellite Ocean Colour (OC) sensors.

Three product types were selected from the CMEMS catalogue:

a) OCEANCOLOUR_MED_BGC_L3_MY_009_143 at a resolution of 300 m (referred to as MY-L3);
b) OCEANCOLOUR_MED_BGC_L4_MY_009_144 at a resolution of 1 km (only at F3000) (MY-L4);
c) OCEANCOLOUR_MED_BGC_HR_L3_NRT_009_205 at a resolution of 100 m (HR-L3).

For the Mediterranean Sea, Chl-a values for products (a) and (b) are estimated using remote sensing reflectance (R_{RS}) values from OC products provided by the various satellite agencies. Chl-a is obtained from two different regionalized, dedicated algorithms: MedOC4.2020 [20] is used for Case I water type (open waters), and the ADOC4 algorithm [21] is used for the Case II water type (as coastal waters). When available, level-4 (L4) gap-filled products were also considered.

High-resolution (HR) products (c) are based on Sentinel-2 MSI data. R_{RS} is first obtained from Level-1 Top of Atmosphere reflectance by combining two atmospheric correction approaches: C2RCC [22] and ACOLITE [23], applied in clear and high turbid waters, respectively. Chl-a is then derived from the R_{RS} using two complementary algorithms: the OC3 empirical blue/green ratio algorithm [24], and a semi-analytic red/near infrared algorithm [25]. The latter is applied to moderate to high-biomass waters and for turbid coastal waters, whereas the former is applied to low to moderate-biomass waters and over clear to moderately turbid waters.

Matchups (i.e., quasi-synchronous satellite and in situ observations of a given area) were obtained by matching daily products with the corresponding in situ data collected on the same day. Chl-a satellite values were extracted from a 3x3 pixel window centered on the two in situ stations. The dimension of the window has been chosen to minimize coast and shallow waters disturbances and taking into account the small scale variability observed during blooming events. The 3x3 pixel values were then averaged and only retained if all 9 pixels contained valid values (i.e., not affected by clouds or other disturbances) and if the coefficient of variation (CV, the ratio of the standard deviation to the average) calculated within the 3x3 window was below 20%.

C. Validation metrics

Once identified, satellite and in situ Chl-a matchups values were compared using a series of statistical metrics. The bias is defined as the mean difference between the satellite and in situ values. The root-mean square difference (RMSD) and the mean absolute difference (MAD) were also used. A linear regression line was fitted to derive the slope, intercept, and coefficient of determination (R^2). This provided an initial assessment of the linearity and overall agreement of the satellite product with respect to in situ data. The statistical metrics were computed as follows (1,2):

$$RMSD = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(sat_i - insitu_i)^2} \quad (1)$$

$$MAD = \frac{1}{n}\sum_{i=1}^{n}|sat_i - insitu_i| \ \#(2) \quad (2)$$

where sat_i is the log-transformed Chl-a satellite value and $insitu_i$ is the log-transformed in-situ Chl-*a*.

Beyond these aggregate statistics, the way error magnitude varied with Chl-a concentration was examined by grouping the paired data into concentration intervals (bins). Within each interval, the bias and RMSD were computed to determine whether satellite performance deteriorates at low or high biomass levels.

III. RESULTS AND DISCUSSION

A. In situ time-series

In situ Chl-a concentrations measured at 3000 m from the mouth of the Foglia River ranged from 0.10 to 50.48 mg m^{-3}, while those at 500 m ranged from 0.10 to 38.30 mg m^{-3}. Both showed high temporal variability. Chl-a concentrations exhibited a marked seasonality, with the highest values typically observed in January and the lowest in August. At 3000 m, the overall mean Chl-a concentration was 3.995 mg m^{-3}, with a median of 1.91 mg m^{-3}. At 500 m, the mean Chl-a concentration was slightly lower at 3.718 mg m^{-3}, with a median of 1.78 mg m^{-3}. Certain years showed notable anomalies; for instance, in 2010, the mean Chl-a concentration at 3000 m was 9.06 mg m^{-3}, with a Z-score of 2.18, indicating a significant deviation from the typical range (Fig. 2, a).

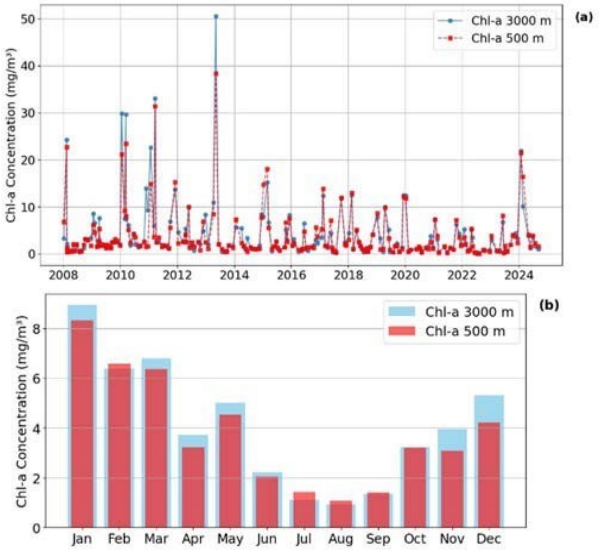


Figure 2. a) Monthly climatologies of in situ Chl-a concentration at 3000 m (blue) and 500 m (red) from the Foglia River mouth (2007–2024) and b) monthly mean Chl-a at station F3000 m and F500 m showing seasonal patterns.

In 2010, diatoms were the most representative phytoplankton taxa at stations F500 and F3000 (Skeletonema marinoi, Pseudonitzschia spp., and Chaetoceros spp.) while no significant concentrations of Dinophyceae were detected [13]. Skeletonema marinoi showed a predominance during January and February, with abundance peaks to 2.17×107 e 1.13 ×107 cell L^{-1} respectively. Pseudo-nitzschia spp. and Chaetoceros spp. were more abundant in April, with concentrations 2.06×106 e 2.92 ×106 cell L^{-1}. While the seasonal distribution of Chl-a generally followed the expected patterns at both distances (Fig. 2, b), the large interannual variability, particularly evident in years such as 2010 suggested the influence of episodic environmental factors of a multifactorial environmental matrix, in which nutrient availability, light, temperature, salinity and hydrodynamics interact on temporal and spatial scales [26].

B. Validation

The uncertainty estimates of satellite-derived Chl-a varied substantially among different products and with the distance from the coastline. It is acknowledged that for values collected at 500 m bottom effects due to relatively shallow waters (6.5 m) may affects the number of available matchups and the results themselves when satellite data are not duly masked. The same may apply to western pixels in the 3x3 window centered at F3000. Adjacency effects may also affect the results at both stations [27].

1) **HR-L3 at 100-m resolution**

The High-Resolution L3 (HR-L3) products with a 100-m resolution showed the poorest validation performance. They had low R^2 values of 0.20 at 500 m and 0.31 at 3 km from the coast and relatively high RMSD values of 0.432 mg m^{-3} and 0.399 mg m^{-3}, respectively. These statistics corresponded to in situ Chl-a concentration ranges of 0.43–11.74 mg m^{-3} (500 m) and 0.31–12.42 mg m^{-3} (3 km) (Fig. 3A, B). The best agreement was observed for in situ concentrations above 1 mg m^{-3}. However, the limited number of matchups—only 10 at 500 m and 11 at 3 km—limited the robustness of these results. Nevertheless, these results were similar to those reported by the QUID of the HR products for the Mediterranean Sea [28] which also showed a slight overestimation in the range of 1-10 mg m^{-3}.

2) **MY-L3 at 300-m resolution**

Matchups between the 300-m resolution MY-L3 products and in situ measurements also showed overestimation, with a few points aligned perfectly on the 1:1 line. This is particularly evident for the smallest in situ values, with an RMSD of 0.351 mg m^{-3}, a MAD of 0.287 mg m^{-3} and an R^2 of 0.39 across 25 matchups (Fig. 3C, D). The best alignment was shown for in situ concentration values falling within the 1-10 mg m^{-3} range. No matchups are available for in situ values greater than 10 mg m^{-3}. However, when considering the point closer to the coast (≤ 500 m), performances improved (RMSD = 0.141 mg m^{-3}, MAD = 0.127 mg m^{-3}, R^2 = 0.75), albeit based on a smaller sample size (N = 6, with Chl-a varying in the range 0.49 - 4.15 mg m^{-3}) (Fig. 3C). The very limited number of matchups (Fig. 3C) and the proximity to the coast suggest the results at 500 m should be considered as indicative rather than conclusive.

3) **MY-L4, at 1-km resolution (only at F3000)**

The MY-L4 1-km resolution gap-free products, which were only used at the F3000 station, showed the largest number of matchups (N = 171), thanks to the gap-filling techniques applied to these products. Validation metrics demonstrated moderate performance (R^2 = 0.44, RMSD = 0.353 mg m^{-3}, Fig. 3E).

TABLE I

STATISTICS ABOUT THE CHL-a (mg m^{-3}) MATCHUP DATASETS.

Products	N	RMSD	MAD	R2	Slope	Intercept	in situ Chl-a (mg m^{-3}) min min	max
(A) l3-hr_insitu500m	11	0.399	0.327	0.20	0.499	-0.002	0.43	11.74
(B) l3-hr_insitu3000m	10	0.426	0.350	0.31	0.745	-0.128	0.31	12.42
(C) l3-olci-res300m_insitu500m	6	0.141	0.127	0.75	0.957	-0.295	0.49	4.15
(D) l3-olci-res300m_insitu3000m	25	0.351	0.287	0.39	0.874	-0.223	0.1	5.3
(F) l4-MY-gapfreeres1km insitu3000m	171	0.353	0.272	0.44	1.054	-0.122	0.1	50.48

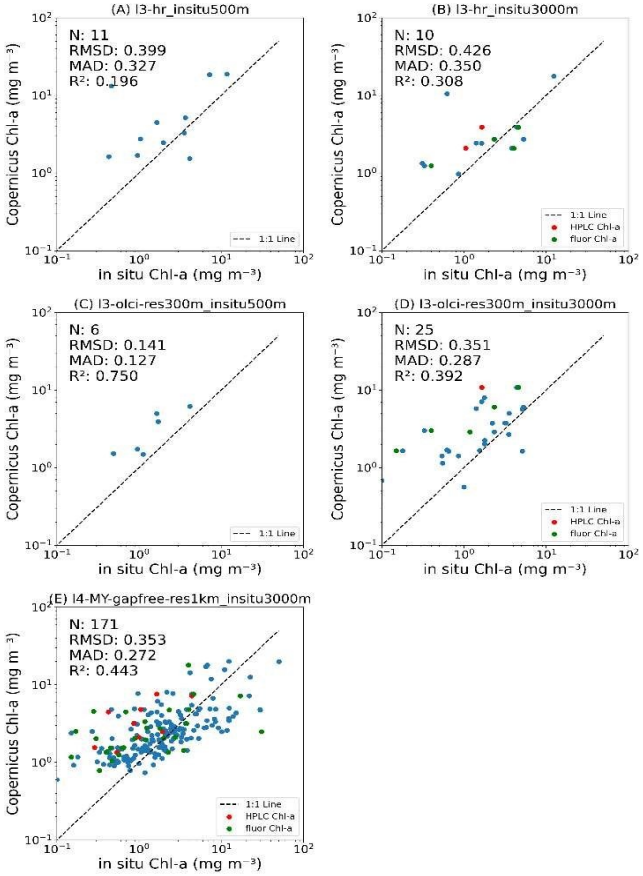


Figure 3. Comparison between satellite (y-axis) and in situ Chla (mg m^{-3}) (x-axis) data at the two points (500 m in A ad C; 3000 m in B, D and E) across multiple satellite products: (A and B) HR L3 at 100-m resolution; (C and D) MY-L3 at 300-m resolution; and (E) MY-L4 gap-free at 1-km resolution. The dashed line is the 1:1 regression line. Blue and red dots show fluorimetric and HPLC Chl-a in situ values respectively.

The analysis of the products revealed notable differences in performance across three concentration ranges. For samples with low Chl-a concentrations (less than 1 mg m^{-3}), the analysis included 46 matchups and showed moderate dispersion, with an RMSD of 0.44 mg m^{-3} and a MAD of 0.34 mg m^{-3} in $\log_{10}$transformed space. In the intermediate range (1–10 mg m^{-3}), representing the majority of the samples (N = 111), results are considerably better, with a RMSD of 0.26 mg m^{-3} and a MAD of 0.21 mg m^{-3}. These results were quite similar to those reported by the QUID for these products [29]. These results indicated quite good reliability for this range of concentrations, which were the conditions most commonly encountered in this coastal region.

Finally, for higher Chl-a concentrations (>10 mg m^{-3}, N = 14), outside the range of the in situ data used to calibrate the ADOC4 algorithm, RMSD and MAD show values of 0.61 and 0.56 mg m^{-3} respectively, with a R^2 of 0.44. This suggested an increasing uncertainties lead by an increasing complexity of the optical properties occurring in bloom conditions with enhanced concentrations of phytoplankton.

IV. Conclusion

Copernicus Marine Service Chl-a products were validated against a long-term in situ dataset from a complex nearshore environment in the Adriatic Sea, characterized by a large variability in Chl-a concentration. Our results provided practical insights into the current capabilities and limitations of these products for operational monitoring and long-term trend analysis. They may help to define the conditions under which these satellite products can be reliably used and where further improvements are needed.

All the products showed significant overestimation of the lowest in situ concentration values, i.e., below 1 mg m^{-3}. Instead, they demonstrated better agreement with in situ data in the 1-10 mg m^{-3} concentration range. The greater number of matchups available for the 1 km resolution MY-L4 gap-filled products allowed to extend their assessment to concentrations greater than 10 mg m^{-3}, not included in previous analysis. However, for these larger concentrations, which were outside the range of values used for the calibration of the Case II water algorithm, performance decreased limiting the application of these products for algal bloom monitoring.

Several factors may explain these discrepancies. First, phytoplankton concentrations ranges included in the regional algorithms calibration may not be able to capture the whole variability occurring along the Adriatic coast. Additionally, not regionalized algorithms applied to HR products may not fully capture the specific optical properties of the Adriatic, which is characterized by strong riverine input, high CDOM, and variable particulate matter. Second, bottom reflectance in optically shallow conditions and adjacency effects from land may affects satellite retrievals and uncertainties, especially within 500 m off the coast. Finally, the strong small-scale heterogeneity of coastal waters, combined with the relatively coarse spatial resolution of the satellite products, may contribute to the misalignment with in situ measurements.

The limited number of matchups obtained for the area closest to the coast, as well as the limited performances obtained in blooming conditions, suggested that further investigation and improvement to the algorithm would be required for these products to be applied efficiently in regional monitoring programs over this area, especially those targeting the surveillance of algal blooms. This further investigation will be supported by a regionally oriented validation effort, which will improve the algorithm performance in optically complex coastal waters through the integration of high-frequency in situ observations and expanded matchup databases. Upcoming activities will include targeted field campaigns and implementing machine-learning-based approaches to refine

satellite retrievals for algal bloom detection and early warning systems.

ACKNOWLEDGMENT

This study has been conducted using E.U. Copernicus Marine Service Information doi.org/10.48670/moi-00299; doi.org/10.48670/moi-00109; doi.org/10.48670/moi-00300

REFERENCES

[1] Y. Huot, M. Babin, F. Bruyant, C. Grob, M. S. Twardowski, and H. Claustre, "Relationship between photosynthetic parameters and different proxies of phytoplankton biomass in the subtropical ocean," Biogeosciences, vol. 4, pp. 853–868, 2007, doi: 10.5194/bg-4-853-2007.

[2] S. Jamshidi and N. Bin Abu Bakar, "A study on distribution of chlorophylla in the coastal waters of Anzali Port, South Caspian Sea," Ocean Sci. Discuss., vol. 8, pp. 435–451, 2011, doi: 10.5194/osd-8-435-2011.

[3] M. Chiara, D. Jean-Noël, A. Palialexis, L. Artigas, L. Boicenco, R. González-Quirós, E. Gorokhova, B. Heyden, A. McQuatters-Gollop, and I. Varkitzi, Pelagic habitats under MSFD D1: Current approaches and priorities— An overview of approaches towards D1C6 assessment. Luxembourg: Publications Office of the European Union, 2021, 45 pp., doi: 10.2760/942589. ISBN: 978-92-76-30988-8. [4] Directive 2000/60/EC of the European Parliament and of the Council of 23 October 2000 establishing a framework for Community action in the field of water policy

[5] E. K. Cherif, P. Mozetič, J. Francé, V. Flander-Putrle, J. Faganeli-Pucer, and M. Vodopivec, "Comparison of in-situ chlorophyll-a time series and Sentinel-3 Ocean and Land Color Instrument data in Slovenian national waters (Gulf of Trieste, Adriatic Sea)," Water, vol. 13, no. 14, p. 1903, 2021, doi: 10.3390/w13141903.

[6] A. N. Tyler, P. D. Hunter, E. Spyrakos, S. Groom, A. M. Constantinescu, and J. Kitchen, "Developments in Earth observation for the assessment and monitoring of inland, transitional, coastal and shelf-sea waters," Science of The Total Environment, vol. 572, pp. 1307–1321, Dec. 2016, doi: 10.1016/j.scitotenv.2016.01.020.

[7] C. Giardino, V. E. Brando, P. Gege, et al., "Imaging spectrometry of inland and coastal waters: State of the art, achievements and perspectives," Surv. Geophys., vol. 40, pp. 401–429, 2019, doi: 10.1007/s10712-018-9476-0.

[8] IOCCG (2019). Uncertainties in Ocean Colour Remote Sensing. Mélin F. (ed.), IOCCG Report Series, No. 18, International Ocean Colour Coordinating Group, Dartmouth, Canada. http://dx.doi.org/10.25607/OBP-696

[9] F. Melin, G. Sclep, T. Jackson, and S. Sathyendranath, "Uncertainty estimates of remote sensing reflectance derived from comparison of ocean color satellite data sets," Remote Sens. Environ., vol. 177, pp. 107–124, 2016, doi: 10.1016/j.rse.2016.02.014.

[10] Brando, V., Santoleri, R., Colella, S., Volpe, G., Di Cicco, A., Sammartino, M., et al. (2024). Overview of Operational Global and Regional Ocean Colour Essential Ocean Variables Within the Copernicus Marine Service. Remote Sens. 16, 4588. doi: 10.3390/rs16234588

[11] Colella, S., Brando, V. E., Cicco, A. Di, Forneris, V., and Bracaglia, M. (2025). Quality Information Document for the OCTAC Products Ocean Colour Mediterranean and Black Sea Observation Product.

[12] C. Marchese, S. Colella, V. Brando, M. L. Zoffoli, and G. Volpe, "Towards accurate L4 ocean colour products: Interpolating remote sensing reflectance via DINEOF," Int. J. Appl. Earth Obs. Geoinf., vol. 135, p. 104270, 2024, doi: 10.1016/j.jag.2024.104270.

[13] S. Casabianca, S. Capellacci, F. Ricci, M. Scardi, and A. Penna, "A phytoplankton time series in the Northwestern Adriatic Sea: Structure and dynamics of the assemblages in a coastal ecosystem," Estuar. Coast. Shelf Sci., vol. 278, p. 108109, 2022, doi: 10.1016/j.ecss.2022.108109.

[14] F. Mélin, V. Vantrepotte, M. Clerici, D. D'Alimonte, G. Zibordi, J.-F. Berthon, and E. Canuti, "Multi-sensor satellite time series of optical properties and chlorophyll-a concentration in the Adriatic Sea," Prog. Oceanogr., vol. 91, no. 3, pp. 229–244, 2011, doi: 10.1016/j.pocean.2010.12.001.

[15] APHA-AWWA-WPCF Spectrophotometric determination of chlorophyll Standard Methods for Water and Wastewater Treatment (sixteenth ed.), American Public Health Association, Washington, D.C., USA10200-H (1985), pp. 1067-1072

[16] F. Ricci, S. Capellacci, A. Campanelli, F. Grilli, M. Marini, and A. Penna, "Long-term dynamics of annual and seasonal physical and biogeochemical properties: Role of minor river discharges in the north-western Adriatic coast," Estuar. Coast. Shelf Sci., vol. 272, p. 107902, 2022, doi: 10.1016/j.ecss.2022.107902.

[17] G. Socal, I. Buttino, M. Cabrini, O. Mangoni, A. Penna, and C. Totti, Eds., Metodologie di studio del plancton marino. Rome, Italy: ISPRA, Manuali e Linee Guida, vol. 56/2010, pp. 365–377, 2010.

[18] E. Canuti, "Phytoplankton pigment in situ measurements uncertainty evaluation: An HPLC interlaboratory comparison with a European-scale dataset," Front. Mar. Sci., vol. 10, p. 1197311, 2023, doi: 10.3389/fmars.2023.1197311.

[19] L. Edler, M. Elbrachter The Utermohl method for quantitative phytoplankton analysis B. Karlson, C. Cusack, E. Bresnan (Eds.), Microscopic and Molecular Methods for Quantitative Phytoplankton Analysis. IOC Manuals and Guides, Unesco, Paris (2010), p. 110

[20] G. Volpe, S. Colella, V. E. Brando, V. Forneris, F. La Padula, A. Di Cicco, M. Sammartino, M. Bracaglia, F. Artuso, and R. Santoleri, "Mediterranean ocean colour Level 3 operational multi-sensor processing," Ocean Sci., vol. 15, pp. 127–146, 2019, doi: 10.5194/os-15-127-2019.

[21] J.-F. Berthon and G. Zibordi, "Bio-optical relationships for the northern Adriatic Sea," Int. J. Remote Sens., vol. 25, pp. 1527–1532, 2004, doi: 10.1080/01431160310001618031.

[22] R. Brockmann, M. Doerffer, K. Peters, S. Stelzer, A. Embacher, and A. Ruescas, "Evolution of the C2RCC neural network for Sentinel-2 and -3 for the retrieval of ocean colour products in normal and extreme optically complex waters," in Proc. ESA Living Planet Symp., Prague, Czech Republic, 2016, ESA SP-740.

[23] Q. Vanhellemont, "Adaptation of the dark spectrum fitting atmospheric correction for aquatic applications of the Landsat and Sentinel-2 archives," Remote Sens. Environ., vol. 225, pp. 175–192, May 2019, doi: 10.1016/j.rse.2019.03.010

[24] J. E. O'Reilly and P. J. Werdell, "Chlorophyll algorithms for ocean color sensors—OC4, OC5, and OC6," Remote Sens. Environ., vol. 229, pp. 32–47, 2019, doi: 10.1016/j.rse.2019.04.021.

[25] H. J. Gons, M. Rijkeboer, and K. G. Ruddick, "Effect of a waveband shift on chlorophyll retrieval from MERIS imagery of inland and coastal waters," J. Plankton Res., vol. 27, no. 1, pp. 125–127, Jan. 2005, doi: 10.1093/plankt/fbh151.

[26] D. Vilicic, M. Kuzmic, I. Tomažić, Z. Ljubešić, S. Bosak, R. Precali, T. Djakovac, D. Marić, and J. Godrijan, "Northern Adriatic phytoplankton response to short Po River discharge pulses during summer stratified conditions," Mar. Ecol., vol. 34, pp. 451–466, 2013, doi: 10.1111/maec.12046. [27] Barbara Bulgarelli and Giuseppe Zibordi, "Adjacency radiance around a small island: implications for system vicarious calibrations," Appl. Opt. 59, C63-C69 (2020)

[28] D. Van der Zande, K. Stelzer, C. Lebreton, A. Dille, R. Shevchuk, J. Santos, M. Böttcher, Q. Vanhellemont, and J. Scholze, Quality Information Document for High Resolution Ocean Colour Product, 2024. [Online]. Available:

[29] Brando, V. E., González Vilas, L., Di Cicco, A., Sammartino, M., Colella, S., D'Alimonte, D., et al. (2025). Quality Information Document for the Baltic Sea Observation Products. Available at: https://documentation.marine.copernicus.eu/QUID/CMEMS-OC-QUID-009131to134.pdf

ANNEX II

This annex was published in Marine Pollution Bulletin

Phytoplankton functional types in the Black Sea: A regionalized algorithm for improving ecosystem monitoring by integrating in-situ HPLC pigment analysis and satellite observations

Elisabetta Canuti

Marine Pollution Bulletin, 220, 118360. https://doi.org/10.1016/j.marpolbul.2025.118360

Marine Pollution Bulletin 220 (2025) 118360

Contents lists available at ScienceDirect

Marine Pollution Bulletin

journal homepage: www.elsevier.com/locate/marpolbu

Phytoplankton functional types in the Black Sea. A regionalized algorithm for improving ecosystem monitoring by integrating In-Situ HPLC pigment analysis and satellite observations

Elisabetta Canuti*

European Commission, Joint Research Centre (JRC), Ispra, Italy

ARTICLEINFO

Keywords:
Phytoplankton functional types (PFTs)
Black sea
Satellite ocean color
Phytoplankton pigments
High-performance liquid chromatography (HPLC)
Machine learning

ABSTRACT

Phytoplankton in the Black Sea plays a key role in regional biogeochemical cycles, but its complex optical environment pose significant challenges for the remote sensing-based monitoring. This study develops a regionalized algorithm for phytoplankton functional types (PFTs) and size classes (PSCs) specifically tailored for the Black Sea. Using high-performance liquid chromatography (HPLC) pigment data from 690 stations collected over 12 bio-optical campaigns (2006–2019), we applied hierarchical clustering, principal component analysis, and network-based community detection to identify dominant phytoplankton groups and characterize their spatial and temporal variability. We derived region-specific coefficients for estimating PFT abundances from pigment signatures and developed algorithms linking chlorophyll-*a* to PFT distributions. Applying these algorithms to a satellite-derived chlorophyll-*a* dataset from 1998 to 2024, we generated long-term climatologies revealing clear spatial patterns: microplankton dominated the nutrient-rich northwestern shelf, comprising up to 70–80 % of total chlorophyll-*a*; nanoplankton exhibited a relatively uniform distribution across the basin (~30–40 %), while picoplankton prevailed offshore, particularly in oligotrophic central and southern regions, contributing over 60 % of chlorophyll-*a* there. Our model outperformed existing global algorithms by reducing estimation errors and bias, particularly for cryptophytes and haptophytes—key functional groups in the Black Sea. Comparison with long-term microscopy data confirmed the model's robustness in capturing seasonal dynamics and ecological gradients. This work provides an improved framework for monitoring phytoplankton functional diversity in optically complex coastal basins like the Black Sea.

1. Introduction

The Black Sea is a unique marine ecosystem characterized by a diverse phytoplankton community (Silkin et al., 2021; Mikaelyan et al., 2024) that plays a fundamental role in regulating regional biogeochemical processes. It is the largest semi-enclosed anoxic basin in the world, with a two-layered structure driven by strong stratification: a well-oxygenated surface layer and an anoxic deep layer rich in hydrogen sulfide. This stratification is maintained by limited vertical mixing and significant freshwater input from rivers such as the Danube, Dniester, and Dnieper (Yakushev et al., 2007; Stewart et al., 2007). These inputs contribute to high nutrient levels, creating favorable conditions for phytoplankton growth but also making the ecosystem sensitive to eutrophication and hypoxic events. The Black Sea's ecology is shaped by its semi-enclosed nature, which restricts water exchange with the Mediterranean through the narrow Bosporus Strait (Yakushev et al., 2007). This limited connectivity influences its salinity, nutrient dynamics, and overall biodiversity. The basin is also subjected to significant anthropogenic pressures, including agricultural runoff, industrial pollution, and overfishing, which further impact its ecological balance (Aubrey et al., 1995; Oguz and Gilbert, 2007). Seasonal variations in light availability, temperature, and nutrient inputs drive pronounced changes in phytoplankton community composition, with small diatoms typically dominating during spring blooms, coccolithophores present from late spring to early summer and large diatoms becoming more prominent in summer and autumn (Silkin et al., 2021). Temporal trends show dramatic shifts in phytoplankton biomass and taxonomic composition. A peak in eutrophication and biomass occurred in the 1990s, followed by a decline as nutrient inputs decreased. Coccolithophore blooms became more prominent from the mid-1980s, coinciding with declining contributions from dinoflagellates and silicoflagellates (Cociasu et al., 2008; Mikaelyan et al., 2011, 2013). However, recent observations suggest increasing nitrate levels, pointing to a potential resurgence of eutrophication. Long-term phytoplankton taxonomic datasets are essential for understanding ecological

* Corresponding author at: European Commission, Joint Research Centre (JRC), via Enrico Fermi, 2749 Ispra, Italy.

E-mail address: elisabetta.canuti@ec.europa.eu. https://doi.org/10.1016/j.marpolbul.2025.118360

Received 12 March 2025; Received in revised form 24 June 2025; Accepted 24 June 2025
Available online 3 July 2025

dynamics and environmental impacts. However, such datasets remain limited due to the logistical and financial constraints of in situ sampling. For example, Mikaelyan et al. (2024) reviewed a long-term microscopic dataset with 523 observations spanning 75 years. Yet, microscopic identification is labor-intensive and requires specialized expertise, limiting its scalability for broad ecological assessments. High-performance liquid chromatography (HPLC) has become a widely used technique for pigment-based phytoplankton classification, offering a detailed assessment of community structure through pigment composition analysis (Mackey et al., 1996; Vidussi et al., 2001). Compared to traditional microscopic identification, HPLC-based chemotaxonomy provides a more comprehensive and scalable approach to phytoplankton taxonomic characterization, reducing the time and expertise required for species identification. This study uses a moderate-sized HPLC dataset from 690 sampling stations collected during 12 oceanographic campaigns (2006–2019) in the western Black Sea by the Joint Research Centre (JRC). The dataset covers spring, early summer, summer, and autumn, providing partial but meaningful seasonal and spatial coverage of the basin. Integrating in situ HPLC data with satellite observations enables to assess phytoplankton dynamics across broader at temporal and spatial scales. Satellite ocean color remote sensing is widely used to monitor chlorophyll-*a* (Chl-*a*) concentrations, a proxy for phytoplankton biomass (Gordon et al., 1988; Sathyendranath et al., 1994, 2001). While early efforts focused on chlorophyll retrieval, increasing attention is now given to phytoplankton functional types (PFTs) and size classes (PSCs), which reflect ecological roles and biogeochemical functions of the classified species (Falkowski et al., 1998, 2003; MacIntyre et al., 2002; Sarthou et al., 2005).

The International Ocean Colour Coordinating Group (IOCCG, 2014) has evaluated various approaches for estimating PSCs in Case-1 waters, including abundance-based algorithms (Uitz et al., 2006; Hirata et al., 2008, 2011; Brewin et al., 2010). These models perform well in oligotrophic oceanic waters but are limited in optically complex regions like the Black Sea. Here, water types are categorized as Case-1 (phytoplankton-dominated), Case-2 (influenced by CDOM and sediments, characteristic of coastal areas), and a distinct Case-3 category introduced specifically for the Black Sea, characterized by additional mineral particles such as mud, sand, and carbonate sediments in river plume zones (Morel and B´elanger, 2006; Dogliotti et al., 2015; Gregoire et al., ´ 2023). The complexity of these waters poses a significant challenge for accurate PFT algorithms development and validation.

In this study, we aim to interpret the ecological structure and variability of phytoplankton communities in the Black Sea using an integrated analysis of HPLC pigment data. Our approach combines multivariate statistics, network theory, and diagnostic pigment modeling. We (i) classify phytoplankton community structures using hierarchical clustering and principal component analysis; (ii) investigate pigment co-variability through network analysis and community detection; (iii) develop regression models for estimating phytoplankton size classes and PFTs, creating a regionally tuned algorithm for the Black Sea; and, (iv) analyze the synoptic distribution of PFTs in relation to chlorophyll-*a*, using a 1998–2024 satellite time series to assess seasonal and regional variability.

Together, these analyses offer a comprehensive view of phytoplankton community dynamics and provide insights into the ecological processes shaping the Black Sea's biogeochemistry.

2. Materials and methods

2.1. Dataset overview

This study analyzed a dataset comprising 690 sampling stations collected by the JRC during twelve oceanographic campaigns in the Black Sea between 2006 and 2019. Sampling was conducted aboard multiple research vessels, spanning various seasons—spring, early summer, summer, and autumn—and covering a wide range of environmental conditions. The campaigns targeted key sub-regions of ecological interest, including the southwestern Bulgarian coast, the northwestern shelf, and central open basins (Fig. 1; Table 1). Chlorophyll-a concentrations across the dataset ranged from 0.112 to 45.854 mg m^{-3}, with spatial coverage between 42.10°N–44.95°N and 27.81°E–33.99°E. The temporal and spatial scope of the campaigns was as follows: campaigns k02 and k07 were conducted on the northwestern shelf in spring and late summer, respectively; k06 (summer) and k09 (early summer) extended into central open waters. The remaining campaigns focused on the southwestern coastal zone: k03, k11, and k12 were in spring; k10 in early summer; k04 and k05 in summer; and k08 in early autumn. Several campaigns had previously been referenced in the literature—k01 in Eker-Develi et al. (2012), k04–k05 in Valente et al. (2022), and k07–k10 in Canuti (2023)—while campaigns k02, k03, k06, k11, and k12 are reported here for the first time.

Surface water samples were collected at all stations using Niskin bottles, and hydrographic parameters were recorded using an SBE-18 CTD (SeaBird, USA). Water volumes ranging from 100 to 2700 mL (guided by onboard fluorometric readings) were filtered through GF/F filters under mild vacuum. Filters were flash-frozen in liquid nitrogen and stored at − 80 °C until laboratory analysis at the JRC. Pigments were extracted following the method described in Canuti (2023), using 2.5 mL of α-Tocopherol (internal standard, Fluka, Germany) in 100 % acetone (HPLC grade, Merck, Germany) with the addition of 0.150 mL distilled water. Samples were sonicated for 90 s (on ice, Bandelin, Germany), incubated for 3 h at − 20 °C, and filtered (0.2 µm PTFE). A volume of 135 µL was injected into an Agilent 1200 HPLC/DAD system (USA). Chromatographic separation was performed on a C8 column (3.5 µm, 4.6 × 150 mm) at 60 °C using a ternary gradient of three mobile phases: (A) 70:30 methanol: 0.028 M TBAA, (B) methanol, and (C) acetone, at a flow rate of 1.1 mL min^{-1}. Detection was carried out at 450 nm for carotenoids and accessory chlorophylls, and at 665 nm for Chl-a and its degradation products. A total of 20 pigments were quantified across all campaigns, including: 19′-hexanoyloxyfucoxanthin (Hex), 19′-butanoyloxyfucoxanthin (But), alloxanthin (Allo), fucoxanthin (Fuco), peridinin (Perid), diatoxanthin (Diato), diadinoxanthin (Diad), zeaxanthin (Zea), chlorophyll-*a* (Chl-*a*), chlorophyll-*b* (Chl-*b*), chlorophyll-*c1* + *c2* (Chl- *c1c2*), chlorophyll-*c3* (Chl-*c3*), neoxanthin (Neo), violaxanthin (Viola), prasinoxanthin (Pras), lutein (Lut), α,β + β,β carotene (Caro), pheophorbid- *a* (Pheo), pheophythin-*a* (Phy) and chlorophyllide-*a* (Chlide-*a*) (Supplementary Table S1). Analytical quality was ensured through JRC's participation in inter-laboratory comparison exercises (Canuti et al., 2022, 2025). In addition, a subset of samples (campaigns k07–k11) was analyzed at an external ISO 17025-certified laboratory (DHI, Denmark) for independent validation (Canuti, 2023). The dataset was further quality-assured in 2023 by applying internal consistency checks in accordance with Hirata et al. (2011) and Aiken et al. (2009). Specifically, the co-variation between log-transformed Chl-*a* and total accessory pigments (TAcc)—defined as the sum of Allo, Diad, Diato, Zea, Caro, But, Fuco, Hex, Perid, Chl-*b*, and Chl-*c* (Trees et al., 2000)—was evaluated using the following criteria: (i) $R^2 > 0.9$ and slope between 0.7 and 1.4; (ii) $|\text{Chl-}a - \text{TAcc}| \leq 0.3 \times (\text{Chl-}a + \text{TAcc})$; and (iii) ≥85 % of samples meeting both criteria. All twelve campaigns satisfied these thresholds, confirming the dataset's suitability for the development of empirical algorithms targeting PFTs (Supplementary Fig. S1).

2.1. Multivariate statistic and network analysis

Hierarchical Cluster Analysis (HCA) was applied using Ward's linkage method, which minimizes the total within-cluster variance. The analysis employed a correlation distance metric defined as 1− R, where R is the Pearson correlation coefficient computed between pigment ratios across all samples. To determine a suitable cutoff for defining clusters in the dendrogram, we applied the elbow criterion, which involves examining the stepwise increases in linkage distance during the agglomerative clustering process. Each cluster was considered a proxy for a specific phytoplankton community structure, shaped by co- occurring pigments.

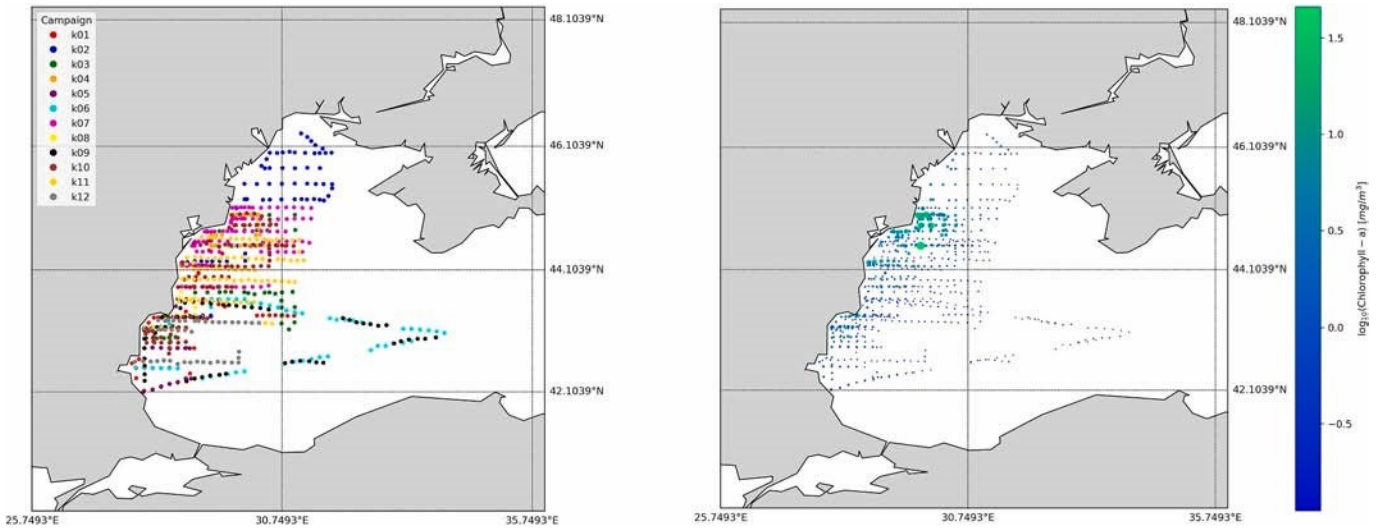


Fig. 1. Samples distribution (left) color coded by campaign name and (right) as a function of Chl-*a* concentrations.

Table 1
Black Sea HPLC pigments dataset spatial-temporal distribution.

Campaign	Days/month	Year	Latitude [˚N]		Longitude [˚E]		Chl-*a*		Sample No.	Data used in
			min	max	min	max	min	max		
k01	2–17 June	2006	42.1205	44.9467	27.7493	30.9916	0.156	28.831	93	Eker-Develi et al., 2012
k02	4–14 May	2009	43.16102	46.30847	28.004	31.75046	0.174	16.322	73	This study
k03	15–20 May	2009	42.99111	45.00444	28.03679	31.01434	0.24	10.597	39	This study
k04	1–5 July	2011	44.16327	45.00352	28.72675	30.49692	0.2537	45.8542	38	Valente et al., 2022 (only Chl-*a*)
k05	9–12 July	2011	42.10387	43.63859	27.98256	29.48383	0.237	0.617	24	Valente et al., 2022 (only Chl-*a*)
k06	13–22 July	2011	42.324	43.63443	27.82215	33.99681	0.112	0.423	59	This study
k07	31 Aug-13 Sept.	2012	43.16146	45.1214	28.0029	31.33388	0.2162	6.3828	91	Canuti, 2023
k08	13–15 Sept.	2012	43.16767	43.7557	28.01047	29.01661	0.1944	1.2529	14	Canuti, 2023
k09	3–11 June	2016	42.27876	43.60222	27.99504	33.82433	0.173	2.6448	53	Canuti, 2023
k10	12–24 June	2016	42.73941	45.02588	28.03655	30.98163	0.173	23.308	82	Canuti, 2023
k11	16–27 May	2019	43.23323	44.63809	28.68773	31.18258	0.13	4.674	80	This study
k12	28 May-4 June	2019	42.56582	43.2914	27.80909	30.28898	0.2301	0.5985	44	This study

Empirical Orthogonal Function (EOF) analysis—commonly known in statistics as Principal Component Analysis (PCA)—was employed to identify the dominant patterns of variability in the pigment dataset. PCA is a statistical technique that decomposes multivariate data into a set of linearly uncorrelated variables called principal components (PCs), which capture the maximum variance in the data along orthogonal directions. In this study, the data matrix X was constructed such that each row represents a sampling station (total of *M* observations), and each column corresponds to one of the pigment concentrations (total of *N* variables), all normalized to Chl-*a*. To remove the influence of scale, the matrix X was standardized (i.e., centered and normalized), and then decomposed using Singular Value Decomposition (SVD), (1):

$$X = U\Sigma V^T \quad (1)$$

This decomposition results in:

- U ($M \times N$): the matrix of principal component scores (u_{ik}), also referred to as loadings, representing the projection of observations onto the principal components;
- ∑ ($N \times N$): a diagonal matrix of singular values (σ_k), which are the square roots of the eigenvalues of the covariance matrix. These quantify the amount of variance captured by each principal component;
- V ($N \times N$): the matrix of principal directions or eigenvectors (v_{kj}), also called component loadings, which indicate the contribution of each original pigment variable to the corresponding principal component.

The equation can be further expanded element-wise as (2):

$$x_{ij} = \sum_{k=1,N} u_{ik}\sigma_k v_{kj}$$

Here, x_{ij} is the value of pigment *j* at station *i*, reconstructed as a weighted sum over the *N* principal modes (Supplementary Fig. S2).

EOF/PCA enables the identification of the most influential pigments associated with major patterns of variability. Typically, the first few components (or modes) explain the bulk of the variance in the data. In our case, the first four principal components were examined to (i) determine which pigments contributed most to each mode (via the loadings in V); (ii) Interpret the taxonomic meaning of each component, (ii) infer dominant phytoplankton functional groups or pigment-based assemblages. It is important to note that PCA/EOF does not assume any specific structure in the covariance of the pigments.

Network analysis was used to explore co-expression relationships among pigments across all sampling sites. First, a similarity matrix s_{ij}. was computed using the absolute value of Pearson correlation between normalized pigment vectors from sites x_i and x_j (3):

$$s_{ij} = |corr(x_i, x_j)| \quad (3)$$

This matrix was then transformed into an adjacency matrix, A = a_{ij} where each element was computed as (4):

$$a_{ij} = (s_{ij})^{\beta} \quad (4)$$

with β = 6, following the method for unsigned networks (Zhang and Horvath, 2005). The resulting weighted network connected all sites, with stronger edges indicating higher similarity. Community detection was performed on the adjacency matrix using the Louvain algorithm (Blondel et al., 2008; Dugue and Perez, 2015´), which optimizes modularity—a measure of the strength of division of the network into communities. A modularity score ≥0.3 indicated significant community structure. Each community was characterized by its dominant pigment and assigned to a phytoplankton group, following established pigment- taxonomic associations (Kramer et al., 2020; Canuti and Penna, 2024).

2.3. Determination of PFT from pigment composition: diagnostic pigment analysis

Diagnostic pigment analysis (DPA) has been employed to estimate the relative abundance of phytoplankton size classes based on specific marker pigments (Vidussi et al., 2001; Uitz et al., 2006; Hirata et al., 2008). In the original method proposed by Vidussi et al. (2001), seven diagnostic pigments were used to classify phytoplankton into three size classes: microplankton (>20 μm), including larger taxa such as diatoms and dinoflagellates; nanoplankton (2–20 μm), typically encompassing haptophytes and some cyanobacteria; and picoplankton (<2 μm), mainly consisting of small cyanobacteria and other unicellular organisms. Uitz et al. (2006) extended this approach by deriving equations to quantify the relative contributions of each size class based on a weighted sum of diagnostic pigments (ΣDP), obtained through multiple regression analysis against Chl-a concentrations. Similar formulations were subsequently developed to estimate the contributions of phytoplankton functional types. Later refinements were introduced by Hirata et al. (2008), who proposed that Chl-b should be considered a picoplankton marker only when Chl-a concentrations were below 0.25 mg m^{-3}; otherwise, it should be attributed to nanoplankton. Further methodological adjustments were made by Brewin et al. (2010) and Hirata et al. (2011), who noted that certain pigments, such as Fuco and Hex, could be associated with multiple size classes, particularly under oligotrophic conditions.

A key limitation of DPA lay in its inability to resolve taxonomic and size class ambiguities when pigments were not exclusive to a single group. For instance, fucoxanthin, while typically attributed to diatoms, did not distinguish between small-sized diatoms (often <20 μm and thus technically nanoplankton) and larger diatoms categorized as microplankton. This ambiguity often led to an overestimation of the microplankton fraction in regions dominated by small diatoms, such as the Black Sea. Similarly, picoplankton assignments were confounded by the presence of Prasinophyta, Haptophyta, and other Chlorophyta, which included picoplanktonic species. These taxa were not readily separable using DPA, which relied on fixed pigment-function-size associations.

To partially mitigate these limitations, diagnostic pigment sets and corresponding coefficients were derived from regional observations. These region-specific formulations reflected average group contributions and helped reduce potential biases. Nevertheless, the underlying assumptions of DPA needed to be considered when interpreting size- based classifications, particularly in complex ecosystems like the Black Sea where small diatoms, picoplanktonic eukaryotes were known to play significant roles and where chloroplast-bearing dinoflagellates constitute only half of the total taxon.

To model and predict phytoplankton sizes classes (nano-, pico-, and micro-phytoplankton, PSCs) and functional types in the Black Sea, we applied a regression-based approach using HPLC pigment measurements. The selection of the optimal ΣDP model for estimating Chl-*a* was guided by two distinct statistical approaches: (a) stepwise selection with standard ordinary least squares (OLS) regression, and (b) gradient descent with non-negativity constraints combined with regularized OLS. In the first approach, bidirectional stepwise selection was employed iteratively refine the predictor set based on statistical significance. Forward selection added variables (i.e., pigments) with *p*-values below 0.01, while backward elimination removed those exceeding 0.05. The selected features were incorporated into an OLS regression model without an intercept, and the final equation was defined by the estimated coefficients. Model accuracy was evaluated by calculating the mean percent difference between the predicted (i.e., ΣDP) and observed Chl-*a* concentrations. This method prioritized interpretability by retaining only statistically significant variables, resulting in a simplified regression equation. However, it did not account for the ecological relevance of features with lower statistical significance. The second approach utilized gradient descent to estimate regression coefficients while enforcing non-negativity constraints, ensuring that pigment contributions remained positive or zero. To enhance model interpretability and robustness, the estimated coefficients were then used to fit an OLS model with L1 regularization (Lasso), which selectively minimized the influence of less informative pigments by shrinking some coefficients to zero. Initial values for gradient descent, were based on the coefficients reported by Uitz et al. (2006), providing a biologically meaningful starting point for optimization. Model performance was assessed by calculating the coefficient of determination (R^2) between the predicted and observed Chl-*a* concentrations. The primary distinction between these methods lay in their strategies for feature selection and coefficient estimation. Stepwise selection emphasized statistical significance, producing a parsimonious model, while gradient descent with regularization enhanced model robustness by reducing the influence of redundant predictors and mitigating overfitting.

Finally, the selected model was evaluated using 5-fold cross- validation to assess its predictive performance and generalizability. The dataset was partitioned into five subsets; in each iteration, four subsets were used for training and one for validation, rotating through all possible combinations. This approach ensured that the model performance was not biased by any particular subset and improved its reliability for broader applications.

2.4. PFT model development

We evaluated several mathematical models commonly used in the literature (Uitz et al., 2006; Brewin et al., 2010; Hirata et al., 2011) to analyze the size distribution and functional composition of phytoplankton groups in relation to Chl-*a*. These included linear, exponential, second-order polynomial, and Michaelis-Menten (MM) equations. The linear model served as a straightforward reference, assuming direct proportionality between variables. Each model was applied to logarithmically transformed Chl-*a* values, with parameters optimized to minimize root mean square error (RMSE). The RMSE was defined as (5):

$$\mathrm{RMSE} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}$$

where: n is the number of observations (data points), y_i is the observed value for the i-th observation. $\hat{y}_i$ is the predicted value for the i-th observation generated by the mathematical model. The best-performing equations were selected based on the lowest RMSE values, allowing for accurate quantification of the relationships between Chl-*a* and size- fractionated or functional phytoplankton abundance.

2.5. PFT climatologies from satellite data

To analyze the spatial and temporal distribution of PFTs and size classes in the Black Sea, we used monthly mean composite Chl-*a* data (1 km Level-4) derived from multiple ocean color sensors (SeaWiFS, MODIS, MERIS, VIIRS, and OLCI-S3A), as provided by provided from European Commission Copernicus Marine Service (CMEMS) (doi:1 0.48670/moi-00302). These satellite products were processed using the empirical PFT model we developed based on in situ pigment data collected mainly during warm months and primarily in the western Black Sea basin.

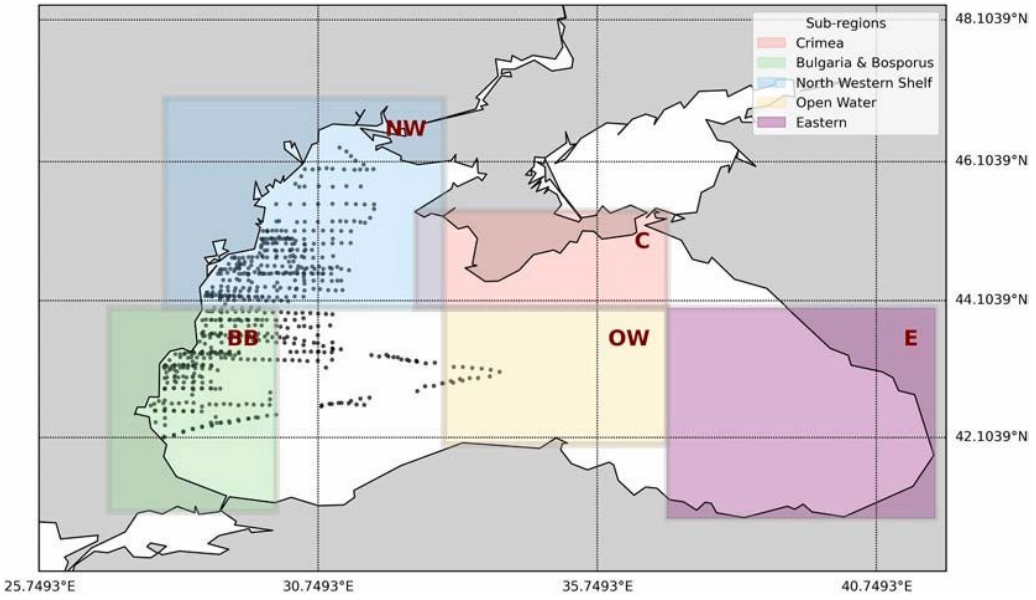


Fig. 2. Five sub-regions Black Sea considered for the seasonal and temporal trends of PFTs: C = Crimea; BB = Bulgaria & Bosporus; NW = North Western Shelf; OW = Open Water: E = Eastern.

It is important to clarify that the in situ dataset was used exclusively for model development and calibration; it was not directly used in the generation of the satellite-based synoptic relationships or climatologies. Instead, the derived PFT model was applied to the satellite Chl-*a* data to generate monthly climatologies and investigate the seasonal and spatial variability of PFTs throughout the basin. This approach also allowed us to test the transferability of the model across seasons and in areas beyond those represented in the original in situ dataset, particularly regions with different hydrographic and ecological characteristics.

We focused our analysis on five hydrographically distinct sub- regions of the Black Sea to evaluate the robustness and applicability of the model: Crimea, the Bosporus and Bulgarian coast, the Northwestern Shelf, the Eastern Basin, and the open waters. The Crimea region (43°–45.4°N; 32.5°–37°E) encompasses coastal waters influenced by strong seasonal variability due to riverine inputs and coastal currents.

The Bosporus and Bulgarian coast area (41°–44°N; 27°–30°E) is located near the Bosporus Strait, where Mediterranean inflow and Black Sea outflow generate complex dynamics. The Northwestern Shelf (44°–47°N; 28°–33°E) is a shallow zone subject to high nutrient loading from the Danube, Dniester, and Dnieper rivers, supporting high biological productivity. The Eastern Basin (37°–41.8°N; 40.9°–44°E) is more isolated and exhibits distinct circulation patterns, with less influence from western riverine sources. The open water region represents the central basin, where oceanographic conditions are relatively stable and nutrient concentrations lower than in coastal zones (Fig. 2).

Visualizations of the climatologies were produced using geospatially referenced mesh grids overlaid with color-coded PFT outputs. These results provided a synoptic view of the seasonal dynamics of phytoplankton groups and allowed for the evaluation of model performance beyond the western and summer-biased conditions of the original in situ observations.

3. Results

3.1. Pigment dataset overview

Pigment analysis across the Black Sea dataset revealed substantial variability in phytoplankton community composition and structure. Table 3 and Fig. 3 summarize descriptive statistics for the main phytoplankton pigments. Chl-*a*, the primary proxy for phytoplankton biomass, exhibited a relatively high mean concentration of 1.52 mg m^{-3} and a pronounced variability (CV: 220.8 %), reflecting strong spatial and temporal heterogeneity. Accessory chlorophylls Chl-*b* and Chl-*c* showed similar patterns, further indicating diverse phytoplankton assemblages across the basin. Among the DPs, Fuco was the most abundant (mean: 0.74 mg m^{-3}), consistent with frequent diatom dominance. Other pigments such as Peri, Hex, Allo, and Zea showed high coefficients of variation, indicating substantial spatiotemporal fluctuations in dinoflagellate, haptophyte, cryptophyte, and cyanobacterial populations, respectively. Neo and Pras, typically associated with green algae, were largely below detection limits, suggesting a minor contribution from these groups during the sampled periods. The average pigment-derived total photosynthetic biomass amounted to 2.87 mg m^{-3}, consistent with previous regional assessments. Summary statistics for individual campaigns are provided in Supplementary Table S2.

Although this study focused solely on biological data, due to the unavailability of concurrent physical and chemical measurements, the pigment composition provided valuable insight into the underlying dynamics of phytoplankton communities and regional ecological variability. In particular, the observed pigment profiles highlighted the complexity of the Black Sea ecosystem, shaped by interacting physical processes and biological adaptations.

Inter-laboratory comparisons of 179 overlapping samples analyzed independently by JRC and DHI showed general agreement but revealed systematic differences. DHI reported lower concentrations for most pigments, particularly Fuco, Peri, and Hex, while Zea concentrations were consistently higher than those from JRC. The median percent differences (e.g., Fuco: − 23.7 %, Hex: − 67 %, Zea: +47 %) suggested methodological or calibration discrepancies that were accounted for during algorithm development (Supplementary Table S3, Fig. S3).

3.2. Multivariate and network analyses

To explore coherent pigment-based groupings that could support functional phytoplankton classification, a combination of unsupervised clustering and dimensionality reduction techniques was applied. The HCA dendrogram was segmented using the elbow criterion, which identified a clear

Table 3
Black Sea dataset summary statistics for the measured pigment concentrations, including the mean, standard deviation (St. dev.), coefficient of variation (CV%), maximum, minimum, and median values. The column "Below LOD" indicates the number of samples below the limit of detection (LOD).

	Mean	St. dev.	CV%	max	min	median	Below LOD
Chl-*a* [mg m-3]	1.523	3.364	220.8	45.854	0.112	0.449	0
Chl-*b* [mg m-3]	0.060	0.111	185.5	1.288	0.000	0.028	26
Chl-*c* [mg m-3]	0.270	0.569	210.9	4.926	0.008	0.062	0
Fuco [mg m-3]	0.744	1.704	229.2	15.780	0.010	0.103	0
Peri [mg m-3]	0.092	0.221	239.4	5.073	0.000	0.039	2
Caro [mg m-3]	0.068	0.169	247.9	2.523	0.007	0.024	0
Zea [mg m-3]	0.037	0.072	193.9	1.071	0.000	0.020	40
Diad [mg m-3]	0.323	0.789	244.2	12.458	0.022	0.117	0
Diato [mg m-3]	0.082	0.181	219.8	2.794	0.002	0.029	0
But [mg m-3]	0.017	0.024	142.2	0.502	0.000	0.012	26
Hex [mg m-3]	0.165	0.159	96.4	1.957	0.000	0.105	10
Allo [mg m-3]	0.050	0.100	199.7	1.289	0.000	0.015	12
Neo [mg m-3]	0.010	0.018	177.3	0.279	0.000	0.005	76
Pras [mg m-3]	0.024	0.038	157.3	0.272	0.000	0.007	189
Viola [mg m-3]	0.020	0.037	182.5	0.529	0.000	0.011	35
Lut [mg m-3]	0.004	0.017	399.5	0.282	0.000	0.000	318
TPig [mg m-3]	3.432	6.845	199.4	92.576	0.246	1.123	
TAcc [mg m-3]	1.909	3.603	188.8	46.721	0.127	0.626	
PSC [mg m-3]	1.018	1.904	187.0	22.809	0.055	0.303	
PPC [mg m-3]	0.561	1.203	214.4	18.137	0.047	0.224	
PSP [mg m-3]	2.871	5.753	200.4	74.438	0.186	0.884	
pPF	0.168	0.125	74.2	0.678	0.000	0.160	
nPF	0.361	0.160	44.5	0.714	0.025	0.387	
mPF	0.471	0.212	45.0	0.963	0.122	0.413	

inflection point at linkage distance of 0.41 (Fig. 4). This threshold was selected to achieve optimal separation between clusters while minimizing within-cluster heterogeneity. Two principal assemblages emerged from the HCA: one characterized by high concentrations of Fuco, Allo, Zea, and Chl-*c3*, indicative of diatoms, cryptophytes, and cyanobacteria; the other characterized by elevated Peri, Chl-*b*, Chl-*c1c2*, But, and Hex, suggesting a prevalence of dinoflagellates, prasinophytes, and haptophytes. These patterns s reflect ecologically and functionally distinct phytoplankton communities likely shaped by regional environmental gradients.

PCA was employed to further resolve the underlying pigment co- variation. The first three principal components explained 70 % of the total variance. PC1, dominated by Fuco and Peri with opposing contributions from green algal markers such as Chl-*b* and lutein (Lut), thereby discriminating bloom-forming large taxa from smaller green algae. PC2 was moderately associated with Hex and Zea, representing a secondary axis of pigment variability, while PC3 captured additional variability from green algae and haptophytes.

To assess higher-order pigment interactions, a correlation-based network analysis was conducted to identify co-occurring pigment modules. Three primary pigment modules emerged: a diatom- dominated cluster (centered on Fuco), a haptophyte-associated group (centered on Hex), and a mixed assemblage including Allo, Zea, and Chl- *b*. The modularity index (0.34) indicated a robust network structure, reflecting ecologically meaningful pigment associations. The spatial and seasonal distribution of these network-defined modules revealed distinct biogeographic patterns. Diatom-rich communities were prevalent in early autumn, particularly in offshore areas aligning with nutrient- enhanced post-stratification conditions. Haptophyte-associated assemblages were more frequent in summer. Mixed communities, indicative of transitional or coastal systems, were common in the eastern basin and Case 1 waters (Fig. 5).

Together, these multivariate and network-based results provided a quantitative foundation for delineating regionally representative PFTs. The identified pigment modules informed the design and structure of the empirical classification algorithm described in the subsequent sections.

3.3. PFT from pigment composition: diagnostic pigment analysis

Two regression models were applied to derive regionalized coefficients for estimating diagnostic pigments under the assumption that ΣDP should equal Chl-a (Uitz et al., 2006). The first model used ordinary least squares (OLS) with a stepwise approach and produced the following equation:

$$\Sigma DP\text{− OLS} = 1.24\cdot \text{Fuco} + 2.15\cdot \text{Peri} + 4.99\cdot \text{Allo} + 2.90\cdot \text{Chl− b} + 3.48\cdot \text{Zea} \quad (6)$$

Hex and But were excluded due to low significance ($p > 0.005$). The model achieved an R^2 of 0.91, with a mean percent difference of 26.82 % between ΣDP and Chl-*a*.

To improve representation of haptophytes, the DESC model (OLS with non-negativity constraints and regularization) was tested. It retained Hex and But in the equation:

$$\Sigma DP - DESC = 1.33\cdot \text{Fuco} + 2.42\cdot \text{Peri} + 2.94\cdot \text{Allo} + 2.78\cdot \text{Chl} - \text{b} + 0.15\cdot \text{Hex} + 0.34\cdot \text{But} + 2.30\cdot \text{Zea} \quad (7)$$

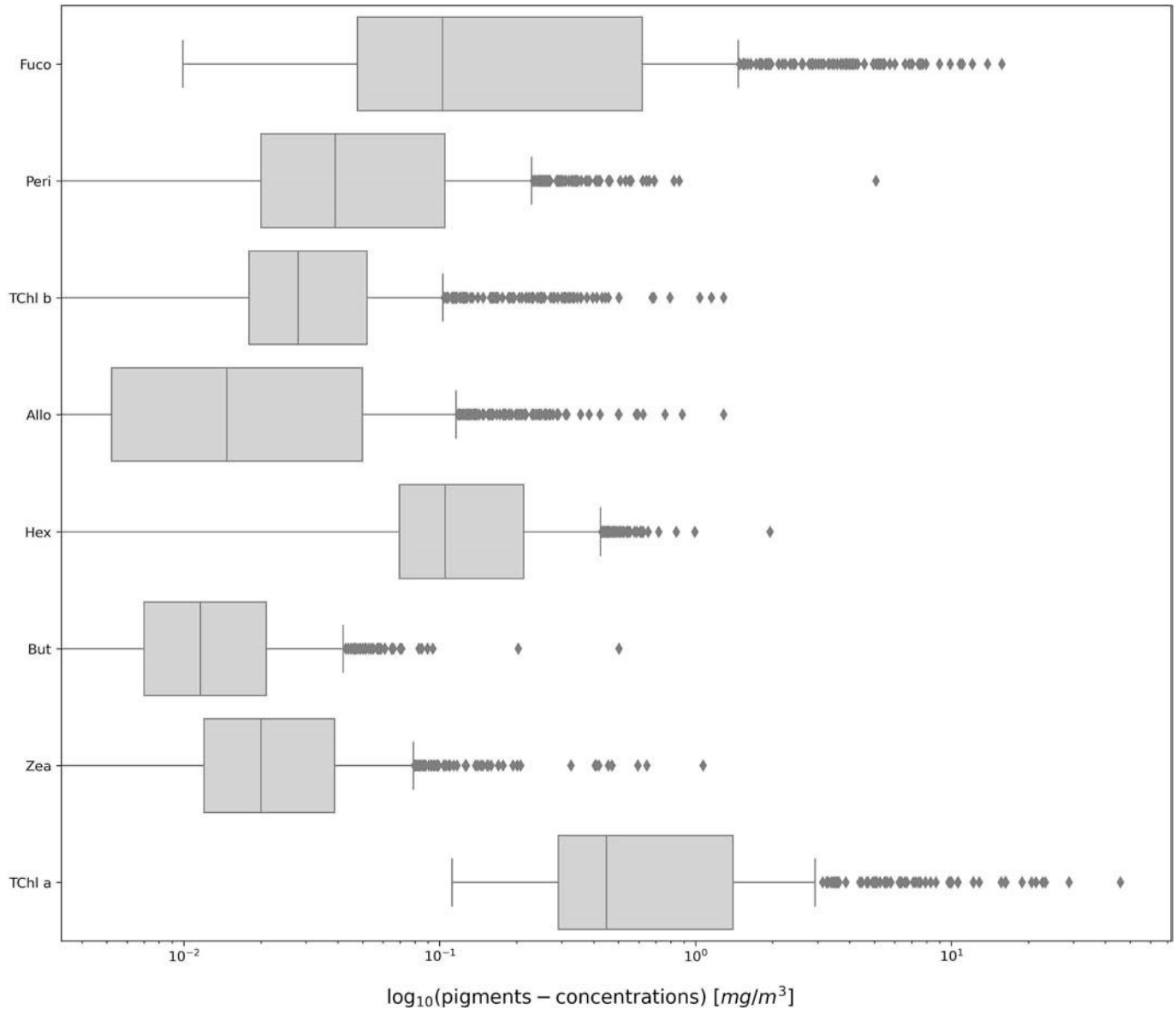


Fig. 3. Boxplot of phytoplankton pigment concentrations (mg m^{-3}) on a $\log_{10}$ scale. The distribution of each pigment is shown with median values, interquartile ranges, and outliers. Pigments include Fuco, Peri, Chl-*b*, Allo, Hex, But, Zea, diagnostic pigments and Chl-*a*. The variability in pigment concentrations reflects differences in phytoplankton community composition across sampled environments.

However, Hex ($p = 0.33$) and But ($p = 0.257$) did not contribute significantly. DESC achieved a similar R^2 of 0.91, with slightly lower RMSE and MAE than OLS, but cross-validation showed a reduced mean R^2 of 0.82. AIC and BIC values were also higher than OLS (Supplementary Table S4).

Both OLS and DESC models were applied to the DHI and JRC datasets. The DHI OLS model included Fuco, Allo, Hex, and Zea, while the JRC OLS model used Fuco, Peri, Allo, Chl-*b*, and Zea. The DHI OLS achieved a higher R^2 of 0.94, compared to 0.912 for the JRC OLS. Model coefficients differed notably between datasets (Supplementary Table S4).

Among various DPA formulations tested, only the original Uitz et al. (2006) model (UITZ) was retained for comparison. Alternative variants yielded similar results (not shown). Both OLS and DESC models provided a closer fit to Chl-*a* than UITZ, with weighted ΣDP values more strongly correlated to Chl-*a* (Supplementary Fig. S4, Table 4).

3.4. PFTs model for the Black Sea

The weighted ΣDP coefficients from OLS, DESC, and UITZ were used to derive regionally tuned equations for estimating PFT fractions (Supplementary Table S5). In model validation, the DESC model achieved high R^2 values for microphytoplankton (0.98) and diatoms (0.95), moderate for cryptophytes (0.36) and haptophytes (0.23), and low for dinoflagellates (0.14) and picoplankton (0.25) (Supplementary Table S5, Fig. 6, bottom panel). The RMSE, which gives a measure of the spread of the estimated values around the in-situ observed ones, goes from 0.018 mg m^{-3} for the "Green algae & Prochlorophytes" to 0.068 mg m^{-3} for the Diatoms in the PFT group and from 0.042 mg m^{-3} for the pico- to 0.070 mg m^{-3} for the micro-phytoplankton in the PSCs. The regional relationships between in-situ Chl-*a* and the fraction of each PFT was presented in Fig. 6 (upper panels) Microplankton and diatoms showed increasing trends with Chl-*a*. Dinoflagellates, cryptophytes, and haptophytes displayed variable or decreasing trends. Picoplankton contributions decreased with increasing Chl-*a*. Predictive performance was strong for microplankton and diatoms, moderate for nanoplankton and green algae, and lower for cryptophytes, haptophytes, and dinoflagellates.

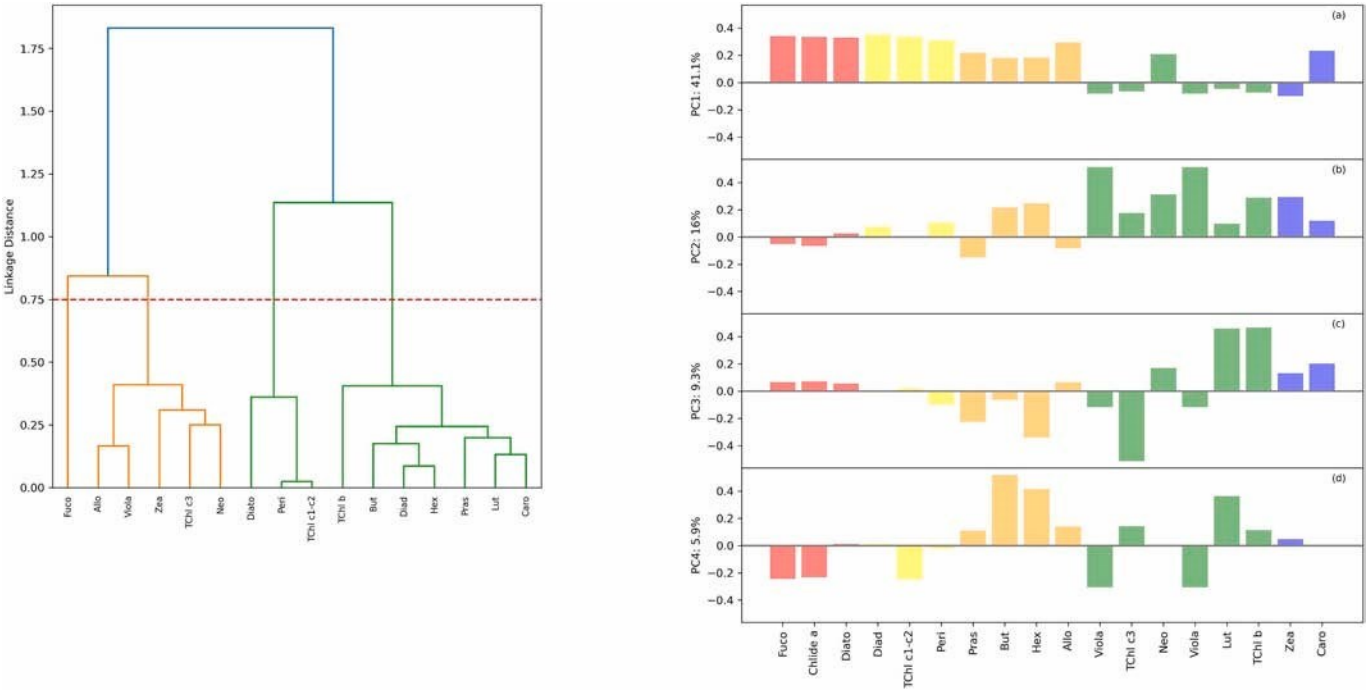


Fig. 4. Multivariate analysis results: (left) Hierarchical clustering dendrogram of phytoplankton pigments based on similarity in their distributions. The red dashed line indicates the cut-off threshold used to define clusters. (right) Principal Component Analysis (PCA) loadings of pigments for the first four principal components (PC1–PC4, letter a-d). The percentage of explained variance for each component is shown on the y-axis. Pigments are color-coded according to their functional grouping: diatoms - micro (red); green algae (green); dinoflagellates (yellow); haptophytes - nano (orange) and pico (blue). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

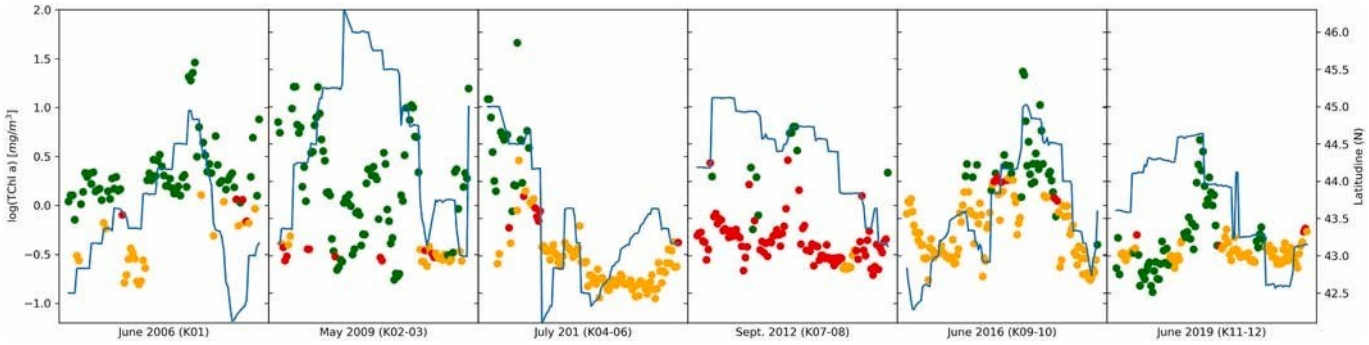


Fig. 5. Latitudinal distribution of log-transformed total chlorophyll *a* ($\log_{10}$[Chl-*a*], mg m^{-3}) across different sampling years, overlaid with community partitioning based on predominant pigment composition and inferred phytoplankton functional types (PFTs). The blue line represents latitude (right y-axis), while data points are color-coded according to the dominant pigment signature: red for Fuco-dominated communities (diatoms), orange for Hex-dominated communities (haptophytes), and green for mixed (green algae and cryophytes). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

Table 4
Equations used to estimate the relative contributions of microplankton (mPF), nanoplankton (nPF), and picoplankton (pPF) to the total diagnostic pigments (ΣDP) using different approaches: UITZ, OLS, and DESC. The coefficients represent the weighting factors assigned to each pigment in the calculation.

		Equations to estimate PSCs		
	Functional Type	UITZ	OLS	DESC
mPF	Microplankton (>20 μm)	1.41· (Fuco+Peri)/ΣDP	(1.24· Fuco+2.15· Peri)/ΣDP	(1.33· Fuco+2.42· Peri)/ΣDP
nPF	Nanoplankton (2–20 μm)	(1.27· Hex+1.01· Chl*b* + 0.35· But+0.60· Allo)/ ΣDP	(4.99· Allo+2.90· Chl*b*)/ΣDP	(2.94· Allo+2.78· Chl*b* + 0.15· Hex+0.34· But)/ΣDP
pPF	Picoplankton (0.2–2 μm)	0.86 Zea/ΣDP	3.48· Zea/ΣDP	2.30· Zea/ΣDP
ΣDP		1.41· Fuco+1.41· Peri+1.27· Hex+1.01· Chl*b* + 0.35· But+0.60· Allo+0.86· Zea	1.24· Fuco+2.15· Peri+4.99· Allo+2.90· Chl*b* + 3.48· Zea	1.33· Fuco+2.42· Peri+2.94· Allo+2.78· Chl*b* + 0.15· Hex+0.34*But+2.30· Zea

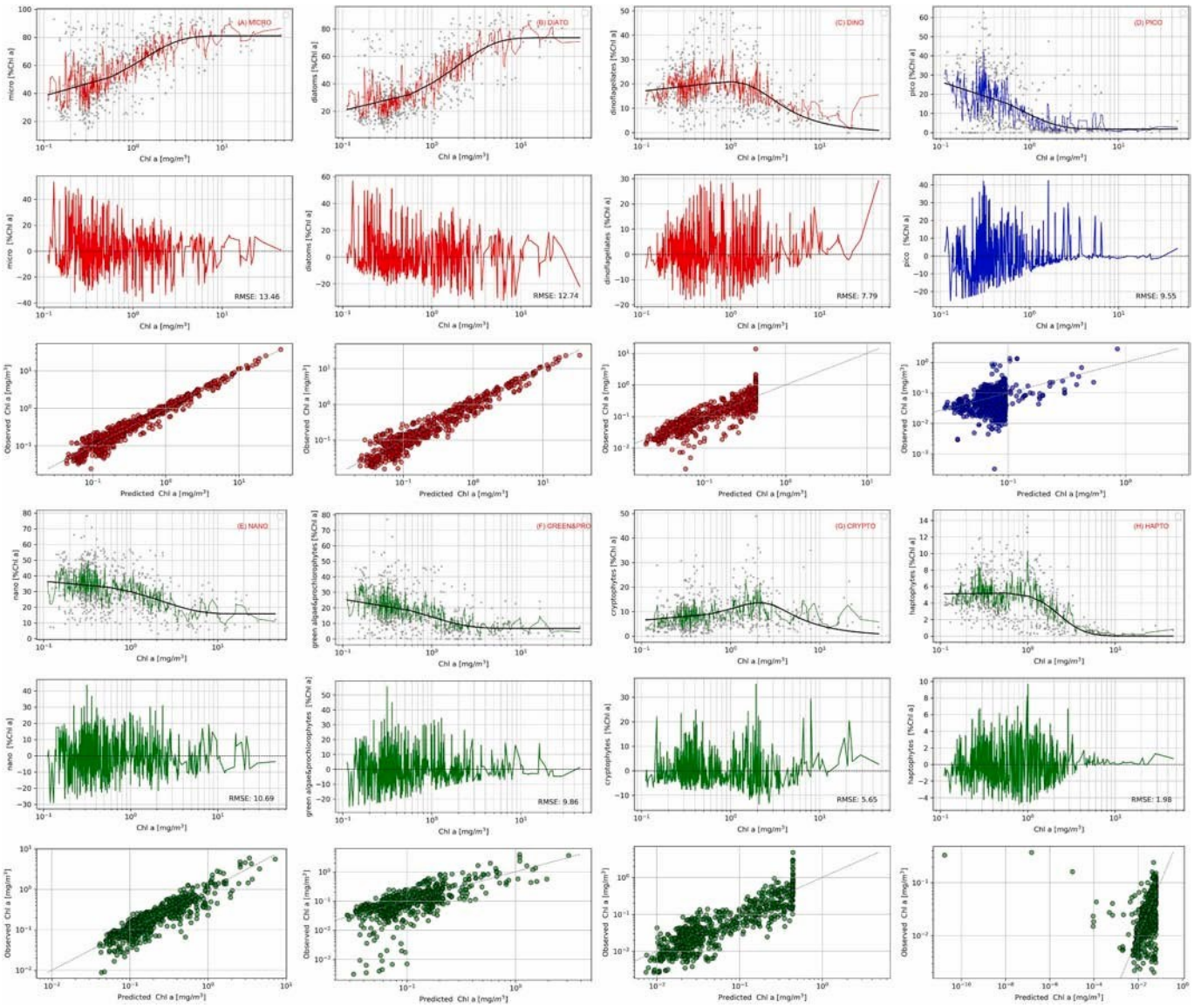


Fig. 6. Regional relationships (DESC-PFT model) between in-situ Chl-*a* and the fraction of each PFT Three panel figures for (A) microplankton, (B) diatoms, (C) dinoflagellates, (D) picoplankton, (E) nanoplankton, (F) green algae & prochlorophytes, (G) cryptophytes, (H) haptohytes, each consists of (upper panel) synoptic relationship PFTs, PSCs, as a fraction of %Chl-*a*, black line representing fitted model, continuous colored line the mean variance.; (middle panel) uncertainties of the synoptic relationship as residual of $X_{obs} - X_{pred}$ for each PFT, and; (lower panel) the results of validation, predicted vs observed %Chl-*a*.

Table 5
Relationships between log-transformed chlorophyll-*a* concentration (x = $\log_{10}$ [Chl-*a*]) and various PSCs and PFTs. The table presents the parameters for each equation, including exponential for micro, dinoflagellates, nano, and green & prochlorophytes, and MM-modified for diatoms, cryptophytes, and pico- phytoplankton.

DESC	Equation	Parameters		
		a	b	c
micro	a· exp.(− b· x) + c	− 0.460	0.785	0.811
diatoms	micro-dino	- -	- -	- -
dinoflagellates	1 / (exp(a· x + b) + c· x)	− 0.921	1.818	2.329
nano	a· exp.(− b· x) + c	0.215	0.427	0.158
green&prochl	a· exp.(− b· x) + c	0.206	1.018	0.067
cryptophytes	1 / (exp(a· x + b) + c· x)	− 0.891	2.805	2.272
haptophytes	1 / (exp(a· x + b) + c· x)	0.558	2.989	− 14.190
pico	a· exp.(− b· x) + c	0.276	1.292	0.018

3.5. Application of the regional algorithms to the climatologies from Satellite Data (1998–2024)

The newly developed regional algorithms (Table 5) were applied to a 26-year time series of satellite-derived Chl-*a* estimates (see Section 2.5) to derive pigment-based PFTs across the Black Sea. This long-term dataset enabled the assessment of spatial and temporal variability in phytoplankton community composition at basin-wide and sub-regional scales. Fig. 7 presents the annual climatologies of PFTs for the period 1998–2024. The spatial distribution of eight PFTs in the Black Sea was shown as a percentage of chlorophyll-*a* (%Chl-*a*). Microplankton (Fig. 7a) dominated the northwestern shelf and coastal areas, particularly around 45°N and 30–33°E, where their contribution exceeded 60 % of Chl-*a*. Dinoflagellates (Fig. 7 b) were concentrated in the northern coastal regions, with maximum values reaching approximately 20 % Chl-*a*.

Diatoms (Fig. 7c) followed a similar pattern to microplankton, with peak values over 20 % in the north and significantly lower values offshore, often falling below 5 %. In contrast, nanoplankton (Fig. 7d) were more evenly distributed, contributing between 35 % and 45 % across much of the basin, with localized depressions in the northern shelf. Cryptophytes (Fig. 7e) exhibited a broad distribution, with contributions ranging from 10 % to 35 %, and maxima occurring between 44°N and 45°N and 30–35°E. Haptophytes (Fig. 7f) were more spatially limited, with values generally below 12 % Chl-*a*, and showed minor presence in northern and central regions. Green algae and prochlorophytes (Fig. 7g) were consistently detected in the southern offshore waters, where their relative abundance hovered around 4–5 %. Picoplankton (Fig. 7h) showed a strong offshore signal, particularly in the southeastern basin between 41°N and 43°N and 36–42°E, where their contribution exceeded 80 % Chl-*a*. Overall, coastal regions, especially the western and northwestern shelves, exhibited the highest phytoplankton biomass levels.

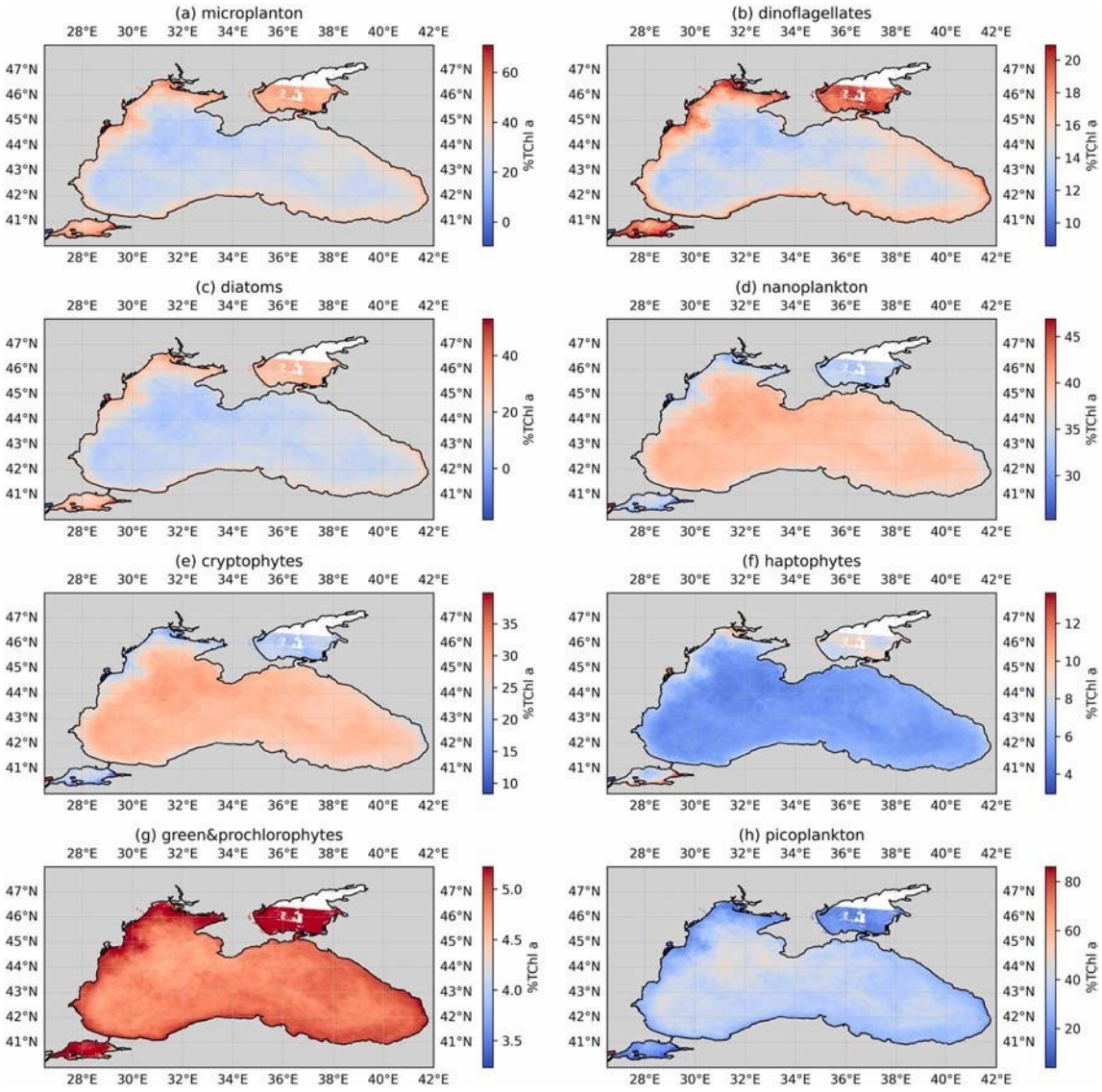


Fig. 7. Synoptic distribution of surface PFTs [% Chl-*a*] and PSC over 1997–2024 derived from CMEMS for the Black Sea for (a) microplanton, (b) dinoflagellates, (c) diatoms, (d) nanoplankton, (e) cryptophytes, (f) haptophytes, (g) green&prochlorophytes, (h) picoplankton.

To complement the spatial climatology, interannual trends and peak years of key PFTs were investigated across the basin and within ecologically relevant sub-regions (Fig. 8). In the Crimea, Open Water, and Eastern sub-regions, peaks in diatoms, green algae, and nanophytoplankton were observed in 1999 and 2001, with additional peals in 2010 in Crimea and Open Water. In the Bosporus region, dinoflagellates concentrations were highly variable, peaking in 1999 (0.212), declining to 0.072 in 2017, and increasing again to 0.130 in 2024. Diatoms concentration in the same area followed a similar trend, with a peak in 1999 (0.307), a decline until 2012, and a gradual rise to 0.153 by 2024. In the Northwestern Shelf, diatom concentrations were particularly high, with sustained elevated values between 2005 and 2009, followed by fluctuations and a marked increase from 2022, reaching 0.475 in 2024. Across the Black Sea, the period from 2003 to 2006 was characterized by widespread elevated concentrations of diatoms, suggesting a basin-wide dominance during this interval.

In addition, monthly climatologies were constructed (Fig. 9) to examine the seasonal dynamics of each PFT across the entire Black Sea and the five sub-regions. These analyses revealed well-defined seasonal succession patterns. Microphytoplankton exhibited stable concentrations in winter and early spring, followed by an increase from June and a peak in October, before declining toward winter. Nanophytoplankton ncreased gradually from winter to a maximum in August, highlighting their prevalence during warmer months. Cryptophytes showed relatively stable levels year-round, with moderate increases between June

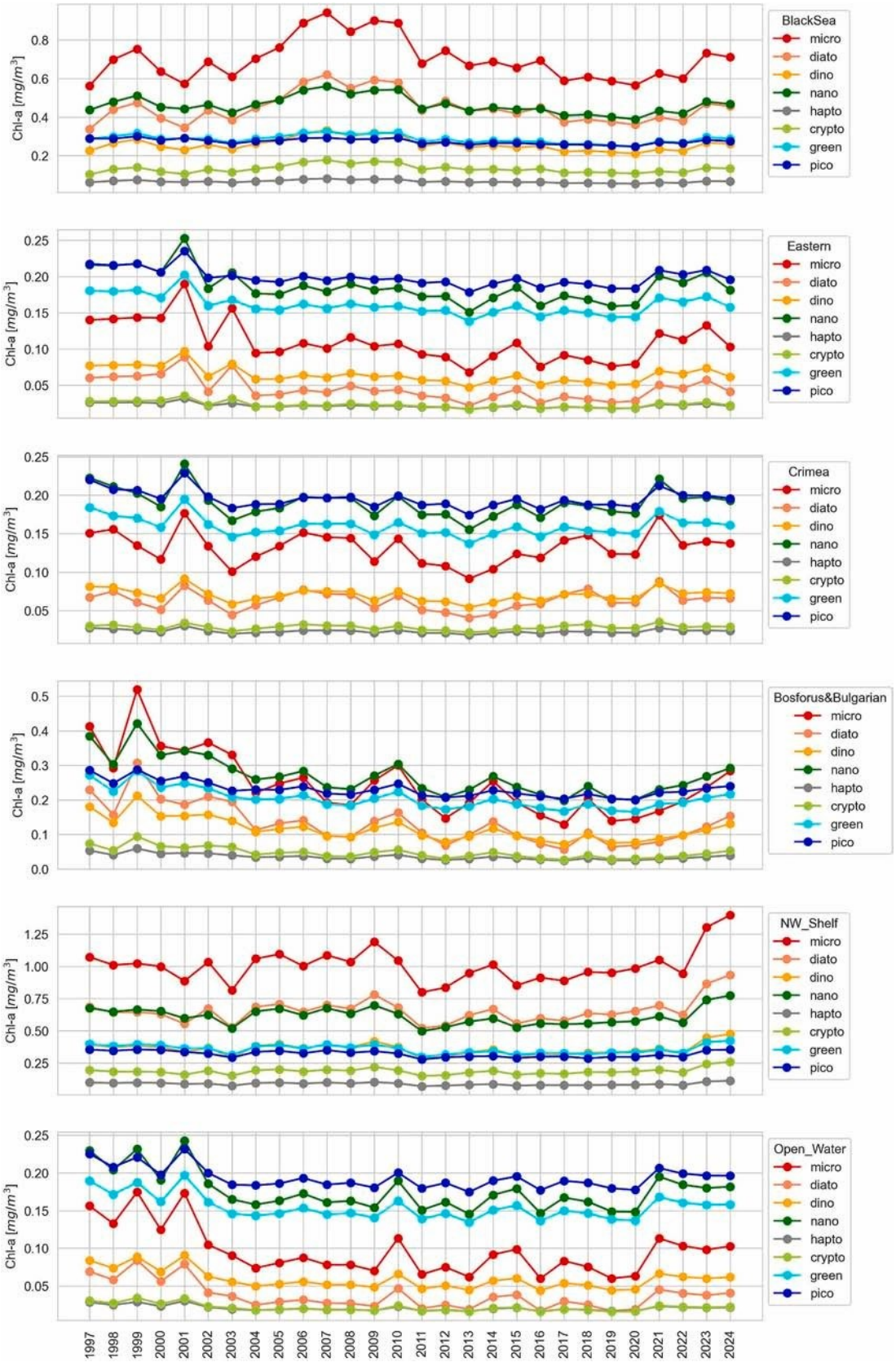


Fig. 8. Yearly contribution of PSCs and PFTs' in the Black Sea (a), with details for the following sub-regions: Eastern (b), Crimea (c), Bosporus (d), NW Shelf (e), and Open Water (f).

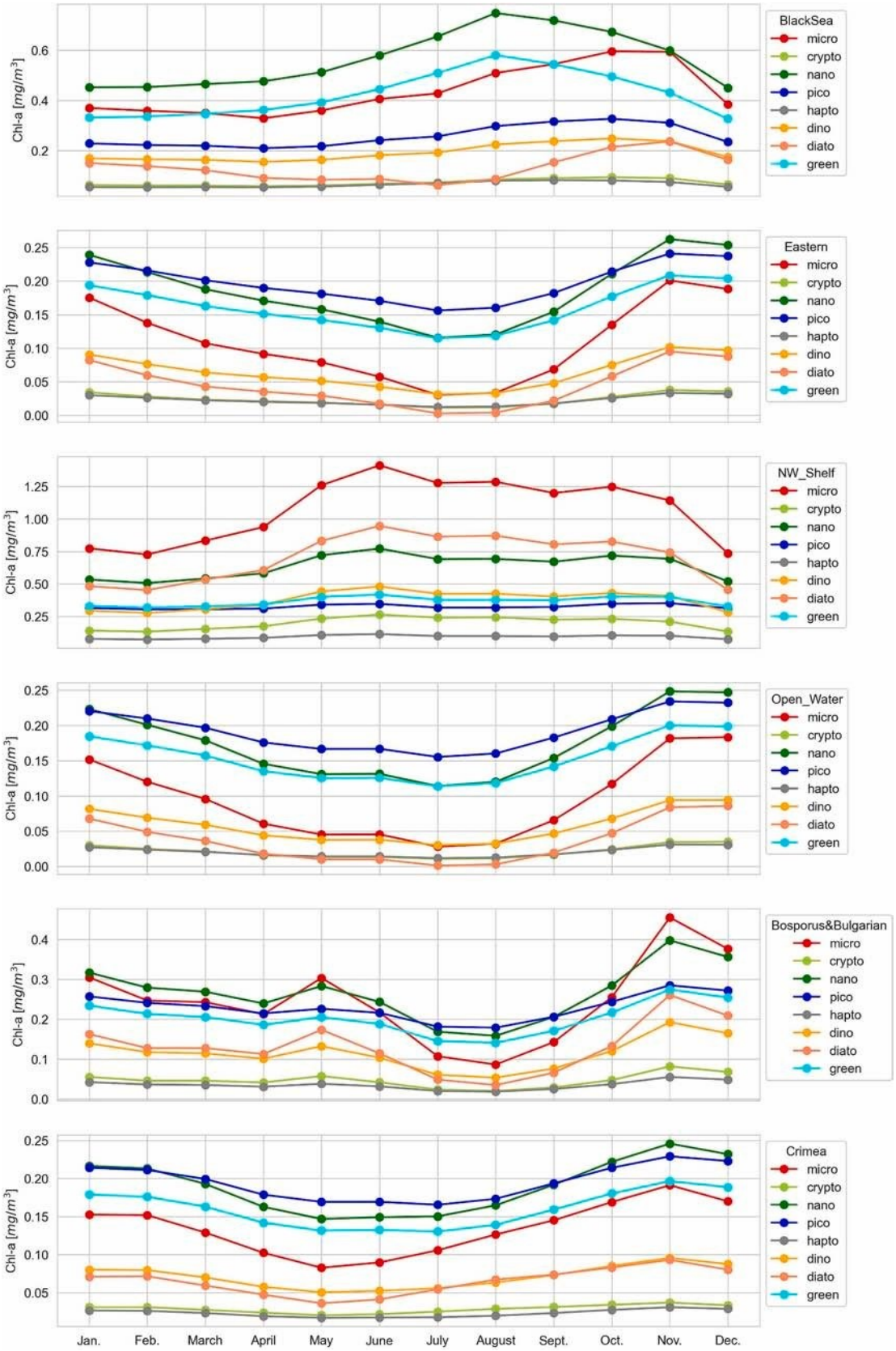


Fig. 9. Monthly climatology for PSCs and PFTs' in the Black Sea (a), with details for the following sub-regions: Eastern (b), Crimea (c), Bosporus (d), NW Shelf (e), and Open Water (f).

and October, peaking in autumn. Haptophytes followed a similar pattern, with a peak in early autumn. Picophytoplankton exhibited a steady rise from winter to a peak in October, before decreasing. Dinoflagellates gradually increased during summer, peaking in October and subsequently declining toward

winter. Spatially-resolved seasonal patterns revealed additional regional differences in community structure. In the Crimea coastal region, microphytoplankton, diatoms, and picophytoplankton peaked in late summer to early autumn, with pico- and microphytoplankton maintaining relatively high levels year-round. Cryptophytes and nanophytoplankton exhibited modest fluctuations, with minor increases in spring. In the Bosporus Strait, seasonal variability was more pronounced. Diatoms and microphytoplankton peaked in May, while nanophytoplankton and cryptophytes increased during summer. Conversely, picophytoplankton and dinoflagellates decreased in summer but increased again in autumn. The Northwestern Shelf, influenced by riverine inputs, consistently supported high concentrations of microphytoplankton and diatoms, with extended peaks from May to October. Picophytoplankton remained relatively stable, with higher concentrations in winter and lower values in summer. Nanophytoplankton, cryptophytes, and haptophytes were present at moderate concentration throughout the year. In the Open Water and Eastern regions, seasonal patterns was less marked. Picophytoplankton and diatoms dominated throughout the year, with gradual increases from summer into autumn. Microphytoplankton showed a secondary peak in late autumn, likely associated with vertical mixing. Haptophytes and cryptophytes remained at lower concentrations compared to coastal regions.

4. Discussion

The Black Sea represents a distinctive semi-enclosed marine ecosystem, characterized by strong vertical stratification, substantial riverine input, and limited water exchange through the Bosporus Strait. These features, shape its distinct optical and biogeochemical characteristics, including elevated concentrations of colored dissolved organic matter (CDOM), suspended particulate matter, and spatially heterogeneous phytoplankton communities. Together, they affect the bio-optical signal, complicating the remote sensing retrieval of standard ocean color products, such as chlorophyll-*a* (Dogliotti et al., 2015). In particular, the northwestern shelf —dominated by coastal and estuarine influences—exhibits pronounced spatial and temporal variability in phytoplankton assemblages and trophic conditions. Here, diatoms, dinoflagellates, and haptophytes alternately dominate, responding dynamically to seasonal cycles and physical forcing (Moncheva et al., 2001; Mikaelyan et al., 2013, 2017, 2024). Global ocean color algorithms generally fail to resolve such complexity, lacking in estimating the contribution of specific PFTs and PSCs. Despite the clear need, regionally-tuned algorithms for PFT and PSC estimation in the Black Sea remain scarce.

In this study, we developed and validated novel empirical algorithms for the estimation of both PFTs and PSCs, using in situ HPLC pigment data in conjunction with satellite-derived remote sensing reflectance (R_{rs}). These algorithms were constructed based on established relationships between diagnostic pigments and R_{rs} across the Black Sea's bio-optical gradients and incorporating ecological distinctions among phytoplankton groups. We compared model performance using different methods for estimating ΣDP coefficients. The UITZ method (representing global coefficients) performed best for microphytoplankton and diatoms, whereas DESC—a regionalized approach developed in this work—demonstrated superior performance for cryptophytes and haptophytes. Among all tested models, DESC provided the most balanced performance across functional groups and, through its inclusion of Hex, proved particularly well-suited for estimating haptophytes—a taxon of ecological significance in the Black Sea. Compared with the CMEMS empirical exponential algorithms (CMEMS-OC-QUID-009-141to144-151to154), which rely on the original PSC classification by Vidussi et al. (2001), our approach demonstrated improved performance for both nano- and pico-sized phytoplankton. For nanophytoplankton, DESC achieved an R^2 of 0.76, exceeding the 0.67 obtained by CMEMS. For picophytoplankton, the improvement was even more pronounced (R^2 = 0.25 vs. 0.16), underscoring the benefit of regionalization.

However, it is important to recognize the inherent limitations of diagnostic pigment-based approaches in resolving the PFTs within the Black Sea context. For example, Fuco, a key pigment associated with diatoms and the microphytoplankton fraction, cannot discriminate diatom size classes—an important factor given the basin's ecological variability.

Similarly, pigments assigned to the picophytoplankton fraction may originate from diverse taxa, reducing classification specificity. Additionally, our pigment-based method does not capture heterotrophic groups such as non-pigmented dinoflagellates. These taxa, lacking chloroplasts, are effectively invisible to pigment-based algorithms. Our classification includes only chloroplast-bearing dinoflagellates, which represent roughly half of the dinoflagellate diversity in the basin (Mikaelyan and Zavyalova, 1999; Mikaelyan et al., 2011; Stoecker et al., 2017). Importantly, heterotrophic dinoflagellates contribute significantly to microbial food webs, preying on a range of taxa including cryptophytes, nanoflagellates, and diatoms (Jeong et al., 2010; Calbet and Saiz, 2005). Their grazing activity, particularly during bloom decline phases, influences nutrient recycling, bloom succession, and carbon fluxes—especially under stratified summer conditions when autotrophic prey is abundant (Mikaelyan and Zavyalova, 1999).

While our HPLC-based PFT estimates do not resolve such trophic interactions, it is important to consider the ecological role of these heterotrophs—especially in nutrient-rich coastal zones where their autotrophic prey (e.g., cryptophytes, nanoplankton) are abundant (Aytan et al., 2018).

The dataset used to construct our algorithm was primarily drawn from the western Black Sea, with limited representation of autumn and winter months. To evaluate broader applicability, we examined seasonal and spatial climatologies derived from long-term satellite data. These were validated against extensive microscopy-based studies (Moncheva et al., 1991, 1995, 2001; Moncheva and Krastev, 1997; Mikaelyan and Zavyalova, 1999; Mikaelyan et al., 2007, 2011, 2015, 2021, 2024; Aytan et al., 2018). Snnual climatology of PFTs as %Chl-*a* showed clear ecological gradient: larger phytoplankton (e.g., diatoms, microplankton, cryptophytes) dominated coastal and shelf waters rich in nutrients, while picoplankton and nanoplankton prevailed offshore under oligotrophic conditions. Dinoflagellates and haptophytes exhibited more localized distributions, possibly governed by stratification and niche- specific conditions such as salinity, or light availability. In the northwest shelf, diatoms and microphytoplankton accounted for over 80 % of Chl-*a*, indicative of eutrophic, high-productivity regimes driven by riverine input (Lazar et al., 2024). Dinoflagellates were also coastal but appeared more sensitive to stratification (Mikaelyan and Zavyalova, 1999). Nanoplankton formed a relatively stable background community, while cryptophytes favored the productive northern regions. Haptophytes showed a more restricted distribution, possibly due to ecological constraints. In contrast, picoplankton dominate in oligotrophic regions, particularly the southeast, where they represent over 80 % of Chl-*a* - consistent with their competitive advantage under nutrient- depleted conditions. Seasonal patterns further supported these spatial dynamics. The relatively persistent presence of nanoplankton throughout the year and across the basin and in sub-regions, aligned with long-term observations in the Black Sea (Mikaelyan et al., 2007). Satellite-derived succession patterns indicated a typical spring diatom bloom, followed by proliferation of nanoplankton (suggesting coccolithophores) and cyanobacteria during summer and autumn, with diatom resurgence in late autumn, likely driven by seasonal mixing and nutrient entrainment (Demidov, 2008; Mikaelyan et al., 2017). These observations align well with established succession patterns in the basin (Kubryakov et al., 2019; Silkin et al., 2021). For example, coccolithophore

blooms—particularly *Emiliania huxleyi*—are routinely observed during May–June (Mikaelyan et al., 2011), coinciding with periods of strong stratification (Oguz and Merico, 2006).

Sub-regional analyses underscored the influence of geography. Coastal regions—including Crimea, the Bosporus, and the northwestern shelf—exhibited strong seasonal variability, with diatoms and microphytoplankton peaking in spring and summer. Offshore areas, by contrast, were more stable and picoplankton-dominated year-round. In the southeastern coastal zone, our findings corroborate Aytan et al. (2018), who documented spring diatom and autotrophic nanoflagellate blooms followed by a summer assemblage of coccolithophores, dinoflagellates, and *Synechococcus* spp. The summer increase in *E. huxleyi* during periods of high stratification is well-documented (Oguz and Merico, 2006), as is the accumulation of *Synechococcus* in oligotrophic surface waters (Uysal, 2001).

Earlier studies, such as Moncheva et al. (2001), also reported diatom and dinoflagellate blooms in Cape Galata and Cape Kaliakra, though they lacked data for small, less conspicuous taxa like cryptophytes—limiting comparisons with our work. Additionally, Aytan et al. (2018) showed that heterotrophic bacteria and nanoflagellates often exceed autotrophic biomass in summer, reflecting enhanced microbial loop activity. While our method does not directly resolve heterotrophs, the elevated satellite-detected picophytoplankton and nanoplankton during these periods indirectly supports this observation.

Long-term trends (Mikaelyan et al., 2024) show increasing diatom biomass post-2000, with major peaks around 2005–2010 and 2020–2022, likely tied to nutrient loading (Mikaelyan et al., 2013, 2024). In contrast, dinoflagellate biomass declined after the 1990s, with a minor peak in 1999–2000, and remained consistently low thereafter—echoing earlier findings (Oguz and Gilbert, 2007). These changes are linked to major climatic phases in the Black Sea: a warm period (pre- 1980s), a cold period (mid-1980s to mid-1990s), and a second warm period from the mid-1990s onward (Oguz and Gilbert, 2007).

Coccolithophore (haptophyte) biomass increased during the cold phase and remained high thereafter, suggesting a sustained ecological shift (Mikaelyan et al., 2011, 2024). Our satellite-based analysis (Fig. 8) aligns with these trends. For instance, Crimea exhibited a delayed dinoflagellate recovery in 2021, whereas diatoms and dinoflagellates continued to decline in the Eastern and offshore regions—where picoplankton dominance persisted. Notably, relatively weak and short-lived spring diatom blooms in satellite records may reflect years with mild winters (e.g., 2010/2011), when elevated sea surface temperatures limited vertical mixing and nutrient entrainment (Oguz and Merico, 2006).

Finally, haptophyte biomass peaked in 2003–2005 and 2015–2017, possibly in response to phosphate fluctuations. Our annual analysis (Fig. 8) reflects these temporal dynamics, including recent peaks in 2020–2024. Regionally, Crimea showed a rebound in dinoflagellates around 2021, while Eastern and offshore regions showed persistent picoplankton dominance—deviating somewhat from Mikaelyan's historical trends.

5. Conclusion

This study applied statistical and multivariate analyses to a comprehensive HPLC pigment dataset (~700 samples) from the Black Sea, primarily collected in the western basin and during warmer seasons over multiple years. We developed a regional algorithm to estimate PFTs from diagnostic pigments, based on empirical pigment–Chl-*a* relationships and tested its applicability across the entire basin. Our results captured broad seasonal trends in community structure, with diatoms dominating in late autumn and increasing contributions from nano- and picophytoplankton in summer, consistent with known Black Sea dynamics. The inclusion of haptophyte-specific pigments, especially Hex, proved important for improving estimation accuracy of this group. While the model performed reasonably well across different sub-regions, some limitations remain—particularly in resolving dinoflagellates and haptophytes under mixed or low-biomass conditions. Future work should focus on expanding the dataset to underrepresented seasons and regions, refining group-specific pigment ratios, and integrating taxonomic or molecular data to improve group resolution. In conclusion, the algorithm developed here offers a coherent and regionally adapted approach to characterizing Black Sea phytoplankton composition from pigment data. Although not exhaustive, it provides a solid foundation for future improvements. The method supports ongoing efforts in satellite remote sensing validation and may contribute to long-term ecological monitoring and modeling of phytoplankton community dynamics in the Black Sea and similar environments.

Supplementary data to this article can be found online at https://doi.org/10.1016/j.marpolbul.2025.118360.

CRediT authorship contribution statement

Elisabetta Canuti: Writing – review & editing, Writing – original draft, Visualization, Validation, Supervision, Methodology, Investigation, Formal analysis, Data curation, Conceptualization.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgment

This study was supported by the European Commission Directorate General Joint Research Centre (JRC) and the Copernicus Program. We gratefully acknowledge the Chief Scientists, the captains and crews of the R/V Mare Nigrum (Romania) and R/V Akademik of the Institute of Oceanology in Varna (Bulgaria), as well as all the scientists and technicians who assisted with sample collection. We also thank the two anonymous reviewers for their valuable comments and constructive suggestions, which significantly improved the quality of this manuscript. This study has been conducted using E.U. Copernicus Marine Service Information doi:10.48670/moi-00302.

Data availability

Data will be made available on request.

References

Aiken, J., Pradhan, Y., Barlow, R., Lavender, S., Poulton, A., Holligan, P., Hardman- Mountford, N., 2009. Phytoplankton pigments and functional types in the Atlantic Ocean: a decadal assessment, 1995-2005. Deep Sea Res. Part II Top. Stud. Oceanogr. 56 (15), 899–917.

Aubrey, D., Moncheva, S., Demirov, E., Diaconu, V., Dimitrov, A., 1995. Environmental changes in the western Black Sea related to anthropogenic and natural conditions. J. Mar. Sys. 7, 411–425.

Aytan, U., Feyzioglu, A.M., Valente, A., Agirbas, E., Fileman, E.S., 2018. Microbial plankton communities in the coastal southeastern Black Sea: biomass, composition and trophic interactions. Oceanologia 60, 139–152.

Blondel, V., Guillaume, J.-L., Lambiotte, R., Lefebvre, E., 2008. Fast unfolding of communities in large networks. J. Stat. Mechanics Theory Exp. 2008. https://doi. org/10.1088/1742-5468/2008/10/P10008.

Brewin, R.J.W., Sathyendranath, S., Hirata, T., Lavender, S.J., Barciela, R.M., Hardman-Mountford, N.J., 2010. A three-component model of phytoplankton size class for the Atlantic. Oc. Ecol. Model. 221, 1472–1483.

Calbet, A., Saiz, E., 2005. The ciliate–copepod link in marine ecosystems. Aquat. Microb. Ecol. 38 (2), 157–167.

Canuti, E., 2023. Phytoplankton pigment *in situ* measurements uncertainty evaluation: an HPLC interlaboratory comparison with a European-scale dataset. Front. Mar. Sci. 10, 1197311. https://doi.org/10.3389/fmars.2023.1197311.

Canuti, E., Penna, A., 2024. Dynamics of phytoplankton communities in the Baltic Sea: insights from a multi-dimensional analysis of pigment and spectral data: part I, pigment dataset. Front. Mar. Sci. Sec. Ocean Obs. 11. https://doi.org/10.3389/ fmars.2024.1425347.

Canuti, E., Artuso, F., Bracher, A., Brotas, V., Devred, E., Dimier, C., Giardina, I., Mendes, C.R., Murawski, S., Peeken, I., Tracana, A., Ras, J., Wiegmann, S., 2022. The

Fifth HPLC Intercomparison on Phytoplankton Pigments (HIP-5) Technical Report, *EUR 31334 EN*. Publications Office of the European Union, Luxembourg. https://doi.org/10.2760/563102.

Canuti, E., Allerup, M., Bracher, A., Brotas, V., Capellacci, S., Carabelli, C., Casabianca, S., Develoouse, I., Dias, A., Dimier, C., Flander-Putrle, V., Laloux, M., Lami, A., Kenemer, C., Kumar, S., Peeken, I., Penna, A., Raman, M., Ras, J., Ricci, F., Salaün, J., Schluter, L., Suzuki, K., Thomas, C., Tirkey, A., Tracana, A., Uitz, J., Vijaya Krishna, A., Vlaemnick, B., Zaccagnini, A., 2025. The 5th workshop on high- performance liquid chromatography (HPLC) inter-comparison on phytoplankton pigments (HIP), Publications Office of the European Union, Luxembourg, JRC140829. https://data.europa.eu/doi/10.2760/1257493.

Catlett, D., Siegel, D.A., 2018. Phytoplankton pigment communities can be modeled using unique relationships with spectral absorption signatures in a dynamic coastal environment. J.Geo.l Res. Oceans 123, 246–264. https://doi.org/10.1002/ 2017JC013195.

Cociasu, A., Lazar, L., Vasiliu, D., 2008. New tendency in nutrient evolution from Romanian coastal waters Cercetari Marine, 38 pp. 7–23. https://www.rmri.ro/Home/Downloads/Publications.RecherchesMarines/2008/paper01.pdf.

Demidov, 2008. Seasonal dynamics and estimation of the annual primary production of phytoplankton in the Black Sea. Oceanology 48, 664–678.

Dogliotti, A.I., Ruddick, K., Nechad, B., Doxaran, D., Knaeps, E., 2015. A single algorithm to retrieve turbidity from remotely-sensed data in all coastal and estuarine waters. Remote Sens. Environ. 156, 157–168. https://doi.org/10.1016/j.rse.2014.09.020.

Dugu´e, N., Perez, A., 2015. Directed Louvain: maximizing modularity in directed networks. Universit´e d'Orleans. hal-01231784. ´ https://hal.archives-ouvertes.fr/hal - 01231784.

Eker-Develi, E., Berthon, J.-F., Canuti, E., Slabakova, N., Moncheva, S., Shtereva, G., Dzhurova, B., 2012. Phytoplankton taxonomy based on CHEMTAX and microscopy in the northwestern Black Sea. J. Mar. Sys. 94, 18–32. https://doi.org/10.1016/j.jmarsys.2011.10.005.

Falkowski, P.G., Barber, R.T., Smetacek, V., 1998. Biogeochemical controls and feedbacks on ocean primary production. Science 281, 200–206.

Falkowski, P.G., Laws, E.A., Barber, R.T., Murray, J.W., 2003. Phytoplankton and their role in primary, new, and export production. In: Fasham, M.J.R. (Ed.), Ocean Biogeochemistry: The Role of the Ocean Carbon Cycle in Global Change. Springer, Berlin, pp. 99–121 (2003).

Gordon, H.R., Brown, O.B., Evans, R.H., Brown, J., Smith, R.C., Baker, K.S., Clark, S.K., 1988. A semianalytic radiance model of ocean color. J. Geophys. Res. 93, 10,909–10,924.

Gr´egoire, M., Alvera-Azcarate, A., Buga, L., Capet, A., Constantin, S., ´ D'ortenzio, F., Doxaran, D., Faugeras, Y., Garcia-Espriu, A., Golumbeanu, M., et al., 2023. Monitoring Black Sea environmental changes from space: new products for altimetry, ocean colour and salinity. Potentialities and requirements for a dedicated in-situ observing system. Front. Mar. Sci. 9, 998970.

Hirata, T., Aiken, J., Hardman-Mountford, N., Smyth, T.J., Barlow, R., 2008. An absorption model to determine phytoplankton size classes from satellite ocean colour. Remote Sens. Environ. 112, 3153–3159.

Hirata, T., Hardman-Mountford, N.J., Brewin, R.J.W., Aiken, J., Barlow, R.G., Suzuki, K., Isada, T., Howell, E., Hashioka, T., Noguci-Aita, M., Yamanaka, Y., 2011. Synoptic relationships between surface chlorophyll-a and diagnostic pigments specific to phytoplankton functional types. Biogeosc 8, 311–327.

IOCCG, 2014. Phytoplankton functional types from space. In: Sathyendranath, S. (Ed.), Reports of the International Ocean-Colour Coordinating Group, No. 15. IOCCG, Dartmouth, Canada.

Jeong, H.J., Yoo, Y.D., Kim, J.S., Seong, K.A., Kang, N.S., Kim, T.H., 2010. Growth, feeding and ecological roles of mixotrophic and heterotrophic dinoflagellates in marine planktonic food webs. Ocean Sci. J. 45 (2), 65–91.

Kramer, S.J., Siegel, D.A., Graff, J.R., 2020. Phytoplankton community composition determined from co-variability among phytoplankton pigments from the NAAMES field campaign. Front. Mar. Sci., Sec. Mar. Ecosys. Ec. 7. https://doi.org/10.3389/ fmars.2020.00215.

Kubryakov, A.A., Mikaelyan, A.S., Stanichny, S.V., 2019. Summer and winter coccolithophore blooms in the Black Sea and their impact on production of dissolved organic matter from Bio-Argo data. J. Mar. Syst. 199, 103220. Available at: https://doi.org/10.1016/j.jmarsys.2019.103220.

Latasa, M., Bidigare, R.R., 1998. A comparison of phytoplankton populations of the Arabian Sea during the Spring Intermonsoon and Southwest Monsoon of 1995 as described by HPLC-analyzed pigments. Deep Sea Res. Part II Top. Stud. Oceanogr. 45 (10–11), 2133–2170. https://doi.org/10.1016/S0967-0645(98)00066-6.

Lazar, L., Vlas, O., Pantea, E., Boicenco, L., Marin, O., Abaza, V., Filimon, A., Bisinicu, E., 2024. Black Sea eutrophication comparative analysis of intensity between coastal and offshore waters. Sustainability 16, 5146. https://doi.org/10.3390/su16125146.

MacIntyre, H.L., Kana, T.M., Anning, T., Geider, R.J., 2002. Photoacclimation of photosynthesis irradiance response curves and photosynthetic pigments in microalgae and cyanobacteria. J. Phycol. 38, 17–38.

Mackey, M., Mackey, D., Higgins, H.W., Wright, S.W., 1996. CHEMTAX - a program for estimating class abundances from chemical markers: application to HPLC measurements of phytoplankton. Mar. Ecol. Prog. Ser. 144, 265–283.

Mikaelyan, A.S., Zavyalova, T.A., 1999. Vertical distribution of heterotrophic phytoplankton in the black sea during the summer period. Oceanology 39 (6), 813–820.

Mikaelyan, A.S., Pautova, L.A., Georgieva, L.V., Dyakonov, V.Yu., 2007. Database on phytoplankton of the Black Sea. Oceanology 47, 445–448. https://doi.org/10.1134/S0001437007030186.

Mikaelyan, A.S., Silkin, V.A., Pautova, L.A., 2011. Coccolithophorids in the Black Sea: their interannual and longterm changes. Oceanology 51, 39–48.

Mikaelyan, A.S., Zatsepin, A.G., Chasovnikov, V.k, 2013. Long-term changes in nutrient supply of phytoplankton growth in the Black Sea. J. Mar. Sys. 117, 117–118.

Mikaelyan, A.S., Pautova, L.A., Chasovnikov, V.K., Mosharov, S.A., Silkin, V.A., 2015. Alternation of diatoms and coccolithophores in the northeastern Black Sea: a response to nutrient changes. Hydrobiologia 755, 89–105. https://doi.org/10.1007/ s10750-015-2219-z.

Mikaelyan, A.S., Shapiro, G.I., Chasovnikov, V.K., Fred Wobus, F., Marcus Zanacchi, M., 2017. Drivers of the autumn phytoplankton development in the open Black Sea. J. Mar. Sys. 174, 1–11.

Mikaelyan, A.S., Pautova, L.A., Fedorov, A.V., 2021. Seasonal evolution of deep phytoplankton assemblages in the Black Sea. J. Sea Res. 178.

Mikaelyan, A.S., Sergeeva, A.V., Pautova, L.A., Chasovnikov, V.K., Gagarin, V.I., 2024. 75-year dynamics of the Black Sea phytoplankton in association with eutrophication and climate change. Sci. Total Environ. 954, 1.

Moncheva, S., Rusev, C., Sophronov, I., 1991. Biological cycle of some blooming phytoplankton species along the Bulgarian Black Sea coast. Oceanology BAS 21, 31–38 (In Bulgarian, English summary).

Moncheva, S., Petrova-Karadjova, V., Palasov, A., 1995. Harmful algal blooms along the Bulgarian Black Sea coast and possible patterns of fish and zoobenthic mortalities. In: Lassus, P., Arzul, G., Denn, E., Gentien, P. (Eds.), Harmful Marine Algal Blooms. Lavoisier Publ. Incorporation, Paris, pp. 193–198.

Moncheva, S., Krastev, A., 1997. Some aspects of phytoplankton long-term alterations off Bulgarian Black Sea shelf. In: Sensitivity to Change: Black Sea, Baltic Sea and North Sea (Oszoy, E. & Mikhaelian, A., eds), NATO ASI Series, 2. Environment—Vol. 27. Kluwer Academic Publishers, New York, pp. 79–94.

Moncheva, S., Gotsis-Skretasb, O., Pagoub, K., Krastev, A., 2001. Phytoplankton blooms in Black Sea and Mediterranean coastal ecosystems subjected to anthropogenic eutrophication: similarities and differences estuarine. Coast. Shelf Sci. 53, 281–295.

Morel, A., B´elanger, S., 2006. Improved detection of turbid waters from ocean color sensors information. Remote Sens. Environ. 102, 237–249. https://doi.org/10.1016/j.rse.2006.01.022.

Oguz, T., Merico, A., 2006. Factors controlling the summer *Emiliania huxleyi* bloom in the Black Sea: a modeling study. J. Marine Syst. 59, 173–188. https://doi.org/10.1016/j.jmarsys.2005.08.002.

Oguz, T., Gilbert, D., 2007. Abrupt transitions of the top-down controlled Black Sea pelagic ecosystem during 1960–2000: evidence for regime shifts under strong fishery exploitation and nutrient enrichment modulated by climate-induced variations. *Deep-Sea Res.* I 54, 220–242.

Sarthou, G., Timmermans, K.R., Blain, S., Treguer, P., 2005. Growth physiology and fate of diatoms in the ocean: a review. J. Sea Res. 53, 25–42.

Sathyendranath, S., Hoge, F.E., Platt, T., Swift, R.N., 1994. Detection of phytoplankton pigments from ocean color – improved algorithms. Appl. Opt. 33 (6), 1081–1089.

Sathyendranath, S., Stuart, G., Maass, H., Platt, T., 2001. Remote sensing of phytoplankton pigments: a comparison of empirical and theoretical approaches. Int. J. Remote Sens. 22, 249–273.

Silkin, V., Mikaelyan, A.S., Pautova, L., Fedorov, A., 2021. Annual dynamics of phytoplankton in the Black Sea in relation to wind exposure. J. Mar. Sci. Eng. 9, 1435. https://doi.org/10.3390/jmse9121435.

Stewart, K., Kassakian, S., Krynytzky, M., DiJulio, D., Murray, J.W., 2007. Oxic, suboxic, and anoxic conditions in the Black Sea. In: Yanko-Hombach, V., Gilbert, A.S., Panin, N., Dolukhanov, P.M. (Eds.), The Black Sea Flood Question: Changes in Coastline, Climate, and Human Settlement. Springer, Dordrecht. https://doi.org/ 10.1007/978-1-4020-5302-3_1.

Stoecker, D.K., Hansen, P.J., Caron, D.A., Mitra, A., 2017. Mixotrophy in the marine plankton. Ann. Rev. Mar. Sci. 9, 311–335. https://doi.org/10.1146/annurev-marine- 010816-060617.

Trees, C.C., Clark, D.K., Bidigare, R.R., Ondrusek, M.E., Mueller, J.L., 2000. Accessory pigments versus chlorophyll a concentrations within the euphotic zone: a ubiquitous relationship. Limnol. Oceanogr. 5.

Uitz, J., Claustre, H., Morel, A., Hooker, S.B., 2006. Vertical distribution of phytoplankton communities in open ocean: an assessment based on surface chlorophyll. J. Geophys. Res. 111, C08005.

Uysal, Z., 2001. Chroococcoid cyanobacteria *Synechococcus* spp. in the Black Sea: pigments, size, distribution, growth and diurnal variability. J. Plankton Res. 23 (2), 175–189. https://doi.org/10.1093/plankt/23.2.175.

Valente, A., Sathyendranath, S., Brotas, V., Groom, S., Grant, M., Jackson, T., et al., 2022. Chlorophyll a concentration obtained by HPLC and fluorometric/spectrophotometric determination (version 3) [dataset]. PANGAEA, doi:10.1594/PANGAEA.941314. In: Valente, A et al. A Compilation of Global Bio-optical In Situ Data for Ocean-colour Satellite Applications - Version 3 [Dataset Bundled Publication]. PANGAEA. https://doi.org/10.1594/PANGAEA.941318.

Vidussi, F., Claustre, H., Manca, B.B., Luchetta, A., Marty, J.C., 2001. Phytoplankton pigment distribution in relation to upper thermocline circulation in the eastern Mediterranean Sea during winter. J. Geophys. Res. 106 (C9), 19939–19956.

Yakushev, E., Chasovnikov, V., Murray, J., Pakhomova, S.V., Podymov, O.I., Stunzhas, P., 2007. Vertical Hydrochemical Structure of the Black Sea. https://doi. org/10.1007/698_5_088.

Zhang, B., Horvath, S., 2005. A general framework for weighted gene co-expression network analysis. Stat. Appl. Genet. Mol. Biol. 4 (1)

ANNEX III

This annex has been submitted and is currently under review in Ocean Modelling

Satellite-Based Estimation of Phytoplankton Functional Types in the Baltic Sea Using a Regionalized Algorithm

Elisabetta Canuti

Satellite-Based Estimation of Phytoplankton Functional Types in the Baltic Sea Using a Regionalized Algorithm

Elisabetta Canuti

Abstract

Remote sensing and diagnostic pigment (DP)-based methods have significantly advanced the estimation of phytoplankton functional types (PFTs). However, most existing models are optimized for open-ocean conditions and are not directly transferable to optically complex coastal or estuarine environments. The Baltic Sea, in particular, presents unique bio-optical challenges that hinder accurate PFT estimation using global algorithms. Although regional DP-based models have been developed, they are primarily limited to the Southern Baltic, restricting their broader applicability. This study introduces a regionalized empirical algorithm for estimating PFTs in the Baltic Sea using High Performance Liquid Chromatography (HPLC) pigment data collected from multiple sub-basins. By refining existing approaches and tailoring the algorithm to the region's distinct optical characteristics, we improved the estimation of phytoplankton size classes (PSCs) and key functional types, including cryptophytes, green algae, and dinoflagellates. Nanoplankton were the dominant size class across the basin, accounting for up to 46% of chlorophyll *a* (Chl *a*), particularly in coastal waters. Picoplankton dominated offshore regions, contributing up to 21% of Chl *a*, while microplankton reached peak proportions (~37.5%) in nearshore areas of the Gulf of Finland. Among functional groups, cryptophytes and dinoflagellates exhibited strong coastal and northern basin dominance, whereas green algae and prochlorophytes were more prevalent offshore. The proposed methodology enables a more accurate and spatially consistent assessment of PFTs across the Baltic Sea, offering a valuable tool for ecosystem monitoring and enhancing our understanding of phytoplankton-driven biogeochemical processes in this dynamic and ecologically sensitive region.

1. Introduction

Phytoplankton are microscopic, photosynthetic organisms that form the base of marine food webs and are key drivers of global biogeochemical cycles. Contributing nearly half of Earth's primary production, they play a vital role in carbon sequestration, nutrient cycling, and climate regulation (Falkowski et al., 1998; Behrenfeld et al., 2001; Le Quéré et al., 2005, Demarcq et al., 2012, Alvain et al., 2013). Understanding phytoplankton dynamics is thus central to marine ecosystem research, particularly in the face of accelerating climate change and human-induced pressures. Traditionally, chlorophyll a (Chl a) has served as a primary proxy for phytoplankton biomass in both field measurements and remote sensing applications. However, Chl a alone offers limited insight into the taxonomic and functional composition of phytoplankton communities, which is critical for assessing ecosystem function and response to environmental change (Harding et al., 2016; Mouw et al., 2017). To overcome this limitation, pigment-based chemotaxonomy has emerged as a robust approach for characterizing phytoplankton functional types (PFTs). This method leverages diagnostic pigments (DPs)—accessory pigments that are specific to certain phytoplankton groups—to infer community composition from pigment profiles (Mackey et al., 1996; Vidussi et al., 2001; Uitz et al., 2006; Brewin et al., 2010; Hirata et al., 2008; Mouw et al., 2017; Sun et al., 2022).

High-Performance Liquid Chromatography (HPLC) remains a reference technique for quantifying phytoplankton pigments with high sensitivity and specificity (Roy et al., 2011). These pigment-based datasets are not only essential for ecological studies, but also provide the foundation for validating and developing satellite-based algorithms for ocean color and phytoplankton composition (Sathyendranath et al., 1994, 2001; IOCCG, 2014). Satellite remote sensing has transformed our capacity to monitor phytoplankton at regional to global scales. While many algorithms have been developed to estimate Chl a and, increasingly, PFTs from space, they are predominantly optimized for Case 1 waters—open-ocean systems where optical properties are primarily influenced by phytoplankton pigments (IOCCG, 2014; Brewin et al., 2010; Sun et al., 2022). In contrast, optically complex environments such as the Baltic Sea present additional challenges. Here, high concentrations of colored dissolved organic matter (CDOM), suspended sediments, and detritus interfere with the optical signal, complicating the remote retrieval of phytoplankton properties (Darecki et al., 2004; Stoń-Egiert et al., 2010; Meler et al, 2016a, 2016b; Woźniak et al., 2022). Recent efforts to develop regionalized empirical algorithms tailored to the Baltic Sea—such as those by Meler et al. (2020)—have improved diagnostic pigment estimations but are often limited in spatial coverage, focusing primarily on the southern basin. Consequently, there remain significant gaps in our understanding of phytoplankton functional composition across the entire Baltic Sea, particularly under its highly variable bio-optical conditions.

This study addresses these gaps by developing a regionally tuned empirical algorithm for estimating PFTs across diverse hydro-optical environments in the Baltic Sea. By integrating an extensive HPLC pigment dataset collected from multiple oceanographic campaigns and refining existing chemotaxonomic models (e.g., Uitz et al., 2006; Hirata et al., 2008; Brewin et al., 2010; Meler et al., 2020), we aim to enhance the accuracy and applicability of satellite-based PFT assessments in this ecologically and optically complex region.

2. Materials and Methods

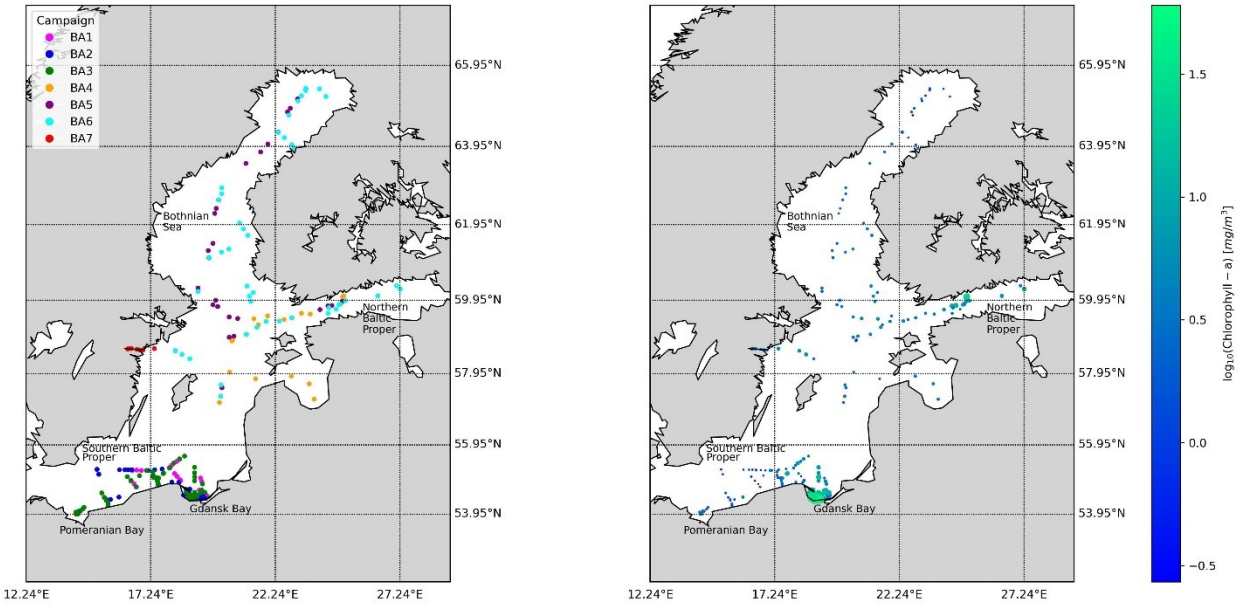
2.1. Study area and dataset

The Baltic Sea is a semi-enclosed, brackish sea in Northern Europe, connected to the North Sea through the Danish Straits. It is characterized by a strong salinity gradient, ranging from near-marine conditions in the southwest to almost freshwater in the northern and eastern basins. (Lehmann et al., 2022) This gradient, combined with limited water exchange, results in pronounced stratification and the formation of hypoxic zones in deeper areas. The Baltic Sea experiences significant seasonal variability, with ice cover in the northern regions during winter and strong thermal stratification in summer. High nutrient inputs from surrounding catchments contribute to frequent phytoplankton blooms, influencing ecosystem dynamics (Kuliński et al., 2022). The HPLC pigments dataset object of the study consisted of 281 samples collected in the Baltic Sea during seven oceanographic campaigns.Three campaigns took place in the Southern Baltic Sea, including Gdańsk Bay and the Pomeranian Bay (May and September 2004, April 2005) and another three were conducted in the Northern Baltic Proper, the Gulf of Finland, and the Bothnian Sea (July 2006, August 2007, August 2008) (Canuti et al., 2024). Additionally, one campaign was carried out in the northwestern Baltic Proper, specifically in Bråviken Bay (BB), in July 2021 (Kratzer et al., 2022). At each station, two samples (or three in the case of B7 campaign) were collected from the surface with 3.5 liter Plexiglas sampling bottle or Niskin bottles operated on a rosette sampler, and filtered onto GF/F filters (filtration volumes from 50 to 4000 mL) after well mixing. The filtration volume was decided based on the absorption coefficient at 412 nm (AC-9, WetLabs, USA) measured meanwhile the water collection. The filters were flash-frozen and stored in liquid nitrogen or -80 C until analysis at JRC laboratory.

Table 1. Dataset spatial-temporal distribution: Southern Baltic Sea and Gdansk Bay (*SB*), Central Baltic (CB), Northern Baltic Proper and Gulf of Finland (*NB*), Bothnian Sea (*BS*),

	Season	N. Samples	Mean Latitude [°N]	Mean Longitude [°E]	Area
BA01	May-04	50	54.7	18.5	SB
BA02	Sep-04	51	54.7	17.7	SB
BA03	Apr-05	62	54.7	17.6	SB
BA04	Jul-06	21	59	22.4	CB, NB
BA05	Aug-07	37	60.9	20.8	NB
BA06	Aug-08	45	61.4	22.1	NB, BS
BA07	Jul-21	8	58.6	16.6	NB

Figure 1. Baltic Sea pigment dataset: (left) spatial distribution of sample collection color coded by campaigns and (right) chlorophyll a concentration at each station



2.2. Analytical method and Diagnostic Pigments

The samples were analyzed at the Joint Research Centre of the European Commission (JRC) following a well-established HPLC method, detailed in Canuti (2023). The JRC adheres to strict quality control protocols for phytoplankton pigment analysis and actively participates in inter-laboratory exercises and inter-comparison activities to evaluate uncertainties in marine pigment measurements (Hooker et al., 2010; Canuti et al., 2016, 2022, 2023, 2025; Kratzer et al., 2022). While the method quantifies multiple phytoplankton pigments, those relevant to this study are 19'-hexanoyloxyfucoxanthin (Hex), 19'-butanoyloxyfucoxanthin (But), alloxanthin (Allo), fucoxanthin (Fuco), peridinin (Perid), diatoxanthin (Diato), zeaxanthin (Zea), chlorophyll a (Chl-a), chlorophyll b (Chl-b). The Baltic Sea dataset underwent a quality check and harmonization beginning of 2023 (Canuti et al., 2024).

2.3. PFT and PSC regression calculation

Diagnostic pigments (DPs) serve as taxonomic markers for phytoplankton functional groups, offering critical insights into community composition and ecological dynamics. The method, initially proposed by Vidussi et al. (2001) and later refined by Uitz et al. (2006), assigns phytoplankton groups based on their characteristic accessory pigments: for example fucoxanthin is considered a proxy for diatoms, peridinin for dinoflagellates, and zeaxanthin for the picoplankton. The basic assumption of these methodologies is that sum of ΣDP should equal Chl-a (Uitz et al., 2006). The original Uitz et al. (2006) equation for ΣDP estimation is as follows:

ΣDP-UITZ= 1.41 * Fuco + 1.41* Peri + 0.6 * Allo + 1.01 * Chl-b + 1.27 * Hex + 0.35*But + 0.86 * Zea

Subsequent modifications by Hirata et al. (2008, 2011) and Brewin et al. (2010) improved the method's applicability across diverse marine environments by accounting for variations in pigment attribution based on pigment concentration and Chl-a levels. For instance, in oligotrophic waters with low phytoplankton biomass (Chl-a < 0.08 mg m^{-3}), Hex was shared between pico- and nanoplankton, while in waters where Chl-a < 0.5 mg m^{-3}, Fuco was distributed between microplankton (diatoms) and nanoplankton (haptophytes). However, a study by Meler et al. (2020) based on an extensive Baltic Sea dataset suggested that these refinements lacked applicability in this region, as they produced results identical to those obtained using the earlier Uitz et al. (2006) and Hirata et al. (2008) methods. Meler et al. (2020) further recalculated the regression parameters (weighting factors) for the DP equation, omitting But and Hex, leading to the following revised expression:

ΣDP-M20 = 3*Fuco + 1.41*Peri+ 4.5*Allo +5.85*Zea* +3.35*Chl-b

In the present study, to assess and predict the size classes (nano-, pico-, and micro-phytoplankton) and PFTs within the Baltic Sea, we utilized a regression-based framework grounded in in situ pigment analysis. We categorized phytoplankton into three size classes—microplankton, nanoplankton, and picoplankton—and four functional type – diatoms, dinoflagellates, green algae & crochlorophytes, cryptophytes - by analyzing seven pivotal pigments: Fuco, Peri, Hex, But, Allo, Chl-b, and Zea (Table 2).

One of the principal constraints of the Diagnostic Pigment Analysis (DPA) approach is its limited capacity to resolve ambiguities in taxonomic attribution and size classification, particularly when diagnostic pigments are shared across multiple phytoplankton groups. For example, fucoxanthin—commonly associated with diatoms—does not discriminate between small-sized diatoms (typically <20 μm, falling within the nanoplankton size class) and larger diatoms classified as microplankton. This limitation can result in a misrepresentation of the microplankton fraction, especially in systems like the Baltic Sea where

nanoplanktonic diatoms are often prevalent. A similar issue arises in the picoplankton size class, where taxa such as Prasinophyta, Haptophyta, and other Chlorophyta may include both pico- and nanoplanktonic representatives. These groups cannot be distinctly resolved through DPA, which depends on fixed pigment-to-group and pigment-to-size associations that do not fully capture the compositional complexity of phytoplankton communities in dynamic environments like the Baltic Sea. To address some of these limitations, regionally optimized pigment combinations and weighting coefficients have been developed based on in situ observations (Meler et al., 2020). These localized formulations better reflect the dominant phytoplankton assemblages and reduce biases inherent in the application of generic DPA schemes. However, careful interpretation remains essential, particularly in the Baltic Sea where picoplanktonic eukaryotes and small diatoms play ecologically significant roles, and where mixotrophic taxa such as dinoflagellates—with variable pigment content—further complicate the assignment of functional types and size classes.

Table 2: Phytoplankton functional types (PFTs) in the Baltic Sea, categorized by size classes (PSCs) and key diagnostic pigments.

PFTs/PSCs	**Key Pigments**
Microplankton (>20 μm)	Fuco, Peri
Diatoms	Fuco
Dinoflagellates	Peri
Nanoplankton (2–20 μm)	Chl-b, Allo, But, Hex
Green Algae & Prasinophytes	Chl-b
Chryptophytes	Allo
Picoplankton (<2 μm)	Zea
C*yanobacteria*	Zea

We derived new ΣDP regression coefficients optimized for the entire Baltic Sea region by applying two distinct statistical methodologies. The first approach was based on gradient descent with non-negativity constraints, combined with regularized ordinary least squares (hereafter DESC), while the second used stepwise selection with standard ordinary least squares regression (hereafter OLS). In the DESC method, gradient descent was used to estimate regression coefficients under non-negativity constraints, ensuring that pigment contributions remained positive. The model was fitted with L1 regularization, which reduces the influence of less significant variables by shrinking some coefficients to zero, thereby improving model generalizability and reducing overfitting. Initial coefficients for the gradient descent algorithm were taken from Uitz et al. (2006), and model performance was evaluated using the coefficient of determination (R^2) between predicted and observed Chl-a concentrations. The OLS method, by contrast, employed bidirectional stepwise selection to identify the set of diagnostic pigments included in the multiple regression model. Forward selection included variables with p-values below 0.01, while backward elimination removed those with p-values above 0.05. The final set of selected predictors was used in an OLS regression without an intercept, and the regression equation was constructed using the estimated

coefficients. The key distinction between these two methodologies lies in their approach to feature selection and coefficient estimation. DESC ensures that all pigment coefficients are non-negative, while L1 regularization helps reduce overfitting. OLS, on the other hand, simplifies the model by retaining only statistically significant features, though this may result in the exclusion of ecologically relevant but less statistically prominent pigments. For both methods, the dependent variable was Chl-a, and predictors were selected from the diagnostic pigment (DP) dataset. Additionally, we applied the OLS model to data from the first three campaigns conducted in the Southern Baltic, where we expected results comparable to those of Meler et al. (2020). We also tested the model on the full dataset, once excluding Hex and once excluding a specific campaign conducted in Bråviken Bay. Finally, for Chl-a concentrations consistently above 0.25 mg m^{-3}, we compared the results from our new DP-based formula with those obtained using the methodological frameworks of Uitz et al. (2006) and Meler et al. (2020).

2.4. **PFTs-to-Chl-a relationship**

We selected different equation having the logarithmically transformed Chl-a value as dependent variable and commonly used for describe the relationship between PFTs and Chl-a (Hirata et al.,2008, 2011; Brewin et al., 2010; Uitz et al., 2006). In the context of our study, RMSE [1] quantifies the difference between the observed phytoplankton group sizes (or functional types) and the values predicted by the various mathematical models (linear, exponential, second order polynomial, and Michaelis-Menten equation). A lower RMSE indicates a better fit of the model to the observed data, allowing for more accurate quantification of the relationships between Chl-a and the abundance of size-fractionated or functional phytoplankton groups.

The RMSE can be calculated using the following formula:

$$\text{RMSE} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2} \qquad [1]$$

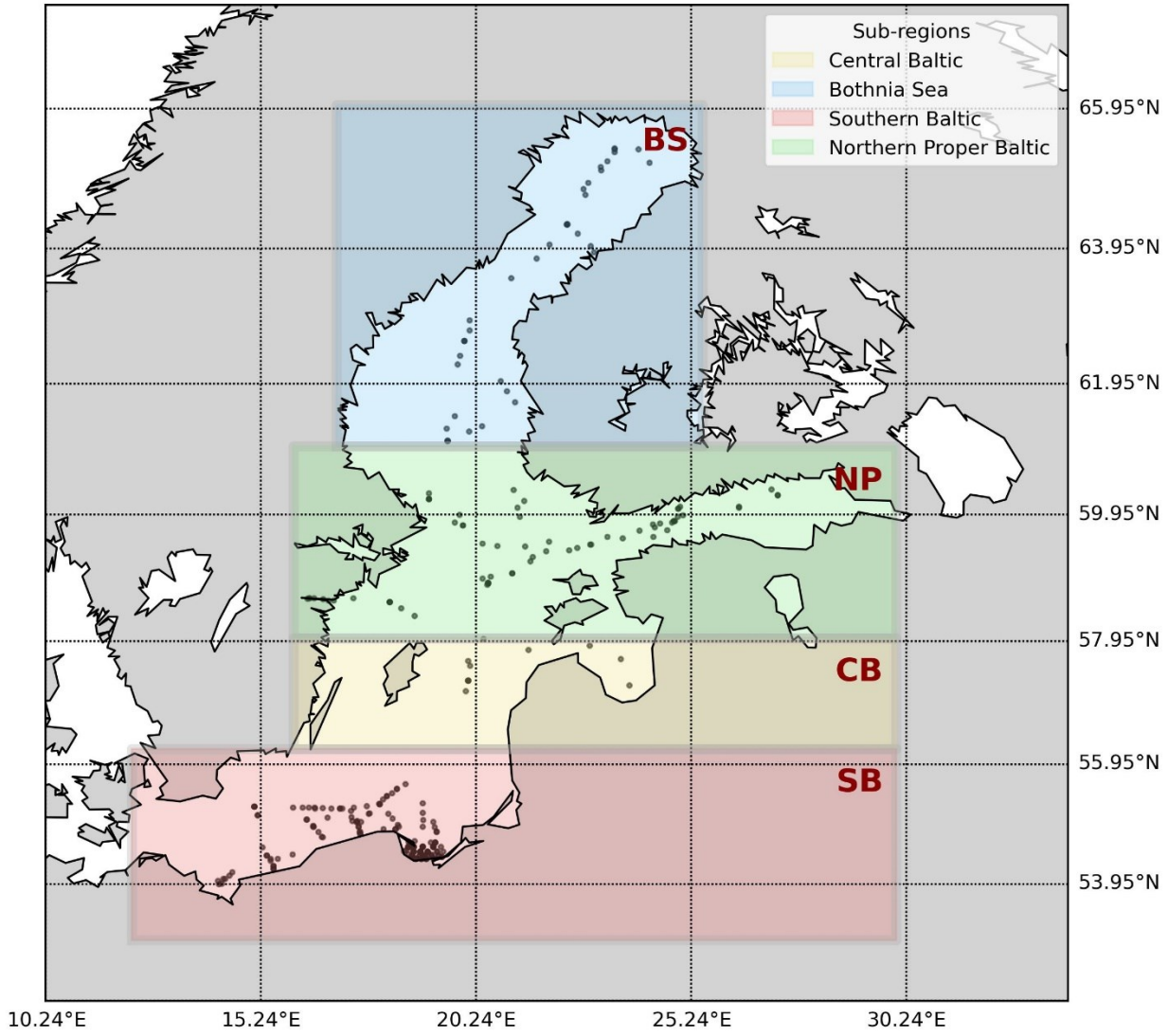
Where: n is the number of observations (data points), y_i is the observed value for the i-th observation. $\hat{y}_i$ is the predicted value for the i-th observation generated by the mathematical model.

Once identified the best describing equation for the PFTs, corresponding maps for the Baltic Sea were created by using the chlorophyll-a data of the multi-years Bio-Geo_Chemical (BGC) regional datasets of Copernicus Marine Environment Monitoring Service (CMEMS) 1 km Level-4 monthly composites of Chl-a data. 1 km spatial resolution (10.48670/moi-00308).

2.5. **PFTs Climatology for the Baltic Sea**

We investigated the seasonal and annual variations of phytoplankton functional types throughout the Baltic Sea basin and within four specific sub-regions (Figure 2). These sub-regions were selected due to their unique hydrographic and ecological characteristics, which play a role in shaping PFT distribution. The Bothnia Sea (BS), in the northernmost part of the study area, experiences lower salinity and strong seasonal variability. The Northern Proper Baltic (NP) is an intermediate region, where seasonal stratification and episodic mixing influence phytoplankton distribution. The Central Baltic (CB) acts as a transitional area with moderate salinity, supporting a diverse phytoplankton assemblage. Lastly, the Southern Baltic (SB) is characterized by significant riverine inputs, leading to higher nutrient loads.

Figure 2. Map of the four regions considered for climatology studies: Bothnia Sea (BS), Northern Proper Baltic (NB), Central Baltic (CB) and Southern Baltic (SB)



3. Results

3.1. Dataset Variability

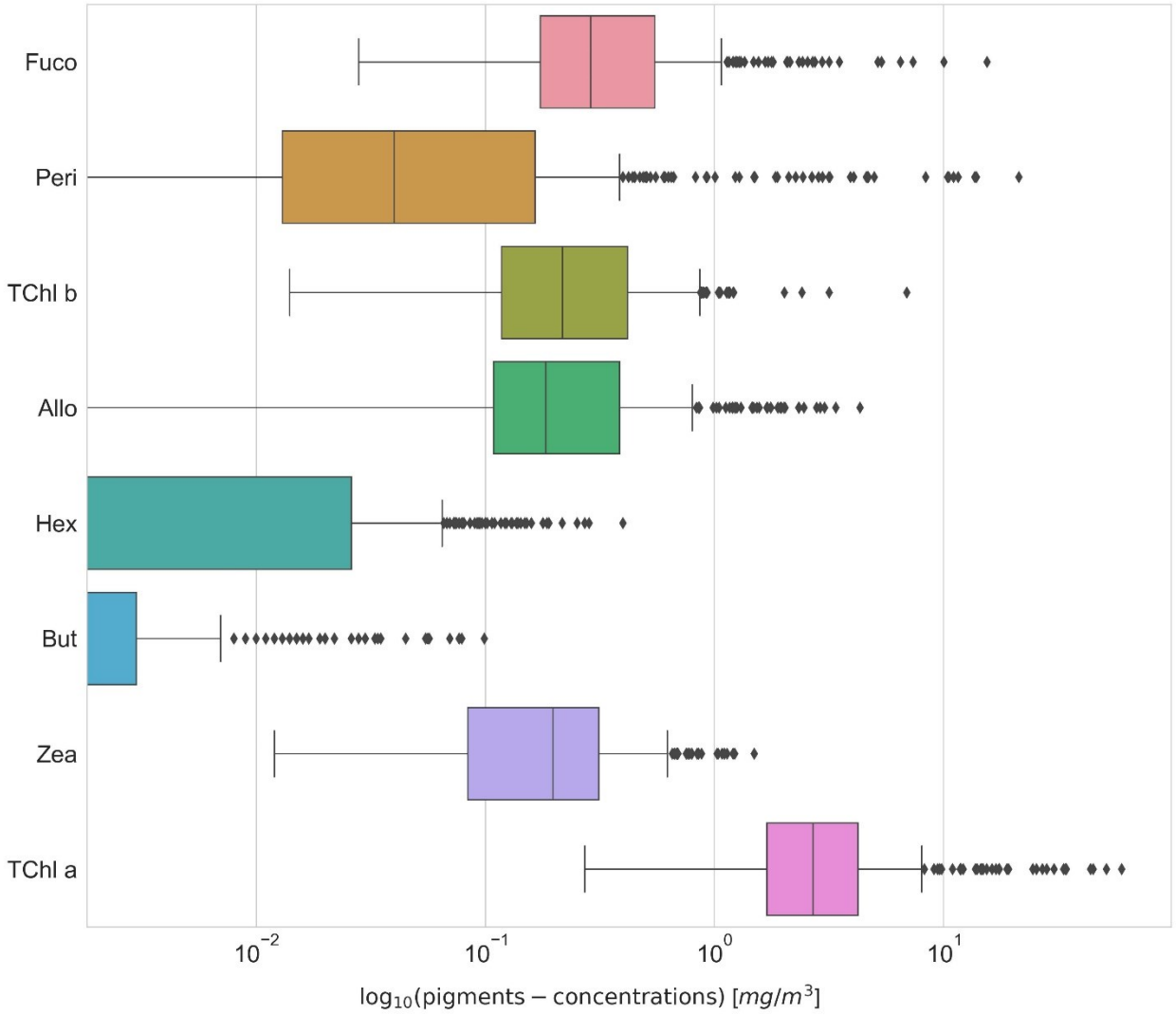
Table 3 and Figure 3 summarize the statistical distribution of DPs and key parameters. Fuco and Peri exhibited high variability, with coefficients of variation (CVs%) exceeding 200%, reflecting heterogeneous contributions from corresponding PFTs. Peri showed the highest CV% (345.5%) and was below the quantification limit in 49 samples. In contrast, Zea had the lowest variability (CV% = 93.9%) and was consistently quantified, indicating a more stable cyanobacterial signal. Chl-*b* and Allo were detected in nearly all samples, with moderate to high variability (CVs% of 150.7% and 116.1%, respectively). Chl-*a* ranged from 0.272 to 60.222 mg m^{-3}, with a mean of 5.08 mg m^{-3} and a CV% of 157.6%, reflecting wide trophic gradients across sampling stations. The PFT indices (pPF, nPF, mPF) showed lower variability (CVs% between 50% and 55%), supporting their use as integrative indicators of phytoplankton composition. But and Hex had the highest proportion of non-quantified values, with 196 and 165 samples respectively. But

was excluded from further analysis. Hex exceeded 0.01 mg m^{-3} in 33% of all samples, but only in 9% of those from the southern Baltic (BA1–BA3), suggesting limited regional occurrence.

Table 3. Summary statistics of diagnostic pigments measured across all samples. The Table includes the mean concentration [mg m^{-3}], standard deviation (St. Dev.), coefficient of variation (CV%), maximum, minimum, and median values for each pigment and the taxonomical significance (as in Meler et al., (2020)). "Not Quantified" indicates the number of samples in which the pigment was below the detection limit.

Diagnostic pigments	mean	St.dev.	CV%	max	min	median	Not Quantified	Taxonomy group*
Fuco	0.6411	1.373	214.2	15.5440	0.0280	0.2890	0	*Diatoms*
Peri	0.6655	2.299	345.5	21.3950	0.0000	0.0400	49	*Dinoflagellates*
Chl-b	0.3506	0.528	150.7	6.9320	0.0140	0.2180	0	*Chlorophytes, prochlorophytes*
Allo	0.4089	0.609	148.8	4.3320	0.0000	0.1840	1	*Cryptophytes*
Hex	0.0264	0.054	203.6	0.4000	0.0000	0.0000	165	*Prymnesiophytes*
But	0.0050	0.013	260.9	0.0990	0.0000	0.0000	196	*Chrysophytes*
Zea	0.2555	0.240	93.9	1.4950	0.0120	0.1980	0	*Cyanobacteria, prochlorophytes*
Other Parameters								
Chl-a	5.0787	8.006	157.6	60.2220	0.2720	2.6960		
pPF	0.402	0.216	53.7	0.845	0.034	0.418		
nPF	0.209	0.115	54.9	0.599	0.017	0.201		
mPF	0.388	0.194	50.1	0.889	0.046	0.365		

Figure 3. Boxplot showing the distribution of various phytoplankton pigment concentrations (log_{10}-transformed, in mg m^{-3}) across samples. Each box represents the interquartile range (IQR), with the median indicated by the central line. Whiskers extend to 1.5 times the IQR, while individual points represent outliers. Pigments analyzed include Fuco, Peri, Chl-b, Allo, Hex, But, Zea, and Chl-a.



3.2. DP coefficients determination

Four regression approaches were tested to estimate the sum of diagnostic pigments (ΣDP): a full OLS model, OLS excluding Hex (OLS-2), a non-negative constrained model (DESC), and a region-specific OLS for the Southern Baltic. In all methods Fuco and Peri were considered as the primary predictors for micro-phytoplankton. Zea was consistently used for pico-phytoplankton, while contributions from Allo, Chl-b, and Hex varied across models for nano-sized groups (Table 4). The model were analyzed for their statistical significance (supplementary Table S1, S2). The full OLS model showed the highest fit (R^2 = 0.99) and low mean percent difference from between ΣDP and Chl-a (1.44%). Excluding Hex (OLS-2) achieved a similar fit (R^2 = 0.99), with higher weighting factor was attributed to Zea. The DESC model slightly reduced model performance (R^2 = 0.98) while ensuring physically meaningful, non-negative coefficients. When applied to the Southern Baltic data subset, the OLS model produced regression coefficients were consistent with those reported by Meler et al. (2020). Across models, pigment weighting differences led to small shifts in

estimated phytoplankton size-class composition, with OLS emphasizing nano-phytoplankton and OLS-2 favoring micro-phytoplankton. ΣDP and PSC from OLS and DESC were compared with those from UITZ and M20 (Supplementary Figure S1 and S2). OLS and DESC produced higher correlations relative to the original coefficients. The ternary plot (Figure S2) showed that OLS emphasized the nano-fraction, particularly in campaigns BA5 and BA7. In contrast, OLS-2 shifted population distributions toward the micro-fraction. DESC and OLS-2 provided more balanced distributions across size classes.

Table 4. Estimation of PSC (mPF, nPF and pPF) for the different models: UITZ, OLS, OLS BA1-3, and DESC. Each equation expresses the contribution of DP to the fractions and should be dived for the corresponding ΣDP.

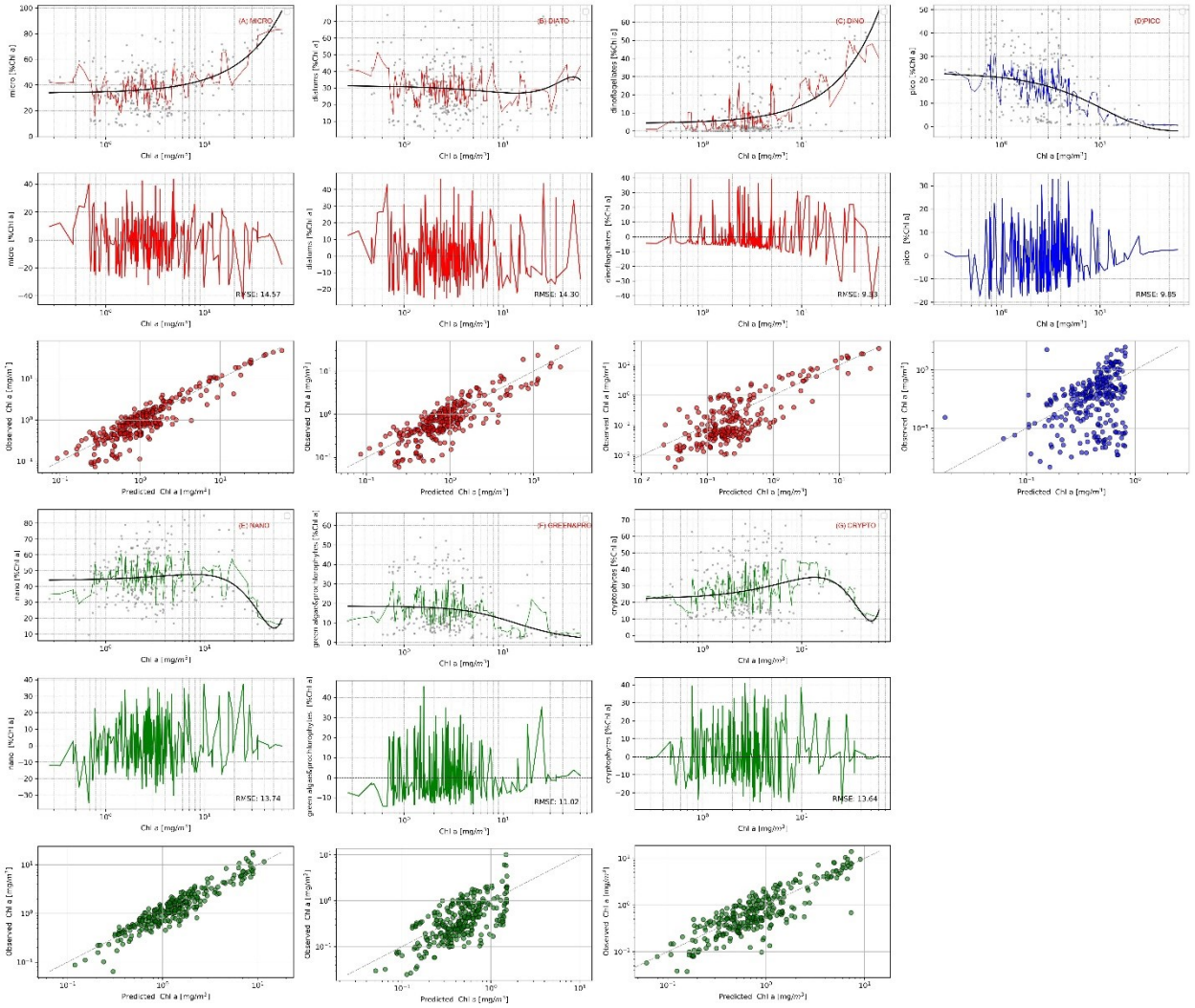
		Equations to estimate PSCs			
	Functional Type	**UITZ**	**OLS**	**OLS (BA1-3)**	**DESC**
mPF	Microplankton (>20 μm)	1.41*(Fuco+Perid)	2.31*Fuco++1.68*Peri	2.15*Fuco+1.69*Peri	2.33*Fuco+1.67*Peri
nPF	Nanoplankton (2–20 μm)	1.27* Hex+1.01* Chl-b + 0.35* But+0.60* Allo	3.26*Allo+1.80*Chlb+5.33*Hex	3.08*Allo+2.87*Chlb	3.28*Allo+1.7*Chlb+0.85*Hex
pPF	Picoplankton (0.2–2 μm)	0.86 Zea	1.33* Zea	1.31*Zea/ΣDP	1.74*Zea

3.3. Synoptic Relation between PSCs/PFTs and Chl-a

PFTs in the Baltic Sea were estimated from in situ HPLC pigment data using four approaches based on diagnostic pigment analysis: UITZ, DESC, OLS, and OLS-2. Each method applied a different set of weighting factors to estimate the fractional contribution of diagnostic pigments to total Chl-a. Among these, the OLS-PFTs algorithm consistently demonstrated the best performance across all evaluated metrics. It achieved the lowest RMSE values, particularly for microphytoplankton (0.16), diatoms (0.15), and dinoflagellates (0.12), and yielded the highest R^2 values (0.94–0.98), indicating strong agreement between predicted and observed pigment-derived PFTs. The OLS-2 approach showed similar, though slightly less consistent, results (Supplementary Table S3). Based on these comparative results, the DESC model was selected as the method of choice for estimating PFTs in the Baltic Sea.

Figure 4 illustrated the relationship between PFTs, PSCs, and their contribution to %Chl-a using the DESC-optimized coefficients (Table 8). The three panels provided complementary perspectives on the model's performance and behavior. In the top panel, the scatter plot showed how the response of different PFTs varied with increasing Chl-a concentrations, reflecting distinct ecological adaptations. The black curve represented the fitted model. The middle panel illustrated the variation of the PFT/PSC fractions as a function of Chl-a, while the bottom panel compared predicted and observed %Chl-a values.

Figure 4. Synoptic relation between PFTs and %Chl-a for the DESC-optimized coefficient. The three panels plots - (top) the synoptic relationship, (middle) the uncertainties as of the synoptic relationship and (bottom) the results of validation - summarize the results for: (A) microplankton, (B) diatoms, (C) dinoflagellates, (D) picoplankton, (E) nanoplankton, (F) green algae & prochlorophytes, (G) cryptophytes.



Microplankton exhibited an increasing trend with Chl-a, particularly at higher concentrations, suggesting a preference for nutrient-rich conditions. Diatoms followed a similar pattern, indicating their significant contribution to the microplankton fraction. Dinoflagellates showed a weak increase at lower Chl-a concentrations, followed by a sharper rise at higher concentrations—albeit with a less steep trend than microplankton and diatoms—suggesting a preference for bloom conditions and a weaker dependency on high-nutrient environments. Nanoplankton showed a moderate increase with Chl-a but with greater variability, implying a widespread presence across diverse trophic conditions. Green phytoplankton maintained a relatively flat trend, indicating stability across varying environments. Cryptophytes displayed high variability, with a peak at intermediate Chl-a concentrations, suggesting they thrived in transitional conditions. Picoplankton demonstrated a decreasing trend at high Chl-a concentrations, consistent with a preference for oligotrophic conditions. The consistently high residuals across the entire range indicated a weaker model fit for this group. The middle panel highlighted variability and RMSE values. High RMSE values were found for diatoms and cryptophytes. In contrast, picoplankton and green phytoplankton

showed lower RMSE values, indicating more stable estimations. In the bottom panel, regression analysis revealed strong linear correlations between observed and predicted values for microplankton, diatoms, and nanoplankton, with data points clustering closely around the 1:1 line—demonstrating robust model performance. Dinoflagellates and cryptophytes exhibited moderate correlations with increased scatter, reflecting greater uncertainty in their prediction. Green phytoplankton and picoplankton had weaker correlations and higher scatter, especially at low concentrations, highlighting challenges in capturing their dynamics at finer scales. Overall, the scatter plots demonstrated how effectively the DESC model predicted observed values for each PFT. Most groups aligned reasonably well with the 1:1 line, though dispersion increased at higher Chl-a concentrations. The strongest correlations were observed for microplankton, diatoms, and dinoflagellates, whereas picoplankton and nanophytoplankton showed greater deviations from predicted values.

Table 8. Relationships between log-transformed chlorophyll-a concentration (x = $\log_{10}$ Chl-a) and various phytoplankton size classes (PSCs) and functional types (PFTs). The Table presents the parameters for each equation, including linear equations for micro, dinoflagellates, nano, and green & prochlorophytes, and nonlinear equations for diatoms, cryptophytes, and pico-phytoplankton.

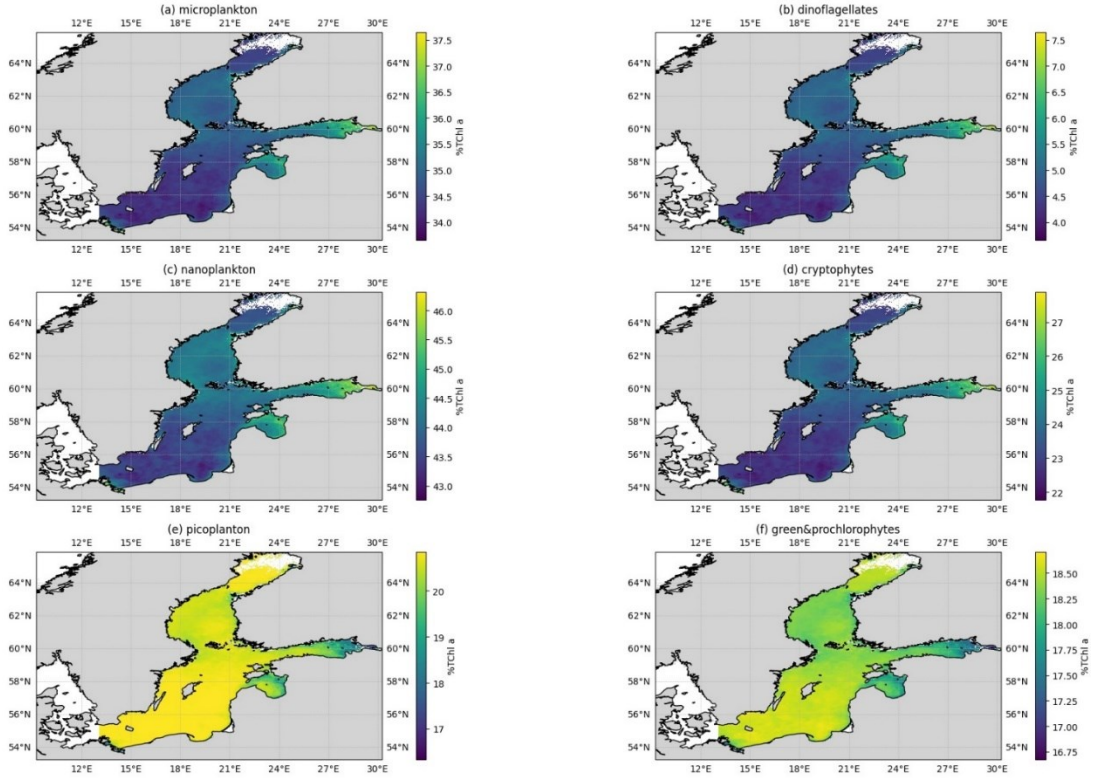
DESC		**Parameters**			**%Chl-a range**	
PSCs/PFTs	**equations**	**a**	**b**	**c**	**min**	**max**
micro	a * x + b	0.01	0.34		4.21	86.03
diatoms	micro - dino				4.14	76.23
dinoflagellates	a * x + b	0.01	0.04		0.00	60.46
nano	1 / (exp(a * x + b) + c*x)	-0.07	0.84	0.10	9.51	84.85
green&prochl	1 / (exp(a * x + b) + c*x)	-0.11	1.68	0.66	2.00	63.55
cryptophytes	1 / (exp(a * x + b) + c*x)	-0.10	1.50	0.13	0.00	72.62
pico	1 / (exp(a * x + b) + c*x)	-0.26	1.57	1.17	0.45	49.39

3.4. Distribution of PFTs

The newly developed regional algorithms (Table 8) were applied to a 26-year time series of satellite-derived chlorophyll-a (Chl-a) estimates to quantify phytoplankton functional types (PFTs) across the Baltic Sea. This reconstruction revealed spatial patterns in phytoplankton distribution and composition, reflecting environmental gradients and ecological niches across the basin (Figure 5). Nanoplankton dominated the phytoplankton community, contributing 43–46% of total Chl-a. Their distribution was widespread, with elevated concentrations in both western and eastern sub-basins, reflecting their adaptability to varying conditions. Microplankton accounted for roughly one-third of total biomass, with

maximum proportions in nutrient-rich regions such as the western Baltic Proper and the coastal areas of the Gulf of Finland and Gulf of Riga, and lower values in central offshore waters. Picoplankton contributed significantly, particularly in the central basin and northernmost areas such as the Bothnian Sea and Gulf of Finland, where their small size likely conferred an advantage under oligotrophic, low-light conditions. Dinoflagellates were most abundant in the northern Baltic Proper and Bothnian Bay, contributing 6.5–7.5% of Chl-a, consistent with their seasonal dominance during spring blooms. Their presence also extended further north compared to diatoms, likely due to their motility and mixotrophic capacity. Cryptophytes contributed 22–28% of Chl-a, with highest values along the coasts of the Gulf of Finland and southeastern Baltic Sea, and decreasing abundance offshore. Green algae and prochlorophytes collectively accounted for 16.75–18.5% of Chl-a, with highest contributions in the southern and central offshore regions and lower presence toward the northern coasts. Their distribution suggested ecological partitioning shaped by gradients in salinity, temperature, and nutrient availability.

Figure 5. Synoptic distribution of surface PFTs [% Chl-a] over 1997-2023 derived from CMEMS for the Baltic Sea for (a) microplankton, (b) dinoflagellates, (c) nanoplankton, (d) cryptophytes, (e) picoplanton, (f) green&prochlorophytes



3.5. PFTs seasonal and yearly trends

Phytoplankton dynamics in the Baltic Sea exhibited distinct seasonal patterns, with all functional groups reaching peak abundance between June and August (Figure 6A). The highest concentrations typically occurred in July, followed by a gradual decline in autumn and winter as light availability and temperature decreased. Microphytoplankton reached its maximum in July, with values around 1.02, while nanophytoplankton followed a similar seasonal pattern but with consistently higher absolute values, peaking at 1.29. Picophytoplankton exhibited the lowest concentrations among the three size classes, with a July maximum of 0.62. Dinoflagellates, present at lower levels, peaked at approximately 0.13 in July. Cryptophytes maintained moderate abundance throughout the year, reaching their highest concentration

in July (0.68), while green algae followed a comparable pattern, peaking at 0.55 in the same month. Dinoflagellates consistently remained the least abundant group, with little variation across the months. Regional differences in PFTs as Chl-a concentrations were evident. The Bothnian region consistently showed the lowest Chl-a levels, while the Southern and Central Baltic followed similar seasonal cycles, with summer peaks slightly lower than those observed in the Northern region. Long-term trends from 1997 to 2023 indicated that microplankton and nanoplankton dominated interannual variability. Their concentrations increased from 1997 to 2005, followed by fluctuations with distinct peaks in 2008, 2011, and 2018. In contrast, picoplankton, cryptophytes, and green algae remained relatively stable, with minor increases during peak years but no significant overall trend. Dinoflagellates exhibited consistently low concentrations with minimal interannual variability. Regional patterns showed that the Northern and Southern Baltic experienced the highest phytoplankton concentrations, particularly during peak years such as 2003, 2008, and 2018, suggesting periodic intensification of blooms. The Bothnian sub-region recorded the lowest values, while the Central Baltic followed the overall Baltic-wide trend, with peak years occurring at slightly lower concentrations.

Figure 6. Seasonal variations in Chl-a concentrations (mg m^{-3}) for different PFTs across the entire Baltic Sea basin (A) and four distinct sub-regions (B) Bothnia Sea, BS; (C) Northern Proper Baltic, NB; (D) Central Baltic, CB; (E) Southern Baltic, SB, from 1997 to 2023) from 1997 to 2023. The PFTs include microphytoplankton (red), cryptophytes (green), green algae & prochlorophytes (cyan), picophytoplankton (blue), nanophytoplankton (dark-green), and dinoflagellates (orange).

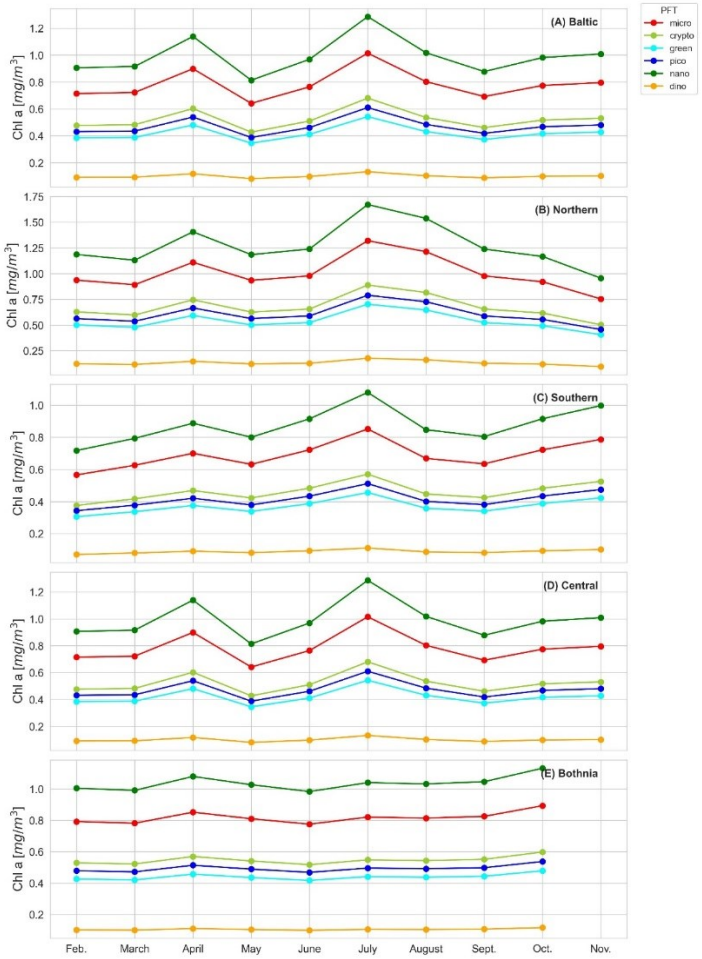
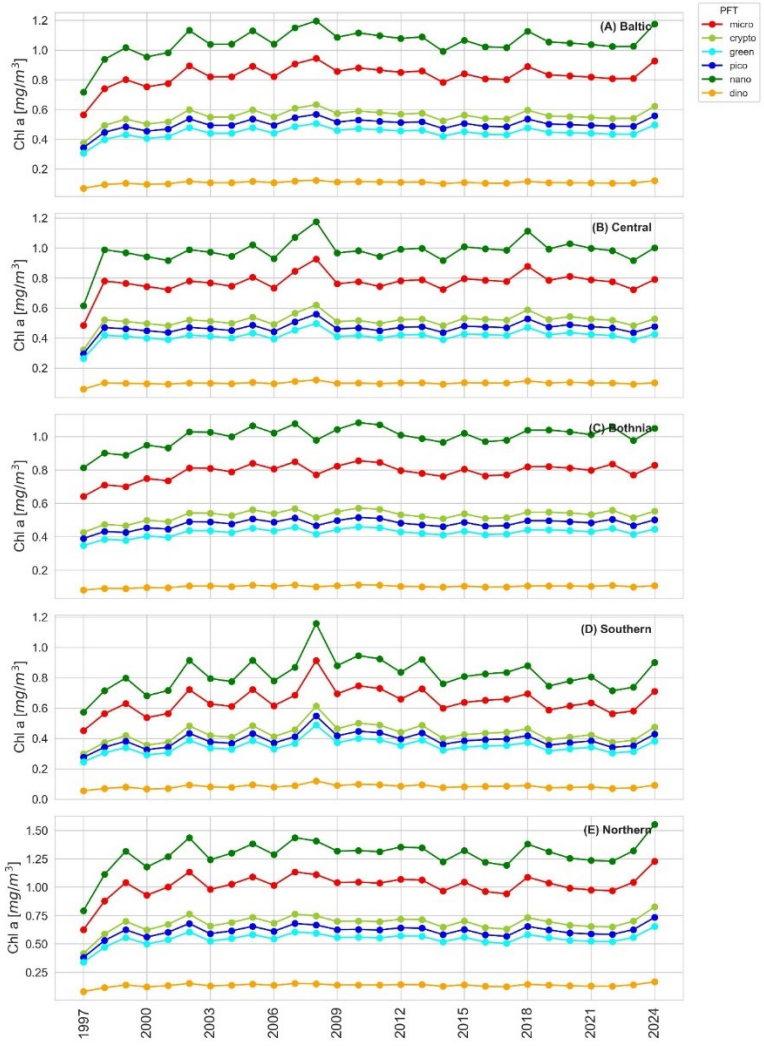


Figure 7. Yearly trends of PFTs in the (A) Baltic and its four sub-regions (B) Bothnia Sea, BS; (C) Northern Proper Baltic, NB; (D) Central Baltic, CB; (E) Southern Baltic, SB, from 1997 to 2023. The PFTs include microphytoplankton (red), cryptophytes (green), green algae & prochlorophytes (cyan), picophytoplankton (blue), nanophytoplankton (dark-green), and dinoflagellates (orange).



4. Discussion

The region-specific PFT algorithm developed in this study was based on 281 measurement points distributed throughout the entire Baltic Sea. This broad spatial coverage ensured the inclusion of diverse environmental conditions, including gradients in salinity, nutrient availability, and hydrodynamic regimes. Such variability was expected to influence pigment–PFT relationships and likely contributed to the observed variation in model performance across phytoplankton groups. In contrast, previous efforts (e.g., Meler et al., 2020; Stoń-Egiert et al., 2022) focused primarily on the southern Baltic, where approximately 1000 sampling events provided a dense and consistent dataset. While this southern dataset offers strong potential for improving regional model robustness, its applicability to the central and northern sub-basins remains uncertain due to environmental heterogeneity. The limited number of pigment observations in those regions hindered comprehensive model validation, potentially explaining the relatively low R^2 values for certain PFT classes. Additionally, the assumption that a single pigment-based algorithm could adequately represent phytoplankton distributions across the entire basin may have overlooked regional differences in community composition and ecological drivers. Analysis of the JRC pigment dataset

confirmed substantial spatial and temporal variability in pigment concentrations, consistent with earlier studies on Baltic Sea phytoplankton dynamics (Wasmund et al., 1998, 2003, 2008, 2011). The persistent detection of chl-b and zea throughout all campaigns indicated a stable presence of green algae, cyanobacteria, and prochlorophytes. In contrast, the sporadic detection of prymnesiophyte and chrysophyte marker pigments—Hex and But—suggested their limited contribution, particularly in offshore or less eutrophic waters.

We optimized the ΣDP coefficient within the DESC model framework using the expanded JRC dataset, and compared PSC estimates with those reported by Meler et al. (2020). The results contrasted with the more homogeneous, nano-dominated size structure reported by Meler et al. (2020), highlighting the importance of incorporating diverse environmental settings in PFT algorithm development. Using these optimized coefficients, we developed regionalized DESC algorithms for PFT estimation. The models demonstrated moderate predictive performance with a linear relationship that may limit the capture of complex ecological feedbacks. Comparison with CMEMS PSC data products (CMEMS-OC-QUID-009-131 to 134) showed improved regression statistics for all size classes using our DESC-based approach (e.g., micro: R^2 = 0.587 vs. 0.315; nano: R^2 = 0.646 vs. 0.246; pico: R^2 = 0.391 vs. 0.286). Finally, the regionalized model proposed here outperformed the global one (Uitz et al., 2006). The application of the DESC model to satellite-derived chlorophyll-a time series provided further insights into spatial and temporal trends (Figure 7). The highest Chl-a concentrations were consistently observed in the western and coastal areas, driven by shallow bathymetry, riverine input, and vertical mixing (Kunlinski et al., 2022). A west–east gradient was evident, with declining concentrations toward the Gulf of Finland and the Bothnian Sea, in agreement with findings by Leman et al. (2022). Seasonal trends were clearly identifiable. Microphytoplankton reached their highest relative abundance during July, likely due to favorable light conditions, stable water columns, and nutrient availability in early summer. This pattern matched the classical seasonal succession in the Baltic Sea, where spring blooms of large diatoms are followed by smaller phytoplankton taxa (Wasmund et al., 1998; HELCOM, 2018). Nanophytoplankton maintained a strong presence year-round, suggesting ecological adaptability, such as mixotrophy and efficient nutrient uptake. Picophytoplankton peaked in the summer months, consistent with their competitive advantage in warm, oligotrophic, and stratified waters. These seasonal patterns in PFT distribution were in line with previous ecological observations (Sommer et al., 2012; Kahru et al., 2007) and demonstrate the model's ability to capture broad phenological dynamics. However, the increased presence of microphytoplankton in northern and central regions highlights the importance of including diverse environmental regimes in model development. Interannual variability was also evident in the satellite-based time series. During the early 2000s, enhanced contributions from micro- and nanophytoplankton were observed, possibly reflecting increased eutrophication, modified nutrient ratios, and shifts in thermal stratification (Klais et al., 2011; Wasmund et al., 2011). Notably, picophytoplankton and cryptophytes exhibited relatively stable contributions across years, suggesting ecological resilience or a broad ecological niche (Hällfors et al., 2013; Olli et al., 2019). Green algae also showed little interannual fluctuation, likely due to their adaptability to variable light, temperature, and nutrient conditions. The resurgence of microphytoplankton in certain years may be attributed to episodic nutrient pulses resulting from vertical mixing, upwelling, or saline inflow events (Naumann et al., 2018). These events are known to temporarily reverse stratification and enhance the growth of large-celled diatoms under well-mixed, nutrient-rich conditions (Wasmund et al., 1998; HELCOM, 2018). Over the longer term, climate-induced regime shifts may explain some of the observed trends. For example, warmer winters and positive North Atlantic Oscillation (NAO) phases have been associated with reduced ice cover, earlier stratification, and increased dominance of dinoflagellates over diatoms in spring (Klais et al., 2013; Wasmund et al., 2017). While our PFT model cannot directly resolve diatom/dinoflagellate dynamics, the relative changes in micro- and nanophytoplankton abundance may reflect similar ecological transitions. These results underscore the value of sustained, long-term satellite observations, especially when integrated with in situ pigment data,

to detect regime shifts and track phytoplankton community responses to climate variability and anthropogenic pressures. Satellite-derived PFT time series offer critical insight into both short-term fluctuations and long-term ecological trends (Wasmund et al., 2003; Kahru et al., 2016; Groetsch et al., 2016), and can help guide adaptive ecosystem management in the Baltic Sea.

5. Conclusion

This study developed a regionally adapted algorithm for estimating phytoplankton functional types (PFTs) in the Baltic Sea, grounded in a comprehensive pigment dataset and applied to multi-decadal satellite observations. The approach enabled the identification of spatial and seasonal dynamics in phytoplankton community structure, highlighting the dominance of nanophytoplankton across the basin, the episodic increase of microphytoplankton during summer, and the distinct ecological niches occupied by smaller functional groups such as picophytoplankton, cryptophytes, and green algae. These patterns reflect the complex biogeochemical gradients that shape phytoplankton distribution in the Baltic Sea, including nutrient availability, salinity, and light conditions. The results reinforce the value of region-specific models, which captured features overlooked by globally derived coefficients and better accounted for local ecological variability. From a management perspective, improved functional group resolution contributes to more informed assessments of ecosystem state, eutrophication dynamics, and bloom development. Understanding the seasonal dominance and spatial distribution of specific phytoplankton groups—particularly those linked to harmful blooms or shifts in food web structure—can support more targeted monitoring and early warning systems under regional environmental policies, such as the EU Marine Strategy Framework Directive. Although empirical algorithms provided reasonable estimates for some PFTs, their limitations—especially for groups like cryptophytes and diatoms—highlight the need for further refinement. Incorporating additional environmental drivers, region-specific optical properties, and advanced statistical or machine learning techniques may improve predictive performance. Ultimately, this work underscores the importance of integrating satellite-based monitoring with ecologically meaningful classifications to better support long-term ecosystem assessment and adaptive management in optically complex, climate-sensitive regions like the Baltic Sea.

Anknoledgments

This study has been supported by the European Commission Directorate General Joint Research Centre (JRC) and the Copernicus Program.

This study has been conducted using E.U. Copernicus Marine Service Information; 10.48670/moi-00308

The author would like to acknowledge Juha Flinkman, Seppo Kaitala, and Jukka Seppälä from the Finnish Environment Institute (former Finnish Institute of Marine Research) for the opportunity to participate in the Oceanographic campaign on-board the R/V “Aranda” and to the crew of R/V “Aranda” for their support during all the cruises. Acknowledgments are due to Dirk Van der Linde from JRC for water sampling and laboratory analyses.

Conflict of Interest

The author declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Author Contributions

EC contributed to this paper by conceptualizing and designing the study and performing analysis and calculations, reviewing and editing the content.

Funding

The study has been supported by the European Commission Directorate General Joint Research Centre (JRC) and the Copernicus Program.

REFERENCES

Aiken, J., Fishwick, J. R., Lavender, S., Barlow, R., Moore, G. F., Sessions, H., Hardman-Mountford, N. J. 2007. Validation of MERIS reflectance and chlorophyll during the BENCAL cruise October 2002: preliminary validation of new demonstration products for phytoplankton functional types and photosynthetic parameters. International Journal of Remote Sensing, 28(3–4), 497–516. https://doi.org/10.1080/01431160600821036

Alvain, S., Le Quéré, C., Bopp, L., Racault, M.F., Beaugrand, G., Dessailly, D., Buitenhuis, E., (2013) Rapid climatic driven shifts of diatoms at high latitudes. *Remote Sens. Env.* 132, 195-201

Behrenfeld, M.J., Randerson, McClain, C.R., Feldman, G.C., Los, S.O., C J Tucker, C.J., Falkowski, P.G., C B Field, C.B., Frouin, R., Esaias, W.E., Kolber, D.D., Pollack, N.H. (2001), Biospheric primary production during an ENSO transition, *Science*, 291, 2594–2597.

Brewin, R.J.W., Sathyendranath, S., Hirata, T. Lavender, S.J., Barciela, R.M., Hardman-Mountford, N.J., 2010. A three-component model of phytoplankton size class for the Atlantic. *Oc. Ecol. Model*., 221 pp. 1472-1483

Canuti, E., Ras, J., Grung, M., Röttgers, R., Costa Goela, P., Artuso, F., Cataldi, D., 2016. HPLC/DAD Intercomparison on Phytoplankton Pigments (HIP-1, HIP-2, HIP-3 and HIP-4), *EUR 28382 EN*, Publications Office of the European Union, Luxembourg

Canuti, E., Artuso, F., Bracher, A., Brotas, V., Devred, E., Dimier, C., Giardina, I., Mendes, C.R., Murawski, S., Peeken, I., Tracana, A., Ras, J. and Wiegmann, S., 2022, The Fifth HPLC Intercomparison on Phytoplankton Pigments (HIP-5) Technical Report, *EUR 31334 EN*, Publications Office of the European Union, Luxembourg, doi:10.2760/563102

Canuti, E., 2023 Phytoplankton pigment *in situ* measurements uncertainty evaluation: an HPLC interlaboratory comparison with a European-scale dataset. *Front. Mar. Sci.* 10:1197311. doi: 10.3389/fmars.2023.1197311

Canuti, E. and Penna, A., 2024, Dynamics of Phytoplankton Communities in the Baltic Sea: Insights from a Multi-dimensional Analysis of Pigment and Spectral Data: Part I, Pigment Dataset, Front. Mar. Sci. Sec. Ocean Observation, Vol. 11 doi: 10.3389/fmars.2024.1425347

Darecki, M., Stramski, D., 2004. An evaluation of MODIS and SeaWiFS bio-optical algorithms in the Baltic Sea. Remote Sens. Environ. 89 (3), 326–350.

Demarcq, H., Reygondeau, G., Alvain, S., Vantrepotte, V., (2012) Monitoring marine phytoplankton seasonality from space, *Remote Sens. Env.*, 117, 211-222, https://doi.org/10.1016/j.rse.2011.09.019.

Falkowski, P.G. Laws, E.A. Barber, R.T. Murray J.W. 2003. Phytoplankton and their role in primary, new, and export production M.J.R. Fasham (Ed.), Ocean Biogeochemistry: The Role of the Ocean Carbon Cycle in Global Change, Springer, Berlin, pp. 99-121

Falkowski, P.G., Barber, R.T., Smetacek, V. 1998 Biogeochemical controls and feedbacks on ocean primary production Science, 281, pp. 200-206

Gordon, H.R., Brown, O.B., Evans, R.H., Brown, J., Smith, R.C., Baker, K.S., Clark, S.K., 1988, A semianalytic radiance model of ocean color *J. Geophys. Res.,* 93, pp. 10,909-10,924

Groetsch, P. M. M., Simis, S. G. H., Eleveld, M. A., and Peters, S. W. M. 2016. Spring blooms in the Baltic Sea have weakened but lengthened from 2000 to 2014, Biogeosciences, 13, 4959–4973, https://doi.org/10.5194/bg-13-4959-2016

Hällfors, G. 2004. Checklist of Baltic Sea Phytoplankton Species including some heterotrophic protistan groups," in Baltic Sea Environment Proceedings, 1–208. Available online at: http://www.helcom.fi/Lists/Publications/BSEP95.pdf.

Harding, J., Mallonee, M.E., Perry, E.S., Miller, W.D., Adolf, J.E., Gallegos, C.L., Paerl, H.W. (2016), Variable climatic conditions dominate recent phytoplankton dynamics in Chesapeake Bay Sci. Rep., 6 (1) p. 2377

Hirata, T., N.J. Hardman-Mountford, R.J.W. Brewin, J. Aiken, R.G. Barlow, K. Suzuki, T. Isada, E. Howell, T. Hashioka, M. Noguci-Aita, Y. Yamanaka. 2011. Synoptic relationships between surface chlorophyll-a and diagnostic pigments specific to phytoplankton functional types. *Biogeosc.*, 8 pp. 311-327

Hooker S.B., Van Heukelem, L., Thomas, C.S., Claustre, H., Ras J., Barlow, R., Sessions ,H., Schluter, L., Perl J., Trees, C., Stuart, V., Head, E., Clemetson, L., Fishwick, J., Llewellyn, C., Aiken, J. 2005. The second SeaWIFS HPLC Analysis Round-Robin Experiemnt /SeaHARRE-2), NASA/TM-2005-212785.

IOCCG (2014). Phytoplankton Functional Types from Space. Sathyendranath, S. (ed.), Reports of the International Ocean-Colour Coordinating Group, No. 15*, IOCCG*, Dartmouth, Canada.

Lehmann, A., Myrberg, K., Post, P., Chubarenko, I., Dailidiene, I., Hinrichsen, H.-H., Hüssy, K., Liblik, T., Meier, H. E. M., Lips, U., and Bukanova, T.: Salinity dynamics of the Baltic Sea, *Earth Syst. Dynam*., 13, 373–392, https://doi.org/10.5194/esd-13-373-2022, 2022.

Le Quéré, C., Harrison, S.P., Prentice, C.I., Buitenhuis, E.T., Aumont, O., Bopp, L., Claustre, H., Cotrim da Cunha, L., Geider, R., Giraud, X., Klass, C., Kohfeld, K.E., Legendre, L., Manizza, M., Platt, T., Rivkin, R., Sathyendranath, R.B., Uitz, J., Watson, A.J., Wolf-Gladrow, D., 2005. Ecosystem dynamics based on plankton functional types for global ocean biogeochemistry models. Glob. Chang. Biol. 11 (11), 2016–2040. https://doi.org/10.1111/j.1365-2486.2005.1004.x

Jaanus, A., Kuprijanov, I., Kaljurand, K., Lehtinen, S., Enke, A., 2017, Optimization of phytoplankton monitoring in the Baltic Sea*, J. of Mar. Sys.,* Vol. 171, 65-72, https://doi.org/10.1016/j.jmarsys.2016.10.009

Kahru, M., Elmgren, R., and Savchuk, O. P.2016. Changing seasonality of the Baltic Sea, Biogeosciences, 13, 1009–1018, https://doi.org/10.5194/bg-13-1009-2016

Klais, R., Tamminen, T., Kremp, A., Spilling, K., Olli, K. 2011. Decadal-Scale Changes of Dinoflagellates and Diatoms in the Anomalous Baltic Sea SpringBloom. PLoS ONE 6(6): e21567. doi:10.1371/journal.pone.002156

HELCOM. 2018. "State of the Baltic Sea – Second HELCOM holistic assessment 2011–2016," in *Baltic Sea Environment Proceedings*, vol. 155. (HELCOM). Available at: https://helcom.fi/media/publications/BSEP155.pdf.

Kratzer, S, Harvey, E. T., Canuti, E., 2022, International Intercomparison of In Situ Chlorophyll-a Measurements for Data Quality Assurance of the Swedish Monitoring Program, *Front. Remote Sens. Sec. Multi- and Hyper-Spectral Imaging*, Vol. 3, https://doi.org/10.3389/frsen.2022.866712

Kuliński, K., Rehder, G., Asmala, E., Bartosova, A., Carstensen, J., Gustafsson, B., Hall, P. O. J., Humborg, C., Jilbert, T., Jürgens, K., Meier, H. E. M., Müller-Karulis, B., Naumann, M., Olesen, J. E., Savchuk, O., Schramm, A., Slomp, C. P., Sofiev, M., Sobek, A., Szymczycha, B., and Undeman, E. 2022. Biogeochemical functioning of the Baltic Sea, Earth Syst. Dynam., 13, 633–685, https://doi.org/10.5194/esd-13-633-2022

Mackey, M., Mackey, D. Higgins, H. W., Wright, S. W., 1996. CHEMTAX - a program for estimating class abundances from chemical markers: Application to HPLC measurements of phytoplankton. *Mar. Ecol. Prog. Ser.* 144: 265– 283

Meler, J., Kowalczuk, P., Ostrowska, M., Ficek, D., Zabłocka, M., Zdun, A., 2016a. Parameterization of the light absorption properties of chromophoric dissolved organic matter in the Baltic Sea and Pomeranian Lakes. *Ocean Sci.* 12, 1013–1032. https://doi.org/10.5194/os-12-1013-2016.

Meler, J., Ostrowska, M., Stoń-Egiert, J., 2016b. Seasonal and spatial variability of phytoplankton and non-algal absorption in the surface layer of the Baltic. *Estuar. Coast. Shelf S*. 180, 123–135. https://doi.org/10.1016/j.ecss.2016.06.012.

Meler, J., Woźniak, S.B., Stoń-Egiert, J., 2020. Comparison of methods for indirectly estimating the phytoplankton population size structure and their preliminary modifications adapted to the specific conditions of the Baltic Sea. *J. Mar. Syst.* 212, 103446. https://doi.org/10.1016/j.jmarsys.2020.103446.

Naumann, M., Mohrholz, V., Waniek, J. J. 2018 Water Exchange and conditions in the Deep Basins, HELCOM Balt. Sea Environ, Fact Sheets, Online, http://www.helcom.fi/baltic-sea-trends/environment-fact-sheets/

Nelson, N.B., Siegel, D. A., 2013. Global distribution and dynamics of chromophoric dissolved organic matter. *Annu. Rev. Mar. Sci.* 5:447–476, doi:10.1146/annurev-marine-120710-100751

Mouw, C.B., Hardman-Mountford, N.J., Alvain, S., Bracher, A. , Brewin, R.J.W. , Bricaud, A., Ciotti, A.M., Devred, E. Fujiwara, A., Hirata, T. Hirawake, T. , Kostadinov, T.S., Roy, S., Uitz J., 2017. A consumer's guide to satellite remote sensing of multiple phytoplankton groups in the Global Ocean. *Front. Mar. Sci.,* 4

Olli, K., Klais-Peets, R., Tamminen, T., Ptacnik, R., Andersen, T., 2011. Long-term changes in the Baltic Sea phytoplankton community. *Boreal Env. Res.* 16 (SUPPL A). 3-14

Olli, K., R. Ptacnik, R. Klais, and T. Tamminen. 2019. Phytoplankton species richness along coastal and estuarine salinity continua. *Am. Nat.* 194: E41–E51. doi:10.1086/703657

Roy, S., Llewellyn, C.A., Egeland, E.S., Johnsen, G., 2011. Phytoplankton Pigments, Characterization, Chemotaxonomy and Applications in Oceanography. Cambridge Univ. Press, p. 845

Sathyendranath, S., Hoge, F.E., Platt, T., Swift, R.N., 1994. Detection of phytoplankton pigments from ocean color – improved algorithms. *Appl. Opt.* 33 (6), 1081–1089.

Sathyendranath, S., Cota, G., Stuart, V., Maass, H., Platt, T., 2001. Remote sensing of phytoplankton pigments: a comparison of empirical and theoretical approaches. *Int. J. Remote Sens.* 22, 249–273.

Sommer, U., Aberle, N., Lengfellner, K. et al. 2012. The Baltic Sea spring phytoplankton bloom in a changing climate: an experimental approach. Mar Biol 159, 2479–2490 https://doi.org/10.1007/s00227-012-1897-6

Stoń-Egiert, J., Kosakowska, A., 2005. RP-HPLC determination of phytoplankton pigments comparison of calibration results for two columns. *Mar. Bio.* 147, 251–260

Stoń-Egiert, J., Łotocka, M., Ostrowska, M., Kosakowska, A., 2010. The influence of biotic factors on phytoplankton pigment composition and resources in Baltic ecosystems: new analytical results *Oceanologia*, 52 (1) pp. 101-125, 10.5697/oc.52-1.10

Stoń-Egiert, J., Ostrowska, M., 2022. Long-term changes in phytoplankton pigment contents in the Baltic Sea: Trends and spatial variability during 20 years of investigations, *Cont. Shelf Res.,* Vol. 236, https://doi.org/10.1016/j.csr.2022.104666

Sun, X., Shen, F., Brewin, R.J.W., Li, M., Zhu, Q., 2022. Light absorption spectra of naturally mixed phytoplankton assemblages for retrieval of phytoplankton group composition in coastal oceans. *Limnol. Oceanogr.,* 67: 946-961. https://doi.org/10.1002/lno.12047

Uitz, J., Claustre, H., Morel, A., Hooker, S.B., 2006. Vertical distribution of phytoplankton communities in open ocean: an assessment based on surface chlorophyll. *J. Geophys. Res.*, 111 p. C08005

Vidussi, F., Claustre, H., Manca, B.B., Luchetta, A., Marty, J.C., 2001. Phytoplankton pigment distribution in relation to upper thermocline circulation in the eastern Mediterranean Sea during winter *J. Geophys. Res.*, 106 (C9) pp. 19939-19956

Wasmund, N., Nausch G., Matthäus , W., 1998. Phytoplankton spring blooms in the southern Baltic Sea — spatio temporal development and long-term trends. *J. Plankton Res.* 20: 1099–1117

Wasmund, N., Uhlig S., 2003. Phytoplankton trends in the Baltic Sea. *ICES J. Mar. Sci.* 60: 177–186

Wasmund N., Göbel J., von Bodungen B., 2008. 100-years-change in the phytoplantkon community of Kiel Bight (Baltic Sea). *J. Mar. Syst.* 73: 300–322

Wasmund, N., Tuimala, J., Suikkanen, S., Vandepitte, L., Kraberg, A., 2011. Long-term trends in phytoplankton composition in the western and central Baltic Sea. *J. Mar. Syst.* 87, 145–159. https://doi.org/10.1016/j.jmarsys.2011.03.010

Wasmund, N., Nausch, G., and Feistel, R. 2013. Silicate consumption: an indicator for long term trends in spring diatom development in the Baltic Sea. J. Plankton Res. 35, 393–406. doi: 10.1093/plankt/fbs101

Wasmund, N., Kownack, J., Göbel, J., Jaanus, A, Johansen, M., Jurgensone, I., Lehtinen, S., Powilleit, M. 2017. The Diatom/Dinoflagellate Index as an Indicator of Ecosystem Changes in the Baltic Sea 1. Principle and Handling Instruction. *Front. Mar. Sci.* 4:22. doi: 10.3389/fmars.2017.00022

Woźniak, S., Meler, J., Stoń-Egiert, J. 2022. Inherent optical properties of suspended particulate matter in the southern Baltic Sea in relation to the concentration, composition and characteristics of the particle size distribution; new forms of multicomponent parameterizations of optical properties*, J. Mar. Sys.*, Vol. 229, https://doi.org/10.1016/j.jmarsys.2022.103720

SUPPLEMENTARY MATERIAL

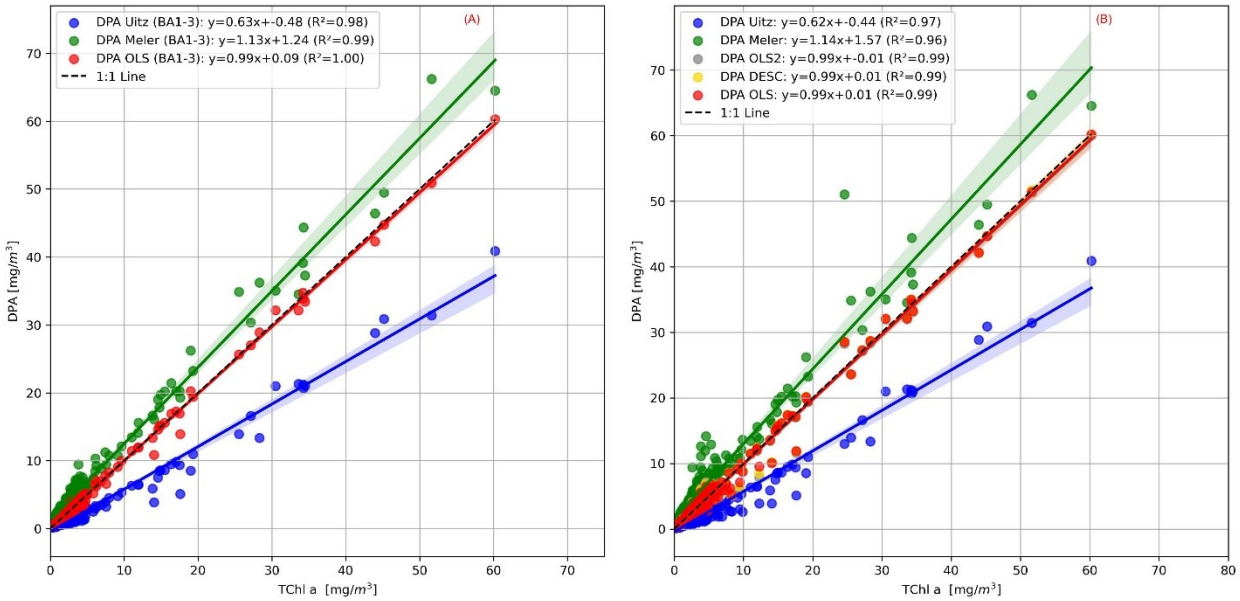
Supplementary Table S1. OLS regression results for Chl-a prediction: Summary of coefficients, standard errors, and statistical significance for the predictor pigments in the regression model.

Predictor	Coefficient	Std. Error	t-value	p-value	95% CI (Lower)	95% CI (Upper)
Peri	1.69	0.032	53.4	<0.001	1.62	1.75
Fuco	2.31	0.045	51.5	<0.001	2.22	2.40
Allo	3.26	0.109	30.0	<0.001	3.05	3.47
Chl-b	1.81	0.12	15.1	<0.001	1.57	2.05
Zea	1.33	0.205	6.5	<0.001	0.93	1.73
Hex	5.33	0.994	5.4	<0.001	3.38	7.29

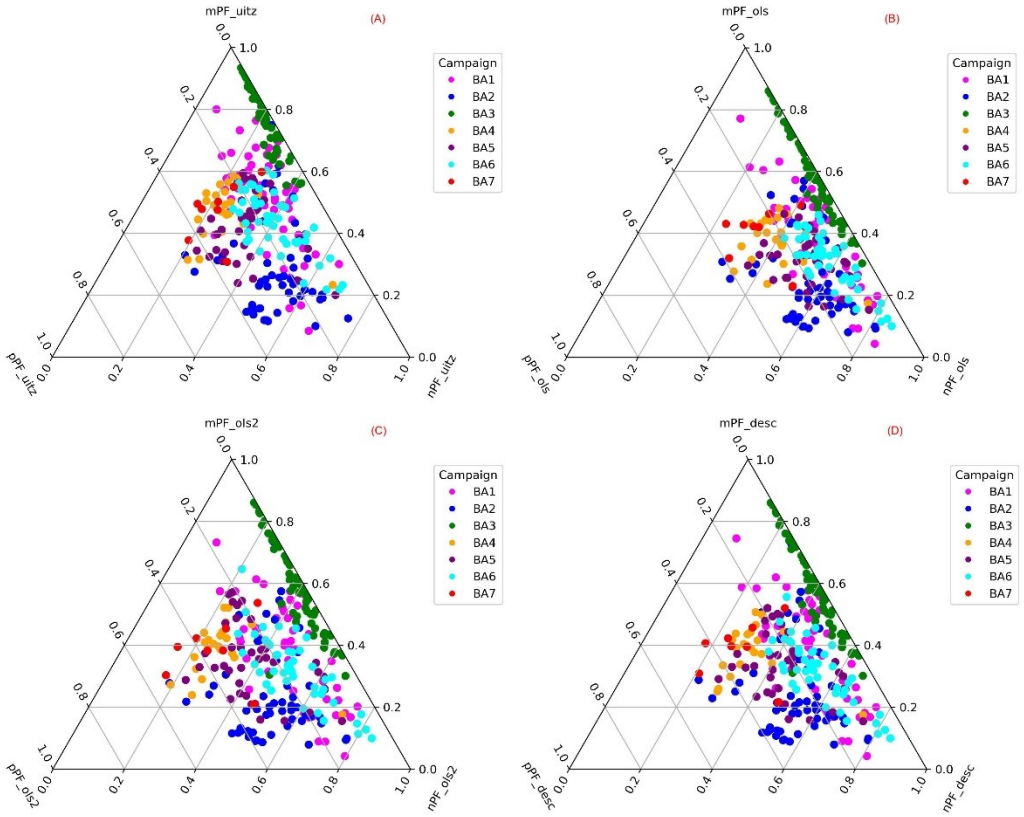
Supplementary Table S2. OLS regression results for Chl-a prediction (excluding Hex) Summary of coefficients, standard errors, and statistical significance for the predictor pigments in the revised regression model.

Predictor	Coefficient	Std. Error	t-value	p-value	95% CI (Lower)	95% CI (Upper)
Peri	1.67	0.033	50.7	<0.001	1.61	1.74
Fuco	2.34	0.047	50.3	<0.001	2.25	2.44
Allo	3.30	0.114	29.0	<0.001	3.08	3.52
Chl-b	1.65	0.121	13.6	<0.001	1.41	1.89
Zea	1.95	0.177	11.0	<0.001	1.60	2.30

Supplementary Figure S1. Relationships between Chl-a and the ΣDP calculated using UITZ, Meler et al. (2020), OLS, OLS-2 and DESC, considering the whole dataset (right) and the campaign on the Southern Baltic (left).



Supplementary Figure S2. Ternary plots illustrating the relative proportions of micro-, nano-, and picophytoplankton as estimated using four different DPA variants: A) based on Uitz et al. (2006), B) derived from the OLS model, C) OLS model excluding Hex and D) obtained using the DESC model.



Supplementary Table S3. Statistical results of the reconstructed relationships between Chl-a and PSCs/PFTs against in situ data using the DPA estimated formula as for the a) OLS; b) OLS-2, c) DESC and d) Uitz et al. 2006 and the derived equation to estimate the fractions of %Chl-a

PSCs/PFTs	Equations	Parameters			RMSE	R^2	R	slope	intercept	std.err.	p-value
OLS		**a**	**b**	**c**							
micro	a * x + b	0.01	0.34	_	0.16	0.96	0.98	1	0.05	0.01	< 0.001
diatoms	a * x + b	0	0.29	_	0.15	0.94	0.97	0.38	0.53	0.01	< 0.001
dinoflagellates	a * x + b	0.01	0.04	_	0.12	0.79	0.89	0.91	0.12	0.03	< 0.001
nano	a * exp(-b * x) + c	-0.55	1.37	0.4	0.15	0.8	0.89	0.74	0.47	0.02	< 0.001
green&prochl	a * exp(-b * x) + c	0.17	0.07	0.04	0.14	0.28	0.53	0.26	0.45	0.02	< 0.001
cryptophytes	1 / (exp(a * x + b) + c*x)	-0.1	1.5	0.13	0.15	0.76	0.87	0.78	0.3	0.03	< 0.001
pico	1 / (exp(a * x + b) + c*x)	-0.23	1.79	1.48	0.08	0.09	0.29	0.15	0.33	0.03	< 0.001
OLS-2											
micro	a * x + b	0.01	0.34	_	0.17	0.96	0.98	1	0.05	0.01	< 0.001
diatoms	1 / (exp(a * x + b) + c*x)	0	1.21	0	0.15	0.66	0.81	0.63	0.6	0.03	< 0.001
dinoflagellates	a * x + b	0.01	0.04	_	0.12	0.79	0.89	0.91	0.12	0.03	< 0.001
nano	1 / (exp(a * x + b) + c*x)	-0.07	0.91	0.1	0.16	0.79	0.89	0.78	0.39	0.02	< 0.001
green&prochl	1 / (exp(a * x + b) + c*x)	-0.12	1.74	0.7	0.12	0.27	0.52	0.21	0.42	0.02	< 0.001
cryptophytes	1 / (exp(a * x + b) + c*x)	-0.1	1.5	0.13	0.14	0.76	0.87	0.78	0.3	0.03	< 0.001
pico	1 / (exp(a * x + b) + c*x)	-0.27	1.48	1.05	0.11	0.1	0.32	0.16	0.47	0.03	< 0.001
DESC											
micro	a * x + b	0.01	0.34	_	0.16	0.96	0.98	1.001	0.0531	0.01	< 0.001
diatoms	micro - dino	_	_	_	0.15	0.66	0.81	0.627	0.594	0.03	< 0.001
dinoflagellates	a * x + b	0.01	0.04	_	0.11	0.79	0.89	0.911	0.123	0.03	< 0.001
nano	1 / (exp(a * x + b) + c*x)	-0.07	0.84	0.1	0.15	0.8	0.89	0.778	0.4064	0.02	< 0.001
green&prochl	1 / (exp(a * x + b) + c*x)	-0.11	1.68	0.66	0.12	0.27	0.52	0.216	0.43	0.02	< 0.001
cryptophytes	1 / (exp(a * x + b) + c*x)	-0.1	1.5	0.13	0.14	0.76	0.87	0.784	0.2966	0.03	< 0.001
pico	1 / (exp(a * x + b) + c*x)	-0.26	1.57	1.17	0.09	0.1	0.31	0.156	0.4218	0.03	< 0.001

UITZ

micro	a * x + b	0.01	0.45	_	0.19	0.96	0.98	1.044	-0.0867	0.01	< 0.001
diatoms	1 / (exp(a * x + b) + c*x)	0	0.98	0.02	0.09	0.1	0.32	0.162	0.4284	0.03	< 0.001
dinoflagellates	a * x + b	0.01	0.08	_	0.17	0.77	0.88	0.946	0.1146	0.03	< 0.001
nano	a * x + b	-0.01	0.36	_	0.15	0.5	0.71	0.512	0.5962	0.03	< 0.001
green&prochl	1 / (exp(a * x + b) + c*x)	-0.14	1.52	0.57	0.15	0.23	0.48	0.19	0.5398	0.02	< 0.001
cryptophytes	1 / (exp(a * x + b) + c*x)	-0.12	2.43	0.38	0.08	0.61	0.78	0.645	0.1878	0.03	< 0.001
pico	1 / (exp(a * x + b) + c*x)	-0.26	1.59	1.13	0.09	0.1	0.32	0.162	0.4284	0.03	< 0.001

ANNEX IV

This annex was published in Microorganism

Characterization of Phytoplankton Composition in Lake Maggiore: Integrated Chemotaxonomy for Enhanced Cyanobacteria Detection

Elisabetta Canuti, Martina Austoni

Microorganisms 2024, 12(11), 2211; https://doi.org/10.3390/microorganisms12112211

microorganisms

MDPI

Article

Characterization of Phytoplankton Composition in Lake Maggiore: Integrated Chemotaxonomy for Enhanced Cyanobacteria Detection

Elisabetta Canuti [1,*] and Martina Austoni [2]

check for updates

Citation: Canuti, E.; Austoni, M. Characterization of Phytoplankton Composition in Lake Maggiore: Integrated Chemotaxonomy for Enhanced Cyanobacteria Detection. *Microorganisms* **2024**, *12*, 2211. https://doi.org/10.3390/microorganisms12112211

Academic Editor: Carlos E. De M. Bicudo

Received: 3 October 2024
Revised: 25 October 2024
Accepted: 28 October 2024
Published: 31 October 2024

[1] European Commission, Joint Research Centre (JRC), 21027 Ispra, VA, Italy

[2] National Research Council of Italy, Water Research Institute, CNR-IRSA, 28922 Verbania Pallanza, VB, Italy; martina.austoni@cnr.it

* Correspondence: elisabetta.canuti@ec.europa.eu

Abstract: Cyanobacterial blooms in lakes have increased in frequency and intensity over the past two decades, negatively affecting ecological and biogeochemical processes. This study focuses on the phytoplankton composition of Lake Maggiore, with a special emphasis on cyanobacteria detection through pigment composition. While microscopy is the standard method for phytoplankton identification, pigment-based methods provide broader spatiotemporal coverage. Between May and September 2023, five measurement campaigns were conducted in Lake Maggiore, collecting bio-geochemical and bio-optical data at 27 stations. The total Chlorophyll-a (TChl *a*) was measured, with concentrations ranging from 1.13 to 6.9 mg/m^3. Phytoplankton pigment composition was analyzed using High-Performance Liquid Chromatography (HPLC) and the CHEMTAX approach was applied for phytoplankton classification. The results were cross-validated using Principal Component Analysis (PCA), Hierarchical Cluster Analysis (HCA), and microscopic counts. Cyanobacteria were identified based on unique pigment markers, such as carotenoids. The HPLC-derived pigment classification results aligned well with both PCA and HCA and microscopic counts verified the accuracy of the pigment-based chemotaxonomy. The study demonstrates that pigment-based classification methods, when combined with statistical analyses, offer a reliable alternative for identifying cyanobacteria and other phytoplankton groups, with potential applications in support of remote sensing algorithm development.

Keywords: CHEMTAX; Lake Maggiore; HPLC pigments phytoplankton; cyanobacteria; bloom; chemotaxonomy

1. Introduction

Harmful algal blooms (HABs), particularly those caused by Cyanobacteria, have been increasing globally in both magnitude and frequency, posing significant risks to aquatic ecosystems and human health [1,2]. Accurately estimating Cyanobacteria presence is critical for mitigating the risks of cyanotoxin exposure and for evaluating both current and historical water quality trends in inland waters [3,4].

In northern Italy, lakes such as Lake Maggiore have experienced a rising frequency of HABs over recent decades [5–7] Lake Maggiore, an oligotrophic lake with a long history of monitoring [8–11], harbors high biodiversity within its phytoplankton community [12]. Recent studies indicate significant seasonal shifts in this community, with diatoms and cyanobacteria groups dominating during the summer months [13,14] In particular, the increasing presence of small cyanobacteria, such as *Aphanothece* and *Aphanocapsa*, during late summer suggests ecological changes potentially driven by climatic factors and increased water column stability. Continuous monitoring of these changes is essential to better understand their ecological implications.

Traditional microscopy, while reliable for species-level identification, is labor-intensive and lacks the spatial and temporal coverage required for comprehensive lake monitoring [15]. In contrast, chemotaxonomy, which classifies phytoplankton based on pigment composition, offers a rapid and cost-effective alternative [16–18]. The detection of cyanobacteria through their unique pigment markers—such as phycobiliproteins and carotenoids (e.g., echinenone, zeaxanthin)—is particularly useful for distinguishing them from other phytoplankton groups [19,20]. Whether the accurate quantification of phycobiliproteins pigments remains challenging due to spectral overlap with chlorophyll a (TChl *a*) [21,22], the quantification of carotenoids and chlorophylls in phytoplankton is a more assessed technique [23]. CHEMTAX algorithm to classify phytoplankton groups based on biomarker pigments [24] has already been successfully applied in marine, estuarine, and freshwater systems [20,25–31]. However, despite its advantages, chemotaxonomy has limitations, including the challenge of accurately describing the phytoplankton community when some pigments, like fucoxanthin, are common across multiple groups. Furthermore, critical challenges include the need to establish diagnostic pigment ratios that are ecosystem-specific and to ensure that samples are grouped based on similar environmental conditions to avoid misrepresentation of phytoplankton diversity [32–34]

This study aims to apply chemotaxonomy to characterize phytoplankton composition and improve cyanobacteria detection in Lake Maggiore. Unlike previous studies on Lake Maggiore [9–11,35–37] that relied primarily on traditional microscopic methods, this research leverages high-performance liquid chromatography (HPLC) alongside the CHEMTAX algorithm to classify phytoplankton groups based on biomarker pigments.

Furthermore, the findings were validated using Principal Component Analysis (PCA), Hierarchical Cluster Analysis (HCA), and microscopic counts, offering a robust and integrated approach for assessing phytoplankton diversity and cyanobacteria prevalence. By enhancing our understanding of the phytoplankton community structure, this research contributes to the development of more effective water quality monitoring tools, particularly in light of increasing HAB occurrences and their associated risks. Furthermore, the study's findings could aid the validation of satellite-based bio-optical algorithms for remote sensing applications in HAB detection [38–40].

2. Methods

Lake Maggiore, the second largest lake in Italy, is a subalpine lake situated at the border between Italy and Switzerland. It stretches over 66 km from the northern Swiss Canton of Ticino to the southern Italian region of Piedmont and Lombardy, with a maximum width of 4.5 km. Lake Maggiore's bathymetry is characterized by its deep glacial origin, with a maximum depth of 383 m. Hydrologically, Lake Maggiore is part of the Ticino River basin, with the Ticino River being its main tributary and outflow. Other significant tributaries include the Toce River in the northern region and several smaller streams along the western and eastern shores. The climate around Lake Maggiore is classified as temperate, with distinct seasonal variations due to the proximity to the Alps. Lake Maggiore experiences stratification due to temperature and density differences between surface and deep waters from late spring to autumn. Because of its great depth and local weather conditions, the lake does not fully mix every year. It is classified as holo-oligomictic, meaning

complete mixing only occurs during very cold and windy winters [41]. The average theoretical renewal time for Lake Maggiore water was estimated to be over 4 years [42]. Oxygenation of the deep waters is partially maintained by oxygen from tributaries. During winter overturns, mixing typically reaches depths of 100–150 m [41]. However, climate change is significantly affecting the lake's thermal and hydrodynamic properties, with major impacts on its oxygen levels [43,44].

2.1. Field Sample Collection

In the period May–September 2023, five measurement campaigns on Lake Maggiore (LM23) were performed for collecting bio-geochemical and bio-optical measurements in support of remote sensing validation activities. The LM23 campaigns consisted of 27 stations (Figure 1), measuring TChl *a*, phytoplankton pigments, and ancillary measurements. In addition, inherent optical properties (particulate absorption coefficient, chromophoric dissolved organic matter) and in-water radiometric measurements (the latter not discussed in the present paper) were collected. The seasons covered in the present study were spring (11 stations, 2 campaigns), summer (11 stations, 2 campaigns), and early autumn (5 stations, 1 campaign). The lake depth in the collection points varied from 60 m to the deepest point of Lake Maggiore (383 m) (Supplementary Figure S1). The LM23 campaigns were designed for satellite validation purposes and sampling points were selected to avoid coastal and bottom influences, ensuring representative mid-lake conditions. Additionally, a station near Ghiffa (LAT 45°.9666, LON 8°.6561), where Ghiffa is considered representative of lake conditions since is the deepest point of the lake (370 m depth), was included in each campaign for consistency across seasons. The time sampling interval was between 9:30 and 15:00 local time, as close as possible to the satellite overpass. The location of the sampling points was selected each time depending on the weather conditions, favoring clear sky and wind-calm conditions, enhancing the possibility of satellite-in situ matchup creation. For each measurement point, 20 L of water was collected 50 cm below the lake's surface in a pre-rinsed polypropylene container. Samples for SPM, TChl *a*, and algal pigments HPLC analysis were preconditioned onboard and stored in a freezer at −80 °C or liquid nitrogen upon the analysis. Ancillary measurements included surface temperature, suspended particulate matter (SPM), and Secchi Disk readings. Matching samples for microscopic determination were collected, preserved in acetic Lugol's solution, and stored in the dark at 4 °C until analysis.

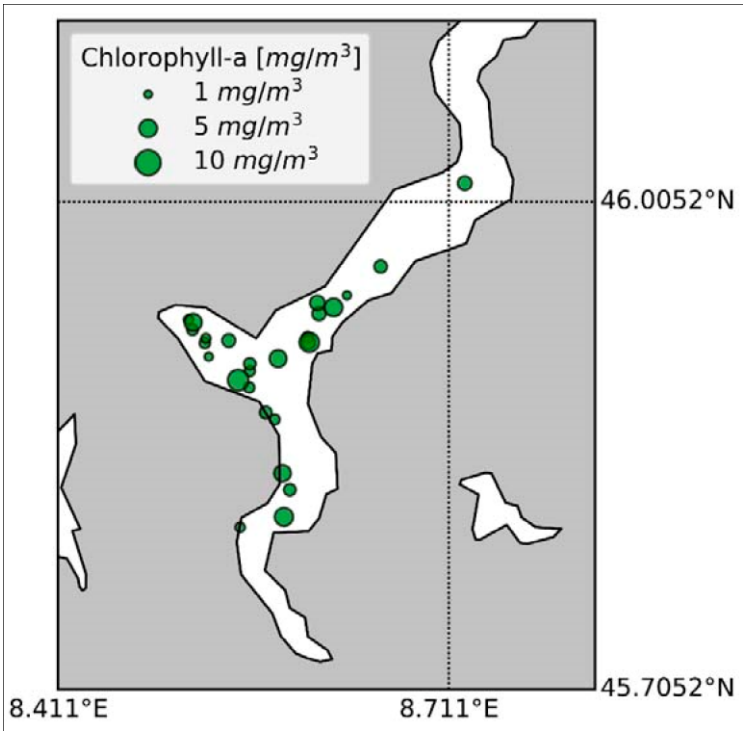


Figure 1. Spatial distribution of the LM23 collecting stations: the size is proportional to the TChl *a* content.

2.2. HPLC Pigment Dataset

The JRC method is a modification of [45] and is described in detail in [46]. The analysis was conducted using an HPLC/DAD system (HPLC 1200 Agilent Tech., USA), an RP-C8 monomeric column (Zorbax Eclipse XDB-C8, 3.5 µm particle size, 150 × 4.6ϕ), and an injection of 150 µL sample mixed in a loop with 365 µL of buffer (tetrabutylammonium acetate, 28 mM). The extraction of the natural samples for the JRC was conducted in 2.5 mL of 0.025 g/L of internal standard (Vitamin E acetate, Fluka v.2024.1.2, USA) dissolved in Acetone (HPLC gradient, Merck, Germany) and 150 µL of MilliQ water, soaked for 1 h at
−20 ◦C, sonicated for 90 s in ice with a sonication probe (Bandelin, Germany), soaked for 3.5–4 h at −20 ◦C, and then clarified by extraction through a 0.2 µm Teflon Syringe prior to injection.

To calibrate the HPLC, pigment standards from DHI Lab Products (Denmark) were used. The calibration curve consisted of nine points, covering a range of concentrations from a dilution close to three times the Signal-to-Noise Ratio (SNR) concentration to the standard concentration, as described by [47]. The linearity of the calibration curves was verified across the entire analysis range. The Low Limit of Detection (LOD) was defined as three times the instrumental SNR at each quantification wavelength. Peaks below the LOD are considered unidentified. The internal standard was detected at 222 nm, while chlorophyll *a* and chlorophyll *a* epymers and allomers were detected at 665 nm. All other pigments were detected at 450 nm.

The pigments that could be identified and quantified by the analysis of the natural samples were 19'-hexanoyloxyfucoxanthin (Hex), 19'-butanoyloxyfucoxanthin (But), alloxanthin (Allo), fucoxanthin (Fuco), peridinin (Perid), diatoxanthin (Diato), diadinoxanthin (Diadino), Zea, divinyl chlorophyll *a*

(DVchl*a*), monovinyl chlorophyll *a* (MVchl*a*), total chlorophyll *b* (TChl *b*), chlorophyll c1 + c2 (Chlc1c2), chlorophyll c3 (Chlc3), neoxanthin (Neo), violaxanthin (Viola), prasinoxanthin (Pras), Lutein (Lut), α,β-carotene (α-caro), β,βcarotene (β-caro), pheophorbide *a* (Pheo), pheophythin *a* (Phy). chlorophyllide *a* (Chlide *a*), Echin, Myxo, Anthe, and Gyro. Following the commonly adopted nomenclature, TChl *a* is intended as the sum of MVchl*a*, DVchl*a,* and Chlide a, and Caro is the sum of α-caro and β-caro [23].

In the chemotaxonomic interpretation via biomarker pigments applied to marine water [16–18,48] Diatoms are represented primarily by pigments such as Fuco, Diad, Diat, and various chlorophyll *c* derivatives. Dinophyceae are most closely associated with Peri and Diad. The phylum of Haptophyta is represented by Hex and TChl c1-c2. Chlorophyceae and Prasinophyceae are indicated by pigments such as Pras, Lut, and TChl *b*. Euglenophyceae, on the other hand, are associated with pigments like TChl b. Cyanobacteria are primarily linked to Zea and DVChl *a* and DVChl *b*, while Cryptophyceae are represented by Allo and TChl *c*. Chlide *a* is identified as a marker for the senescence of diatoms, indicating the aging and breakdown of these algae [23]. Zea and Echin are found to be representative of the pico-fraction and Cyanobacteria as secondary pigments, respectively [49]. In lakes [20], Myxo, Zea, Gyro, and Anthe were also associated with the Cyanobacteria population. Echin was found in *Anabaena* and *Dolichospermum* genera and *Anabaena variabilis* Kützing ex Bornet and Flahault [50]. *Dolichospermum lemmermannii* (Richter), pertaining to P.Wacklin, L.Hoffmann, and J.Komárek, has also been recognized in Lake [4,5,8–11,14,51]. In addition to *Dolichospermum lemmermannii*, the chroccocales cyanobacteria *Microcystis aeruginosa* (Kützing) Kützing, which has a high content of Echin and Myxo, was also identified in Lake Maggiore.

Pigment analysis also differentiated between photoprotective carotenoids (PPC) and photosynthetic carotenoids (PSCs). PPC, comprising Allo, Diad, Diato, Zea, and Caro, plays a crucial role in protecting phytoplankton from excessive light and oxidative stress. On the other hand, PSCs, including But, Fuco, Hex, and Peri, are primarily involved in the light-harvesting process for photosynthesis. In Lake Maggiore samples, But was not present and Hex could be considered as not significant due to its low concentration. The total accessory pigments (TAcc) were calculated as the sum of PPC, PSC, TChl b, and TChl c.

The phytoplankton proportion factor was introduced by Hooker et al., 2005, as a proxy of the different phytoplankton cell size fractions. The Microplankton Proportion Factor (mPF) is determined by the combined concentrations of Fuco and Peri, both of which are pigments commonly associated with larger phytoplankton, such as diatoms and dinoflagellates. This index provides an estimate of the relative abundance of microplankton, organisms typically larger than 20 µm. A higher mPF value indicates a greater presence of these larger phytoplankton, suggesting conditions favorable for species that thrive in environments with ample nutrients and light availability. The Nanoplankton Proportion Factor (nPF) focuses on pigments associated with smaller phytoplankton, specifically Hex, But, and Allo. Nanoplankton, ranging from 2 to 20 µm, often include species belonging to the phyla of Haptophyta and the class of Cryptophyceae, which, in the case of Lake Maggiore, are indicative of oligotrophic conditions especially during periods of low water mixing [7]. The nPF value reflects the relative contribution of these intermediate-sized phytoplankton to the total phytoplankton community. Lastly, the Picoplankton Proportion Factor (pPF) is derived from the concentrations of Zea and TChl *b*. Zea is

a key pigment in Cyanobacteria and Prochlorophytes, while chlorophyll *b* is found in green algae. These pigments are indicative of picoplankton, the smallest phytoplankton group, usually less than 2 µm in size. A higher pPF value suggests a predominance of these tiny, often photosynthetically efficient, organisms, which are well-suited to low-nutrient environments or deeper water layers where light is limited.

The HPLC dataset included the pigment characterization of the twenty-seven LM23 stations and of pure cultures of *Anabaena* sp. PCC 7120 and *Microcystis aeruginosa.*

2.3. Methods in Data Analysis

2.3.1. HPLC Dataset Quality

The criteria established by Aiken et al. [52] and used for assessing the quality of datasets for bio-optical algorithm development [18] (were applied to evaluate the internal consistency of the LM23 database. The co-variation in log-transformed total chlorophyll a (TChl *a*) and the sum of accessory pigments (TAcc) [53] was verified. The other criteria included i. the correlation between TChl *a* and TAcc, with a variance ($r^2 > 0.9$) and a slope within the range of 0.7–1.4; ii. the difference between TChl *a* and TAcc being less than 0.3 TPig; and iii. 85% of the collected stations that satisfy the first two criteria.

2.3.2. Biomass of Individual Plankton Groups

CHEMTAX, developed by [24], is a powerful tool for estimating the biomass of the individual phytoplankton groups as TChl *a*. The composition of phytoplankton communities is based on pigment concentrations measured by HPLC. The CHEMTAX algorithm utilizes an iterative process to refine the initial ratio matrix (F0), adjusting it based on the actual pigment concentrations observed in the samples. Through this iterative adjustment, CHEMTAX minimizes the discrepancy between measured and estimated pigment concentrations, ultimately converging on a solution that best represents the phytoplankton community structure. Our analysis included 12 key pigments: TChl *a*, TChl *b*, TChl *c*, Fuco, Zea, Peri, Allo, Diad, Echin, β,β-Caro, Chlide *a*, and Chl *c3*. These pigments were selected based on their specific association with different phytoplankton groups and their relevance in the ecological study of Lake Maggiore. The 6 phytoplankton groups considered in our analysis were *Bacillariophyceae* (Diatoms), *Cyanobacteria*, *Chlorophyceae*, *Cryptophyceae*, *Dinophyceae*, and *Chrysophyceae*. These groups included a broad range of the primary producers present in Lake Maggiore, providing a comprehensive overview of the phytoplankton community structure. The F0 matrix (Supplementary Table S1) was built using the available lake literature [20]. To account for the inherent variability and potential inaccuracies in the pigment ratio matrix, we set the ratio limit matrix to the default value of 500%. This conservative approach allowed for a broad range of potential ratios, thereby accommodating natural variability and reducing the risk of introducing bias from overly restrictive limits.

2.3.3. Hierarchical Cluster Analysis (HCA)

A hierarchical cluster analysis was performed on the HPLC pigment dataset, with pigment values normalized to TChl *a* (e.g., Fuco a ratios). This method, utilizing Ward's linkage approach (minimizing inner squared distances), was based on correlation (R, Pearson correlation between phytoplankton pigment ratios) distance calculations. This approach aligns with methodologies used by Latasa et al. [54] and

Catlett et al. [55]. A dendrogram was constructed with a cutoff distance of 0.5 to segment the dataset into distinct phytoplankton community clusters. Each sample was subsequently assigned to one of these clusters based on the calculated correlation distances

2.3.4. Principal Component Analysis (PCA)

The Principal Component Analysis (PCA) was applied to the HPLC pigment dataset to decompose the data into orthogonal modes, which represent distinct patterns within the dataset. This method leverages eigenvalues to quantify the variance captured by each mode, where the majority of the dataset's variability is often explained by the first few components. Previous studies [26–59] have successfully employed PCA to reveal spatial and temporal patterns in pigment data. For this analysis, the dataset comprised 16 variables
(pigments normalized to TChl *a*), and the decomposition was conducted using a Singular Value Decomposition (SVD), as follows:

$$\mathrm{X} = \mathrm{U}\Sigma\mathrm{V}^{\mathrm{T}},\ \mathrm{where} x_{ij} = \Sigma_{k-l,N} u_{ik}\sigma_k v_{kj} \quad (1)$$

The matrix U contains the principal components (loadings), and the eigenvalues in Σ represent the variance explained by each mode. The PCA modes allow for a clear visualization of the dominant structures in the data, but it is important to note that this approach does not make assumptions about the covariance between pigments. Though already used, further comparisons with alternative methods are needed to assess its full potential of application of PCA analysis in this context.

2.3.5. Network Community Detection Analysis (NCA)

Network-based community detection was performed on the phytoplankton pigment data using techniques from Kramer et al. [58] and Canuti et al. [59]. In this approach, pigments were represented as nodes in a network, with edges depicting the co-occurrence or similarity between pigments across different stations. Edge weights, calculated using Pearson's correlation coefficients of normalized pigment ratios, were used to create a similarity matrix, as follows:

$$s_{ij} = |corr\ x_i, x| \quad (2)$$

This similarity matrix was transformed into an adjacency matrix, which formed the basis of the Weighted Gene Network Community Analysis (WGNCA) [60]. WGNCA uses a beta parameter (set to 6 in this study) to control the influence of edge weights, enhancing the detection of meaningful pigment communities, as follows:

$$a_{ij} = (s_{ij})^{\beta} \quad (3)$$

The Louvain method [61] was employed to maximize network modularity, grouping pigments into distinct communities based on their connectivity. A modularity score greater than 0.3 indicated strong clustering, with well-defined internal community structures and weak external connections.

2.3.6. Microscopy Phytoplankton Determination

In this study, phytoplankton samples were taken at 50 cm below the lake surface. Phytoplankton determinations were carried out on subsamples preserved

in acetic Lugol's solution by placing a defined volume (between 5 mL and 25 mL) into a sedimentation chamber, following the Utermöhl technique [62,63], according to CEN 15204 [64] and classifying the taxa to the species level, where possible. Phytoplankton biovolume was calculated by approximating the shape of the algae to rotational solids; therefore, relative phytoplankton biomass was estimated from the density data with the original measurements of the species biovolume, where each taxon is associated to a geometric shape following Hillebrand et al. [65] and Sun and Liu [66] according to CEN 16695 [67].

3. Results

Dataset Overview

In our analysis, we examined pure species of *Anabaena* sp. PCC 7120 and *Microcystis aeruginosa* (Kützing), reporting their pigment concentrations in [ng/inj] and ratios to TChl *a* and β-caro (Table 1). For *Microcystis aeruginosa*, Myxo was found at 6.3 ng/inj, accounting for 3.0% of TChl a and 37.3% of β-caro. Zea was present at 27.1 ng/inj, representing 12.9% of TChl *a* and 160.4% of β-caro. Gyro was detected at 2.1 ng/inj, making up 1.0% of TChl *a* and 12.4% of β-caro. Echin showed a concentration of 5.2 ng/inj, constituting 2.5% of TChl *a* and 30.8% of β-caro. Lastly, β-caro itself was measured at 16.9 ng/inj, equating to 8.1% of TChl *a*. The relatively high concentration of Zea suggests its critical role in the lightharvesting process and protection against oxidative stress. The high concentration of Zea in particular highlights its significant presence and role within *Microcystis aeruginosa*, while the levels of other pigments like Myxo, Gyro, and Echin provided a detailed understanding of the pigment profile associated with this cyanobacterial species. From the microscopy analysis, *Microcystis aeruginosa* were present in the samples collected in July–August 2023. In contrast, *Anabaena* sp. PCC 7120 showed a much higher concentration of β-caro at 29 ng/inj, which constitutes 15.2% of TChl *a*. This higher proportion indicated a greater reliance on β-caro, potentially for photoprotection or structural purposes. Interestingly, while *Anabaena* sp. PCC 7120 exhibits a lower amount of Zea (2.1 ng/inj), it also had a markedly higher concentration of Echin at 22.4 ng/injection, making up 11.7% of TChl *a* and 77.2% relative to β-caro. This suggested that Echin plays a more significant role in *Anabaena* sp. PCC 7120 compared to *Microcystis aeruginosa.*

Table 1. Pigment concentrations (ng/inj) and ratios of *Anabaena* sp. PCC 7120 and *Microcystis aeruginosa*.

	Microcystis aeruginosa			***Anabaena PCC 7120***		
	Amount [ng/inj]	**Ratio%: TChl *a***	**Ratio%: β Caro**	**Amount [ng/inj]**	**Ratio%: TChl *a***	**Ratio%: β-caro**
Myxo	6.3	3.0	37.3	2.4	1.3	8.3
Zea	27.1	12.9	160.4	2.1	1.1	7.2
Gyro	2.1	1.0	12.4	3.1	1.6	10.7
Echin	5.2	2.5	30.8	22.4	11.7	77.2
β-caro	16.9	8.1	100	29	15.2	100.0
TChl *a*	209.2	100		190.8	100	

The variability of pigments in Lake Maggiore was analyzed by considering various statistical metrics including average concentrations, coefficients of variation (CV%), maximum and minimum values, and the number of measurements lying under the detection limit (Table 2). The results showed significant variability across different pigments (Figure 2), which provided insights into the phytoplankton dynamics and ecological conditions of the lake. Moreover, of the biomarker pigments considered, the Fuco, Allo, and TChl *b* were isolated in almost all the samples, Peri in 76% and Zea in 68% of the samples, while Hex and But were identified in 11% and 5% of the samples, respectively. Both Zea and Echin are markers for cyanobacteria, providing information on their abundance and distribution. The documented pathway involving β-caro as a precursor for Zea [47] highlights the importance of carotenoids in cyanobacterial physiology. The detection of Echin was particularly noteworthy due to its association with cyanobacteria *Dolicospermum* and *Anabaena* species, a cyanobacterial population present in Lake Maggiore. In contrast, the analysis did not detect Myxo and Anthe in natural samples, pigments also linked to cyanobacteria. Their absence may point to either a low abundance of these pigments or particular cyanobacterial species present in the Lake Maggiore.

Table 2. Selected statistical data characterizing the variability of the LM23 dataset: the average (mean) value along with the variation coefficient. The number of measurements falling below LOD for the pigments was also reported.

Parameters	Mean	CV%	Max	Min	Below Detection Limit
TChl *a* [mg/m^3]	2.933	54.7	6.901	1.133	-
TChl *b* [mg/m^3]	0.078	42.	0.155	0.024	-
TChl *c* [mg/m^3]	0.197	55.2	0.442	0.077	-
Fuco [mg/m^3]	0.513	55.9	1.101	0.145	-
Zea [mg/m^3]	0.199	42.6	0.404	0.097	-
Peri [mg/m^3]	0.136	95.2	0.491	0	1
Allo [mg/m^3]	0.146	52.3	0.336	0.023	-
Diato [mg/m^3]	0.040	58.1	0.08	0	4
Caro [mg/m^3]	0.195	55.9	0.509	0.076	-
Echin [mg/m^3]	0.038	54.2	0.071	0	2
SPM [g/m^3]	0.011	38.9	0.025	0.007	
T [°C]	20.53	17.02	25.62	15.5	
Depth [m]	222.6	49.7	383	60	
LAT			45.559	45.48186	
LON			8.347	8.30507	

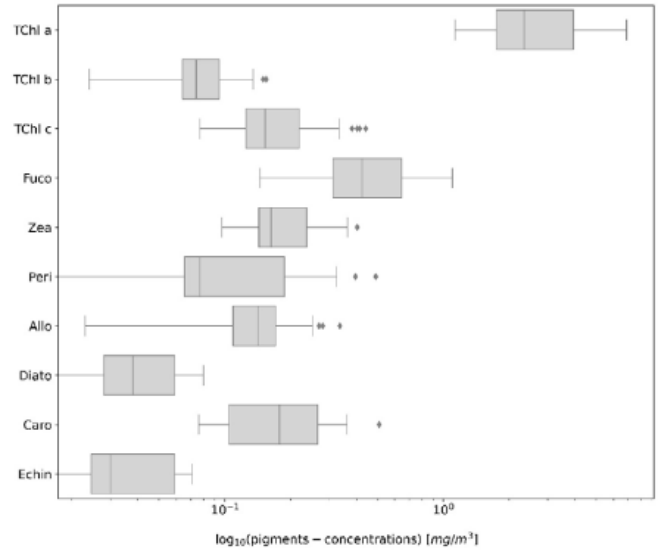


Figure 2. Boxplot of the pigment concentrations log$_{10}$-transformed) of Lake Maggiore 2023 (LM23) campaigns. Each box shows the interquartile range (IQR), with the central line indicating the median concentration. Whiskers extend to 1.5 times the IQR, with outliers represented by dots beyond this range.

TChl *a* exhibited an average concentration of 2.933 mg/m^3, with a CV% of 54.79%. This moderate variability indicated relatively consistent primary productivity levels across different sampling sites. The concentrations ranged from 1.133 mg/m^3 to a maximum of 6.901 mg/m^3. The TChl *a* concentration was higher in the samples collected in the afternoon compared to the station in the morning (Supplementary Figure S2). TChl *b* had a lower average concentration of 0.0789 mg/m^3 and a CV% of 42.6%, indicating less variability compared to TChl *a*. The values ranged from 0.024 mg/m^3 to 0.155 mg/m^3 and were present in all of the samples analyzed, which points to a consistent but lower abundance of green algae in the lake. TChl *c* showed an average concentration of 0.1976 mg/m^3 and a CV% of 55.2%,

with values ranging between 0.077 mg/m^3 and 0.442 mg/m^3. Fuco, indicative of diatom species, had an average concentration of 0.5131 mg/m^3 and a CV% of 55.9%. Its concentration varied from 0.145 mg/m^3 to 1.101 mg/m^3, reflecting significant spatial variability. No measurements lower than the detection limit were recorded, highlighting the consistent presence of Diatoms. Zea, associated with Cyanobacteria, had an average of 0.1991 mg/m^3 and a CV% of 42.6%. The concentrations ranged from 0.097 mg/m^3 to 0.404 mg/m^3, indicating moderate variability and ubiquitous detection in the samples, pointing to a consistent presence of prokaryotic phytoplankton. Peri, linked to Dinophyceae, exhibited substantial variability with an average concentration of 0.1365 mg/m^3 and a high CV% of 95.2%. Its values ranged to 0.491 mg/m^3, with one measurement lower than the limit of detection, indicating patchy distribution and episodic blooms. Allo, indicative of Cryptophyceae, had an average of 0.1467 mg/m^3 and a CV% of 52.3%, with concentrations between 0.023 mg/m^3 and 0.336 mg/m^3. The pigment had a widespread presence across the lake. Diato showed an average of 0.0405 mg/m^3 and a CV% of 58.1%. The max concentration was at 0.08 mg/m^3, with four samples where Diato was not identified, indicating a more sporadic distribution of certain phytoplankton groups. Caro had an average concentration of 0.1956 mg/m^3 and a CV% of 55.9%, with values between 0.076 mg/m^3 and 0.509 mg/m^3. Echin had an average of 0.0381 mg/m^3 and a CV% of 54.2%. The concentrations ranged to 0.071 mg/m^3 and were not quantified in two stations, indicating some spatial heterogeneity. SPM had the lowest average concentration of 0.0119 g/m^3 and the lowest CV% of 38.93%, with values ranging from 0.007 g/m^3 to 0.025 g/m^3, indicating relatively stable particulate levels in the lake. Geographical coordinates (LAT and LON) for the sampling stations were also recorded, showing a range of latitudes from 45.482 to 45.559 and longitudes from 8.305 to 8.347, covering the spatial extent of the Italian side of Lake Maggiore. The water temperature at the surface varied from a maximum of 25.6 °C to a minimum in April of 15.5 °C. The sampling station corresponds to the shore (60 m) to the deepest bottom (383 m) of Lake Maggiore.

The HPLC dataset was tested for quality assurance following the Aiken criteria. A strong log-correlation relationship was observed between TChl *a* and TAcc (Figure 3), with a correlation coefficient (r^2) of 0.88. The slope of this relationship (0.76) fell within the Aiken criteria (range 0.7–1.4), and the mean differences between TChl *a* and TAcc were lower than 30%, which indicates the consistency and reliability of the measurements and their interpretations (Figure 3).

The analysis of variance (ANOVA) and Tukey tests (Table 3) were conducted on LM23 samples to analyze seasonal variations in measured parameters. Among the pigments, TChl *a* exhibited a notable seasonal effect (p = 0.006), with higher concentrations observed in spring compared to fall (Tukey, p = 0.0045), although no significant differences were detected between summer and the other seasons. TChl *b* showed no significant seasonal variation (p = 0.420). For the total TChl *c*, significant seasonal differences were observed (p = 0.0003), with meaningful pairwise differences across all seasons (spring vs. fall, summer vs. fall, and spring vs. summer). Fuco exhibited moderate seasonal variation (p = 0.045), with significantly higher values in spring compared to fall (Tukey, p = 0.036). Zea showed a strong seasonal effect (p = 0.0017), with concentrations in spring significantly higher than those in fall (Tukey, p = 0.0015). Similarly, Peri displayed highly significant seasonal differences (p = 0.0001), with fall consistently exhibiting lower concentrations than both spring and summer. Allo also demonstrated strong seasonal variation (p = 0.002), with significantly higher concentrations in spring compared to both fall and

summer (p = 0.0041). Caro followed a similar pattern, with strong seasonal effects (p = 0.0017) and lower concentrations in fall relative to spring (p = 0.0014). Echin showed significant seasonal differences, specifically between fall and spring (p = 0.025). However, parameters such as diatom-specific pigments (Diato) and SPM showed no significant seasonal variation.

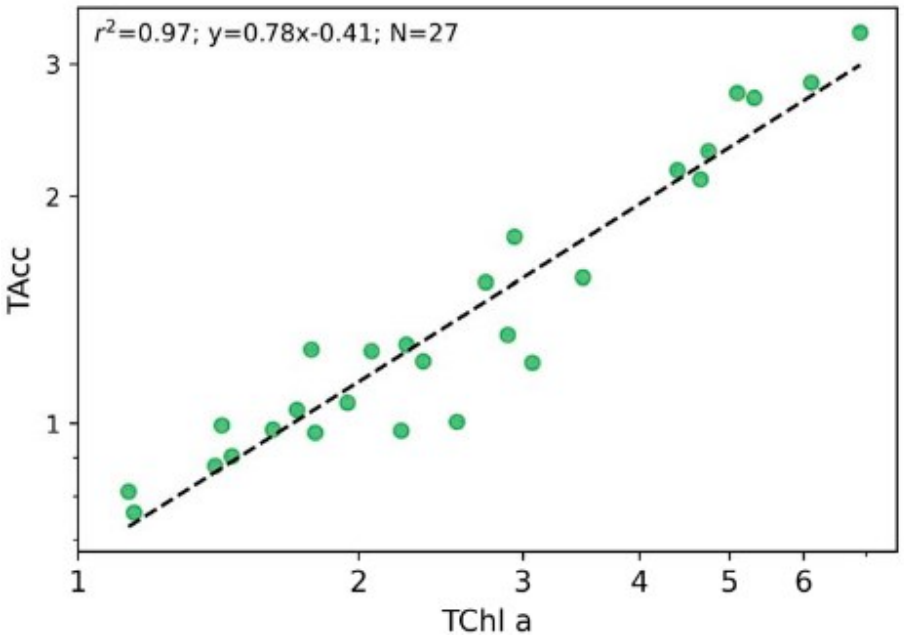


Figure 3. The log–log correlation TAcc/TChl *a* along the five Lake Maggiore Campaigns; the axis in log scale (mg/m^3). The green dots are the 27 stations of the five campaigns on Lake Maggiore.

Table 3. ANOVA analysis and Tukey test for the analysis of seasonal variation in Lake Maggiore.

Variable	**ANOVA F-Value**	**ANOVA *p*-Value**	**Tukey Significant Differences ($p < 0.05$)**
TChl *a*	6.31	0.0063	Fall vs. Spring (p = 0.0045)
TChl *b*	0.9	0.4204	None
TChl *c*	11.38	0.0003	Fall vs. Spring (p = 0.0003), Fall vs. Summer (p = 0.0492), Spring vs. Summer (p = 0.0297)
Fuco	3.54	0.045	Fall vs. Spring (p = 0.0357)
Zea	8.39	0.0017	Fall vs. Spring (p = 0.0015)
Peri	13.28	0.0001	Fall vs. Spring (p = 0.0001), Fall vs. Summer (p = 0.0216), Spring vs. Summer (p = 0.0267)
Allo	7.92	0.0023	Fall vs. Spring (p = 0.0165), Spring vs. Summer (p = 0.0041)
Diato	1.19	0.3221	None
Caro	8.44	0.0017	Fall vs. Spring (p = 0.0014)

Echin	4.18	0.0276	Fall vs. Spring (p = 0.0250)
SPM	2.16	0.1369	None

From the ternary plot of pPF, nPF, and mPF (Figure 4) it emerged that most of the samples cluster toward the middle-to-upper portion of the triangle, indicating that microplankton (mPF) generally has a significant presence in these samples. The points were concentrated around the 0.4 to 0.8 range on the mPF axis, suggesting that microplankton often constitute a substantial proportion of the phytoplankton community in Lake Maggiore. However, there is also a notable spread toward the lower part of the plot along the nPF and pPF axes, particularly for the m4 group (green dots). This suggests that in some samples, nanoplankton and picoplankton are more dominant, with these points reflecting a more balanced or even pico-/nanoplankton-dominated community. The color-coded points (m1 to m5) indicated different monthly campaigns, showing that there were distinct patterns in the phytoplankton community structure. For example, the m4 group (green, the month of July) tends to have lower mPF values, indicating a lower contribution of microplankton and a higher presence of either nanoplankton or picoplankton. In contrast, the other groups (e.g., m1, m2, m3, and m5) show varying degrees of microplankton dominance. Overall, the plot reflected the diversity in phytoplankton size classes within Lake Maggiore, with certain samples showing a strong microplankton presence while others demonstrate a more balanced or smaller-sized phytoplankton community.

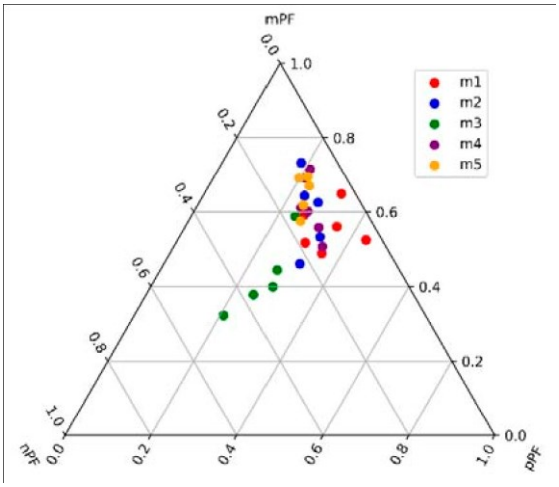


Figure 4. Ternary plot of functional Indices pPF, nPF, and mPF for LM23 campaigns.

The dendrogram provided a visual representation of the hierarchical clustering of various phytoplankton pigments based on their similarity in taxonomic distribution (Figure 5). The *y*-axis, labeled “Linkage Distance”, indicated the degree of similarity or difference between the pigments, with shorter linkage distances representing closer associations. The dendrogram showed a primary split into three major clusters at a linkage distance of 0.5, as marked by the red dashed line. The first cluster on the left included pigments that were closely associated with micro-phytoplankton and diatom senescence, with Chlide *a* being a prominent representative. This cluster also included Fuco, which, like Chlide *a*, was a pigment strongly linked to diatoms [23]. The second and third major clusters on the right were more diverse, comprising pigments that were distributed across a broader range of phytoplankton groups. Within these clusters, smaller sub-groups can be identified.

For example, pigments such as Pras, Viol, Hex, and Neo were linked at a closer distance, indicating that they were predominantly associated with Chlorophyceae. Another distinct grouping included Peri, which was uniquely representative of Dinophyceae and Zea, which is usually associated with prokaryotes such as pico-phytoplankton.

The analysis of PCA (Supplementary Figure S3) showed how different pigments, associated with specific phytoplankton groups, contributed to the overall composition and variance in the dataset. TChl *a*, TChl *b*, Fuco, and Zea were positioned near the origin, suggesting that these pigments were consistently present across all sampling campaigns, with limited variation across different environmental conditions.

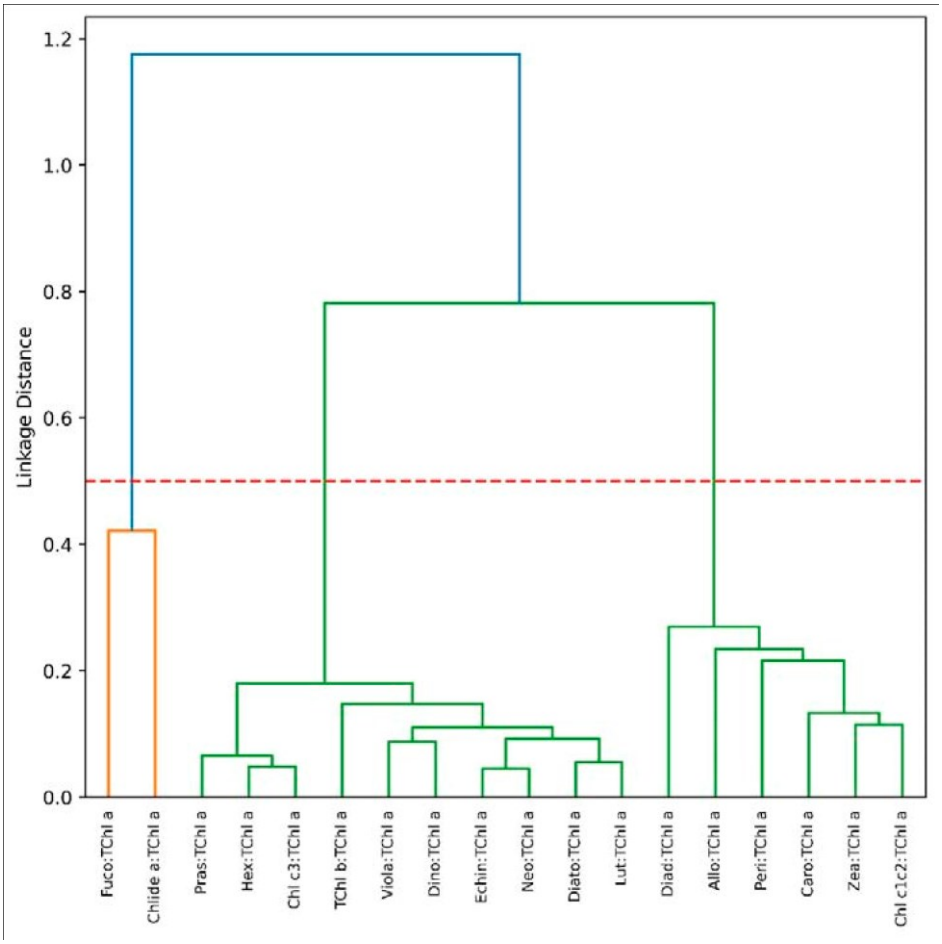


Figure 5. Hierarchical clustering of phytoplankton pigment ratios to TChl *a* for the LM23 dataset. The three-seizes major pigment communities (micro-, nano-, and pico-phytoplankton, from left to right) are identified based on a linkage distance cutoff of 0.5 (red dashed line).

Surface temperature (TSurf) and Allo showed a strong positive correlation along PC1, indicating that higher surface temperatures were associated with increased concentrations of these pigments, particularly during the m5 campaign. In contrast, the depth exhibited a negative correlation with both PC1 and PC2, linking it with lower pigment concentrations, particularly in samples from the m1 campaign, which experienced cooler water conditions. The PCA also applied to the pigment ratio, with the TChl *a*: the contribution of each ratio being considered in terms of loadings (Figure 6). The Principal Component 1 (PC1) was largely influenced by the

ratios of Diad:TChl *a*, Hex:TChl *a*, and Chl c1c2:TChl *a*. These pigments were significant markers for various phytoplankton groups, with Diad being indicative of diatoms and Hex-Fuco often associated with certain flagellates. Additionally, the ratios of Zea and Allo further underscored the contribution of Cyanobacteria and Cryptophyceae, respectively. On the other hand, negative loadings for Chlide *a* and Fuco suggested an inverse relationship, where the senescence of diatoms (indicated by Chlide *a*) and the presence of fucoxanthin-rich diatoms were less dominant in explaining PC1. PC2 was primarily driven by the ratio of TChl *b*:TChl *a*, highlighting its role in differentiating green algae, particularly chlorophytes, within the phytoplankton community. Diatomassociated pigments like Diato:TChl *a* and Fuco:TChl *a* also contribute positively to PC2, emphasizing their importance in this secondary source of variance. Conversely, Peri and Caro, the latter being a precursor of Zea and Echin, showed negative loadings. This suggested that while green algae and diatoms were prominent in PC2, dinoflagellates marked by Peri and Caro pigments were inversely related to this component. PC3 highlights the ratio of Neo to TChl *a* as a critical factor, marking the presence of certain chlorophytes.

The positive loading of Echin:TChl *a*, a secondary pigment for cyanobacteria, along with Lut:TChl *a*, suggested an important role for these groups in PC3. The negative loadings for Chlide *a* and Peri further delineated the distinct variance pattern, where diatom senescence and dinoflagellate presence are inversely related to the positively loading pigments. PC4 was significantly characterized by the ratio of Pras:TChl *a*, highlighting prasinophytes as a crucial component of the phytoplankton community. The positive contributions of Lut:TChl *a* and TChl *b*:TChl *a* indicated the role of green algae and related groups. Conversely, negative loadings for Hex:TChl *a* and Allo suggested that while Chlorophyceae were prominent, the presence of certain flagellates and Cryptophyceae was less influential in this component.

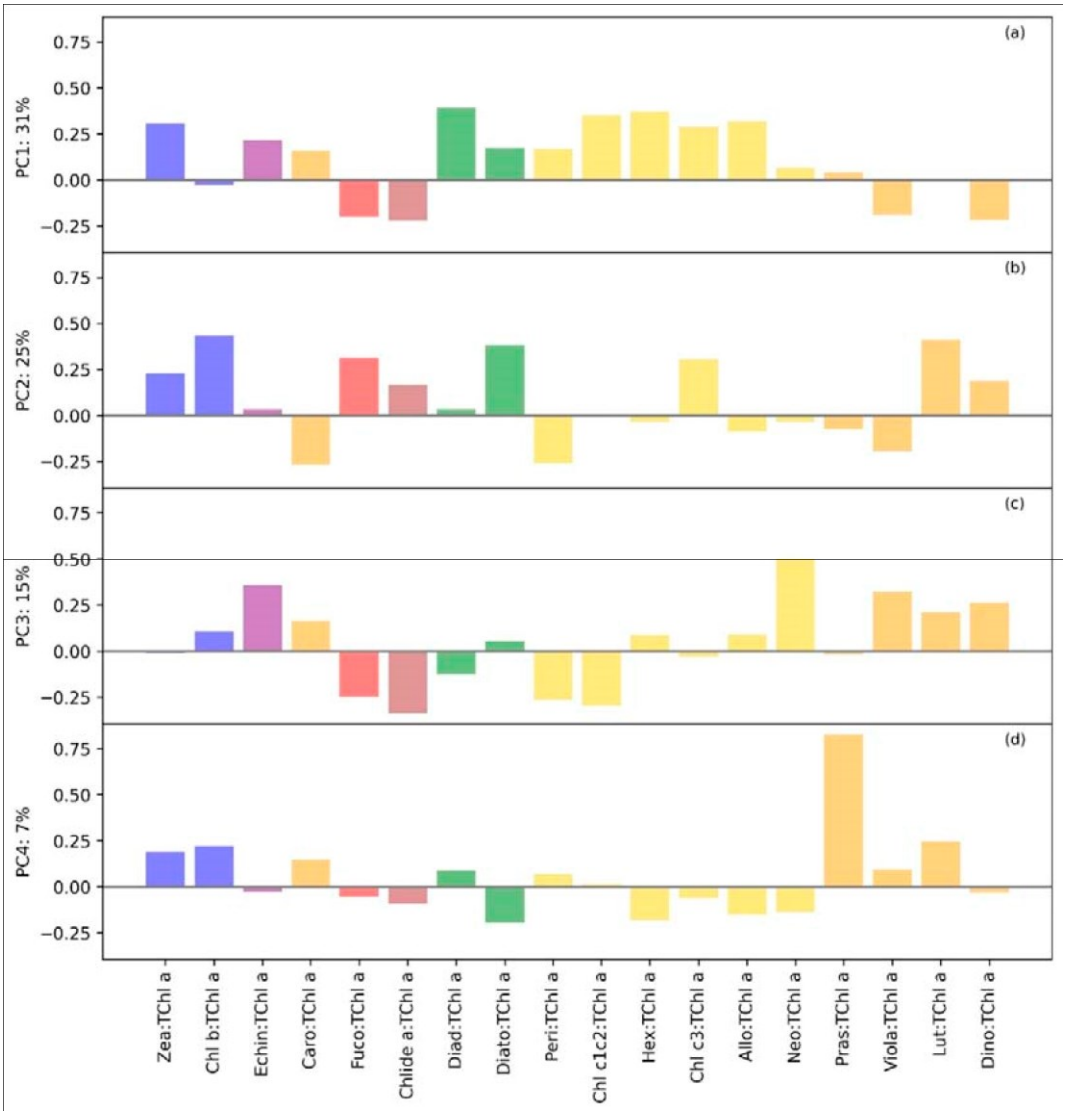


Figure 6. The loadings corresponding to the principal component modes for the pigments ratio to the TChl *a* are shown in panels (**a**–**d**) for the LM23 dataset.

The findings of unsupervised machine learning (HCA and PCA) were compared with the findings of the CHEMTAX analysis (Figure 7). The final matrix (F1, Table 4) was used for analyzing the whole LM23 pigments dataset. The dominant communities found through CHEMTAX analysis were Diatoms (Bacillariophyceae), Cryptophyceae, and Chrysophyceae, followed by Chlorophyceae and Cyanobacteria. The TChl *a* concentration showed a clear temporal variation across the months (Supplementary Figure S2). Early months (May to June) exhibited lower TChl *a* concentrations compared to later months (July to October), with an increase starting in July. Regarding the algal group CHEMTAX composition, diatoms (represented in red) dominate the phytoplankton community throughout the sampling period, especially in the early months, contributing significantly to the TChl *a* concentration. Chlorophyceae (green) were present in all months but showed only a minor contribution, with their presence remaining relatively stable across different stations. Cyanobacteria (blue) showed a more substantial presence from August to October, indicating a possible seasonal bloom. Dinophyceae (light yellow)

contributed minimally across most stations but showed slight increases in the later months. Cryptophyceae (light brown) exhibit significant contributions starting in July and continue to be prominent in the later months, suggesting favorable conditions for their growth during this period.

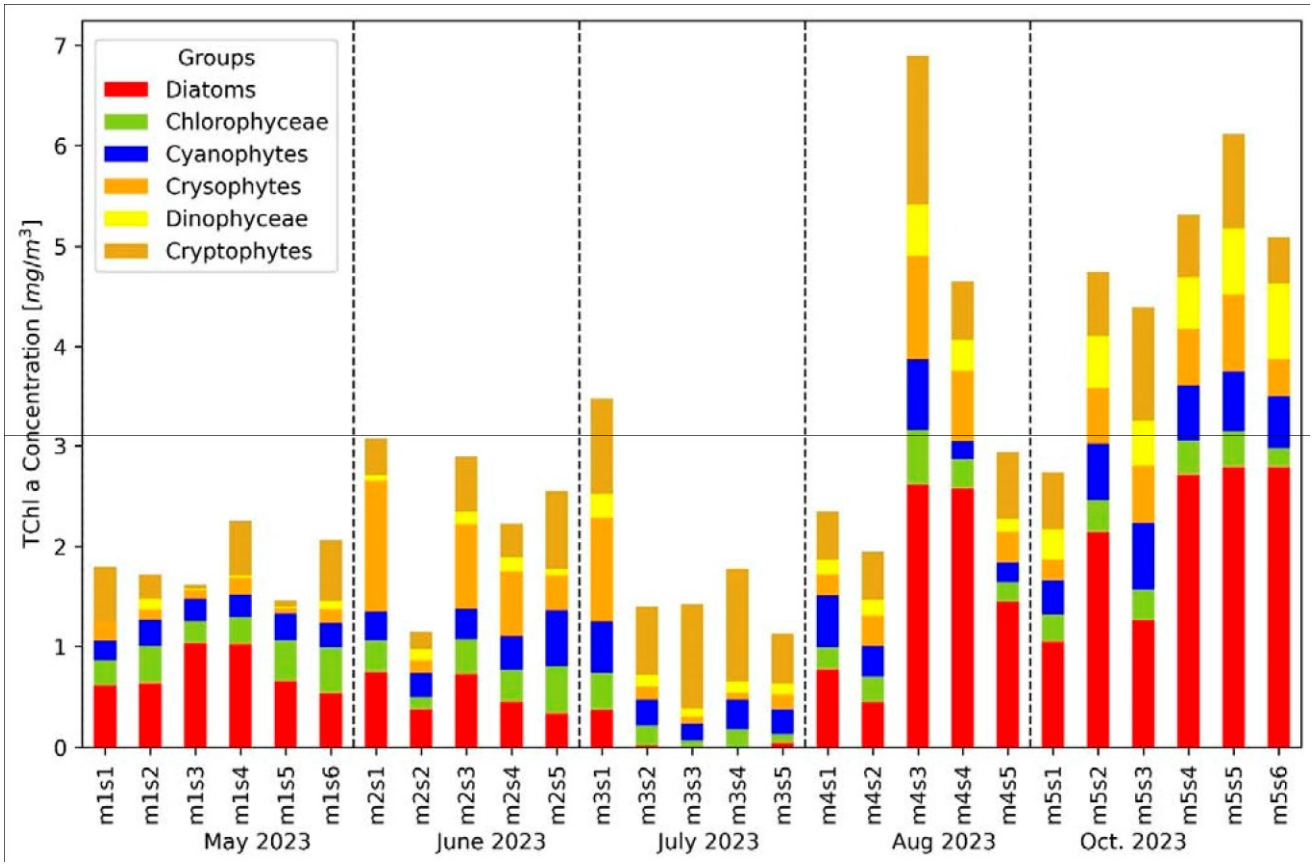


Figure 7. The algal group composition at various stations in Lake Maggiore is determined by CHEMTAX analysis, with each bar representing a specific station. The height of each bar indicates the TChl *a* concentration in mg/m³, and the stations are organized chronologically from May to October, as indicated by the vertical dashed lines separating each month.

Table 4. F1 Refined CHEMTAX matrix (from initial F0, Schlüter et al. [20]).

	Chl *c*1	**Peri**	**Fuco**	**Neo**	**Viola**	**Allo**	**Lut**	**Zea**	**Echin**	**TChl_*b***
Diatoms	0.021	0.000	0.039	0.000	0.000	0.000	0.000	0.005	0.000	0.000
Chlorophyceae	0.000	0.000	0.000	0.030	0.025	0.000	0.099	0.001	0.000	0.172
Cyanophytes	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.284	0.055	0.000
Crysophytes	0.000	0.000	0.207	0.000	0.094	0.000	0.000	0.001	0.000	0.000
Dinophyceae	0.000	0.334	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
Cryptophytes	0.000	0.000	0.000	0.000	0.000	0.143	0.000	0.000	0.000	0.000

Upon examining the monthly trends, May (stations m1s1 to m1s6) was characterized by lower TChl *a* concentrations with a strong dominance of diatoms. June (stations m2s1 to m2s5) showed similar patterns with low overall TChl *a* concentrations and diatom dominance. In July (stations m3s1 to m3s5), there was a

marked increase in TChl *a* concentrations, with a more diverse algal community, including significant contributions from Dinophyceae and Cryptophyceae, while Diatoms were almost absent. August (stations m4s1 to m4s5) saw further increases in TChl *a* concentrations, with continued high contributions from Diatoms and an increase in Cyanobacteria. October (stations m5s1 to m5s6) maintained high TChl *a* concentrations, with a notable presence of Cyanobacteria and Dinophyceae.

The phytoplankton composition in Lake Maggiore was analyzed by microscopic determination (Figure 8b) and was compared with the CHEMTAX analysis for matching stations (Figure 8a). Each bar represents a station, with the algal group composition shown as a percentage of total biomass for the microscopy determination and as TChl *a* concentration for CHEMTAX.

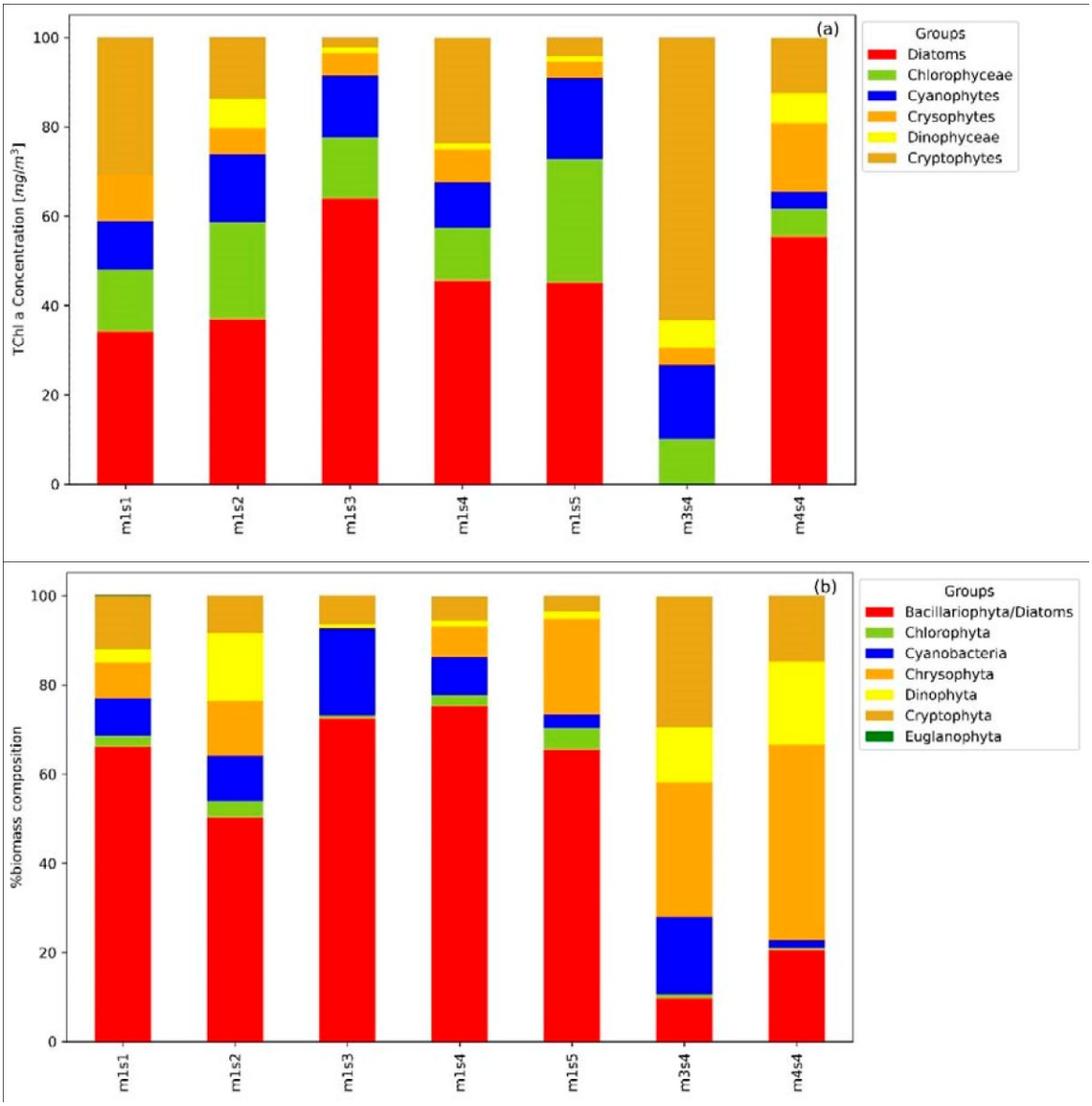


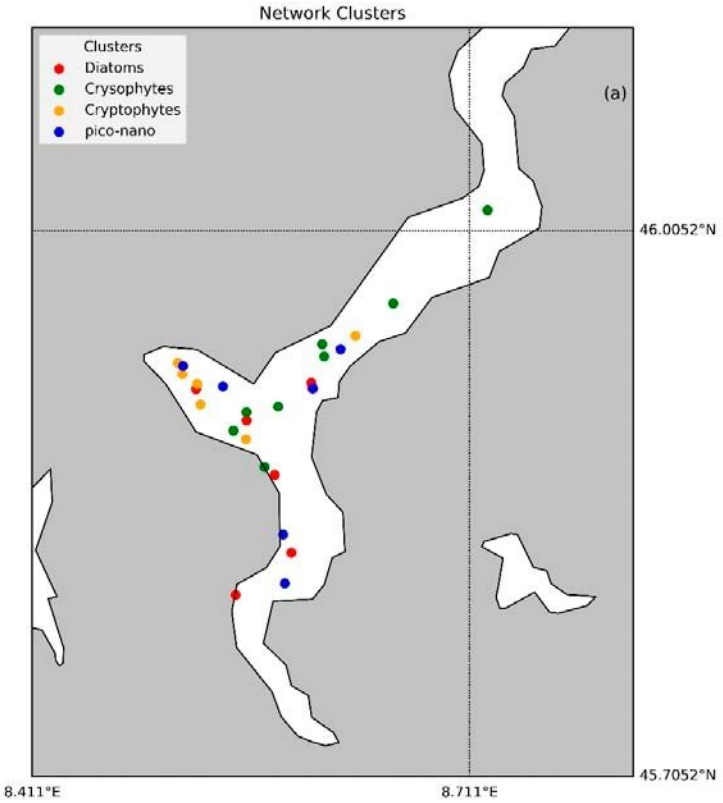
Figure 8. CHEMTAX distribution (**a**) and corresponding microscopy determination (**b**) for matched stations.

In the microscopy analysis, the dominant groups across most stations were Bacillariophyceae (Diatoms, in red) and Chrysophyta (in brown). Notably, at all the stations of the m1 campaign, Bacillariophyta makes up the majority of the biomass,

whereas stations m1s2 showed significant contributions from Cyanobacteria (blue) and m1s5 Chrysophyta (orange). Station m3s4 was characterized by a substantial presence of Dinophyceae (yellow) and Cryptophyceae (orange), indicating a diverse phytoplankton community. Station m4s4 shows a similar trend with a dominance of Chrysophyta and a notable increase in Dinophyta.

In the CHEMTAX analysis of the algal group composition for matching LM23, stations showed that the Diatoms were consistently dominant across all stations except m3s4, corroborating the findings from the microscopy analysis. The CHEMTAX results shown a higher contribution of Chlorophyceae (green) and Cryptophytes (brown) across several stations, particularly at station m1s5 and m3s4. Cyanobacteria (blue) shown significant presence at stations m1s3 and m3s4, similar to the microscopy results. Dinophyceae (yellow) and Crypsophytes (orange) were also prominent, particularly in the later stations (m3s4 and m4s4), aligning with the microscopy determinations observations.

The network community analysis of phytoplankton pigments yielded a modularity score of 0.18, indicating a relatively weak community structure. Modularity measures the strength of division within a network into distinct communities, with higher values reflecting a clearer separation. Four primary pigment communities were identified through this analysis. However, Fuco emerged as the dominant pigment at nearly all sampling stations. To refine the community identification, the results were re-evaluated using CHEMTAX.
According to CHEMTAX, the refined community structure includes Diatoms, Chrysophyceae, Cryptophyceae, and a fourth community not clearly distinguished. Specifically, Diatoms were identified as the primary community due to the prevalence of Fuco, while chrysophytes were recognized for their characteristic pigments. Cryptophyceae, on the other hand, did not show a clear single pigment correlation, leading to the absence of a distinct classification for this group. The spatial distribution of dominant phytoplankton groups at different stations in Lake Maggiore can be analyzed by using two distinct analytical methods: CHEMTAX analysis and network analysis (Figure 9). Accordingly, in the CHEMTAX classification, Diatoms dominate most stations, particularly in the southern and central parts of the lake, which corresponds with their typical prevalence in nutrient-rich environments. In contrast, Cryptophyceae are more dominant in certain northern stations, suggesting variations in local environmental conditions that favor different phytoplankton communities. The results of network analysis, clusters stations based on similarities in phytoplankton community structure, were categorized with the same color scheme as the CHEMTAX map but include an additional group, pico-nano (blue). The network clusters provide a more nuanced perspective, grouping stations together based on the overall similarity of their phytoplankton communities rather than the dominance of a single group. Comparing the two maps, it is evident that Diatoms were a predominant group across many stations in Lake Maggiore. However, the maps also highlight the spatial heterogeneity of the phytoplankton community.


Network Clusters
Clusters
Diatoms
Crysophytes
Cryptophytes
pico-nano
(a)
46.0052°N
45.7052°N
8.411°E
8.711°E

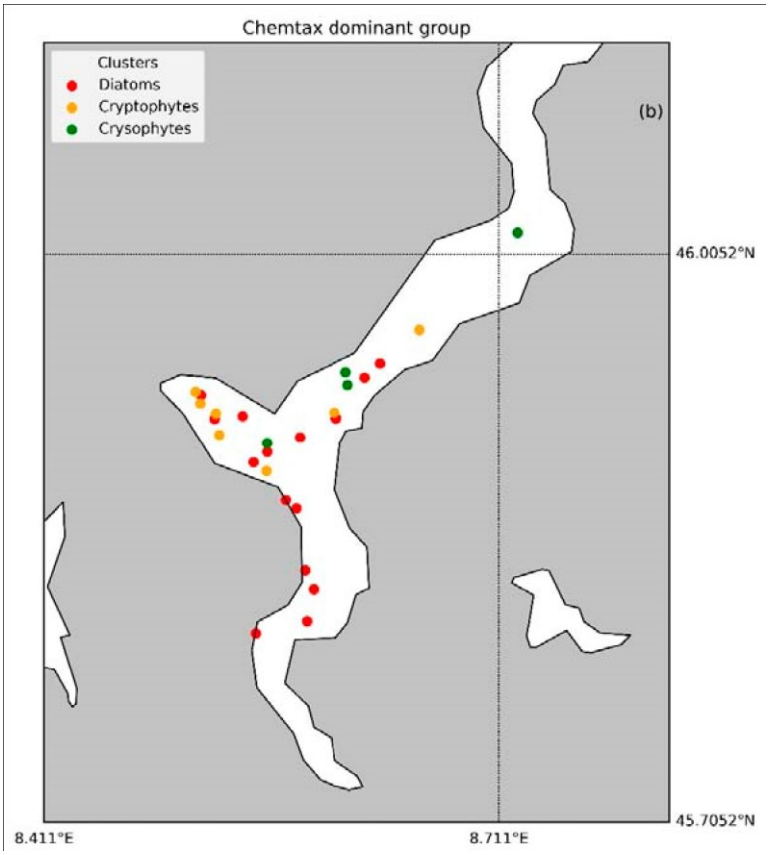

Chemtax dominant group
Clusters
Diatoms
Cryptophytes
Crysophytes
(b)
46.0052°N
45.7052°N
8.411°E
8.711°E

Figure 9. Dominant phytoplankton groups at each station as identified by Network (**a**) and CHEMTAX

(**b**) analysis. Stations are color-coded to indicate the prevailing algal groups: red for diatoms, green for Chrysophyceae, yellow for Cryptophyceae, and blue for pico-nano mixed fraction (only in Network analysis).

4. Discussion

The analysis of LM23 pigments dataset of Lake Maggiore provided insights into the dynamic and diverse nature of the lake's phytoplankton community, reflecting its oligotrophic state and seasonal variations. The observed variability in key pigments, such as TChl *a* and Fuco, points to significant spatial and temporal shifts in primary productivity and community composition. The current study's findings align with those from the 2022 CIPAIS report and the previous literature, confirming that diatoms dominate the phytoplankton community during different times of the year [12,68]. This seasonal succession is typical of oligotrophic lakes, where Diatoms thrive during cooler nutrient-rich conditions, while Cyanobacteria and Dinophyceae take over in warmer stratified waters [13,67]. The ANOVA analysis highlights the seasonal effect, with TChl *a* concentrations being significantly higher in spring compared to fall. This finding aligns with our observations of diatom dominance during the spring months, emphasizing how nutrient availability during this period fosters robust primary productivity. Additionally, the significant variations in Echin across seasons further supported the seasonal shifts in community composition, highlighting how environmental conditions affect phytoplankton dynamics. The detection of Zea, a cyanobacteria marker, in 68% of the samples suggests a consistent presence of small cyanobacteria species throughout the year, while analysis based only on in our superficial samples did not capture the cyanobacterial dominance that can occur during late summer. The presence of *Anabaena* species, as indicated by Echin, further reinforces the notion of a diverse cyanobacterial community. However, the absence of pigments such as Myxo and Gyro, typically associated with *Microcystis* species, despite microscopy findings, suggests that these cyanobacteria may be present in low abundance [4].

The spatial analysis showed that diatom-related pigments like Fuco exhibited significant variability, likely due to localized environmental conditions such as nutrient availability and water column stability [11]. The patchy distribution of Peri, a dinoflagellate marker, suggests episodic blooms driven by specific triggers, such as increased stratification and light availability in summer [5]. The PCA results provided additional context, revealing shifts in phytoplankton community health, particularly the inverse relationship between Chlide *a*, a marker of senescence, and Fuco, indicating changes in diatom vitality throughout the sampling period. The absence of significant markers for prasinophytes and haptophytes, such as But and Hex, suggests that these groups play a lesser role in Lake Maggiore compared to other taxa, reinforcing the findings of earlier studies [9,13]. In terms of phytoplankton size fractions, the dominance of micro-phytoplankton in high-TChl *a* samples highlights the importance of larger phytoplankton in driving productivity, especially during periods of increased nutrient availability. Meanwhile, nanoand pico-phytoplankton maintained a stable presence, underscoring their adaptability to nutrient-poor conditions and their contribution to the lake's overall biodiversity [12]. The CHEMTAX analysis largely confirmed the results from PCA and HCA, with diatoms emerging as the dominant group. The diatom bloom observed in early spring, followed by an increase in cryptophytes and large dinoflagellates by mid-summer, mirrors seasonal trends noted in the Ghiffa station data [11].

Buchaca et al. [25] suggested that in oligotrophic systems like Lake Maggiore, caution was needed when using CHEMTAX, particularly under low-irradiance or nutrient-replete conditions, such as those found beneath ice cover or, in our case, during the stratified summer months. CHEMTAX may overestimate the biovolume of certain taxa, particularly chlorophytes, when their abundance is low and colony-forming species are present. This insight mirrors the discrepancies observed in our study, where CHEMTAX and microscopy findings for chlorophytes diverged in late summer-autumn samples. CHEMTAX's pigment-based classification offers a broader perspective on community composition, while microscopy provides more detailed taxonomic insights, a combination that provided better understanding in spatio-temporal phytoplankton dynamics in complex ecosystems like Lake Maggiore [68,69]. While the CHEMTAX map gives a direct view of the dominant phytoplankton group at each station, the network analysis adds depth by showing how stations with similar community structures were related, revealing potential underlying ecological or environmental gradients. This highlights the dynamic nature of the lake's ecosystem and the influence of local conditions on phytoplankton communities. This analysis, paired with previous studies, suggests that the lake's phytoplankton composition was heavily influenced by environmental drivers such as nutrient fluxes, water temperature, and light availability, with spatial and temporal variations adding further complexity [9].

To improve monitoring and understanding of phytoplankton dynamics in Lake Maggiore, a combination of in situ and satellite-based approaches should be considered. In situ methods, such as pigment analysis and the application of CHEMTAX algorithm, offer detailed insights into the composition of phytoplankton communities, but they can be resource-intensive and provide only limited spatial and temporal coverage. Satellite remote sensing, on the other hand, can offer continuous large-scale observations, particularly of surface chlorophyll concentrations, which can serve as a proxy for phytoplankton biomass. Recent advances in hyperspectral sensors allow for the detection of specific phytoplankton groups, including diatoms and cyanobacteria, making remote sensing an increasingly valuable tool for ecosystem monitoring. Combining these approaches would provide a more comprehensive assessment of phytoplankton variability and productivity [38,39]. In situ data could be used to calibrate and validate satellite-derived estimates, ensuring greater accuracy in detecting phytoplankton blooms and understanding their drivers. Additionally, integrating physical and chemical data, such as water temperature, nutrient levels, and stratification patterns, would help to link phytoplankton dynamics with environmental conditions, offering insights into the ecological mechanisms that determine community composition.

5. Conclusions

This study applied chemotaxonomy through HPLC pigment analysis and the CHEMTAX algorithm to provide a detailed characterization of the phytoplankton community in Lake Maggiore. The findings revealed a dynamic seasonal succession dominated by diatoms in cooler nutrient-rich periods and cyanobacteria and dinoflagellates during warmer stratified conditions. Despite the presence of cyanobacteria throughout the year, 2024 did not exhibit the late-summer dominance seen in previous years, suggesting interannual variability influenced by environmental factors.

The integration of PCA, HCA, and microscopy with pigment data allowed for a comprehensive assessment of phytoplankton diversity and confirmed the

suitability of chemotaxonomy as a complementary method to traditional microscopic techniques. However, discrepancies in chlorophyte abundance between CHEMTAX and microscopy highlight the need for cautious interpretation, particularly under conditions of low abundance. The study also underscores the importance of continuous monitoring, especially for detecting subtle shifts in community composition due to climate-driven changes in lake conditions.

Moving forward, the combination of in situ methods with satellite remote sensing holds promise for more comprehensive and efficient monitoring of phytoplankton dynamics and HABs in Lake Maggiore [70]. Future efforts should focus on calibrating satellite-derived estimates with in situ data to enhance the accuracy of bio-optical algorithms for detecting and monitoring phytoplankton blooms and their ecological drivers. These insights are crucial for developing strategies to mitigate the impacts of HABs and maintain the ecological integrity of Lake Maggiore.

Supplementary Materials: The following supporting information can be downloaded at: https://www.mdpi.com/article/10.3390/microorganisms12112211/s1, Figure S1: TChl a concentrations (mg/m^3) and the correspondent bottom-depth for the LM23 stations; Figure S2: Temporal variations in TChl a concentrations (mg/m^3) across different months with reference (dashed grey vertical line) to the station collected close to the 10:00 UTC time; Figure S3: Principal Component analysis of pigments and main environmental variables; Table S1: CHEMTAX F0 matrix from Schlüter et al. [20].

Author Contributions: E.C. contributed to this paper by writing the original draft, conceptualization and investigation, data curation, formal analysis, methodology, validation, reviewing and editing of the content. M.A. contributed to the methodology, formal analysis, and reviewing and editing of the content. All authors have read and agreed to the published version of the manuscript.

Funding: This study has been supported by the European Commission Directorate General Joint Research Centre (JRC) and the Copernicus Program.

Data Availability Statement: The data will be made available on a reasonable request.

Acknowledgments: The authors would like to thank Frédéric Melin and Ilaria Cazzaniga from Joint Research Centre, European Commission for their support during the field campaigns on Lake Maggiore.

Conflicts of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflicts of interest.

References

1. Mishra, S.; Stumpf, R.; Schaeffer, B.; Werdell, P.; Loftin, K.; Andrew, M. Measurement of Cyanobacterial Bloom Magnitude using Satellite Remote Sensing. *Sci. Rep.* **2019**, *9*, 18310. [CrossRef] [PubMed]
2. Havens, K.E. Cyanobacteria blooms: Effects on aquatic ecosystems. In *Cyanobacterial Harmful Algal Blooms: State of the Science and Research Needs*; Hudnell, H.K., Ed.; Advances in Experimental Medicine and Biology; Springer: New York, NY, USA, 2008; Volume 619. [CrossRef]
3. Dörnhöfer, K.; Klinger, P.; Heege, T.; Oppelt, N. Multi-sensor satellite and in situ monitoring of phytoplankton development in a eutrophic-mesotrophic lake. *Sci. Total Environ.* **2018**, *612*, 1200–1214. [CrossRef] [PubMed]
4. Bresciani, M.; Giardino, C.; Lauceri, R.; Matta, E.; Cazzaniga, I.; Pinardi, M.; Lami, A.; Austoni, M.; Viaggiu, E.; Congestri, R.; et al. Earth observation for monitoring and mapping of cyanobacteria blooms. Case studies on five Italian lakes. *J. Limnol.* **2016**, *76*. [CrossRef]
5. Mosello, R.; Ambrosetti, W.; Arisci, S.; Bettinetti, R.; Buzzi, F.; Calderoni, A.; Carrara, E.; de Bernardi, R.; Galassi, S.; Garibaldi, L.; et al. Trend of water quality of the deep subalpine lakes in relation to anthropic pressure and climate - Evoluzione recente

della qualità delle acque dei laghi profondi sudalpini (Maggiore, Lugano, Como, Iseo e Garda) in risposta alle pressioni antropiche e alle variazioni climatiche. *Biol. Ambient.* **2010**, *24*, 167–177.

6. Salmaso, N.; Capelli, C.; Shams, S.; Cerasino, L. Expansion of bloom-forming *Dolichospermum lemmermannii* (Nostocales, Cyanobacteria) to the deep lakes south of the Alps: Colonization patterns, driving forces and implications for water use. *Harmful Algae* **2015**, *50*, 76–87. [CrossRef]
7. Salmaso, N.; Boscaini, A.; Pindo, M. Unraveling the Diversity of Eukaryotic Microplankton in a Large and Deep Perialpine Lake Using a High Throughput Sequencing Approach. *Front. Microbiol.* **2020**, *11*, 789. [CrossRef]
8. Mosello, R.; Ruggiu, D. Nutrient load, trophic conditions and restoration prospects of Lake Maggiore. *Int. Rev. Ges. Hydrobiol. Hydrogr.* **1985**, *70*, 63–75. [CrossRef]
9. Ruggiu, D.; Morabito, G.; Panzani, P.; Pugnetti, A. Trends and relations among basic phytoplankton characteristics in the course of the long-term oligotrophication of Lake Maggiore (Italy). *Hydrobiologia* **1998**, *369*, 243–257. [CrossRef]
10. Marziali, L.; Guzzella, L.; Salerno, F.; Marchetto, M.; Valsecchi, L.; Tasselli, S.; Roscioli, C.; Schiavon, A. Twenty-year sediment contamination trends in some tributaries of Lake Maggiore (Northern Italy): Relation with anthropogenic factors. *Environ. Sci. Pollut. Res.* **2021**, *28*, 38193–38208. [CrossRef]
11. CIPAIS (Commissione Internazionale per la Protezione delle Acque Italo-Svizzere); CNR-IRSA. *Ricerche Sull'evoluzione del Lago Maggiore*; Aspetti Limnologici. Programma Triennale 2022–2024. Campagna 2022; Commissione Internazionale per la Protezione Delle Acque Italo-Svizzere, Ed.; Sede di Verbania: Verbania Pallanza, VB, Italy, 2023; p. 109. (In Italian)
12. Kamenir, Y.; Morabito, G. Lago Maggiore oligotrophication as seen from the long-term evolution of its phytoplankton taxonomic size structure. *J. Limnol.* **2009**, *68*, 146–161. [CrossRef]
13. Morabito, G.; Oggioni, A.; Austoni, M. Resource ratio and human impact: How diatom assemblages in Lake Maggiore responded to oligotrophication and climatic variability. In *Phytoplankton Responses to Human Impacts at Different Scales*; Developments in Hydrobiology; Salmaso, N., Naselli-Flores, L., Cerasino, L., Flaim, G., Tolotti, M., Padisák, J., Eds.; Springer: Dordrecht, The Netherlands, 2012; Volume 221. [CrossRef]
14. Austoni, M.; Marchetto, A. *Struttura delle Associazioni Fitoplanctoniche nel Lago Maggiore e loro Modificazioni in Relazione a Fattori di Controllo Trofici e Climatici in CNR IRSA*; Sede di Verbania. 2019. Ricerche Sull'evoluzione del Lago Maggiore. Aspetti Limnologici. Programma Triennale 2016–2018. Campagna 2018 e Rapporto Triennale 2016–2018; Commissione Internazionale per la Protezione Delle Acque Italo-Svizzere, Ed.; Sede di Verbania: Verbania, Italy, 2019; pp. 55–59.
15. HELCOME. *Guidelines for Monitoring of Phytoplankton Species Composition, Abundance and Biomass*; HELCOM: Helsinki, Finland, 2023; p. 22. [CrossRef]
16. Vidussi, F.; Claustre, H.; Manca, B.B.; Luchetta, A.; Marty, J.C. Phytoplankton pigment distribution in relation to upper thermocline circulation in the eastern Mediterranean Sea during winter. *J. Geophys. Res.* **2001**, *106*, 19939–19956. [CrossRef]
17. Uitz, J.; Claustre, H.; Morel, A.; Hooker, S.B. Vertical distribution of phytoplankton communities in open ocean: An assessment based on surface chlorophyll. *J. Geophys. Res.* **2006**, *111*, C08005. [CrossRef]
18. Hirata, T.; Hardman-Mountford, N.J.; Brewin, R.J.W.; Aiken, J.; Barlow, R.G.; Suzuki, K.; Isada, T.; Howell, E.; Hashioka, T.; Noguci-Aita, M.; et al. Synoptic relationships between surface chlorophyll-a and diagnostic pigments specific to phytoplankton functional types. *Biogeosciences* **2011**, *8*, 311–327. [CrossRef]
19. Schlüter, L.; Garde, K.; Kaas, H. Detection of the toxic cyanobacteria *Nodularia spumigena* by means of a 4-keto-myxoxanthophylllike pigment in the Baltic Sea. *Mar. Ecol. Prog. Ser.* **2004**, *275*, 69–78. [CrossRef]
20. Schlüter, L.; Behl, B.; Striebel, M.; Stibor, H. Comparing microscopic counts and pigment analyses in 46 phytoplankton communities from lakes of different trophic state. *Freshw. Biol.* **2016**, *61*, 1627–1639. [CrossRef]
21. Lauceri, R.; Bresciani, M.; Lami, A.; Morabito, G. Chlorophyll a interference in phycocyanin and allophycocyanin spectrophotometric quantication. *J. Limnol.* **2018**, *77*, 169–177. [CrossRef]
22. Wojtasiewicz, B.; Ston´-Egiert, J. Bio-optical characterization of selected cyanobacteria strains present in marine and freshwater ecosystems. *J. Appl. Phycol.* **2016**, *28*, 2299–2314. [CrossRef]
23. Roy, S.; Llewellyn, C.A.; Egeland, E.S.; Johnsen, G. *Phytoplankton Pigments, Characterization, Chemotaxonomy and Applications in Oceanography*; Cambridge University Press: Cambridge, UK, 2011; p. 845.
24. Mackey, M.; Mackey, D.; Higgins, H.W.; Wright, S.W. CHEMTAX—A program for estimating class abundances from chemical markers: Application to HPLC measurements of phytoplankton. *Mar. Ecol. Prog. Ser.* **1996**, *144*, 265–283. [CrossRef]
25. Buchaca, T.; Felip, M.; Catalan, J. A comparison of HPLC pigment analyses and biovolume estimates of phytoplankton groups in an oligotrophic lake. *J. Plankton Res.* **2005**, *27*, 91–101. [CrossRef]
26. Guisande, C.; Barreiro, A.; Acuna, A.; Marciales, L.J.; Hernandez, E.; Torres, A.M.; Aranguren, N.; López, W.; Duque, S.R.; Gallo, L.J.; et al. Testing of the CHEMTAX program in contrasting Neotropical lakes, lagoons, and swamps. *Limnol. Oceanogr.-Methods*

2008, *6*, 643–652. [CrossRef]
27. Lauridsen, T.L.; Schlüter, L.; Johansson, L.S. Determining algal assemblages in oligotrophic lakes and streams: Comparing information from newly developed pigment/chlorophyll a ratios with direct microscopy. *Freshw. Biol.* **2011**, *56*, 1638–1651. [CrossRef]
28. Goela, P.C.; Danchenko, S.; Icely, J.D.; Lubian, L.M.; Cristina, S.; Newton, A. Using CHEMTAX to evaluate seasonal and interannual dynamics of the phytoplankton community off the south-west coast of Portugal. Estuarine. *Coast. Shelf Sci.* **2014**, *151*, 112–123. [CrossRef]
29. Agirbas, E.; Feyzioglu, A.M.; Kopuz, U.; Llewellyn, C.A. Phytoplankton community composition in the south-eastern Black Sea determined with pigments measured by HPLC-CHEMTAX analyses and microscopy cell counts. *J. Mar. Biol. Assoc. UK* **2015**, *95*, 35–52. [CrossRef]
30. Lee, M.; Won, N.-I.; Baek, S.H. Comparison of HPLC Pigment Analysis and Microscopy in Phytoplankton Assessment in the Seomjin River Estuary, Korea. *Sustainability* **2020**, *12*, 1675. [CrossRef]
31. Saggiomo, M.; Bolinesi, F.; Brunet, C.; Passarelli, A.; Margiotta, F.; Saggiomo, V.; Mangoni, O. A CHEMTAX-derived phytoplankton community structure during 12-year observations in the Gulf of Naples (LTER-MC). *Mar. Ecol.* **2023**, *44*. [CrossRef]
32. Irigoien, X.; Meyer, B.; Harris, R.; Harbour, D. Using HPLC pigment analysis to investigate phytoplankton taxonomy: The importance of knowing your species. *Helgol. Mar. Res.* **2004**, *58*, 77–82. [CrossRef]
33. Tamm, M.; Freiberg, R.; Tõnno, I.; Nõges, P.; Nõges, T. Pigment-based chemotaxonomy—A quick alternative to determine algal assemblages in large shallow eutrophic lake? *PLoS ONE* **2015**, *10*, e0122526. [CrossRef] [PubMed]
34. Simmons, L.; Sandgren, C.; Berges, J. Problems and pitfalls in using HPLC pigment analysis to distinguish Lake Michigan phytoplankton taxa. *J. Great Lakes Res.* **2016**, *42*, 397–404. [CrossRef]
35. Hernández-Avilés, J.S.; Callieri, C.; Bertoni, R.; Morabito, G.; Leoni, B.; Lepori, F.; Buzzi, F.; Salmaso, N. Prokaryoplankton and phytoplankton community compositions in five large deep perialpine lakes. *Hydrobiologia* **2018**, *824*, 71–92. [CrossRef]
36. Morabito, G.; Austoni, M. *Caratteristiche Strutturali delle Associazioni Fitoplanctoniche nel Lago Maggiore ed Evoluzione Stagionale dei Popolamenti-C.N.R.-I.S.E.*; Ricerche Sull'evoluzione del Lago Maggiore. Aspetti Limnologici. Programma Triennale 2013–2015. Campagna 2015 e Rapporto triennale 2013–2015; Commissione Internazionale per la Protezione Delle Acque Italo-Svizzere, Ed.; Sede di Verbania: Verbania Pallanza, VB, Italy, 2016; pp. 41–45.
37. Rogora, M.; Austoni, M.; Caroni, R.; Giacomotti, P.; Kamburska, L.; Marchetto, A.; Mosello, R.; Orru', A.; Tartari, G.; Dresti, C. Temporal changes in nutrients in a deep oligomictic lake: The role of external loads versus internal processes. *J. Limnol.* **2021**, *80*. [CrossRef]
38. IOCCG. *Observation of Harmful Algal Blooms with Ocean Colour Radiometry*; Bernard, S., Kudela, R., Robertson Lain, L., Pitcher,
G.C., Eds.; IOCCG Report Series, No. 20; International Ocean Colour Coordinating Group: Dartmouth, NS, Canada, 2021.
39. Kutser, T.; Metsamaa, L.; Strömbeck, N.; Vahtmäe, E. Monitoring cyanobacterial blooms by satellite remote sensing. *Estuar. Coast. Shelf Sci.* **2006**, *67*, 303–312. [CrossRef]
40. Simis, S.G.H.; Peters, S.; Gons, H.J. Remote sensing of the cyanobacterial pigment phycocyanin in turbid inland water. *Limnol. Oceanogr.* **2005**, *50*, 237–245. [CrossRef]
41. Ambrosetti, W.; Barbanti, L. Deep water warming in lakes: An indicator of climatic change. *J. Limnol.* **1999**, *58*, 1–9. [CrossRef]
42. Ambrosetti, W.; Barbanti, L.; Rolla, A.; Castellano, L.; Sala, N. Hydraulic paths and estimation of the real residence time of the water in Lago Maggiore (N Italy): Application of massless markerstransported in 3D motion fields. *J. Limnol.* **2012**, *71*, 23–33. [CrossRef]
43. Fenocchi, A.; Rogora, M.; Sibilla, S.; Ciampittiello, M.; Dresti, C. Forecasting the evolution in the mixing regime of a deep subalpine lake under climate change scenarios through numerical modelling (Lake Maggiore, Northern Italy/Southern Switzerland). *Clim. Dyn.* **2018**, *51*, 3521–3536. [CrossRef]
44. Rogora, M.; Buzzi, F.; Dresti, C.; Leoni, B.; Lepori, F.; Mosello, R.; Patelli, M.; Salmaso, N. Climatic effects on vertical mixing and deep-water oxygen content in the subalpine lakes in Italy. *Hydrobiologia* **2018**, *824*, 33–50. [CrossRef]
45. Van Heukelem, L.; Thomas, C.S. Computer-assisted high-performance liquid chromatography method development with applications to the isolation and analysis of phytoplankton pigments. *J. Chromatogr. A* **2001**, *910*, 31–49. [CrossRef]
46. Canuti, E. Phytoplankton pigment in situ measurements uncertainty evaluation: An HPLC interlaboratory comparison with a European-scale dataset. *Front. Mar. Sci.* **2023**, *10*, 1197311. [CrossRef]
47. Hooker, S.B.; Van Heukelem, L.; Thomas, C.S.; Claustre, H.; Ras, J.; Barlow, R.; Sessions, H.; Schluter, L.; Perl, J.; Trees, C.; et al.
The Second SeaWIFS HPLC Analysis Round-Robin Experiment/SeaHARRE-2; NASA/TM-2005-212785. 2005. Available online: https://oceancolor.gsfc.nasa.gov/fsg/hplc/SH2_TM2005_212785.pdf (accessed on 25 October 2024).

48. Brewin, R.J.W.; Sathyendranath, S.; Hirata, T.; Lavender, S.J.; Barciela, R.M.; Hardman-Mountford, N.J. A three-component model of phytoplankton size class for the Atlantic Ocean. *Ecol. Model.* **2010**, *221*, 1472–1483. [CrossRef]
49. Punginelli, C.; Wilson, A.; Routaboul, J.M.; Kirilovsky, D. Influence of zeaxanthin and echinenone binding on the activity of the Orange Carotenoid Protein. *Biochim. Biophys. Acta (BBA)-Bioenerg.* **2009**, *1787*, 280–288. [CrossRef]
50. Takaichi, S.; Mochimaru, M. Carotenoids and carotenogenesis in cyanobacteria: Unique ketocarotenoids and carotenoid glycosides. *Cell. Mol. Life Sci.* **2007**, *64*, 2607–2619. [CrossRef] [PubMed]
51. Callieri, C.; Bertoni, R.; Contesini, M.; Bertoni, F. Lake level fluctuations boost toxic cyanobacterial "oligotrophic blooms". *PLoS ONE* **2014**, *9*, e109526. [CrossRef] [PubMed]
52. Aiken, J.; Pradhan, Y.; Barlow, R.; Lavender, S.; Poulton, A.; Holligan, P.; Hardman-Mountford, N. Phytoplankton pigments and functional types in the Atlantic Ocean: A decadal assessment, 1995–2005. *Deep. Sea Res. Part II Top. Stud. Oceanogr.* **2009**, *56*, 899–917. [CrossRef]
53. Trees, C.C.; Clark, D.K.; Bidigare, R.R.; Ondrusek, M.E.; Mueller, J.L. Accessory pigments versus chlorophyll a concentrations within the euphotic zone: A ubiquitous relationship. *Limnol. Oceanogr.* **2000**, *45*, 1130–1143. [CrossRef]
54. Latasa, M.; Bidigare, R.R. A comparison of phytoplankton populations of the Arabian Sea during the Spring Intermonsoon and Southwest Monsoon of 1995 as described by HPLC-analyzed pigments. *Deep. Sea Res. Part II Top. Stud. Oceanogr.* **1998**, *45*, 2133–2170. [CrossRef]
55. Catlett, D.; Siegel, D.A. Phytoplankton pigment communities can be modeled using unique relationships with spectral absorption signatures in a dynamic coastal environment. *J. Geophys. Res. Ocean.* **2018**, *123*, 246–264. [CrossRef]
56. Anderson, C.R.; Siegel, D.A.; Brzeniski, M.A.; Guillocheau, N. Controls on temporal patterns in phytoplankton community structure in the Santa Barbara Channel, California. *J. Geophys. Res.* **2008**, *113*, C04038. [CrossRef]
57. Kramer, S.J.; Siegel, D.A. How can phytoplankton pigments be best used to characterize surface ocean phytoplankton groups for ocean color remote sensing algorithms? *J. Geophys. Res. Ocean.* **2019**, *124*, 7557–7574. [CrossRef]
58. Kramer, S.J.; Siegel, D.A.; Graff, J.R. Phytoplankton Community Composition Determined From Co-variability Among Phytoplankton Pigments From the NAAMES Field Campaign. *Front. Mar. Sci.* **2020**, *7*. [CrossRef]
59. Canuti, E.; Penna, A. Dynamics of Phytoplankton Communities in the Baltic Sea: Insights from a Multi-dimensional Analysis of Pigment and Spectral Data: Part I, Pigment Dataset. *Front. Mar. Sci. Sec. Ocean. Obs.* **2024**, *11*, 1425347. [CrossRef]
60. Zhang, B.; Horvath, S. A general framework for weighted gene co-expression network analysis. *Stat. Appl. Genet. Mol. Biol.* **2005**, *4*, 17. [CrossRef] [PubMed]
61. Blondel, V.; Guillaume, J.-L.; Lambiotte, R.; Lefebvre, E. Fast unfolding of communities in large networks. *J. Stat. Mech. Theory Exp.* 2008. [CrossRef]
62. Utermöhl, H. Zur vervollkommnung der quantitativen phytoplankton-methodik. *Mitt. Int. Verein. Limnol.* **1958**, *9*, 1–38. [CrossRef]
63. Lund, J.W.G.; Kipling, C.; Le Cren, E.D. The inverted microscope method of estimating algal number and the statistical basis of estimations by counting. *Hydrobiologia* **1958**, *11*, 143–170. [CrossRef]
64. *CEN/TC 15204:2006*; Water Quality-Guidance Standard on the Enumeration of Phytoplankton Using Inverted Microscopy (Utermöhl Technique). CEN: Brussels, Belgium, 2006; p. 46.
65. Hillebrand, H.; Dürselen, C.C.-D.; Kirschtel, D.; Pollingher, U.; Zohary, T. Biovolume calculation for pelagic and benthic microalgae. *J. Phycol.* **1999**, *35*, 403–424. [CrossRef]
66. Sun, J.; Liu, D. Geometric models for calculating cell biovolume and surface area for phytoplankton. *J. Plankton Res.* **2003**, *25*, 1331–1346. [CrossRef]
67. *CEN/TC 230. EN 16695:2015*; Water Quality-Guidance on the Estimation of Phytoplankton Biovolume. CEN: Brussels, Belgium, 2015.
68. Salmaso, N. Factors affecting the seasonality and distribution of cyanobacteria and chlorophytes: A case study from the large lakes south of the Alps, with special reference to Lake Garda. *Hydrobiologia* **2000**, *438*, 43–63. [CrossRef]
69. Austoni, M.; Giacomotti, P.; Kamburska, L.; Lami, A.; Manca, D.; Marchetto, A. *Struttura delle Associazioni Fitoplanctoniche nel Lago Maggiore e loro Modificazioni in Relazione a Fattori di Controllo Trofici e Climatici in CNR IRSA*; Sede di Verbania. 2023. Ricerche Sull'evoluzione del Lago Maggiore. Aspetti Limnologici. Programma Triennale 2022–2024. Campagna 2022; Commissione Internazionale per la Protezione Delle Acque Italo-Svizzere, Ed.; Sede di Verbania: Verbania Pallanza, VB, Italy, 2022; pp. 68–75.
70. Giardino, C.; Bresciani, M.; Stroppiana, D.; Oggioni, A.; Morabito, G. Optical remote sensing of lakes: An overview on Lake Maggiore. *J. Limnol.* **2014**, *73*, 210–214. [CrossRef]

ANNEX V

This annex was published in Frontiers in Marine Science

Dynamics of phytoplankton communities in the Baltic Sea: Insights from a multi-dimensional analysis of pigment and spectral data, Part I: Pigment dataset

Elisabetta Canuti, Antonella Penna

Front. Mar. Sci., 11, 1425347. https://doi.org/10.3389/fmars.2024.1425347

Dynamics of Phytoplankton Communities in the Baltic Sea: Insights from a Multi-dimensional Analysis of Pigment and Spectral Data: Part I, Pigment Dataset

Elisabetta Canuti[1, 2], Antonella Penna [3,4,5]

[1]European Commission, Joint Research Centre, Ispra, Italy

[2]Department of Pure and Applied Sciences, University of Urbino "Carlo Bo", Urbino, Italy

[3]Department of Biomolecular Sciences, University of Urbino, 61029, Urbino, Italy

[4]CoNISMa, National Inter-University Consortium for Marine Sciences, 00196, Rome, Italy

[5]Fano Marine Center, The Inter-Institute Center for Research on Marine Biodiversity, Resources and Biotechnologies, 61032, Fano, Italy

Corresponding author: elisabetta.canuti@ec.europa.eu

Keywords: Phytoplankton Pigments, HPLC, Phytoplankton Community, Baltic Sea, Clustering

Abstract:

The study investigated the distribution of surface phytoplankton communities in the Baltic Sea using datasets from different seasons and areas. Data collected during six oceanographic campaigns conducted between 2005 and 2008, included high-performance liquid chromatography (HPLC) pigment characterization, and hydrological parameters. The first part of this comprehensive study is focused on the HPLC phytoplankton pigments dataset in relation to hydrological conditions. The research highlighted the importance of high quality input data for accurate taxonomic analysis. The results evidenced a seasonal pattern of phytoplankton succession in the Baltic Sea, with diatom blooms in spring, cyanobacterial blooms in mid-summer, and haptophyte and dinoflagellate peaks in late summer and autumn. Several unsupervised machine learning approaches, including Hierarchical Cluster Analysis (HCA), Principal Component Analysis (PCA) and Network-Based Community Detection Analysis (NCA), were used to analyze the data and identify phytoplankton communities. Network analysis revealed significant connectivity among communities, which helped to understand the phytoplankton spatio-temporal distribution. Five distinct phytoplankton communities were identified based on biomarker pigments and PCA analysis, including diatoms, dinoflagellates, cryptophytes, green algae, and cyanobacteria. Where the results from PCA and NCA were in good agreement, the outcome from the HCA analysis was less helpful in elucidating

the phytoplankton structure. The results of the statistical analysis were then compared with traditional approaches such as CHEMTAX and region-specific bio-optical algorithms, providing new perspectives on the taxonomic composition of phytoplankton groups, functional types (PFTs) and size classes (PSCs). Overall, the study provided valuable insights into phytoplankton dynamics and the effectiveness of different analytical approaches in understanding community structure, providing metrics that can enhance current and future advancements in remote sensing, including support for hyperspectral ocean color remote sensors such as NASA's Plankton, Aerosol, Cloud, and Ocean Ecosystem (PACE) mission.

1. INTRODUCTION

Phytoplankton play a crucial role in the marine ecosystem, acting as primary producers that form the base of the aquatic food web. The phytoplankton community composition influences the marine food chain and the regional carbon cycle. Monitoring phytoplankton distribution is important not only for aquatic sciences but also due to growing societal needs. Human activities, such as overfishing, pollution, and nutrient loading, severely threaten aquatic ecosystems (Lotze et al., 2006). Environmental factors such as temperature, light availability, nutrient concentrations, and water currents significantly influence phytoplankton dynamics. The unique brackish water conditions of the Baltic Sea, with salinity levels ranging from almost fresh in the Bothnian Bay (Schmelzer et al., 2008) to more saline in the southern regions (Ax and Sahlsten, 2001), create a diverse environment that supports a wide variety of phytoplankton species. Seasonal variations lead to distinct phytoplankton blooms, particularly in spring and summer, which are crucial periods for primary productivity. Increased eutrophication has led to more frequent harmful algal blooms in coastal areas, raising public concern as these blooms affect recreation, fisheries, and drinking water supplies.

Recent technological advancements, including satellite remote sensing and high-performance liquid chromatography (HPLC), have enhanced the monitoring and understanding of phytoplankton communities. This approach includes abundance-based, spectral-based, and ecological-based methods (Nair et al., 2008; Brewin et al., 2010; IOCCG, 2014). In recent years, there has been encouraging progress in investigating the composition of biomarker pigments and the correlation between pigments and cell size. This research has been carried out using extensive datasets of phytoplankton pigments, which have been characterized by high-performance liquid chromatography (HPLC) (Vidussi et al., 2001; Uitz et al., 2006; Brewin et al., 2010; Hirata et al., 2008; Mouw et al., 2017; Sun et al., 2022). HPLC is a powerful analytical technique used to separate, identify, and quantify analytical component in a mixture. In the context of phytoplankton research, HPLC allows for the precise characterization of pigment composition

in discrete natural water samples, distinguishing between a wide range of pigment molecules. The relative abundance of various phytoplankton groups can be inferred by analyzing the specific pigment composition (as detailed in Supplementary Table S1), providing insights into the presence and abundance of different phytoplankton groups and to ecological dynamics of phytoplankton communities. These approaches facilitate the categorization of phytoplankton into distinct size groups and are often based on the widely used CHEMTAX tool. CHEMTAX employs matrix factorization and known pigment ratios to determine the composition of taxonomic groups and has been already used in the Baltic Sea context (Schlüter et al., 2000, 2004, 2016).

Numerous studies have investigated the taxonomic composition of phytoplankton communities in the Baltic Sea (Wasmund et al., 1998, 2003, 2008, 2011; Olli et al., 2011; HELLCOM, 2018). They have traditionally relied on methods such as light microscopy and flow cytometry. However, it has become clear that relying solely on discrete, in situ observations to monitor phytoplankton communities is not sufficient. This is mainly due to the lack of comprehensive and extensive spatio-temporal data sets. Tools based on biomarker pigments can help tracking phytoplankton distribution and composition, providing essential data for managing the health of the Baltic Sea ecosystem and addressing issues such as harmful algal blooms. While several studies have investigated the chemotaxonomic characterization of Baltic phytoplankton communities based on pigment relationships (Schlüter et al., 2000; Stoń-Egiert et al., 2022) and the development of bio-optical algorithms (Meler et al., 2018), most of these investigations have been tailored to the southern Baltic Sea. There remains a noticeable gap in research on the wider Baltic Sea basin.

The present study aimed to evaluate data analysis methodologies that can be applied to investigate the relationship between phytoplankton community structure and pigment composition in the whole Baltic Sea. The study was based on a complete dataset that provided an adequate spatial-temporal coverage of this basin and included measurements of HPLC pigments and physical variables. The identification of phytoplankton communities through various statistical data analysis methods was applied to the HPLC pigment dataset, representative of different sub-regions of the Baltic Sea, and collected over five years, in conjunction with optical properties measurements. The dataset was subjected to three types of statistical analysis: Hierarchical Cluster Analysis; Principal Component Analysis (PCA) and Network-Based Community Detection Analysis (NCA). The insights gained from these analyses were then compared with those derived from traditional approaches, such as CHEMTAX (Mackey et al., 1996) and region-specific bio-optical algorithms for estimating phytoplankton functional types (PFTs) and sizes (PSCs) composition (Brewin et al., 2010; Hirata et al., 2011). Stoń-Egiert

2. MATERIALS AND METHODS

2.1. Field dataset

The data sets were collected between 2004 and 2008, during 6 bio-optical oceanographic campaigns, 3 covering the Southern Baltic Sea, Gdansk Bay, and the Pomeranian Bay (May and September 2004, April 2005), and 3 in the Northern Baltic Proper, Gulf of Finland and the Bothnian Sea (July 2006, August 2007, August 2008) (Fig. 1). The whole Baltic Sea dataset (BA) is composed of 273 stations.

Sea-surface Temperature (SST) and Salinity were measured by a SBE 911 CTD (Conductivity Depth Temperature) system (SeaBird, Alifax, USA). The water was sampled by a Niskin bottle at surface depth (1 meter below the sea surface) and pre-filtered through a 150 µm mesh (Kartel, Germany). The filters (GF/F filters, Φ 25 mm, 0.7 µm pore size, Whatmann, Germany) for HPLC measurements, were preconditioned under constant mild vacuum (not exceeding 0.5 bar), flash freeze in liquid nitrogen and successively stored a -80 °C

2.2. Phytoplankton pigment dataset

HPLC samples were analyzed at the Joint Research Centre of the European Commission (JRC).

The JRC method is described in detail in Canuti (2023). The HPLC was calibrated with pigment standards (DHI Lab Products, Denmark). The calibration curves and consequently the compound quantification cover a range of concentrations from a dilution close to three times the Signal to Noise Ratio (SNR) concentration to the standard concentration, as described by Hooker et al. (2005). The compounds below the Limit of Detection (LOD) are considered unidentified. The JRC follows strict quality control measures for the analysis of phytoplankton pigments using HPLC and regularly participates in inter-laboratory exercises and inter-comparison activities to assess the uncertainties associated with marine pigments (Hooker et al., 2010; Canuti et al., 2016, 2022, 2023).

Twenty-two pigments were quantified for all stations (Supplementary Table S1): 19'-hexanoyloxyfucoxanthin (Hex), 19'-butanoyloxyfucoxanthin (But), alloxanthin (Allo), fucoxanthin (Fuco), peridinin (Perid), diatoxanthin (Diato), diadinoxanthin (Diadino), zeaxanthin (Zea), divinyl chlorophyll *a* (DVChl *a*), monovinyl chlorophyll *a* (MVChl *a*), monovinyl chlorophyll *b* (MVChl *b*), divinyl chlorophyll *d* (DVChl *b*) chlorophyll c1 + c2 (TChl c1c2), chlorophyll c3 (TChl c3), neoxanthin (Neo), violaxanthin (Viola), prasinoxanthin (Pras), lutein (Lut), carotene (Caro), pheophorbide *a* (Pheo), pheophythin *a* (Phy) and chlorophyllide *a* (Chlide *a*).

The pigments considered for the clustering statistical analysis were 16 of those determined by HPLC and the total Chlorophyll *a* (TChl *a*) that is the sum of MVChl *a*, DVChl *a* and Chlide *a.* Of the other pigments, the Pheo, Phy, Chlide *a,* and DVChl *a,* were excluded because detected in concentration lower than the LOD for more than 95% of the stations. MVChl *a* was not included among pigments object of statistical analysis, because is considered a redundant accessory pigment (i.e., is a component of the TChl *a*). Similarly, for total Chlorophyll b (TChl *b*), the choice of considering the sum of MVChl *b* and DVChl *b* pigments (i.e., the TChl *b*) instead of their separate contribution was due to the lack of chromatographic separation: it was not possible to discriminate the MV and DV components for all the sampling stations.

The pigment compositions and their relative proportions in phytoplankton cells are distinctive features of different classes of algae and cyanobacteria , so that pigments may serve as unique taxonomic identifiers for phytoplankton (Wright et al. 1991). Certain carotenoids, which are quantitatively dominant, are considered taxonomic markers of phytoplankton. Fuco is a marker for diatoms, Zea for blue-green algae (cyanobacteria), Allo for cryptophytes, Hex for prymnesiophytes, Pras for prasinophytes, Per for dinophytes and TChl *b*, Neo, and Lut for green algae (Chlorophyceae). Determining the phytoplankton community structure based on pigment compositions and concentrations has become a standard practice (Mackey et al. 1996, Rodriguez et al., 2002). However, some pigments (i.e., Fuco) are common to multiple phytoplankton groups (Supplementary Table S1) and experiments have shown that the qualitative and quantitative proportions of pigments can vary even within cells of organisms from the same class, so that it is acknowledged that using a pigment as a marker for a specific group remains a simplification.

2.1. Methods in Data Analysis

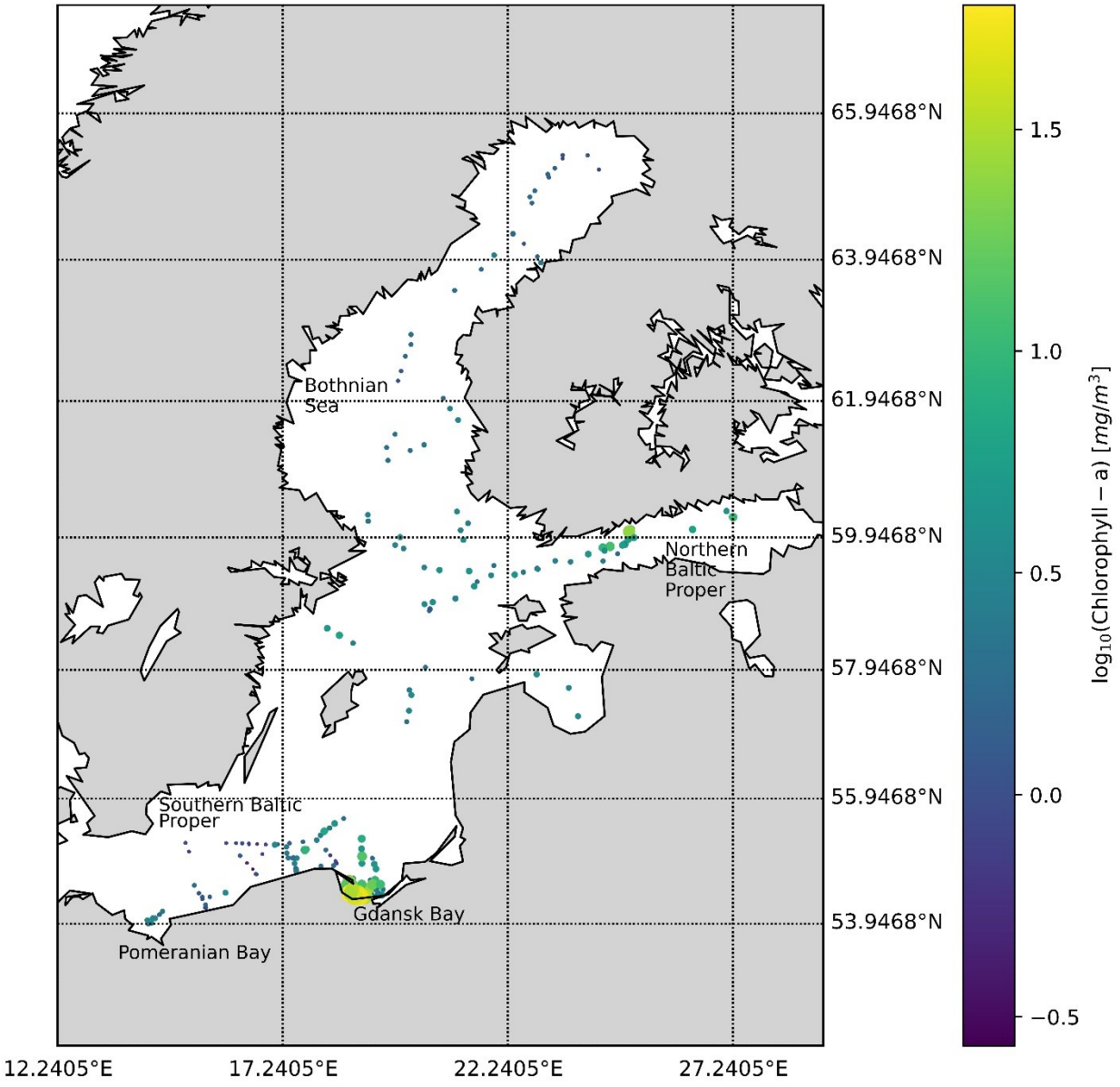
2.1.1. Hierarchical Cluster Analysis (HCA)

A hierarchical cluster analysis was performed on the HPLC pigment dataset, using all sixteen pigments described above after normalization to TChl *a* (i.e., using ratios Fuco:TChl*a*, etc.). This method uses Ward's linkage method (the inner squared distance), based on the correlation distance (1-R, where R is Pearson's correlation coefficient between phytoplankton pigment ratios), as in Latasa and Bidigare (1998) and Catlett and Siegel (2018). A linkage cutoff distance of 1 is used to divide the resulting dendrogram into distinct phytoplankton community clusters. The correlation distances between samples were then used to assign each sample to one of the resulting clusters.

2.1.1. Principal Component Analysis (PCA)

EOF (empirical orthogonal function) analysis serves as a valuable tool for exploring potential spatial patterns of variability and their temporal evolution (Anderson et al., 2008; Barron et al., 2011; Bracher et

Figure 1. Spatial distribution of measuring stations in 2004-2008 shows the location and the associated TChl *a* [mg m^{-3}] concentration.



al., 2015; and Kramer et al., 2020). In the field of statistics, this analysis is recognized as Principal Component Analysis (PCA). Essentially, PCA analysis decomposes a dataset into mathematically orthogonal (independent) modes, which can be interpreted as distinctive patterns or structures within the data. We defined X as the pigment matrix, where each row (M) corresponds to a sampling station and each column (N) corresponds to the 16 variables (i.e., the 16 pigments normalized by TChl a). The standardized matrix X underwent Singular Value Decomposition (SVD) to derive the PCA modes:

$X = U\Sigma V^{T}$, where $x_{ij}=\sum_{k=1,N} u_{ik}\sigma_k v_{kj}$ [1]

In this equation, V is an N × N matrix containing the pigment concentration data, U is an M × N matrix containing the principal components, Σ is an N × N matrix containing the singular values along the diagonal, and k represents the index of the PCA mode (with a length of N)

The columns of X are the principal components. We refer to the columns of U as the loadings, representing the directions (or weights) in the original variable space that defines each principal component. Eigenvalues and eigenvectors are used to compute the principal components. The eigenvalues represent the amount of variance explained by each principal component, and the corresponding eigenvectors (columns of U) represent the direction of maximum variance in the original variable space. Typically, the majority of variance, or "power," is captured by the first few modes. We focus on visualizing and summarizing the key features of these original 16 variables through the presentation of the top principal components (PCs). It is worth noting that PCA analysis does not impose any preconceived assumptions about the underlying covariance of pigments.

2.1.2. **Network-Based Community Detection Analysis (NCA)**

The initial step in constructing a network within the pigment's dataset involves defining a similarity matrix between the pigment ratios, which are treated as the network's features in this context. We adopted the absolute value of the Pearson's correlation coefficient, a common method for assessing co-expression (Allocco et al., 2004), as our evaluation metric. This approach results in the creation of a weighted, nearly fully connected graph with minimal zero values in the similarity matrix, thus limiting the identification of specific interaction clusters.

To establish a graph based on pairwise similarities, a straightforward strategy was used: we connected all pairs of nodes with non-zero similarity values and assigned edge weights corresponding to these similarity values. In our specific case, we first transformed the HPLC pigment dataset into a symmetrical adjacency matrix. Each node, or vertex, represents a sampling station, while edges connecting two nodes indicate the correlation across pigments between these stations—essentially, the relationship between any two of the 273 sampling sites. The edge weights provide insight into the strength of these connections, with Pearson's correlation coefficients serving as the means to describe the associations between nodes, primarily based on the normalized pigment ratios to TChl *a*. Based on this assumption, we can describe the similarity matrix as:

$$s_{ij} = \left|corr\left(x_i, x_j\right)\right| \qquad [2]$$

where s_{ij} is the similarity matrix, corr(x_i,x_j) is the Pearson correlation coefficient between nodes (sampling sites), and x_i and x_j are concentrations of the pigments at the different sampling sites.

However, Mason et al. (2009) has evidenced that utilizing the absolute value of the correlation may obscure biologically significant insights in unsigned network.To quantify the strength of connections between features (i.e., pigment ratios), we employed an adjacency matrix, denoted as A = [a_{ij}]. This matrix A was constructed by applying a threshold to the similarity matrix S = [s_{ij}].

In reference to the weighted gene co-expression network analysis (WGCNA) algorithm for network construction (Zhang and Horvath, 2005), which was designed for datasets of similar dimensions, we employed the following power function to evaluate the strength of connections between nodes:

$a_{ij} = (s_{ij})^{\beta}$ [3]

where the power β is the soft thresholding, with the default value β = 6 for the unsigned network.

Subsequently, the analysis of the network's community detection was performed on the adjacency matrix denoted as a_{ij}. This analysis was conducted using the undirected modularity method from NetworkX (Hagberg et al., 2008) in Python. To partition the community effectively, we employed Louvain's algorithm (Blondel et al., 2008). This algorithm identifies the optimal number and type of communities that maximize the network's modularity score.

Modularity, in this context, is a metric that falls within the range of -0.5 to 1, indicating the density of edges within communities compared to edges that extend outside of communities. It quantifies the level of connectedness within the network's communities. A modularity score of 0.3 or higher is considered substantial, signifying strong interconnections among sites within each community and relatively weaker connections between different groups.

The modularity output provides a community assignment for every sampling site within the matrix, reflecting the interrelatedness of the considered pigment ratios. To assess the taxonomic relevance of each community, we relied on the mean ratios of biomarker pigments within these communities: for the samples in the community, we considered the pigment with the highest ratio to TChl *a*, then we assigned the community to the taxa where this pigment is predominant (i.e. where Fuco: TChl*a* is the highest ratio, the sample is assigned to Diatoms).

3. RESULTS

The HPLC pigment dataset for surface samples for cruises BA01 to BA06 covered a wide range of environmental and ecological conditions (Table 1). The lowest average surface temperature and the highest average surface concentration of TChl *a* (11.5 mg m^{-3}) were found in April (BA03). Additionally, a contemporary highmean ratios of both Fuco:TChl*a* and Peri:TChl*a* were observed in these surface samples, indicating a relatively higher presence of diatoms and dinoflagellates compared to other research cruises. July and August (BA04-05 in 2006 and 2007) had the warmest mean surface waters of the cruises, together with the second highest mean TChl *a* concentration. During this period, mean Zea:TChl*a* ratio were at their highest, suggesting a higher proportion of pico-phytoplankton, which includes cyanobacteria. September (BA02) and late August (BA06) had intermediate mean surface water temperatures. Notably, BA02 exhibited the highest mean TChl*b*:TChl*a* ratio, a biomarker pigment indicative of all green algae. Conversely, BA01, BA03 and BA04 had the lowest mean TChl*b*:TChl*a*ratios, suggesting a reduced presence of green algae during these cruises.

Table 1. Summary of the relevant variables and diagnostic pigments: average TChl *a* ratios for surface samples of BA01-BA06 campaigns. Southern Baltic Sea (*SB*), Gdansk Bay (*GB*), Pomeranian Bay (*PM*), Northern Baltic Proper (*NP*), Gulf of Finland (*GF*), Bothnian Sea (*BS*)

	Campaign			**BA01**	**BA02**	**BA03**	**BA04**	**BA05**	**BA06**
	Sampling period			May 2004	Sept. 2004	April 2005	July 2006	Aug. 2007	Aug. 2008
	Sampling area			*SB, GB, PM*			*NP, GF*	*NP, GF*	*BS*
Parameters	**max**	**mean**	**min**						
N. Samples				50	51	62	21	37	45
Latitude [°N]	65.4	57.1	53.9	54.7	54.7	54.7	59.0	60.9	61.4
Longitude [°E]	27.2	19.5	14.2	18.5	17.7	17.6	22.4	20.8	22.1
Temperature [°C]	20.7	12.8	3.4	9.0	16.9	5.5	19.4	17.1	15.7
Salinity [PSU]	7.8	6.2	1.7	6.9	7.0	6.9	5.6	5.1	4.6
TChl a [mg m^{-3}]	60.222	5.107	0.272	2.846	2.894	11.514	4.025	3.129	2.929
Zeax:TChl a				0.071	0.141	0.014	0.164	0.139	0.089
TChlb:TChl a				0.070	0.216	0.043	0.069	0.084	0.104
Fuco:TChl a				0.141	0.091	0.135	0.125	0.143	0.134
Allox:TChl a				0.093	0.045	0.105	0.045	0.081	0.109
Peri:TChl a				0.011	0.015	0.180	0.041	0.010	0.016
Diad: TChl a				0.029	0.023	0.124	0.072	0.069	0.049
Hex: TChl a				0.004	0.002	0.000	0.024	0.028	0.013

Besides the quality control measures introduced in Section 2.2, the criteria established by Aiken et al. (2009) for evaluating the quality of datasets used in the development of bio-optical algorithms (Hirata et

al., 2011) were also applied to ensure internal consistency within each individual oceanographic campaign. The relationship between log-transformed TChl *a* and the sum of accessory pigments (TAcc, Trees, et al., 2000) was independently verified for each of the six oceanographic campaigns before proceeding with the data analysis and indicated determination coefficients above 0.97 (Supplementary Fig. S1). The multivariate statistical and network analysis on the HPLC datasets was mainly based on the correlation among the pigments of this dataset. The Pearson coefficient (R values) was selected as the correlation coefficient (Kramer et al., 2019) and the correlation matrix, associated with both absolute concentrations and ratios to TChl *a* was calculated for the 16 pigments chosen for further evaluation (Supplementary Fig. S2). The ratio correlations, which had a normalized coefficient, were chosen as they minimized the variance among the pigment concentration magnitude.

3.1. Hierarchical Cluster Analysis (HCA)

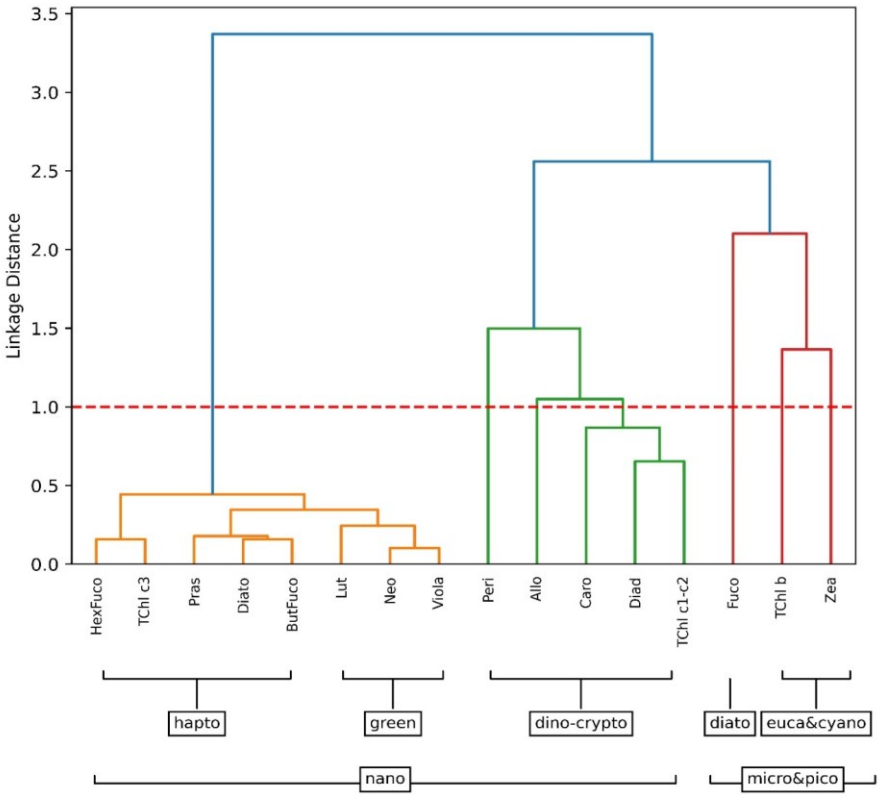
The pigments used for the hierarchical clustering were the 16 selected from the HPLC dataset (see section 2.2.). In the present study and in agreement with Catlett et al. (2018) the addition of Caro did not significantly change the cluster assignment in the HCA analysis.

Hierarchical cluster analysis of the BA dataset illustrated the presence of five distinct groups of phytoplankton pigments that significantly influence the co-variability of the dataset (Fig. 2). These dominant groups were diatoms, dinoflagellates, haptophytes, green algae, and cyanobacteria. Within the mixed nano-phytoplankton community, three sub-clusters were identified: haptophytes, a combination of cryptophytes – dinophyta, and green algae. The identification of the haptophyte community cluster was based on observations of But, Hex, and TChl c3. This pigment group was believed to represent a haptophyte community based on previous observations during early spring and late autumn blooms in the Baltic Sea (Blanz et al., 2005; Hällfors, 2004). The pigments Neo, Lut, and Viola were predominantly associated with green algae, with Viola specifically found in the euglenophytes. A third cluster, including Allo, representative of the red algae, and Peri, characteristic of dinophyta suggested the coexistence of cryptophytes – dinophyta in the nano-size fraction. It was noteworthy that there was a substantial linkage distance between these clusters and Zea, TChl *b*, and Fuco. Based on previous findings (Vidussi et al., 2001) the Fuco was assumed representative of all diatoms (microfraction), while Zea and TChl *b* were considered representative of the pico-eucaryotes fraction (cyanobacteria).

Previous studies (Stoń-Egiert et al., 2010, 2022) in the southern Baltic region have identified the dominance of various phytoplankton groups recognizable in our cluster analysis, including cyanobacteria, dinophyceae, cryptophyceae, chlorophyceae, and euglenophyceae.

A separate analysis was performed for each campaign (Supplementary Fig. S3). The prevalence of Fuco and Zea, which are indicative of euglenophyceae, was most pronounced during campaign B05 which extensively covered the Gulf of Finland. In the first three campaigns conducted in spring (BA01 and BA03) and early autumn (BA02) that were dedicated to the Southern Baltic Sea, Pomeranian Bay and Gdansk Bay, the Fuco was in a separate and dominant cluster, suggesting the dominance of diatoms in this area . The 2010 study by Stoń-Egiert, conducted on phytoplankton pigment data and microscopy samples collected during 12 campaigns in the Southern Baltic Sea and Gdansk Bay regions between 1999 and 2005, confirmed that diatoms are the primary components of phytocenoses in this area of the Baltic Sea. The proportion of Dinophyceae in the total biomass ranged from 7.5% in autumn within the gulfs to 59.2% in early summer in the open Baltic. The contribution of diatoms to the total phytoplankton biomass in early summer in open waters was of 23.5%.

Figure 2. Hierarchical clustering of phytoplankton pigment ratios to TChl *a* for the Baltic dataset. The major pigment communities (micro-, nano-, and pico-phytoplankton) are identified based on a linkage distance cutoff of 1.0 (red dashed line). The suggested phytoplankton cell size classes for each group are delineated with brackets.

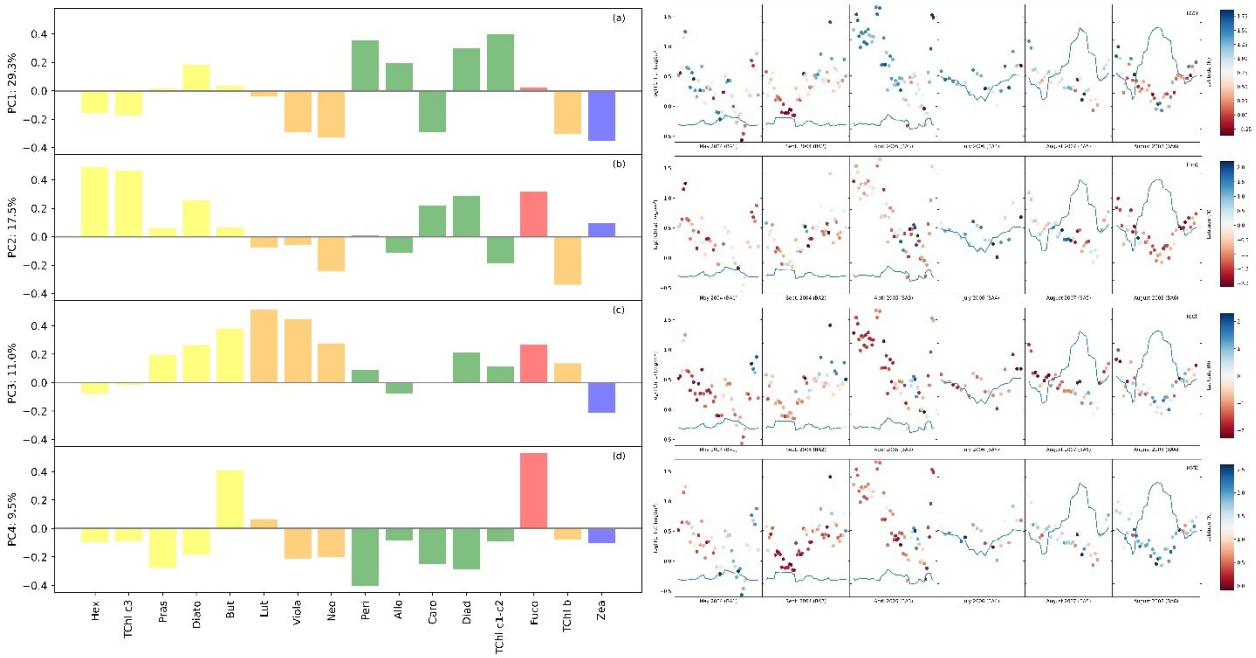


3.2. Principal Component Analysis (PCA)

The composition of the phytoplankton community was evaluated via a Principal Component Analysis (PCA) on the phytoplankton pigment concentrations, which were normalized by TChl *a* concentration. PCA, combined with Hierarchical Cluster Analysis (HCA), enabled us to establish connections among the identified groups and the spatiotemporal variations observed in the BA dataset. The eigenvector loadings for the most influential PCA modes were displayed in a bar plot (Fig. 3). We examined the Amplitude Function (AF) associated with every mode. The Amplitude Function, as previously utilized by Anderson et al. (2008), served as an indicator of the community structure pattern intensity both in spatial and temporal dimensions (Fig. 3). When the AF values were near zero, it indicated that the PCA mode was less important for the current time and location. The present study examined the leading quartile of modes for evaluating community composition in the BA dataset as they collectively explained 67.4% of the community structure's variability. Our observations of the first two modes' high variability were akin to results by Kramer et al. (2020) on a dataset with a comparable number of observations collected during 4 oceanographic campaigns in the North Atlantic Ocean. However, the distinctive features of the Baltic Sea, including a high presence of humic acids from rivers, were likely to result in greater variability within the dataset than would be observed in more homogeneous aquatic systems, such as the North Atlantic Ocean. In Mode 1 (Fig. 3a), which accounts for 29.3% of the variance, there was a strong positive correlation between TChlc1-c2 associated with dinoflagellates and cryptophytes. Conversely, pigments associated with green algae, cyanobacteria (Zea), and pico-phytoplankton exhibited a strong negative correlation with Mode 1. The spatial distribution showed negative patterns associated with cyanobacteria at mid-range latitudes and positive patterns associated with dinoflagellates at southern latitudes, thereby supporting this interpretation. Mode 2 (Fig. 3b) explained 17.5% of the dataset variance and was positively correlated with pigments related to diatoms (Fuco) and haptophytes and moderately correlated with cyanobacteria. Mode 3 (Fig. 3c), was negatively correlated with cyanobacteria, and strongly correlated with green algae, and nano- fraction pigments. Finally, PCA Mode 4 explained 9.5% of the total variance and was negatively correlated with nearly all pigments except Fuco and But, which served as markers for diatoms. Mode 4 (Fig. 3d) indicated a dominance of diatoms when the amplitude function was positive and was specifically found in the Gulf of Finland. This result suggests that Mode 4 has the potential to separate diatoms from other groups. Furthermore, this mode showed a negative correlation with all other groups. The PCA analysis suggested a similarity with one aspect of the HCA, namely the prevalence of Diatoms in campaigns BA01-03. However, identifying the dominant community was not consistently possible. In the HCA the TChl

b was associated mostly with the pico-fraction, while in the PCA analysis, the TChl *b* was more correlated with the green algae, in terms of variations.

Figure 3. The loadings corresponding to the principal component modes are shown in panels (a), (b), (c), and (d) for the Baltic dataset. The pigment order was as for the HCA analysis, to facilitate comparisons. The mode number is shown above each plot, together with the percentage of variance explained by that mode. The loadings are color-coded based on the main taxonomic groupings: blue for cyanobacteria-pico, red for diatoms-micro, orange for green algae-nano, green for haptophytes-dinoflagellates-nano, and orange for euglanophytes-nano. In panels (as), (bb), (cc), and (dd) corresponding to each mode, are presented the Amplitude Function (Positive in blu, negative in red) corresponding to the $\log_{TChl\ a}$ concentration (scale left side) by each campaign. The latitude is represented by the continuous line on each plot (scale right side).



2.1. Network-Based Community Detection Analysis (NCA)

The Louvain partition method (Dugue' et al., 2015; Traag et al., 2019) was utilized to identify the phytoplankton communities within the BA dataset network by optimizing its modularity. Modularity acts as a metric for differentiating intra-community connections. In the case of the pigment ratio network of the BA HPLC dataset, the modularity score reached 0.4. This emphasized the significant similarity observed

among samples grouped within the same community and denoted a robust differentiation between various community types using this approach.

The employed network-based method for detecting communities successfully revealed the existence of five prominent phytoplankton pigment communities. To determine the taxonomic classification of each major phytoplankton pigment community, we assessed the mean pigment-to-TChl *a* ratio of five key biomarker pigments for each community (Table 2). The most relevant community exhibited the highest average Zea ratios, indicating elevated concentrations of picoplankton and cyanobacteria, which were predominant across most of the stations. The second community registered the highest average Fuco ratio. In the third community, the highest ratio of TChl *b* was detected, typically associated with green algae. The fourth community known for the highest Peri ratio, was representative of dinoflagellates. Lastly, the fifth group presented the highest Allo ratio, which was considered as representative of the Cryptophyta (nano-fraction). The classification attained through community analysis was subsequently compared with the one derived from the implementation of the PFTs (Phytoplankton Functional Types) algorithm on the Baltic Sea dataset (Brewin et al., 2010; Hirata et al., 2011; Meler et al., 2022) that partially confirmed the results of the network-based analysis. The equations used for the PFTs calculations are summarized in Table 2. It has to be recalled that one of the assumptions of the PFT model is that the Diagnostic Pigments Sum (ΣDP) is considered as equal to the TChl a. It has to be noticed that, while in the PFT classification TChl b was assigned to two different groups (i.e. green algae and Pico-eukaryotes), in the network-community classification the partition with predominance of TChl b was assigned to green algae, consequently in network-community classification no Pico-eukaryotes fraction was present. In the PFT classification, Allo was considered only within the nanoplankton size class, while no functional group was associated specifically with cryptophytes. These are the main differences that could be found between the network-community partition and the PFT approach. Ultimately, it is noteworthy that, adhering to the PFT analysis, no station was identified as dominated by Haptophytes PFT. The spatiotemporal arrangement of the samples in the NCA (Fig. 4a) illustrated that dinoflagellate and diatom communities were predominantly present in the Baltic Sea with a significant concentration observed in the Southern Baltic Sea and Gdansk Bay. Notably, the prevalence of these two communities varied with the changing seasons. However, prokaryotic and pico-eukaryotic communities were distributed over a wider range of latitudes, including different temperature and salinity conditions. Comparing the outcomes of the NCA analysis with the communities identified through PFT analysis (Fig. 4b), a remarkable similarity in the spatiotemporal distribution of the samples within each community was observed for all the campaigns except for BA02 and BA05 campaigns. For the BA02 campaign, the PFT analysis assigned part of the stations to the pico-

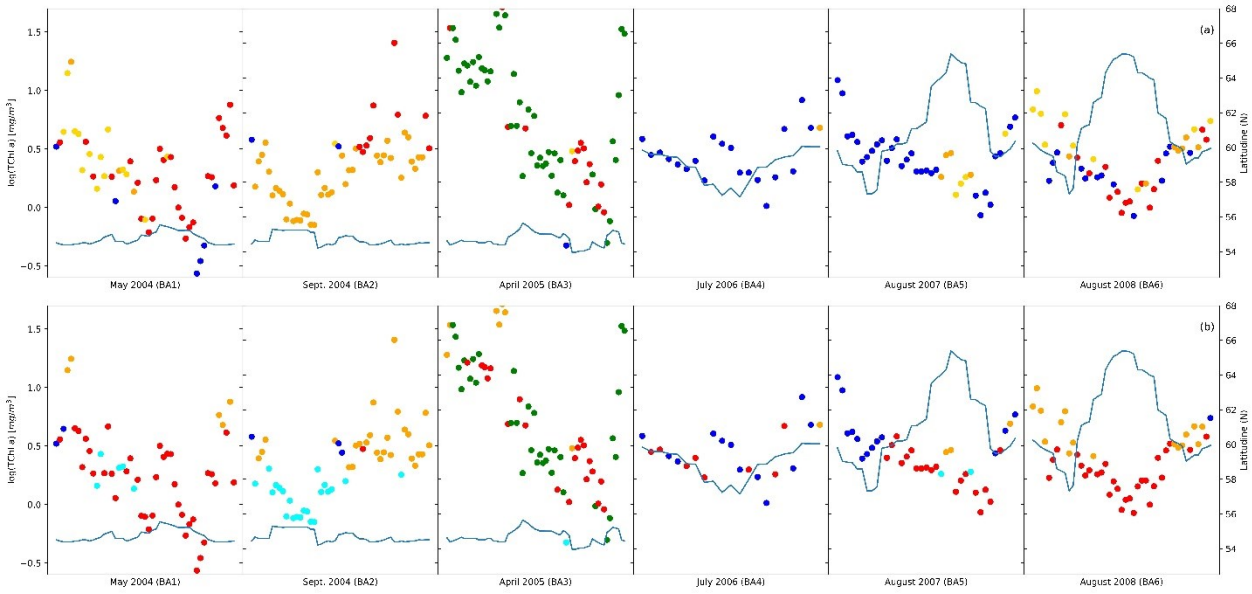
eucaryotes instead of to the cryptophytes, and in the BA05 campaign, the NCA analysis identified no diatoms among the sampling stations, while the PFT

Table 2. Phytoplankton Functional Types (PFTs), diagnostic pigments, and their taxonomic association.

PSCs/PFTs	Diagnostic Pigments	Estimation Formula
Micro (> 20 µm)	Fuco, Peri	1.41 (Fuco + Peri)/ ΣDP
Nano (2-20 µm)	Hex	(X*n* 1.27 Hex + 1.01 TChl *b* + 0.35 But + 0.60 Allo)/ ΣDP
Pico (0.2-2 µm)	Zea, Hex, TChl *b*	(0.86 Zea + Y*p* 1.27 Hex)/ ΣDP
Diatoms	Fuco	1.41 Fuco/ ΣDP
Dinoflagellates	Peri	1.41 Peri/ ΣDP
Haptophytes	But, Hex	(Nano-Green Algae)
Green algae	TChl *b*	1.01 TChl *b*/ ΣDP
Prokaryotes	Zea	0.86 Zea/ ΣDP
Pico-eukaryotes	Hex, TChl *b*	(Pico - Prokaryotes)
Prochlorococcus sp.	DVChl *a*	0.74 DVChl *a*/TChl *a*
	ΣDP	1.41 (Fuco + Peri) + 1.27 Hex + 1.01 TChl *b* + 0.35 But + 0.60 Allo + 0.86 Zea

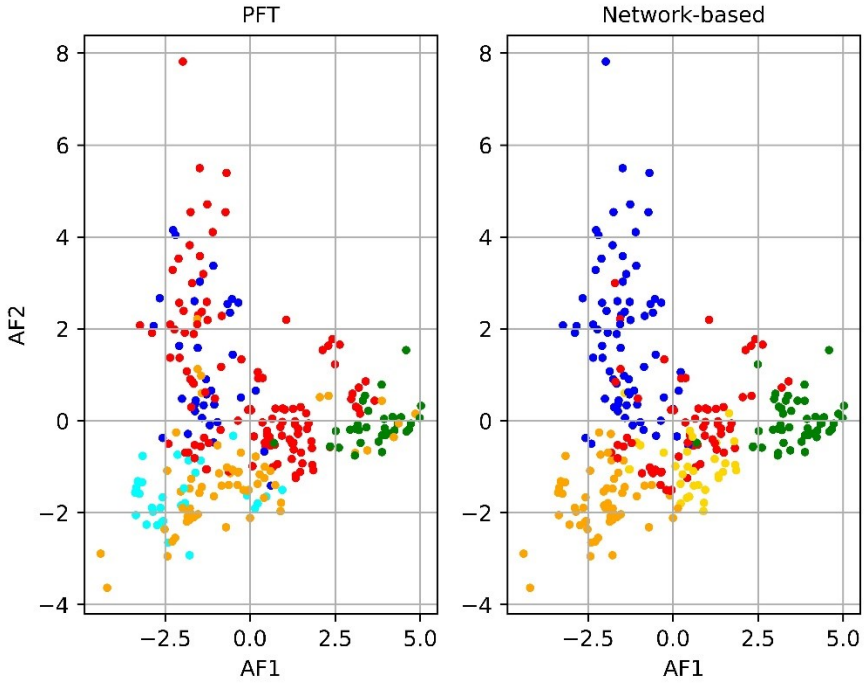
*Xn=0.5 and Yp=0.5 for TChl *a* 0.04 mg/m^3; Yp increasing at 1 at TChl a = 0.01 mg/m^3; Xn increasing at 1 at TChl a = 0.08 mg/m^3

Figure 4. The group distribution from NCA analysis (top) compared with PFTs distribution (bottom). Colours indicate major taxonomic groups: Diatoms (red), Dinoflagellates (green), Green algae (orange), Prokaryotes (cyan), Pico-eukaryotes (blue), Cryptophytes (gold); the continuous line indicates the latitude (right axis) and the position related to TChl a concentration (left axis).



analysis did. Moreover, the number of samples assigned to each community was quite consistent, except diatoms (Supplementary Table S1). To enhance the differentiation of communities identified by PCA analysis, the results of the PCA analysis were integrated with the communities recognized through the NCA analysis (Fig. 5). Comparing the amplitude function of the first two modes, we can separate the pico-eucaryotes community in the case of the NCA analysis, while the outcome was not so evident in case of the PFT application, where the pico-eucaryotes were mixed with diatoms. Similarly, the dinoflagellates community looks better isolated in the NCA analysis than in the PFT.

Figure 5. PCA analysis compared to network partitioning results (shown in color): Diatoms (red), Dinoflagellates (green), Green algae (orange), Procaryotes (cyan), Pico-eucaryotes (blu), Cryptophytes (gold).



In network analysis, particularly in ecological studies, understanding the underlying structure and relationships within a community is crucial. One method to achieve this is through the computation of a minimum spanning tree (MST). The outcomes of a spanning tree resulting from a matrix modularity analysis can provide insights into the community structure and relationships within the phytoplankton community as well, revealing hierarchical or structural aspects of the network (Supplementary Fig. S4). In terms of community detection, this could indicate that the central community played a more significant role or stronger connections with other communities compared to the peripheral community. In our case, the community (2) that dominated the center was the one with predominant Allo pigments. This was

interpreted as the dominance of cryptophytes. At the center, the network showed more nodes and denser connections, while the sparser community in the peripheral (community (0)) was dominated by TChl *b*. However, it should be noticed that the other three communities – procaryotes, dinoflagellates, and diatoms - were widespread among the leaf, in between the center and the peripheral, thus suggesting the cross-diffusion of these pigments among various phytoplankton species present in the Baltic Sea.

We compare the results of machine learning analysis with the outcome of the CHEMTAX (Mackey et al., 1996) community classification. As the initial input of the CHEMTAX matrix, we used the pigment ratios developed for the southern Baltic (Schlüter et al., 2000). As the southern Baltic populations could not be considered as representative of the whole basin, we established a different matrix for each campaign (Supplementary Figure S5) considering each campaign belonging to a separate cluster. However, the ratios used in the generated matrix applied to the Northern and Central Baltic campaigns, were not verified by microscopy counts, thus lacking robustness. Applying CHEMTAX to our dataset and accordingly to the matrix used for the community identification, we obtained information on seven groups (Dinophyceae, Diatoms, Cryptophytes, Cyaonophytes, Chlorophyceae, Euglenophytes, Prasinophytes) instead of the five identified through our statistical approaches. According to the CHEMTAX analysis, the cryptophytes' fraction dominated in three campaigns (BA01, BA03, BA05) and this partially confirmed the analysis through network community. However, the cyanobacteria were present in a smaller fraction of stations, while in all other analyses they were predominant in BA05 and BA06 campaigns. The diatoms, that are the most representative community according the PFT analysis, are present in percentages lower than 17%.

3. DISCUSSION

The present study examined the surface phytoplankton community distribution in the Baltic Sea, derived from HPLC datasets covering different seasons and different areas of the Baltic Sea. The dataset underwent different statistical analyses to assess the phytoplankton community composition based on the predominance of diagnostic pigments representative of a phytoplankton group or species."

In the Baltic Sea, the typical seasonal cycle of phytoplankton species succession follows a well-established pattern. When comparing our results with previous findings, the BA cruises reflected this seasonal progression of phytoplankton communities. A spring diatom bloom and a mid-summer cyanobacteria bloom were followed by a late summer to autumn peak of haptophytes and dinoflagellates. In the BA campaigns, spring and early summer cruises showed an abundance of samples from the diatom community, coinciding with the spring phytoplankton bloom. Stoń-Egiert Stoń-Egiert(2010) reported that

diatoms accounted for about 50% of the total phytoplankton biomass in the Gdansk Bay in spring and that the composition of the phytoplankton community varied with increasing distance from the river mouth, with a notable presence of dinophytes (ranging from 10% in the vicinity of the river mouth to 40% in more distant regions of the Gulf). The aforementioned studies have demonstrated the presence of diatoms (Fuco pigment) in the initial BA01-03 campaigns that focused on the southern Baltic Sea and Gdansk Bay. Both the analysis by single campaign and a combined analysis of these data sets have shown the presence of green algae during this period. In early summer, dinoflagellates became an important part of the community alongside diatoms. The transition from late summer to early autumn (BA05-06 and B04 respectively) was dominated by samples from the haptophyte community, with some cyanobacteria present as the bloom wanes. These results are in agreement with previous findings on cyanobacteria cycle in the Baltic Sea.

Conventional pigment-based methods such as CHEMTAX assume linear independence of pigments and require pre-defined knowledge of pigment contributions to individual phytoplankton groups. In the context of the Baltic Sea study, the dataset's collinearity and dynamic conditions challenge the linear independence of pigment assumptions, making the methods used here more suitable for capturing the phytoplankton community composition. In our initial data analysis, we applied different statistical tools and unsupervised machine learning techniques, including Hierarchical Cluster Analysis (HCA), Principal Component Analysis (PCA), and network-community (NCA) analysis on the HPLC dataset. The aim was to assess the consistency and coherence of the results derived from these analyses. To validate the robustness of our chemotaxonomic data analysis, we compared the outcomes with alternative models and algorithms commonly used in characterizing chemotaxonomic composition (PFT). Additionally, we cross-referenced our findings with results from prior studies conducted in the Baltic region.

TChl *b* is a pigment present in Euglenophyta, Chlorophyta, Prasinophytes, and, as DVChl *b*, can contribute to pico-eukaryotes (Supplementary Table S1). In the HCA, TChl *b* clustered with Zea but not with other pigments representative of green algae (i.e., Pras, Lut, Neo, and Viol). This suggested that TChl *b* primarily contributed to the pico-eukaryote fraction rather than to green algae. However, in the PCA analysis, TChl *b* followed the behavior of Neo and Viol in all four modes, and Lut in the first three modes, while Pras followed a distinctive path compared to TChl *b* in the first two modes. The PCA interpretation suggested that TChl *b* is associated as a biomarker pigment with green algae since it follows the path of other pigments commonly present in green algae. The distinctive behavior of Pras in the first two modes can be interpreted as a characteristic of Prasinophytes, distinguishing the Prasinophytes from the other green algae. In the NCA analysis, and similarly to the PCA outcome, the community linked to TChl *b* was

composed predominantly of green algae pigments, while Zea is associated with a different community. In light of these considerations, the HCA clustering appears to be the least explanatory regarding both TChl *b* and green algae. Ultimately, the cryptophytes were not clearly identified in either the HCA or PCA analysis, while in the NCA a community associated with Allo was identified.

In the NCA, the initial step was the analysis of the adjacency correlation matrix, which evidenced the presence of five areas of strong correlation. The Network was employed to identify communities through the Louvain partition method. This revealed a modularity value above 0.3 (0.39 in our case), indicating a significant level of community interconnectedness, as noted by Newman (2006), where significant interconnectedness within a community indicated that the population assigned to each community had many common traits. We also examined whether the communities identified through network analysis corresponded with those identified using other methods such as PCA and HCA, with the limitation mentioned above regarding the TChl *b* for HCA analyses. In the network analysis, we assigned each sample to a specific community in the network-based community detection analysis, thus allowing us to consider the spatio-temporal distribution of the five communities. The principal phytoplankton community identified by PCA were composed by diatoms, dinoflagellates, cyanobacteria, cryptophytes and green algae, which was consistent with the results of the network analysis (Supplementary Figure S6). Upon comparison of HCA with the network analysis, a notable distinction was raised: HCA failed to assign a specific cluster to TChl *b*, whereas the NCA recognized TChl *b* as representative of a distinct community. Furthermore, the presence of Zea was identified differently in each analysis, with network analysis highlighting as a key feature. It is worth noticing that, while PCA and HCA offer an overarching perspective of the dataset, the network analysis discreetly assigns a dominant phytoplankton community to each station. Therefore, it appeared more beneficial to compare the results of network analysis with PFT analysis, which similarly assigns a dominant community to each observation (Fig. 4 and 5). Comparing the two methodologies, the difference between NCA and PFT analysis was more evident for the BA02 and BA06 campaigns. If in BA02we compare the outcome of the PFT with the NCA analysis, the PFT assigned twentytwo stations to pico-eucaryotes whether the NCA assigned them to cryptophytes: considering the season and the temperature condition (Table 1 and Fig. 6) these twenty-two stations are more likely to associate these stations with cyanobacteria (i.e., PFT assignation). In the case of B02, the PFT assignation are more coherent with the environmental conditions. Conversely, in the BA06 the PFT assigns to diatoms community, stations that for the environmental conditions (temperature and season) are more likely to be representative of a Prokaryotes dominance (NCA assignation). In B06 the station assignation to phytoplankton group of NCA has to be preferred to the PFT. Overall, it has to be recognized that a limit of

this approach was that in the network analysis –and the PFT analysis - each sample was associated with a specific category (color) discretely, whereas reality is more complex than clear-cut categories. In addition, the PCA analysis proved to be a valuable tool in analyzing seasonal variations in phytoplankton composition. The two approaches, PCA and NCA, could be complementary to develop a holistic view of the dataset. We extended our comparison between PFT and network analysis to hydrological conditions that can influence the observed variability in phytoplankton community structure, such as temperature and salinity (Table 3). A parallel analysis using the PFT algorithm yielded similar conclusions (Fig. 6). In this context, the spatio-temporal distribution of phytoplankton communities inferred from HPLC pigments in the Baltic Sea through Network analysis aligned well with the phytoplankton community composition that was found in previous analysis (Stoń-Egiert et al., 2010, 2022), evidencing the presence of cyanobacteria at higher temperature and salinity (Pli nski et al., 2007). Our data-driven statistical analyses on the HPLC pigment dataset were able to identify five distinct taxonomically defined phytoplankton communities in the Baltic Sea, characterized by five biomarker pigments: diatoms (Fuco), dinoflagellates (Peri), cryptophytes (Allo), green algae (TChl *b*) and cyanobacteria-pico-plankton (Zea). Notably, dinoflagellates were found to be distinct from diatoms in this regional context, a differentiation that was not commonly observed globally (Kramer and Siegel, 2019). Samples from the six BA cruises had sufficient concentrations of dinoflagellate pigments to allow for their clear separation from diatoms and other red algal pigments in PCA analysis and network-based community detection. However, this distinction was less clear in a hierarchical cluster analysis.

Figure 6. Regression of physical parameter temperature and salinity for PFT analysis and NCA analysis, all colored with the dominant community (Diatoms (red), Dinoflagellates (green), Green algae (orange), Procaryotes (blue), Pico-eucaryotes (cyan), Cryptophytes (gold).

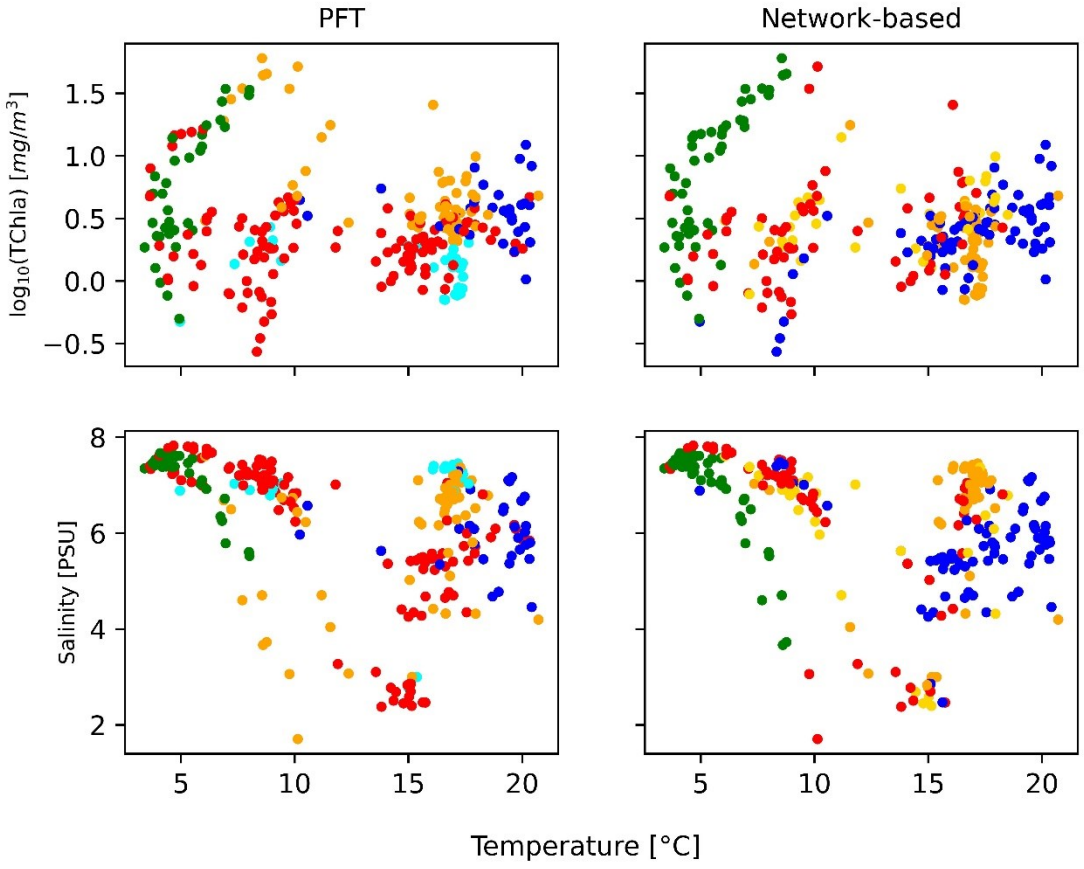


Table 3. Summary of the relevant variables and averaged diagnostic pigments: TChl *a* ratios resulting from Network-based analysis

Parameters	**Pico-eukaryotes**	**Diatoms**	**Green algae**	**Dinoflagellates**	**Cryptophytes**
Number of Stations	73	66	55	47	32
Latitude [°N]	59.4	56.6	56.2	54.8	57.7
Temperate [°C]	16.9	10.8	16.3	5.4	12.8
Salinity [PSU]	5.7	6.2	6.5	6.9	5.7
TSM [g L^{-1}]	1.138	1.685	0.964	2.628	1.397
TChl *a* [mg m^{-3}]	2.914	4.222	2.456	13.241	3.861
TChl *a* (min-max)	0.49-60.22	0.78-14.05	0.27-12.24	0.71-17.58	0.54-51.6
Peri:TChl*a*	0.009	0.008	0.01	0.055	0.005
Fuco:TChl*a*	0.088	0.112	0.038	0.026	0.032
Hex:TChl*a*	0.012	0.002	0.002	0	0.001
Allo:TChl*a*	0.037	0.044	0.033	0.033	0.057
Zea:TChl*a*	0.089	0.039	0.097	0.006	0.022
TChl*b*:TChl*a*	0.039	0.043	0.135	0.015	0.027

4. CONCLUSION

In our investigation, we used multivariate statistics, and unsupervised and supervised machine learning methods to analyze a diverse dataset of HPLC pigment observations obtained from Baltic Sea surface samples. We aimed to extract key insights from these pigment observations, considering various spatial and temporal dimensions as well as available hydrological variables. We addressed the selection process of appropriate statistical methods and underscored the importance of the data quality of the pigment dataset. We also performed a comparative analysis, comparing our results with those of alternative models and algorithms commonly employed to characterize chemotaxonomic composition, such as PFTs and CHEMTAX. Additionally, we compared our results with findings from previous studies conducted in the Baltic Sea region (Wasmund et al., 2011; HELCOM, 2018).

Our findings suggested that the network-based community identification alongside PCA analysis on the HPLC dataset holds promise for effective interpretation of phytoplankton community composition. This combined approach demonstrated the potential to identify phytoplankton communities, even within the complexities of basins such as the Baltic Sea. However, it has to be acknowledged that the data-driven statistical analyses employed in this study had some limitations. In particular, pigment-based methods are bound by the specific conditions under which the data were collected. In all these approaches, we referred to the diagnostic pigments or to the known pigment ratios to reconstruct the phytoplankton community. As it has been already remarked by Meler et al. (2020), all these approaches, based on diagnostic pigments, are simplified with statistical error in the order of 20%. A step forward in the research is integrating the analysis of the pigment dataset with concomitant measurements of optical features. Ultimately, these methods did not directly measure phytoplankton biomass or productivity, which limited both the derived phytoplankton communities and the potential development of satellite algorithms. Despite these limitations, the insights gained from these methods offered valuable metrics and datasets that can contribute to both current and future advances in remote sensing technologies.

REFERENCES

Aiken, J., Pradhan, Y., Barlow, R., Lavender, S., Poulton, A., Holligan, P., Hardman-Mountford, N., 2009. Phytoplankton pigments and functional types in the Atlantic Ocean: A decadal assessment, 1995-2005. *Deep Sea Res. Part II: Topical Studies in Oceanography*, 56(15), 899– 917

Allocco, D. J, Kohane, I. S., Butte, A. J., 2004. Quantifying the relationship between co-expression, co-regulation, and gene function. *BMC bioinformatics*, 5(1):18

Anderson, C. R., Siegel, D. A., Brzeniski, M. A., Guillocheau, N., 2008. Controls on temporal patterns in phytoplankton community structure in the Santa Barbara Channel, California. *J. Geophys. Res.*

Axe P., Sahlsten E., 2001. Hydrografi/hydrokemi. Bottniska Viken 2001. Available from http://www.havet.nu/dokument/Bv2001hydro.pdf

Barrón, R. K., Siegel, D. A., Guillocheau, N., 2014. Evaluating the importance of phytoplankton community structure to the optical properties of the Santa Barbara Channel, California, *Limn. Oceanogr.,* 59, doi: 10.4319/lo.2014.59.3.0927

Blanz, T., Emeis, K., Siegel, H., 2005. Controls on alkenone unsaturation ratios along the salinity gradient between the open ocean and the Baltic Sea. *Geochim. Cosmochim. Acta*. 69. 3589-3600. 10.1016/j.gca.2005.02.026

Bianchi, T., Kautsky, L. Argyrou, M., 1997. Dominant chlorophylls and carotenoids in macroalgae of the Baltic Sea (Baltic proper): Their use as potential biomarkers. *Sarsia.* 82. 55-62. 10.1080/00364827.1997.10413637

Blondel, V., Guillaume, J-L., Lambiotte, R., Lefebvre, E., 2008. Fast Unfolding of Communities in Large Networks. *J. Stat. Mechanics Theory Exp*. 2008. 10.1088/1742-5468/2008/10/P10008

Bracher, A., Taylor, M. H., Taylor, B., Dinter, T., Rottgers, R., Steinmetz, F., 2015. Using empirical orthogonal functions derived from remote-sensing reflectance for the prediction of phytoplankton pigment concentrations. *Oc. Sci.*, 11(1), 139–158. https://doi.org/10.5194/os-11-139-201

Brewin, R.J.W., Sathyendranath, S., Hirata, T. Lavender, S.J., Barciela, R.M., Hardman-Mountford, N.J., 2010. A three-component model of phytoplankton size class for the Atlantic. *Oc. Ecol. Model.*, 221 pp. 1472-1483

Canuti, E., Ras, J., Grung, M., Röttgers, R., Costa Goela, P., Artuso, F., Cataldi, D., 2016. HPLC/DAD Intercomparison on Phytoplankton Pigments (HIP-1, HIP-2, HIP-3 and HIP-4), *EUR 28382 EN*, Publications Office of the European Union, Luxembourg

Canuti, E., Artuso, F., Bracher, A., Brotas, V., Devred, E., Dimier, C., Giardina, I., Mendes, C.R., Murawski, S., Peeken, I., Tracana, A., Ras, J. and Wiegmann, S., 2022, The Fifth HPLC Intercomparison on Phytoplankton Pigments (HIP-5) Technical Report, *EUR 31334 EN*, Publications Office of the European Union, Luxembourg, doi:10.2760/563102

Canuti, E., 2023 Phytoplankton pigment *in situ* measurements uncertainty evaluation: an HPLC interlaboratory comparison with a European-scale dataset. *Front. Mar. Sci.* 10:1197311. doi: 10.3389/fmars.2023.1197311

Catlett, D., Siegel, D. A., 2018. Phytoplankton pigment communities can be modeled using unique relationships with spectral absorption signatures in a dynamic coastal environment. *J. Geophys. Res.: Oc.*, 123, 246– 264. https://doi.org/10.1002/2017JC013195

Dugué, N., Anthony Perez, A., 2015. Directed Louvain: maximizing modularity in directed networks. Université d'Orléans. hal-01231784. https://hal.archives-ouvertes.fr/hal-01231784

Hagberg, A., Swart, P. and S Chult, D., 2008. NetworkX: High Productivity Software for Complex Networks, Los Alamos National Laboratory (LANL), https://networkx.org/

HELCOM, 2018. State of the Baltic Sea – Second HELCOM holistic assessment 2011–2016. In: Baltic Sea Environment Proceedings 155. Available online at: https://helcom.fi/media/publications/BSEP155.pdf

Hirata, T. , N.J. Hardman-Mountford, R.J.W. Brewin, J. Aiken, R.G. Barlow, K. Suzuki, T. Isada, E. Howell, T. Hashioka, M. Noguci-Aita, Y. Yamanaka. 2011. Synoptic relationships between surface chlorophyll-a and diagnostic pigments specific to phytoplankton functional types. *Biogeosc.*, 8 pp. 311-327

Hällfors, G. 2004. Checklist of Baltic Sea phytoplankton species (including some heterotrophic protistan groups). Helsinki Commission, *Baltic Marine Environment Protection Commission*. https://www.helcom.fi/wp-content/uploads/2019/10/BSEP95.pdf

Hooker, S.B., Thomas, C.S., Van Heukelem, L., Schluter, L., Ras, J., Claustre, H.,Russ, M.E., Clementson, L., Canuti, E., Berthon, J-F., Perl, J., Normandeau, C., Cullen, J., Kienast, M., Pickney J.L., 2010. The Forth SeaWiFS HPLC Analysis Round-Robin Experiment (SeaHARRE-4) *NASA Tech. Memo. 2010-215857*, NASA Goddard Space Flight Center, Greenbelt, Maryland, 74 pp.

IOCCG (2014). Phytoplankton Functional Types from Space. Sathyendranath, S. (ed.), Reports of the International Ocean-Colour Coordinating Group, No. 15*, IOCCG*, Dartmouth, Canada.

Latasa, M., Bidigare, R.R., 1998. A comparison of phytoplankton populations of the Arabian Sea during the Spring Intermonsoon and Southwest Monsoon of 1995 as described by HPLC-analyzed pigments, Deep Sea Res.Part II: Topical Studies in Oceanography, Vol. 45, Issues 10–11, 2133-2170, https://doi.org/10.1016/S0967-0645(98)00066-6

Lotze, H.K., Lenihan, H.S., Bourque, B.J., Bradbury, R.H., Cooke, R.G., Kay, M.C., Kidwell, S.M., Kirby, M.X., Peterson, C.H., Jackson, J.B.C.,2006. Depletion, degradation, and recovery potential of estuaries and coastal seas. Science 312: 1806-1809

Kramer, S.J., Siegel, D.A., 2019. How can phytoplankton pigments be best used to characterize surface ocean phytoplankton groups for ocean color remote sensing algorithms?. *J. Geophys. Res.: Oc.*, 124, 7557–7574. https://doi.org/10.1029/2019JC015604

Kramer, S.J. Siegel D.A., Graff, J.R., 2020. Phytoplankton Community Composition Determined From Co-variability Among Phytoplankton Pigments From the NAAMES Field Campaign. *Front. Mar. Sci.,* Sec. Mar. Ecosys. Ec., Vol. 7 https://doi.org/10.3389/fmars.2020.00215

Mackey, M., Mackey, D. Higgins, H. W., Wright, S. W., 1996. CHEMTAX - a program for estimating class abundances from chemical markers: Application to HPLC measurements of phytoplankton. *Mar. Ecol. Prog. Ser.* 144: 265– 283

Mason, M.J., Fan, G., Plath, K. Signed weighted gene co-expression network analysis of transcriptional regulation in murine embryonic stem cells. *BMC Genomics* **10**, 327 (2009). https://doi.org/10.1186/1471-2164-10-327

Mouw, C.B., Hardman-Mountford, N.J., Alvain, S., Bracher, A. , Brewin, R.J.W. , Bricaud, A., Ciotti, A.M., Devred, E. Fujiwara, A., Hirata, T. Hirawake, T. , Kostadinov, T.S., Roy, S., Uitz J., 2017. A consumer's guide to satellite remote sensing of multiple phytoplankton groups in the Global Ocean. *Front. Mar. Sci.,* 4

Meler, J., Wo´zniak, S.B., Stoń-Egiert, J., 2020. Comparison of methods for indirectly estimating the phytoplankton population size structure and their preliminary modifications adapted to the specific conditions of the Baltic Sea. *J. Mar. Syst.* 212, 103446. https://doi.org/10.1016/j.jmarsys.2020.103446.

Nelson, N.B., Siegel, D. A., 2013. Global distribution and dynamics of chromophoric dissolved organic matter. *Annu. Rev. Mar. Sci.* 5:447–476, doi:10.1146/annurev-marine-120710-100751

Newman, M.E.J., 2004. Detecting community structure in networks. *Eur. Phys. J.* B **38**, 321–330 https://doi.org/10.1140/epjb/e2004-00124-y

Newman, M.E.J., 2006. Modularity and community structure in networks, *App. Math.* 103 (23) 8577-8582 https://doi.org/10.1073/pnas.0601602103

Olli, K., Klais-Peets, R., Tamminen, T., Ptacnik, R., Andersen, T., 2011. Long-term changes in the Baltic Sea phytoplankton community. *Boreal Env. Res.* 16 (SUPPL A). 3-14

Olofsson, M., Suikkanen, S., Kobos, J., Wasmund, N., Karlson, B., 2020. Basin-specific changes in filamentous cyanobacteria community composition across four decades in the Baltic Sea. *Harmful Algae* 91, 101685. https://doi.org/10.1016/j.hal.2019.101685

Pli nski, M., Mazur-Marzec, H., Jo´zwiak, T., Kobos, J., 2007. The potential causes of cyanobacterial blooms in Baltic Sea estuaries. *Oceanol. Hydrobiol. Stud.* 36 (1), 125–137. https://doi.org/10.2478/v10009-007-0001-x

Rodriguez F., Varela M., Zapata M., 2002. Phytoplankton assemblages in the Gerlache and Bransfield Straits (Antarctic Peninsula) determined by light microscopy and CHEMTAX analysis of HPLC pigment data, Deep-Sea Res. Pt. II, 49 (4–5), 723–747

Roy, S., Llewellyn, C.A., Egeland, E.S., Johnsen, G., 2011. Phytoplankton Pigments, Characterization, Chemotaxonomy and Applications in Oceanography. Cambridge Univ. Press, p. 845

Schlüter, L., Møhlenberg, F., Havskum, H., and Larsen, S., 2000. The use of phytoplankton pigments for identifying and quantifying phytoplankton groups in coastal areas: testing the influence of light and nutrients on pigment/chlorophyll a ratios. *Mar. Ecol. Prog. Ser.* 192, 49–63. doi: 10.3354/meps192049

Schlüter, L., Garde, K., Kaas, H., 2004. Detection of the toxic cyanobacteria Nodularia spumigena by means of a 4-keto-myxoxanthophyll-like pigment in the Baltic Sea. *Mar. Ecol. Prog. Ser.*. 275. 69-78. 10.3354/meps275069

Schlüter, L., Behl, B., Striebel, M., Stibor, H., 2016. Comparing microscopic counts and pigment analyses in 46 phytoplankton communities from lakes of different trophic states. *Freshw. Biol.* 61, 1627–1639. https://doi.org/10.1111/fwb.12803

Seppälä J., 2009. Fluorescence properties of Baltic sea phytoplankton. Monographs of the Boreal environment research (34) Helsinki: Edita Prima Ltd. p. 83

Stoń-Egiert, J., Łotocka, M., Ostrowska, M., Kosakowska, A., 2010. The influence of biotic factors on phytoplankton pigment composition and resources in Baltic ecosystems: new analytical results *Oceanologia*, 52 (1) pp. 101-125, 10.5697/oc.52-1.10

Stoń-Egiert, J., Ostrowska, M., 2022. Long-term changes in phytoplankton pigment contents in the Baltic Sea: Trends and spatial variability during 20 years of investigations, *Cont. Shelf Res.,* Vol. 236, https://doi.org/10.1016/j.csr.2022.104666

Sun, X., Shen, F., Brewin, R.J.W., Li, M., Zhu, Q., 2022. Light absorption spectra of naturally mixed phytoplankton assemblages for retrieval of phytoplankton group composition in coastal oceans. *Limnol. Oceanogr.,* 67: 946-961. https://doi.org/10.1002/lno.12047

Traag, V.A., Waltman, L., van Eck, N.J., 2019. From Louvain to Leiden: guaranteeing well-connected communities. *Sci Rep* 9, 5233 https://doi.org/10.1038/s41598-019-41695-z

Trees, C. C., Clark, D. K., Bidigare, R.R., Ondrusek, M.E., Mueller, J. L., 2000. Accessory pigments versus chlorophyll a concentrations within the euphotic zone: A ubiquitous relationship, *Limnol. Oceanogr.*, 5.

JGOFS, 1994: Protocols for the Joint Global Ocean Flux Study Core Measurements. Intergovernmental Oceanographic Commission, Scientific Committee on Oceanic Research. Manual and Guides, UNESCO, 29, 91–96

Jeffrey, S.W., Mantoura, R.F.C., Wright, S.W., 1997. Phytoplankton Pigments in Oceanography, *Monogr. Oceanogr. Meth*., UNESCO

Uitz, J., Claustre, H., Morel, A., Hooker, S.B., 2006. Vertical distribution of phytoplankton communities in open ocean: an assessment based on surface chlorophyll. *J. Geophys. Res.*, 111 p. C08005

Vidussi, F., Claustre, H., Manca, B.B., Luchetta, A., Marty, J.C., 2001. Phytoplankton pigment distribution in relation to upper thermocline circulation in the eastern Mediterranean Sea during winter *J. Geophys. Res.*, 106 (C9) pp. 19939-19956

Zhang, B., Horvath, S., 2005. A general framework for weighted gene co-expression network analysis. Statistical applications in genetics and molecular biology, 4(1)

Wasmund, N., Nausch G., Matthäus , W., 1998. Phytoplankton spring blooms in the southern Baltic Sea — spatio temporal development and long-term trends. *J. Plankton Res.* 20: 1099–1117

Wasmund, N., Uhlig S., 2003. Phytoplankton trends in the Baltic Sea. *ICES J. Mar. Sci.* 60: 177–186

Wasmund N., Göbel J., von Bodungen B., 2008. 100-years-change in the phytoplantkon community of Kiel Bight (Baltic Sea). *J. Mar. Syst.* 73: 300–322

Wasmund, N., Tuimala, J., Suikkanen, S., Vandepitte, L., Kraberg, A., 2011. Long-term trends in phytoplankton composition in the western and central Baltic Sea. *J. Mar. Syst.* 87, 145–159. https://doi.org/10.1016/j.jmarsys.2011.03.010

SUPPLEMENTARY

Supplementary Figure S1. The log-log co-variance TAcc/TChl *a* along the six Baltic Oceanographic Campaigns, axis in log scale (mg m^{-3}).

Supplementary Figure S2. The correlation matrix (Pearson correlation coefficient) is associated with both absolute concentrations (bottom left) and ratios to TChl *a*(top right), for the 16 pigments.

Supplementary Figure S3. Hierarchical clustering of phytoplankton pigment ratios to TChl *a* for each campaign of the Baltic dataset.

Supplementary Figure S4. Spanning tree resulting from a matrix modularity analysis: Green Algae (community 0), Procaryotes (community 1), Cryptophytes (community 2), Dinoflagellates (community 3), Diatoms (community 4)

Supplementary Figure S5. CHEMTAX applied to the six oceanographic campaigns (initial matrix provided in Schlüter et al., 2000). In the pies the relative biomass abundance correspond to Dinophycacee (blue), Diatoms (orange), Cryptophytes (green) , Cyaonophytes (red), Chlorophyceae (purple), Euglenophytes (brown) and Prasinophytrs (pink)

Supplementary Table S1. Phytoplankton Pigments Acronym (Jeffrey et al., 1997; Roy et al., 2011), description and taxonomic relevance of the main algal divisions/classes (see tables in Catlett and Siegel, 2018; Kramer et al., 2019 and Seppälä, 2009 for more detail). Only pigments used in the present study are shown. As in Seppälä (2009), a single "x" indicate secondary pigment or pigment representative of few species/strains.

Supplementary Table S2. Stations assigned to different functional types by PFT and Network analysis

Acknowledgments

The author would like to acknowledge Juha Flinkman, Seppo Kaitala, and Jukka Seppälä from the Finnish Environment Institute (former Finnish Institute of Marine Research) for the opportunity to participate in the Oceanographic campaign on-board the R/V "Aranda" and to the crew of R/V "Aranda" for their support during all the cruises. Acknowledgments are due to Dirk Van der Linde from JRC for water sampling and laboratory analyses.

Conflict of Interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Author Contributions

First authorship: Elisabetta Canuti.

EC contributed to this paper by conceptualizing and designing the study and performing analysis and calculations. AP contributed to the paper by reviewing and editing the content.

Funding

The study has been supported by the European Commission Directorate General Joint Research Centre (JRC) and the Copernicus Program.

ANNEX VI

This annex was published in Frontiers in marine Science

Dynamics of Phytoplankton Communities in the Baltic Sea: Insights from a Multidimensional Analysis of Pigment and Spectral Data: Part II, Spectral Dataset

Elisabetta Canuti, Antonella Penna

Front. Mar. Sci. 12:1518057.https//doi.org/10.3389/fmars.2025.1518057

Dynamics of Phytoplankton Communities in the Baltic Sea: Insights from a Multidimensional Analysis of Pigment and Spectral Data: Part II, Spectral Dataset

Authors: Elisabetta Canuti[1, 2], Antonella Penna [3,4,5]

[1] European Commission, Joint Research Centre (JRC), Ispra, Italy

[2]Department of Pure and Applied Sciences, University of Urbino "Carlo Bo", Urbino, Italy

[3]Department of Biomolecular Sciences, University of Urbino, 61029, Urbino, Italy

[4]CoNISMa, National Inter-University Consortium for Marine Sciences, 00196, Rome, Italy

[5]Fano Marine Center, The Inter-Institute Center for Research on Marine Biodiversity, Resources and Biotechnologies, 61032, Fano, Italy

Corresponding author: elisabetta.canuti@ec.europa.eu

Keywords: inherent optical properties, phytoplankton pigments, spectral decomposition, bio-optics, Baltic Sea

Abstract:

The use of hyperspectral satellite missions opens new opportunities for integrated approaches to the study of phytoplankton communities. The Baltic Sea, with its distinct mixture of marine and freshwater characteristics, is a natural laboratory for understanding marine ecosystems.

In this study, we analyzed a dataset from the Baltic Sea containing simultaneous phytoplankton pigment concentrations and absorption spectra. We applied spectral derivative analysis and unsupervised machine learning techniques to identify the unique statistical relationships among phytoplankton pigments and inherent optical properties. The statistical analysis of the absorption spectra provides the basis for a predictive model to assess pigment concentrations from optical measurements. Additionally, we compare our results to know assessment methods, such as Gaussian spectral decomposition, that link the spectral analysis with phytoplankton pigment content.

This study investigates the potential of statistical, data-driven analytical approaches in the development and validation of models for retrieving phytoplankton community composition. The integration of these findings with existing research contributes to the advancement of remote sensing capabilities for monitoring marine ecosystems in the Baltic Sea.

1. INTRODUCTION

The Baltic Sea, connected to the North Sea and the Atlantic Ocean via the Skagerrak, is a shallow intra-continental shelf sea with an average depth of 52 meters. It has low salinity and significant land-derived substance inflows (Blanz et al., 2005), resulting in optical properties distinct from those of ocean waters (Meler et al., 2017, 2018, 2020, 2023; Woźniak et al., 2022). Phytoplankton, the main producers of organic matter in this marine ecosystem, form the basis of marine life. Extensive research has studied the taxonomic structure of phytoplankton communities in the Baltic Sea (Wasmund et al., 2003, 2008; Olli et al., 2011, HELCOM, 2018).

In our previous study (Canuti & Penna, 2024), we used various statistical analyses and machine learning techniques to assess the composition of the phytoplankton community in the Baltic Sea from a phytoplankton pigment dataset. This investigation included a dataset of 273 samples, ensuring sufficient spatio-temporal representation across different regions of the Baltic Sea, including the central and northern Baltic Proper, the Gulf of Gdansk, the Gulf of Finland, and the Bothnian Sea (Fig. 1). In the present study, we investigate phytoplankton pigment concentrations and spectral coefficients (i.e., light absorption spectra) and the relationship between optical properties and the dominant phytoplankton community (Anderson et al., 2008; Sun et al., 2022).

The deployment of hyperspectral satellite such as the German Environmental Mapping and Analysis Program (EnMAP, Guanter et al., 2015) and the Italian Precursore IperSpettrale of Missione Applicativa (PRISMA, Candela et al., 2016) as well as the launch of the NASA Plankton, Aerosol, Cloud, ocean Ecosystem (PACE, Meister et al., 2024) mission, opens new possibilities for an integrated approach to phytoplankton studies. Several studies have focused on modeling or parameterizing light absorption by phytoplankton indirectly through diagnostic pigment analysis (DPA) and categorizing phytoplankton into phytoplankton functional types (PFT) and phytoplankton size classes (PSC) (Vidussi et al., 2001; Uitz et al., 2006, 2008; Brewin et al., 2010, 2012; Hirata et al., 2011, Mouw et al., 2017). IOCCG (2014) discusses various approaches to identify phytoplankton size structure from satellite data, and explores the potential for developing algorithms to remotely determine the contribution of different PFTs in water bodies. The use of biomarker pigments as representative of individual taxa, such as fucoxanthin (Fuco) for diatoms, has become a common method to infer phytoplankton community composition. However, this approach represents a simplification of the ecological and biochemical reality. While fuco is predominantly associated with diatoms, it is also present in other taxa such as haptophytes and some raphidophytes. This overlap limits the ability to accurately assign pigment concentrations to specific groups, thereby introducing potential biases and uncertainties into assessment of phytoplankton diversity and their use as

proxies for individual taxa. Nevertheless, the PFT concept is relevant to the study of ecological and biogeochemical processes, particularly in model studies, as it helps to understand the role of phytoplankton in global cycles of key chemical elements and primary production. However, these methods are not effective for the Baltic Sea, which is an optically complex water body (Meler et al., 2020).

Due to significant freshwater inputs, the upper layer of the Baltic Sea contains a notable amount of chromophoric dissolved organic matter (CDOM), making CDOM absorption the dominant optical factor both in open water and along coastal regions. Particularly in the southern Baltic Sea, the inflow of large rivers contributes to increased optical variability (Harvey et al., 2015). CDOM is responsible for a significant fraction of the absorption of ultraviolet (UV) and blue spectral light in the ocean (Nelson and Siegel, 2013). Previous studies have explored the relationship between phytoplankton structure and CDOM content, correlating measurements of CDOM with phytoplankton composition (Barrón et al., 2014).

The phytoplankton absorption coefficient is a critical parameter in several applications, such as pigment biomass remote sensing or light attenuation in the ocean (Chase et al., 2013, 2017). The pattern of absorption spectrum is influenced by factors such as chlorophyll *a* concentration, algal species, cell size, and changes in pigment composition, including accessory pigments such as chlorophyll *b* and *c*, carotenoids, and phycobiliproteins. Similar to the chemotaxonomic characterization of Baltic Sea phytoplankton communities, which depends on diagnostic pigments (Schlüter et al., 2000; Stoń-Egiert et al., 2022), the available models for retrieving phytoplankton biomarker pigments composition from absorption spectra for the Baltic Sea have focused on the southern Baltic Sea (Ficek et al., 2004, Kowalczuk et al., 2005). However, a significant research gap remains in the study of the wider Baltic basin. Additionally, most algorithms for interpreting remote sensing data in marine and oceanic environments are unsuitable for the Baltic Sea and lead to significant errors due to its unique characteristics (Meler et al., 2020).

This study aims to assess data analysis methods to investigate the relationship between phytoplankton pigment composition determined by high-performance liquid chromatography (HPLC) technique and optical properties throughout the Baltic Sea. The spectral dataset was subjected to different analysis techniques, including hierarchical cluster analysis (HCA) and principal component analysis (PCA). The results of the spectral analysis are compared with those of the corresponding phytoplankton pigment dataset studied previously (Canuti & Penna, 2024). Derivative analysis was applied to the pigmented absorption spectra dataset to obtain a predictive model for assessing the phytoplankton biomarker pigments from optical data. The proposed predictive model was then compared with the Gaussian spectral decomposition model (Hoepffner et al., 1991, 1993, Ficek et al., 2004, Chase et al., 2013) to evaluate the advantages and limitations of different approaches in a complex water basin such as the Baltic Sea.

MATERIALS AND METHODS

a. Field dataset

The dataset for the Baltic Sea was collected during six ocean colour validation campaigns conducted in May and September 2004, April 2005, July 2006, August 2007, and August 2008. These campaigns focused primarily on the summer period, which is characterized by the dominance of filamentous cyanobacteria. However, they also included periods with low phytoplankton standing stock, typically from mid-May to mid-June, and the autumn diatom bloom. The first three campaigns covered the southern Baltic Sea, the Gulf of Gdansk, and the Pomeranian Bay, while the latter three campaigns focused on the northern Baltic Proper, the Gulf of Finland, and the Bothnian Sea (Fig. 1). The complete Baltic Sea dataset (BA) comprises 273 stations.

Seasurface temperature (SST) and salinity were measured using the SBE 911 CTD (conductivity, temperature, depth) system (SeaBird, Alifax, USA). Water samples were collected at a surface depth of 1 m using a Niskin bottle and pre-filtered through a 150 µm mesh (Kartel, Germany). Filters for HPLC and spectrophotometric absorption measurements (GF/F filters, Φ 25 mm, 0.7 µm pore size, Whatmann, Germany) were pre-conditioned under constant mild vacuum (not exceeding 0.5 bar), flash-frozen in liquid nitrogen, and subsequently stored at -80 °C. Additionally, 150 ml of water samples were filtered in the field (GSWP filters, Φ 47 mm, 0.22 µm pore size, Whatmann, Germany) and stored at 4 °C in amber bottles until analysis.

b. Phytoplankton Pigments dataset

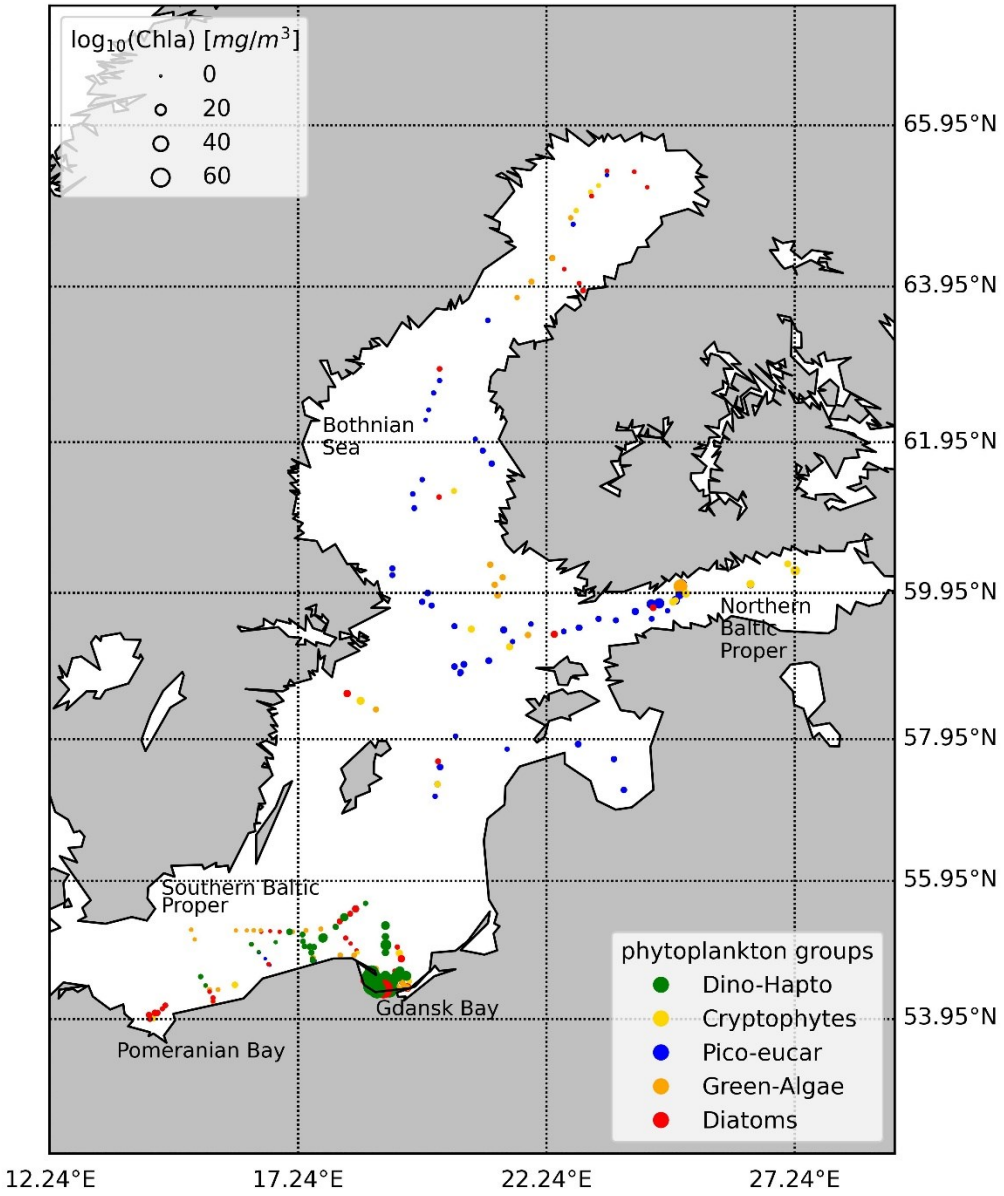
Phytoplankton pigments were quantified at the Joint Research Centre of the European Commission (JRC) using the HPLC method described in Canuti (2023). The data set was subjected to quality control including participation in inter-laboratory exercises (Hooker et al., 2010; Canuti et al., 2016).

Of all the quantified pigments, those found below the detection limit in more than 95% of the stations were excluded from the study, as were redundant pigments (e.g., monovinyl chlorophyll already included in TChl *a*). The remaining eighteen pigments were the focus of the present study: 19'-hexanoyloxyfucoxanthin (Hex), 19'-butanoyloxyfucoxanthin (But), alloxanthin (Allo), Fuco, peridinin (Perid), diatoxanthin (Diato), diadinoxanthin (Diad), zeaxanthin (Zea), total chlorophyll *b* (TChl *b*), chlorophyll *c1* + *c2* (Chl*c1c2*), chlorophyll c3 (Chlc3), total chlorophyll c (TChl *c*), neoxanthin (Neo), violaxanthin (Viola), prasinoxanthin (Pras), lutein (Lut), α,β-carotene + β,β-carotene (Caro), and total chlorophyll *a* (TChl *a*). The composition of the total chlorophylls is described in Table S1 (supplementary

material). The photoprotective carotenoids, PPC, were defined as the sum of Caro, Zea , Allo, and Diad, while the phosynthetic carotenoid, PSC, was considered as the sum of Hex, Fuco, But and Peri.

In the use of phytoplankton pigments for categorizing phytoplankton into community and functional types, we referred to Table S1, that was derived from Jeffrey et al. (1997), Seppälä, (2009) and Roy et al. (2011).

Figure 1. Spatial distribution of the sampling points in the years 2004-2008 colour corresponds to dominant species present according to the network-community classification (NCA) in Canuti & Penna (2024).



c. Absorption Coefficient dataset $a_p(\lambda)$

Similar to the HPLC analysis, the absorption coefficient, $a_p(\lambda)$, was quantified at the JRC.

The method used to determine the absorption coefficient, $a_p(\lambda)$, was a modified protocol of Ferrari and Tassan (1995). Transmittance–reflectance measurements were performed using a Perkin Elmer 950 (Perkin Elmer, USA) equipped with a 150 mm integrating spectralon sphere (Labsphere, USA). To calculate the absorption coefficients from the optical density, $OD_s(\lambda)$, an appropriate correction must be applied to compensate for the elongation of the optical path due to multiple scattering in the material collected on the filter. This is achieved by using the dimensionless path length amplification factor, the β-factor, which converts the measured optical density of particles collected on the filter, $OD_s(\lambda)$, to the optical density of these particles in solution, $OD_{sus}(\lambda)$, (Mitchell 1990). The β-factor formula for the T-R method used in the present work is:

$OD_{sus}(\lambda)=0.719\ OD_s^{1.2287}$ [1]

The coefficient of light absorption by all suspended particles was then calculated using the formula:

$a_p(\lambda) =(\ln(10)\cdot OD_{sus}(\lambda))/l$ [2]

where l [m] is the hypothetical optical path in solution, determined as the ratio of the volume of filtered water to the effective area of the filter. The discrimination due to the pigmented, $a_{ph}(\lambda)$, and the non-pigmented, $a_{NAP}(\lambda)$, fractions of the particle absorption coefficient, is obtained by bleaching the sample on the filter by adding a solution of sodium hypochlorite (NaClO, Merck, Germany) as an oxidizing agent. The filter depigmentation is obtained by placing the filter horizontally against a filter holder (PALL, USA) and adding 1 mL of a 3.3% vol. of NaClO (4% active Cl) in Milli-Q water solution as a bleaching agent. In the original method, the distinction between the pigmented fraction, $a_{ph}(\lambda)$, and the non-pigmented fraction, $a_{NAP}(\lambda)$, of the particle absorption coefficient was monitored visually. In the modified protocol, the depigmentation process was monitored with a radiometer measurement (FieldSpec FS VNIR, Analytical Spectral Devices, USA) and a light source (Schott halogen fiber optic lamp, type 6423FO, Philips, Germany) between 300 and 1000 nm. The disappearance of the reflectance peak corresponding to TChl *a* at 675 nm was considered evidence of complete depigmentation. The bleaching process was monitored using the VNIR software (AnalyticalSpectral Devices, USA) (Canuti & van der Linde, 2006).

d. Chromophoric dissolved organic matter (CDOM) and suspended particulate matter (SPM) dataset

Additionally, the dataset includes matched samples of suspended particulate matter (SPM) concentrations and the spectra of the light absorption coefficient of CDOM, $a_g(\lambda)$. CDOM and SPM samples were analyzed at the JRC. Concentrations of SPM [g m^{-3}] were determined by the gravimetric method combined with the loss-on-ignition technique. 47 mm GF/F filters were pre-combusted at 450°C for 4.5 h, rinsed in 1 liter of pure deionized water, and dried at 75°C for 2.5 h. After 4 hours in a desiccator, the filters were weighed and labeled using an analytical balance (Sartorius, Germany, precision 0.001 mg). In the field, depending on the concentration of suspended particles, 150 to 1500 mL of seawater was filtered through the filters, which were then rinsed with 30 mL of deionized water to remove surface salts and stored at -20°C until analysis. In the laboratory, the filters were thawed and dried at 75°C for 2.5 h, kept in a desiccator for 4 h, and then reweighed to determine the SPM value.

The $a_g(\lambda)$, (m^{-1}) measurement was carried out in the spectral range of 350 and 750 nm in a double beam double ray spectrophotometer (Perkin Elmer, USA), measuring the sample against double distilled water as a reference, using 100 mm optical path length quartz cuvettes. The optical density OD(λ) was converted to CDOM absorption by multiplying OD(λ) by 2.303 and then dividing by the path length l (m) (Mannino et al., in contribution).

e. Methods in Data Analysis

i. Statistical analysis of the phytoplankton pigments dataset

The phytoplankton pigment dataset was subjected to statistical analysis to evaluate the pigment distribution in the different campaigns. In our previous study (Canuti & Penna, 2024), we applied a hierarchical clustering analysis (HCA) and a multivariate analysis (Principal Component Analysis, PCA), together with community partitioning by networkcommunity detection analysis (NCA), to the pigment dataset. These statistical analyses aimed to assess the phytoplankton community in the Baltic Sea in an alternative approach to widely used chemotaxonomic approaches, such as CHEMTAX (Mackey et al., 1996) or PFTs analysis. Here, the results of HCA and PCA applied to the pigments dataset were used in the organization and interpretation of the absorption coefficient dataset analysis.

ii. Spectral decomposition of the absorption coefficient dataset

The spectral decomposition was applied with the aim of evaluating the agreement between our proposed predictive model (see section 2.5.4) and the Gaussian decomposition approach. To analyze the absorption

coefficient, $a_p(\lambda)$, dataset, we applied the Gaussian spectral decomposition proposed by Chase et al. (2013) and originally applied to AC-S (WetLabs, USA) data. The Chase et al. (2013) decomposition has shown a small residual error (0.001 m^{-1}), comparable to other Gaussian decompositions (Huping et al., 2019). In an attempt to use an algorithm not specialized for the southern Baltic Sea, we prefer the model of Chase et al. (2013) model to Ficek (2004). In our study, we used the measured $a_{NAP}(\lambda)$ value instead of approximating it with an exponential function, as in the original work of Chase et al. (2013). The initial wavelengths for the decomposition were those proposed by Chase et al. integrating the previous work of Hoepffner (1991, 1993) and Bricaud et al. (2004, 2007) (Table 1). To determine the optimal set of functions to approximate the decomposition, we used a leastsquares minimization (cost function, in equation [3], Python3) to compare the reconstructed curve with the original curve:

$$\chi^2 = \sum_{\lambda=410}^{720} \frac{\left(a_p(\lambda) - \left(\sum_{i=1}^{12} a_{gauss}(\lambda_i)\left(\exp\left\{-0.5\left[\frac{\lambda-\lambda_i}{\sigma_i}\right]^2\right\}\right) + a_{NAP}(\lambda_i)\right)\right)}{\sigma_{SD}^2(\lambda)} \qquad [3]$$

where $a_p(\lambda)$, is the measured particle absorption, $a_{gauss}(\lambda_i)$ is the magnitude of the i-Gaussian function, λ_i is the centre wavelength of the i-Gaussian function, σ_i, is the width of the Gaussian. The equation is normalized by the standard deviation of the measured absorption spectra at each wavelength $\sigma_{SD}^2(\lambda)$. Following the approach of Chase et al. (2013), we identify correlations at specific wavelengths between HPLC pigments and Gaussian amplitude functions by checking the Spearman's rank correlation coefficient. We then performed a linear regression analysis (least-square best fit) between the HPLC pigment concentration, assumed as a reference value, and the phytoplankton peak absorbance values, $a_{gauss}(\lambda_i)$: the fit was calculated individually for each log-normalized pigment (pig_j) and absorbance peak according to the equation [4]:

$$log_{10}\left(a_{gauss}(\lambda_i)\right) = \mathrm{A}_{ij} + B_{ij} * log_{10}(pig_j) \qquad [4]$$

The pigment concentration can be solved from the previous equation as:

$$pig_j = \{a_{gauss}(\lambda_i)/exp * (\mathrm{A}_{ij})\}^{1/B_{ij}} \qquad [5]$$

A_{ij} and B_{ij} describe the relation between the jth pigment, pig_j , derived from the absorption spectra and the $HPLC_j$ corresponding pigments**.**

Table 1. Center wavelengths of Gaussian functions used for the spectral decomposition as in Chase et al. (2013). PPC = photoprotective carotenoids = Caro + Zea +Allo +Diad; PSC = phosynthetic carotenoids = Hex + Fuco + But +Peri

	TChl a + TChl c	TChl a	TChl b + TChl c	TChl b	PPC	PSC	TChl c	TChl a	TChl c	TChl b	TChl a
peak locations [nm]	409	437	457	467	491	527	585	620	639	658	678
peak std. dev [nm]	17	13	12	14	17	15	17	15	15	15	13
σ std. dev. [nm]	±3.8	±1.9	±1.9	±2.8	±1.7	±2.5	±2.6	±2.9	±1.4	±1.5	±0.6

PPC = Caro + Zea +Allo +Diad; PSC =Hex + Fuco + But +Peri

iii. Statistical and derivative analysis of absorption coefficient dataset

The correlation between $a_{ph}(\lambda)$ (the absorption coefficient of pigmented particles) and changes in the phytoplankton community composition and abundance has been stidued extensively (Bricaud et al., 2004, 2007, Chase et al., 2013). It is well known that the presence of different accessory pigments, such as diagnostic pigments found in different species, and differences among phytoplankton groups can influence the spectral characteristics of $a_{ph}(\lambda)$ (Ciotti et al., 2002; Bricaud et al., 2004).

In our examination, we will consider both $a_{ph}(\lambda)$ and the chlorophyll specific absorption coefficient of phytoplankton $a^*_{ph}(\lambda)$. The choice to also investigate $a^*_{ph}(\lambda)$ was based on the consideration that the package effect and the proportion of accessory pigment compositions are two major factors influencing the changes in the pattern and magnitude of $a^*_{ph}(\lambda)$. In addition, in the previous analysis of the HPLC pigment dataset, all the pigments were normalized against the ubiquitous TChl *a*. Consideration of $a^*_{ph}(\lambda)$ allowed us to compare the results of our derivative analysis and clustering more coherently with previous findings. We then considered the Pearson correlations (R) among spectral wavelengths, in particular those related to $a_{ph}(\lambda)$, $a_{ph}'(\lambda)$ and $a_{ph}''(\lambda)$ (where the prime signs indicate the first and second spectral derivatives). We repeated the exercise for the $a^*_{ph}(\lambda)$, $a^*_{ph}'(\lambda)$ and $a^*_{ph}''(\lambda)$. Overall, we expected strong correlations among the pigment concentrations and the absorption spectra at all wavelengths, based on the previous findings by Catlett & Siegel (2018). Catlett observed significant multicollinearity problems in the analysis of the correlation between the pigments and $a^*_{ph}(\lambda)$, which could lead to an ill-conditioned model. To address this, a derivative transformation was applied to reduce the multicollinearity across different wavelengths.

Following the methodology outlined by Catlett & Siegel (2018) and Teng (2022), we used derivative analysis to identify meaningful absorption characteristics and explore their associations with phytoplankton pigments and community structure. We focused on the spectral range between 400 and

750 nm, considering both the first derivative spectra, $a_{ph}'(\lambda)$ and $a^*_{ph}'(\lambda)$, and the second derivative spectra, $a_{ph}''(\lambda)$ and $a^*_{ph}''(\lambda)$:

$$a'_{ph}(\lambda) = \frac{a_{ph}(\lambda+\Delta\lambda) - a_{ph}(\lambda-\Delta\lambda)}{2*\Delta\lambda}; \qquad a^{*\prime}_{ph}(\lambda) = \frac{a^*_{ph}(\lambda+\Delta\lambda) - a^*_{ph}(\lambda-\Delta\lambda)}{2*\Delta\lambda} \quad [6]$$

$$a^{"}_{ph}(\lambda) = \frac{a_{ph}(\lambda+\Delta\lambda)+a_{ph}(\lambda-\Delta\lambda)-2*a_{ph}(\lambda)}{\Delta\lambda^2}; a^{*"}_{ph}(\lambda) = \frac{a^*_{ph}(\lambda+\Delta\lambda)+a^*_{ph}(\lambda-\Delta\lambda)-2*a^*_{ph}(\lambda)}{\Delta\lambda^2} \quad [7]$$

We applied a finite band separation (i.e., $\Delta\lambda$) of 1 nm at the second order. To mitigate the effect of noise on our analysis, we implemented a noise reduction technique (Teng et al., 2022, Catlett & Siegel 2018). In signal processing, we used the Savitzky-Golay filter to reduce signal noise and enhance the signal trend smoothness (Tsai & Philipot, 1998). This filter (Python3, svgolay) computed a polynomial fit for each running spectral window, employing a polynomial degree of 2 in our case and a window size of 15 nm. Before performing derivative analysis, we applied the smoothing filter to the spectral data to optimize the linear relationships between selected pigments and their respective absorption maxima, which were identified in the second derivative spectra.

iv. Empirical Orthogonal Function (EOF) Model to assess pigment concentrations from optical measurements

We developed and evaluated a novel model to predict pigment concentrations in biological samples using the distinct correlation patterns observed between pigments and $a_{ph}(\lambda)$, $a_{ph}'(\lambda)$, $a^*_{ph}(\lambda)$ and $a^*_{ph}'(\lambda)$. Ideally, the model could be used to assess pigment concentrations from continuous optical measurements. To identify relevant features and reduce dimensionality, we used PCA for finding patterns in the data and transform them into a set of orthogonal variables called principal components. We will refer to this as an Empirical Orthogonal Function (EOF) model to avoid confusion with the PCA analysis of the HPLC pigment dataset.

The EOF analysis on the absorption dataset determines the wavelengths that most effectively capture the variance within the dataset. Here, we called X the absorption matrix, where each row (M) corresponds to an observation and each column (N) corresponds to a wavelength. The pigment dataset was represented by the matrix Y of dimensions MxP, where each row (M) represents an observation and each column (P) represents the concentration of pigments.

The matrix X was then subjected toa Singular Value Decomposition (SVD) to derive the EOF modes:

$X = U\Sigma V^T$, where $x_{ij} = \sum_{k=l,N} u_{ik}\sigma_k v_{kj}$ [8]

In this equation, V is an N × N matrix containing the absorption data, U is an M × N matrix containing the principal components, Σ is an N × N matrix containing the singular values along the diagonal, and k represents the index of the EOF mode (with a length of N). We used a general linear model to forecast pigment concentrations (log-transformed) for each pigment, denoted as y_p. This model used a subset of principal components (PCs), represented as **U**, as covariates [Eq. 8]. We used both $a_{ph}(\lambda)$ and $a^*_{ph}(\lambda)$ as well as $a_{ph}'(\lambda)$ and $a^*_{ph}'(\lambda)$ PC decompositions, and in the case of $a^*_{ph}(\lambda)$ we also considered a linear regression with the non-log-transformed value of the pigments [Eq. 7]. The multiple regression model took the form:

$$\log(y_p) = a + b_1 u_1 + b_2 u_2 + \cdots + b_n u_n \quad [9]$$

$$y_p = c + d_1 u_1 + d_2 u_2 + \cdots + d_n u_n \quad [10]$$

Here, $\log(y_p)$ represented the log(10) transformed concentration of pigment p, while u_1, u_2 … u_n represented the leading n PC scores of **U**. The model incorporated an intercept a and regression coefficients b_1, b_2, ..., b_n. When the log-transformed pigment values were used as the target, a concentration of 0.00001 mg m^{-3} was added to all values. When the non-log -transformed y_p concentration of pigment p, was used, the model intercept is denoted c and the regression coefficients d_1, d_2, ..., d_n.

To define the subset of PCs, we excluded all those with standard deviations less than or equal to 0.0001 times the standard deviation of the first component. We then used multiple linear regression (MLR) to build predictive models using selected principal components. Stepwise feature selection, implemented through ordinary least squares (OLS) regression, facilitated the iterative inclusion and exclusion of features based on statistical significance. This process aimed to optimize model performance by selecting the most informative features, defining the number of n for each prediction. The selection of the best linear models was based on minimizing the Akaike Information Criterion (AIC). Once the optimal linear model was identified, the significance of the included terms was assessed by measuring the change in AIC (ΔAIC) when each term was removed.

Hence, we assumed that principal components explaining a larger proportion of the variability would be the first 100 principal components of $a^*_{ph}(\lambda)$ or of $a'^*_{ph}(\lambda)$, $a_{ph}(\lambda)$ and $a'_{ph}(\lambda)$, thus containing all the relevant information for modelling all pigments. To evaluate the robustness of this assumption, we performed cross-validation using 10, 25, 50, 100, and 200 principal components in the model formulation, respectively. The PCs with values too close to noise were excluded from the model. We tested this sensitivity using six distinct biomarker pigments: TChl *b*, Fuco, Peri, Allo, Zea and Peri. To evaluate the robustness and generalization of our models, we performed cross-validation with 500 permutations. This

involved randomly splitting the data into training and testsets multiple times and calculating the Root Mean Squared Difference (RMSD) for each permutation. Scatter plots with logarithmic scales were generated to visualize the relationship between predicted and observed values, providing insight into the model's predictive accuracy.

Metrics for the modelled $a^*_{ph}(\lambda)$, $a'^*_{ph}(\lambda)$, $a_{ph}(\lambda)$ and $a'_{ph}(\lambda)$, such as the coefficient of determination (R^2), root mean square difference (RMSD), slope (S) and intercept (a) of the linear regression were derived from the log predicted values, $\log(y_p^P)$, compared to the log observed pigment concentration data, $\log(y_p)$. Conversely, metrics asuch as mean percent difference (MPD), percent bias (PB), and median percent difference (MDPD) were always computed based on non-log-transformed pigment concentrations.

$$\text{RMSD} = \sqrt{\frac{1}{N}\sum_{i=1}^{N}[\log(y_p^P) - \log(y_p)]^2} \quad [11]$$

$$\text{MPD} = \frac{100}{N}\sum_{i=1}^{N}[(y_p^P - y_p)/y_p] \quad [12]$$

$$\text{MDPD} = median(100\ \frac{(y_{i,p}^P - y_{i,p}^V)}{y_{i,p}^V}) \quad [13]$$

Where $y_{i,p}^V$ is the observed pigment concentration for observation *i*, $y_{i,p}^P$ is the predicted pigment concentration for observation *i* and N is the number of observations.

In evaluating the models, we performed cross-validations with 500 permutations to rigorously assess their performance. This process involved iteratively partitioning the dataset into training (X_{train}) and (X_{val}) test subsets to endure robustness in model evaluation. In the first approach, the data was randomly split into two subsets, with 80% of the data used for model fitting/training, including X_{train}, and the remaining 20% was used for prediction validation, including X_{val}. In the second approach, we isolated a specific campaign as the training dataset and fine-tuned the method on the remaining data. From each permutation, we calculated the RMSD and its variation coefficient, providing insight into the prediction accuracy and consistency across different data splits.

To gain a visual understanding of model performance, we generated scatter plots with logarithmic scales. These plots allowed us to observe the relationship between predicted and original values, providing a graphical representation of the model's predictive ability. Additionally, regression lines, representing the relationship between predicted and observed values were superimposed on the scatterplots. These lines were accompanied by R-squared values, providing a quantitative measure of the goodness of fit of the models.

2. RESULTS

a. Variability and clustering of the dataset

Our dataset was examined in terms of variability. The key statistical data highlighting the variability within our dataset are summarized in Table 2 and Fig. 2 (Table 2 and Fig. 2). It is noteworthy that the yellow substance shows a high variation in the whole dataset. The range of TChl *a* varies from a concentration typical of oligotrophic water (0.272 mg m^{-3}) to concentration characteristic of eutrophic conditions (60.222 mg m^{-3}). In our dataset, Peri, the diagnostic pigment representing the dinoflagellate community, exhibited the highest range of concentration variation: the range was from the instrumental limit of detection to 21.395 mg m^{-3} reached in the BA03 campaign (supplementary material, Table S2). The salinity varies from 7.83 to 1.70 PSU. The maximum and minimum values were recorded in the BA03 campaign (Southern Baltic), while for the other campaigns the salinity value was always higher than 2.3 PSU (supplementary material, Table S2). The concentrations of Hex, But, Peri, and Zea contained several null values, which may later affect the model, as the computation was based on few measurements. It is important to point out that a dataset with such extreme variability (where key features have a CV greater than 100%) can make it difficult to build a stable model. This level of variability can lead to an ill-conditioned model, where small changes in the input data cause large, unpredictable changes in the output, reducing reliability and accuracy of the model.

We considered the different contributions to light absorption at 412 and 443 nm from detritus, CDOM and plankton and their relation with environmental variable: temperature and salinity (Fig. 3). All stations showed a relevant light absorption (higher than 60%) from the dissolved organic matter. The light absorption by phytoplankton, a_{ph} was less than 50% for most of the points: the highest relative absorption was found in the BA03 campaign in the southern Baltic Sea in April (supplementary material, Table S2). The largest contribution to the total light absorption was made by CDOM: 52 % ± 20 %. a_g is more dominant in the Bothnian Sea (BA06), where the ternary plot shows a strong clustering toward the CDOM-rich end, consistent with high terrestrial input from surrounding rivers, while the contribution of a_{NAP} was the lowest (less than 20% for all the stations) (supplementary material, Table S2). In contrast, the southern Baltic (BA01-BA03) exhibits a more balanced distribution among phytoplankton a_{ph}, a_g and detrital material a_{NAP}, reflecting a more mixed optical regime with a significant phytoplankton contribution. The Gulf of Finland and the Northern Baltic Proper (BA4, BA5) show higher variability, likely driven by episodic riverine inputs and hydrodynamic mixing events. The environmental conditions further support these findings. The inverse relationship between salinity and a_g confirms that CDOM originates primarily from terrestrial sources, with the highest concentrations found in the low-salinity waters of the Bothnian Sea and Gulf of

Finland while the Southern Baltic shows lower a_g values at higher salinity (~6-7 PSU), indicating more marine-influenced waters with less terrestrial CDOM input. Phytoplankton absorption a_{ph}, on the other hand, follows seasonal trends, with higher values occurring in cooler waters during the spring months (BA01, BA03), indicative of phytoplankton blooms. Late summer campaigns (BA02, BA05) show a shift toward higher CDOM absorption, likely due to increased riverine discharge and organic matter degradation. The spectral dependency of optical properties, evident in the shifts from 412 nm to 443 nm, suggests variability in phytoplankton composition and CDOM characteristics. The ternary plot patterns align with these trends, as campaigns with high a_{ph}, correspond to periods of active phytoplankton growth, while those with dominant a_g reflect enhanced CDOM inputs. Overall, the separation of data points across different campaigns underscores the diverse optical regimes within the Baltic Sea, shaped by riverine influence, seasonal biological productivity, and basin-specific hydrodynamic conditions.

Both methods, HCA and PCA analyses, were used to analyze the relationships between different pigments across various phytoplankton groups, highlighting distinct subgroups and their correlations (Fig. 4). The HCA analysis of the Baltic (BA) pigment dataset (HCA_{pig}) identified five distinct phytoplankton pigment groups. Marker pigments helped to identify specific phytoplankton groups. The orange cluster (haptophytes, nanoflagellates) included pigments such as Hex, TChl *c*3, Pras, and Diad, which are typically found in haptophytes and nano-sized phytoplankton. Pigments such as Allo, Caro, and Lut fell into the Green Cluster (green algae, dinophytes, cryptophytes) representing a mix of pigments found in green algae, dinophytes, and cryptophytes. The presence of peri, often specific to dinophytes, strengthened the grouping of this taxonomic class. The red cluster (micro/pico, diatoms, eukaryotic/cyanobacteria) consisted of the pigments Fuco, TChl *b*, and Zea, indicating different phytoplankton size classes, with Fuco representing diatoms, and Zea and TChl *b* representing picoeukaryotes. PCA on the normalized phytoplankton pigment concentrations highlighted spatiotemporal variations in the BA dataset. The four leading PCA modes toghether explained 67.4% of the variability. Mode 1 showed a strong positive correlation with dinoflagellates and cryptophytes, and a negative correlation with green algae and cyanobacteria. Mode 2 correlated positively with diatoms and haptophytes, while Mode 3 had strong correlations with green algae and nanofraction pigments. Mode 4 discriminated diatoms from other groups, especially in the Gulf of Finland. The PCA results recalled some aspects of the HCA, such as the dominance of diatoms in certain campaigns. The HCA_{pig} was shown here to see how well it matched the clustering of the absorption spectra ($HCA_{spectra}$).

Figure 2. Boxplot of pigment concentrations in the Baltic Sea on a logarithmic scale. Each box represents the distribution of pigment concentrations (mg m^{-3}) measured across sampling stations. The x-axis uses a log_{10} scale to highlight differences in magnitude among pigments. Outliers are displayed as individual points beyond the whiskers.

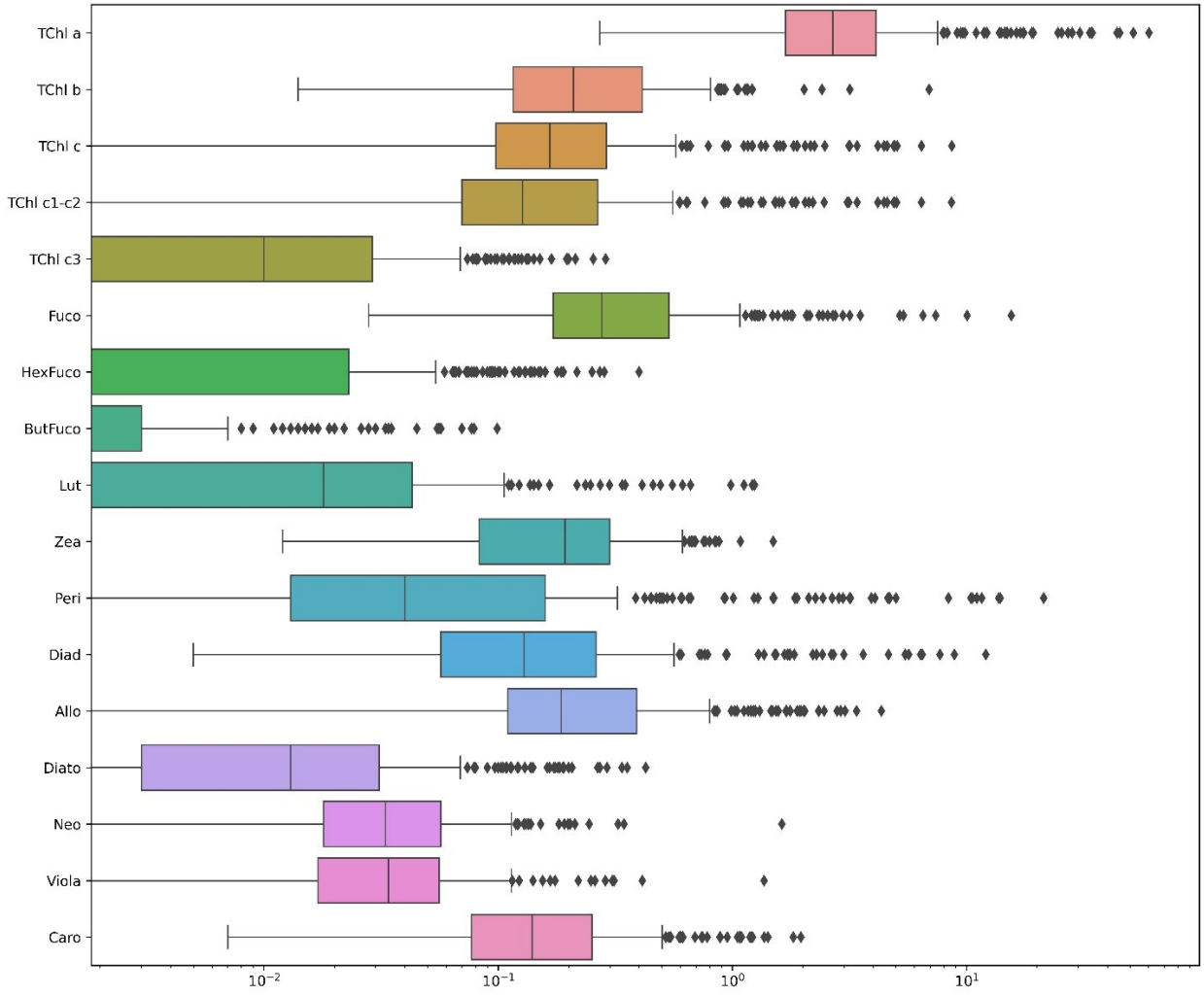


Figure 3. Multivariate analysis of absorption properties a_g, a_{ph} and a_{NAP} at 412 nm and 443 nm and in relation to environmental parameters (temperature and salinity). Top row: Ternary plots showing the relative contributions of different absorption coefficients. Middle and bottom rows: Scatter plots illustrating the variations in absorption coefficients as functions of temperature and salinity across different oceanographic campaigns (BA1–BA6).

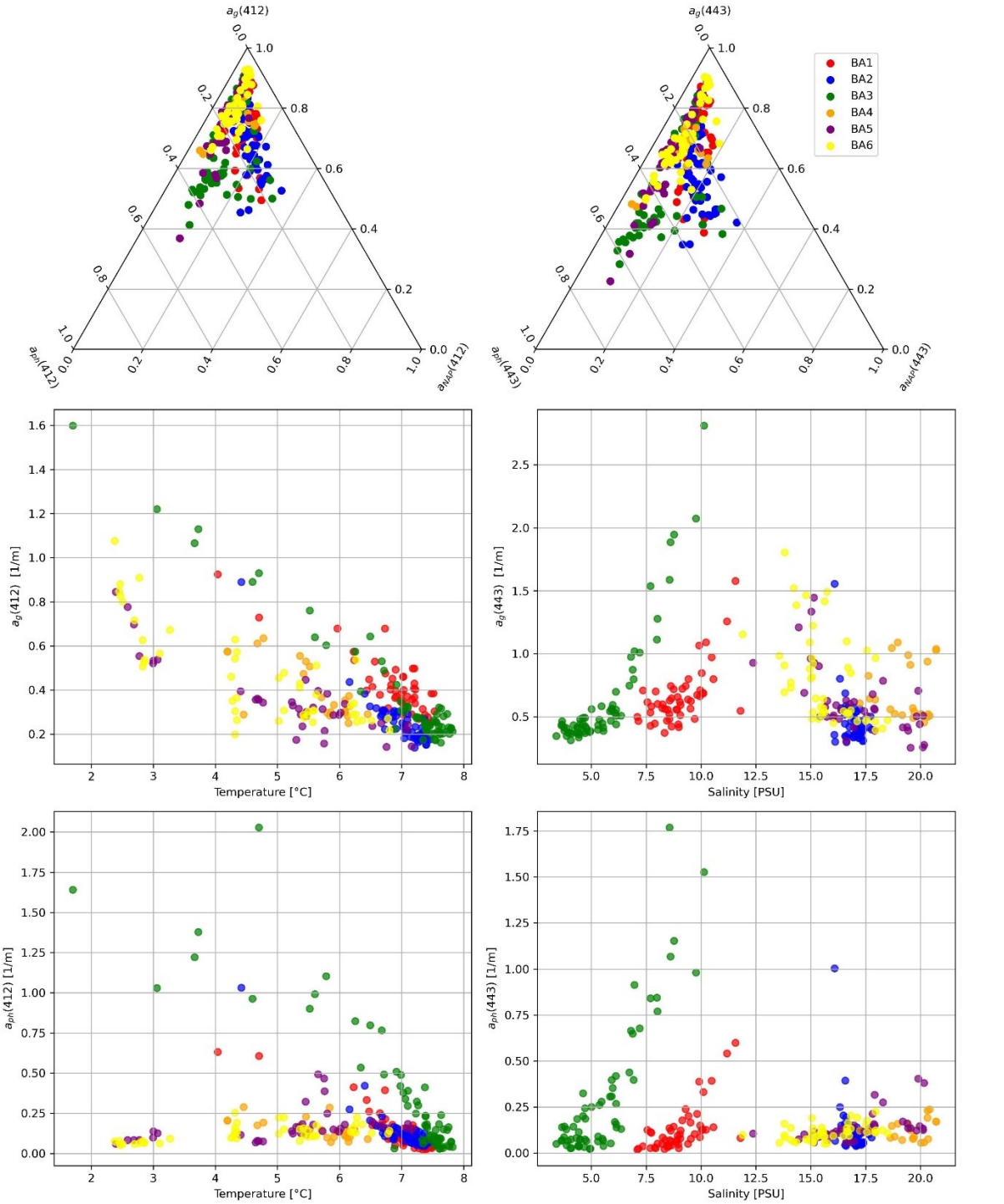


Table 2. Summary statistics for the entire data set: the pigments and the pigments sum concentrations and the spectral at specific absorption ($a_g(443)$, $a_{ph}(443)$), coefficient of variation (CV%), mean and range (max-min). The number of stations where the compound was below the limit of detection (LOD) is given in the last column.

	Average	CV%	max	min	No. stations under LOD
TChl *a* [mg m^{-3}]	5.107	159.0	60.222	0.272	-
TChl *b* [mg m^{-3}]	0.347	154.2	6.932	0.014	-
TChl *c* [mg m^{-3}]	0.469	218.3	8.67	LOD	2
Fuco [mg m^{-3}]	0.640	217.5	15.544	0.028	-
Zea [mg m^{-3}]	0.234	86.4	1.495	0.012	-
Peri [mg m^{-3}]	0.676	345.2	21.395	LOD	48
Allo [mg m^{-3}]	0.415	148.4	4.332	LOD	1
Diato [mg m^{-3}]	0.034	178.7	0.427	LOD	62
Hex [mg m^{-3}]	0.026	209.3	0.4	LOD	162
But [mg m^{-3}]	0.005	270.1	0.099	LOD	192
PPC [mg m^{-3}]	1.419	151.2	17.568	0.07	
PSC [mg m^{-3}]	1.346	245.0	26.812	0.037	
PSP [mg m^{-3}]	7.269	172.9	96.918	0.343	
DP [mg m^{-3}]	2.342	168.3	31.113	0.128	
$a_g(443)$ [m^{-1}]	0.370	55.1	1.599	0.138	
$a_{ph}(443)$ [m^{-1}]	0.191	128.8	2.028	0.023	
SPM [g m^{-3}]	1.524	106.9	12.875	0.32	
SST [°C]	12.8	40.8	20.7	3.4	
Salinity [PSU]	6.18	22.82	7.83	1.70	

Figure 4. HCA_{pig} dendrogram (left) and PCA biplots (right) of phytoplankton pigment ratios to TChl *a* for the Baltic dataset. In the HCA_{pig} (left figure), the major pigment communities (micro- , nano- and pico-phytoplankton) were identified based on a linkage distance cutoff of 1.0 (red dashed line). The proposed phytoplankton cell size classes for each group are shown in brackets. In the right figure, the loadings corresponding to the principal component (PC) modes are shown in panels (a), (b), (c), and (d) for the Baltic dataset. The pigment order was the same as in the HCA analysis, to facilitate comparison. The mode number was shown above each plot, together with the percentage of variance explained by that mode. The loadings were colour coded according to the main taxonomic groups: blue for cyanobacteria-pico, red for diatoms-micro, orange for green algae-nano, green for haptophytes-dinoflagellates-nano, and orange for euglanophytes-nano.

b. Spectral decomposition of the phytoplankton absorbance

The spectral decomposition approach gave a reliable representation of the pigment composition (Chase et al., 2013, Hoepffner et al., 1991, 1993, Ficek et al., 2004), however, it presented the limit of omitting some DPA pigments without considering carotenoids and xanthophylls individually. The omitted pigments included Peri, Allo and Zea that were the DPA of a representative phytoplankton population commonly found in the Baltic ecosystem. In our dataset, the Gaussian decomposition of a_{ph} at the selected wavelengths (Table 3) produced residuals within ± 0.001 m^{-1} for all stations (Fig. 5), confirming its robustness for most pigments. The pigments were correlated with different a_{gauss} as in Chase et al. (2013). The linear regression emphasized the relationship between pigment concentrations and their absorption characteristics for the selected wavelengths (Fig. 6): strong linear fit indicated that Gaussian decomposition estimated well the pigment.

The Table 4 reported Spearman's rank correlation coefficients (r^2_s) and correlation coefficients R^2 along with the coefficients A and B determined by the equation [4] described in section 2.5.2. The coefficients A_{ij} and B_{ij} describe the relation between the j-th pigment, pig_j , derived from the absorption spectra and the $HPLC_j$ corresponding pigments, providing insight into the effectiveness of our model in predicting pigment concentrations. The Gaussian decomposition at wavelengths 437 nm and 620 nm, show high correlation with TChl *a* obtained by HPLC: we obtained a r^2_s of 0.90 and an R^2 values of 0.92 between for the a_{gauss} at 437 nm and TChl *a* and we had r^2_s of 0.84 and R^2 of 0.88 for a_{gauss} 620 nm and TChl *a*. These correlations indicated a strong relationship between HPLC concentrations and optical measurements at these wavelengths. At 678 nm, the correlation of TChl *a* and a_{gauss} was slightly lower but still substantial, with r^2_s of 0.80 and R^2 of 0.82. For TChl *b*, we considered the a_{gauss} at 467 and 557 nm: the correlation was weaker with respect what we had with TChl a. TChl *b* exhibited an r^2_s values of 0.60 at 467 nm and 0.77 at 557 nm, and R^2 values of 0.65 and 0.84, respectively. At 557 nm The sum of TChl *b* and TChl *c* showed a better correlation with the a_{gauss} than TChl *b* alone (Table 4). TChl *c* had r^2_s values of 0.64 (R^2 = 0.64) at a_{gauss} 585 nm and of r^2_s of 0.48 (R^2 = 0.44), at a_{gauss} 639 nm, showing moderate correlations. For the pigments PPC and PSC, the correlations were relatively strong with r^2_s values of 0.81 and 0.64 and R^2 values of 0.86 and 0.75, respectively. The coefficients A and B vary, indicating different degrees of sensitivity and offset in the correlation. Regarding the aggregated pigments, the photoprotective carotenoids (PPC) correlated better (0.81) than the photosynthetic carotenoids (PSC, 0.64), i.e. in the range of TChl *b*. It should be noticed that the PSC included the fuco, a pigment that was ubiquitous present in the dataset.

Figure 5. Observed absorption spectra (in blue), best Gaussian fitting (in grey) and the fitted curves (in orange dashed lines) for the BA01_12 station, and (bottom plot) the residual between observed and fitted spectra (in red).

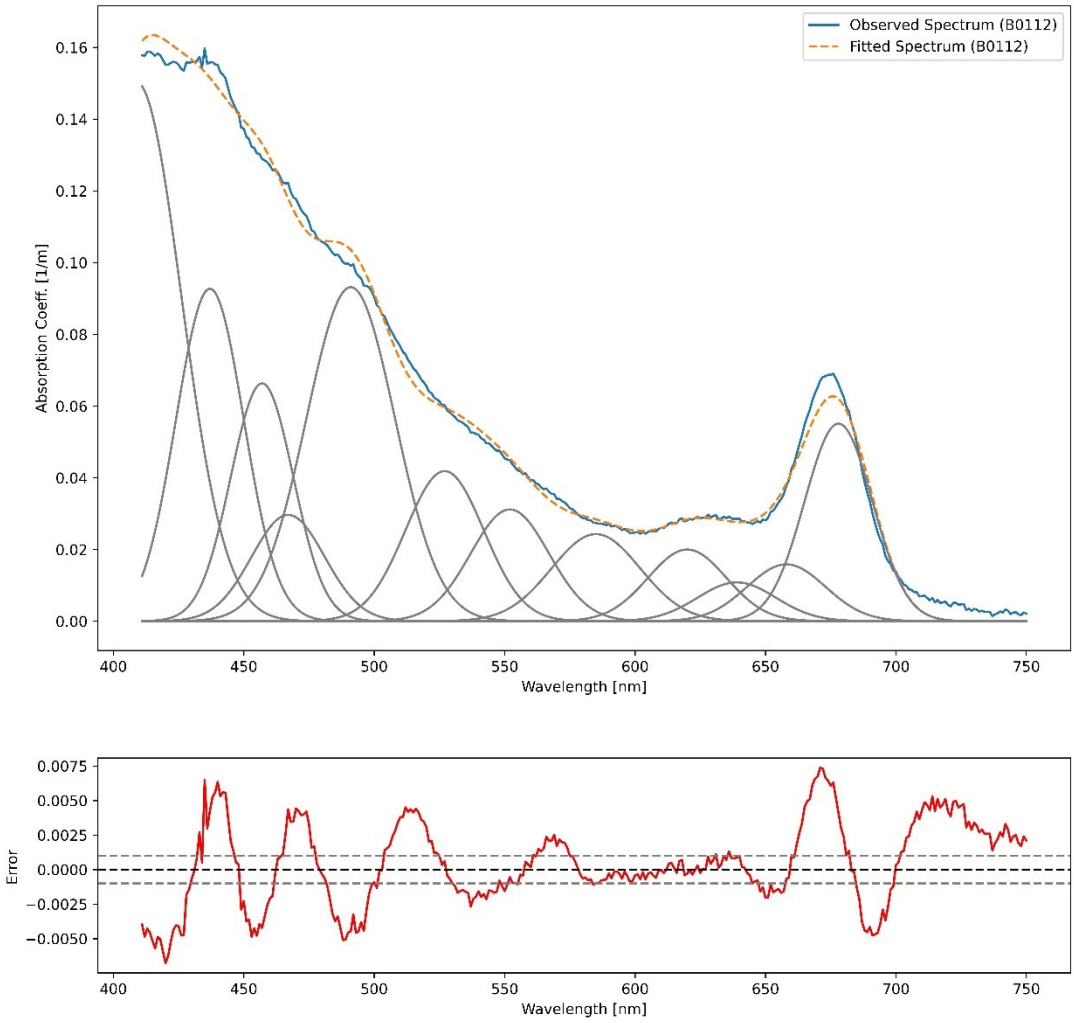


Figure 6. HPLC-measured chlorophyll, photoprotective carotenoids (PPC) and photosynthetic carotenoids (PSC) concentrations (on the y-axis) compared to the magnitudes of Gaussian peak absorption (on the x-axis) at eight distinct wavelengths. The continuous line represents the best-fit linear regression on the logarithmic transformations of both datasets.

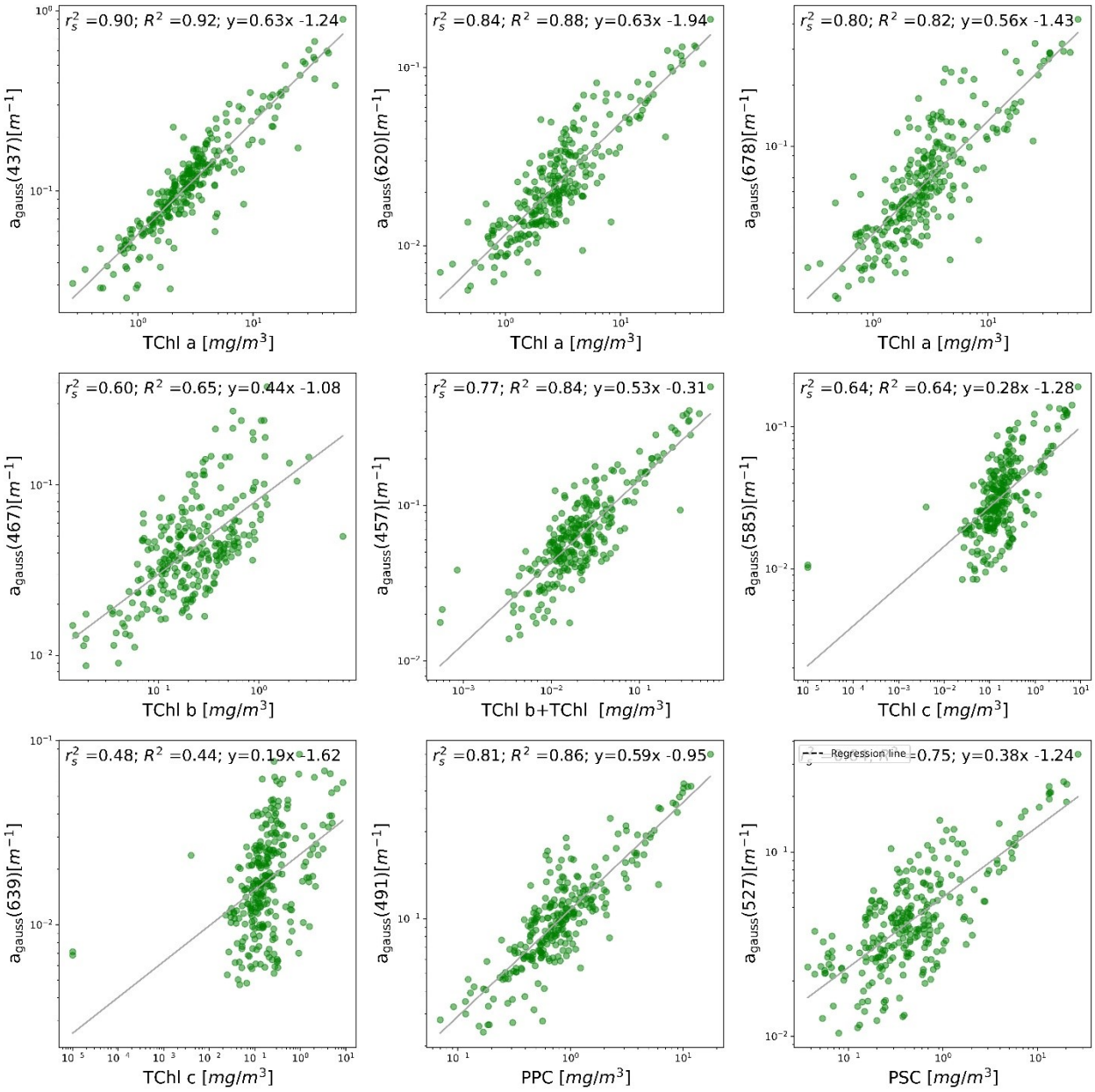


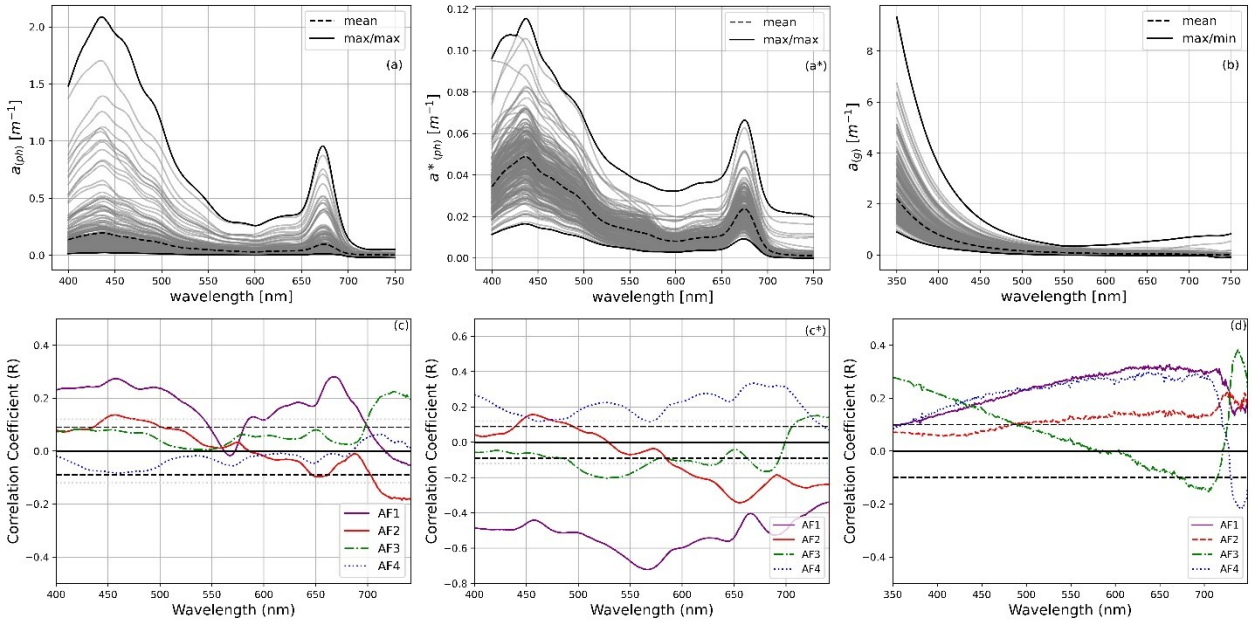
Table 4. Correlations between HPLC pigment concentrations at ten different a_{gauss} wavelenghts. Correlation values are Spearman's rank correlation coefficient (r^2_s) and correlation coefficient (R^2). A and B are coefficients determined by eq. [4], section 2.5.2.

Pigment [mg m^{-3}]	Wavelength (λ)	r^2_s	R^2	A	B
TChl a	437	0.90	0.92	0.63	-1.24
TChl a	620	0.84	0.88	0.63	-1.94
TChl a	678	0.80	0.82	0.56	-1.43
TChl b	467	0.60	0.65	0.44	-1.08
0.03*TChl b + 0.07*TChl c	557	0.77	0.84	0.53	-0.31
TChl c	585	0.64	0.64	0.28	-1.28
TChl c	639	0.48	0.44	0.19	-1.62
PPC	491	0.81	0.86	0.59	-0.95
PSC	527	0.64	0.75	0.38	-1.24

c. Correlation of phytoplankton pigments with CDOM and absorption spectra

Correlation analyses were carried out to determine the associations between the data series of each inherent optical property (IOP) and the first four amplitude functions (AFs) representing various phytoplankton communities, as determined by PCA analysis of the HPLC pigment datasets. The AF values were used as proxies for these communities. The inclusion of different optical properties in our analysis provided a more comprehensive understanding of the relationships between pigments and optical properties. Correlations with CDOM absorption values were computed for the first four modes (Fig. 7c). Modes 1 and 4, associated with dinoflagellates and diatoms, showed positive associations with $a_g(\lambda)$. Conversely, the other modes did not show significant correlations with $a_g(\lambda)$. The $a_{ph}(\lambda)$ correlation with the PCA results was significant in the case of the first mode, i.e. the mode associated with predominantly dinoflagellates. The AFs included in the analysis are derived from the analysis of the pigment to TChl *a* ratio.

Figure 7. Average absorption properties and their correlation with the first four modes derived from the PCA analysis for **(a, c)** phytoplankton absorption, $a_{ph}(\lambda)$; **(a', c')** phytoplankton absorption ratio to TChl *a*, $a^*_{ph}(\lambda)$; **(b, d)** CDOM absorption, $a_g(\lambda)$. In **c**, **c'** and **d** the 95% confidence interval (CI) is shown by the black dashed line.



Further insights were gained by analysing $a_{ph}'(\lambda)$ and $a_{ph}''(\lambda)$, and the distinct spectral relationships between the concentrations of selected biomarker pigments and values within specific wavelength intervals (Fig. 8). The same exercise was repeated for the $a^*_{ph}'(\lambda)$ and $a^*_{ph}''(\lambda)$ (Fig. 10). The pigments were grouped into the following categories based on the outcomes of the HCA_{pig} analysis: (a, e) dinoflagellates, (b, f) chlorophytes, (e, f) haptophytes, and (d, h) diatoms and picoplankton. It is noteworthy that Lut and Diato were omitted – for clarity in the representation- , although they showed analogous patterns to the chlorophytes (b, f) and haptophytes pigments, respectively.

Figure 8. Correlations of selected phytoplankton pigments with $a_{ph}'(\lambda)$ and $a_{ph}''(\lambda)$ grouped as the results of the HCA_{pig} analysis ((a, b, c, d) for $a_{ph}'(\lambda)$ and (e, f, g, h) for $a_{ph}''(\lambda)$). The dashed and dotted black lines indicate the magnitude of the significant correlation coefficients at 95% and 99% confidence, respectively.

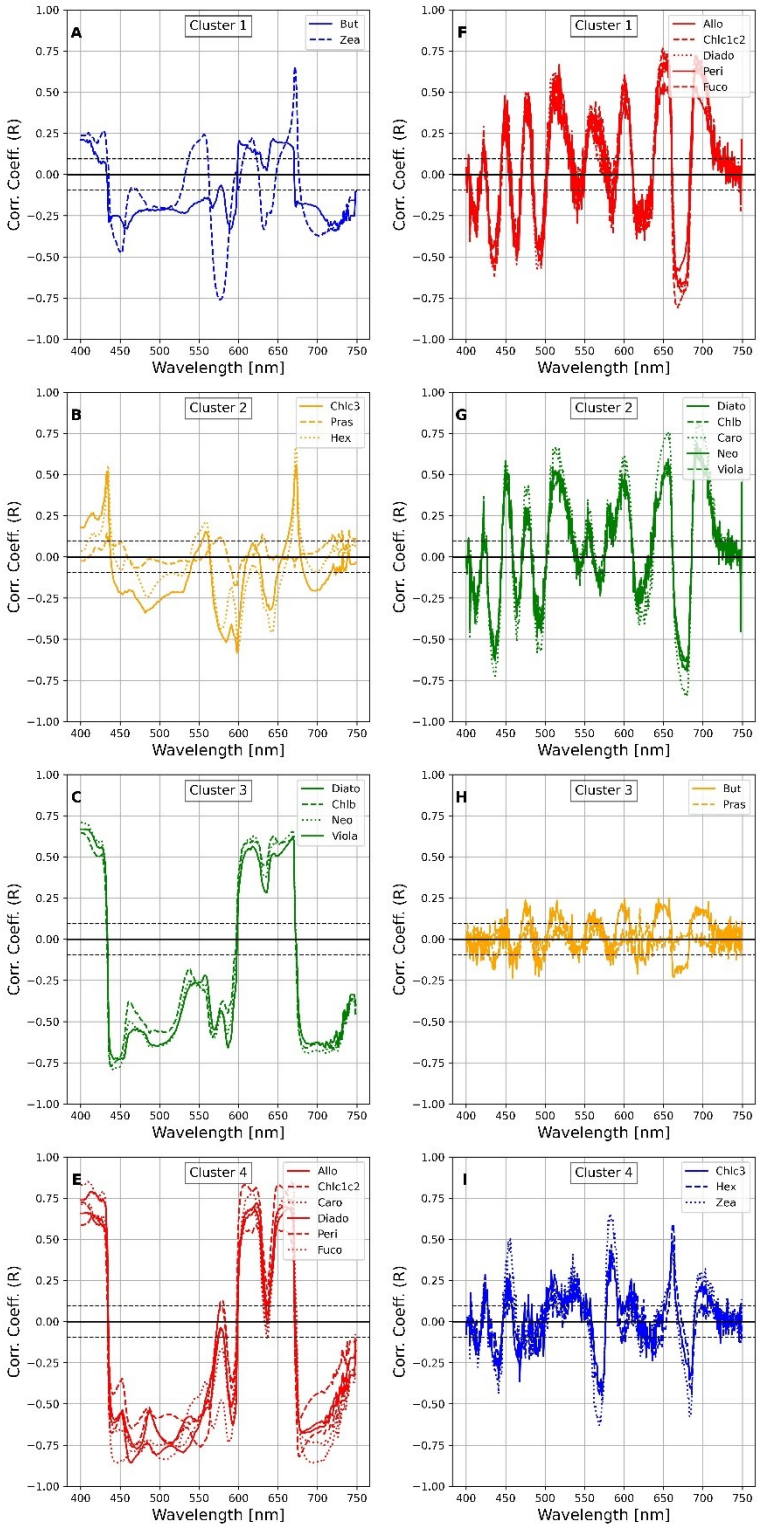
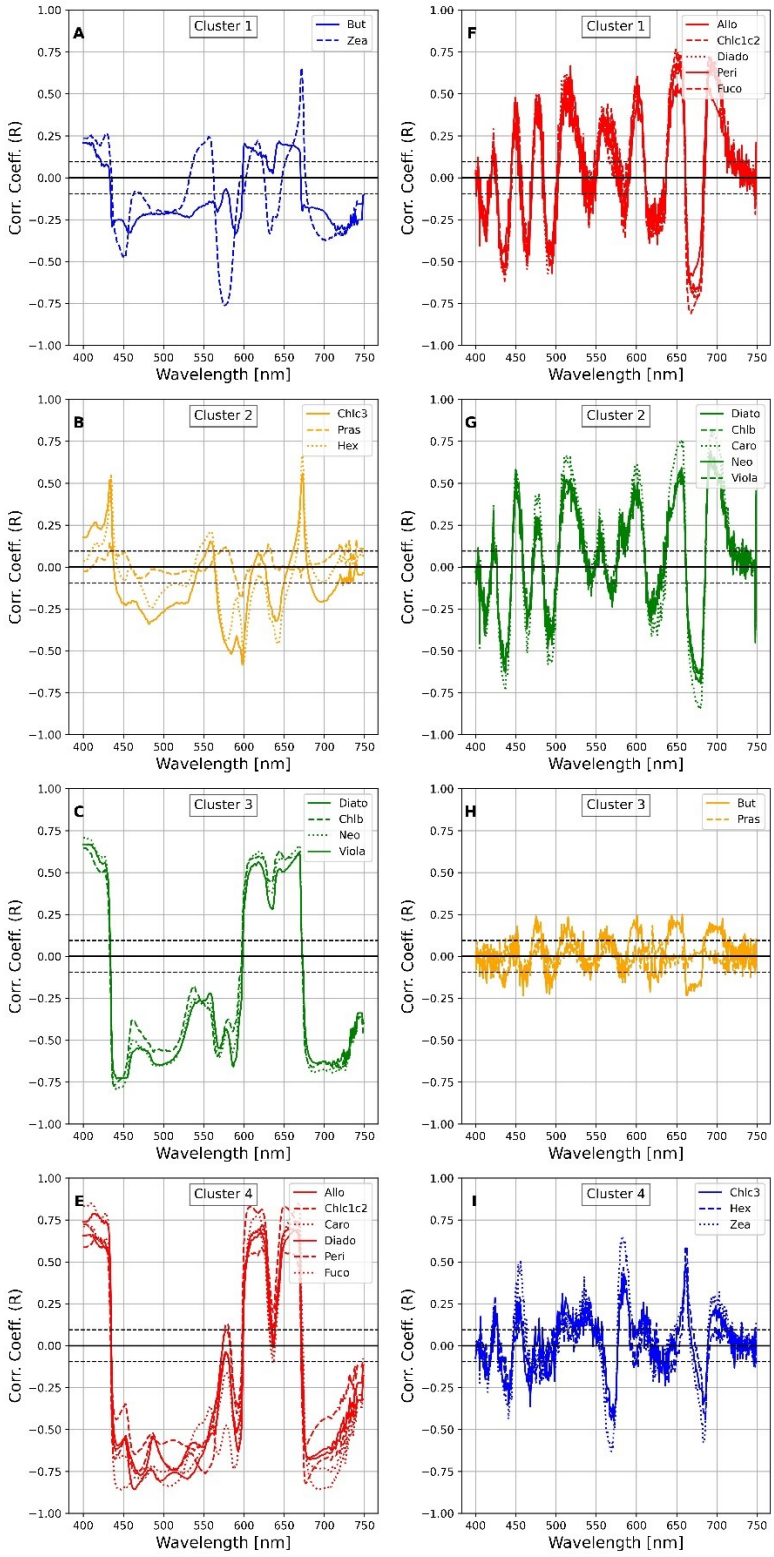


Figure 9. Correlations of selected phytoplankton pigments with $a_{ph}'(\lambda)$ and $a_{ph}''(\lambda)$ grouped as the results of the $HCA_{spectra}$ ((A, B, C, D) for $a_{ph}'(\lambda)$ and (E, F, G, H) for $a_{ph}''(\lambda)$). The dashed and dotted black lines show the magnitude of the significant correlation coefficients at 95% and 99% confidence, respectively.



The HCA spectra clustering analysis was conducted to explore the relationships between pigment concentrations and spectral absorption features across different wavelength intervals. Using Ward's HCA and Euclidean distance, the clustering was applied to the first and second derivatives of the phytoplankton absorption coefficients ($a_{ph}'(\lambda)$ and $a_{ph}''(\lambda)$) collectively referred to as $HCA_{spectra}$ (Supplementary material, Fig. S1). This approach allowed for the identification of distinct clusters that represent groups of pigments based on their spectral behavior. To maintain consistency with the HCA_{pig} results, a distance threshold of 5 was used, producing four clusters for both $a_{ph}'(\lambda)$ and $a_{ph}''(\lambda)$. The $HCA_{spectra}$ resulting clusters showed that Fuco grouped with Allo, Peri, and Caro, despite Fuco forming a distinct cluster in the pigment-based HCA_{pig}, analysis. The overall clustering pattern highlighted consistent pigment distributions, particularly in the largest cluster, between the first and second derivatives. The number of clusters obtained for $HCA_{spectra}$ was four for both $a_{ph}'(\lambda)$ and $a_{ph}''(\lambda)$, and the pigment distribution among the biggest cluster was consistent between $a_{ph}'(\lambda)$ and $a_{ph}''(\lambda)$. In Fig. 10 we reported the correlation between the pigments and the spectra following the clusters obtained from the $HCA_{spectra}$ analysis for $a_{ph}'(\lambda)$ and $a_{ph}''(\lambda)$. We repeated the exercise for the $a^*_{ph}(\lambda)$ and in Fig. 11 we reported the correlation between the pigments and the spectra following the clusters obtained from the $HCA_{spectra}$ analysis for $a^*_{ph}'(\lambda)$ and $^*a_{ph}''(\lambda)$: the number of clusters obtained in these case were five for the first derivative, and three for the second derivative.

Figure 10. Correlations of selected phytoplankton pigments with $a^*_{ph}'(\lambda)$ and $a^*_{ph}''(\lambda)$ group as the results of the HCA_{pig} analysis [(a, b, c, d) for $a^*_{ph}'(\lambda)$ and (e, f, g, h) $a^*_{ph}''(\lambda)$]. The dashed and dotted black lines show the magnitude of significant correlation coefficients at 95% and 99% confidence, respectively.

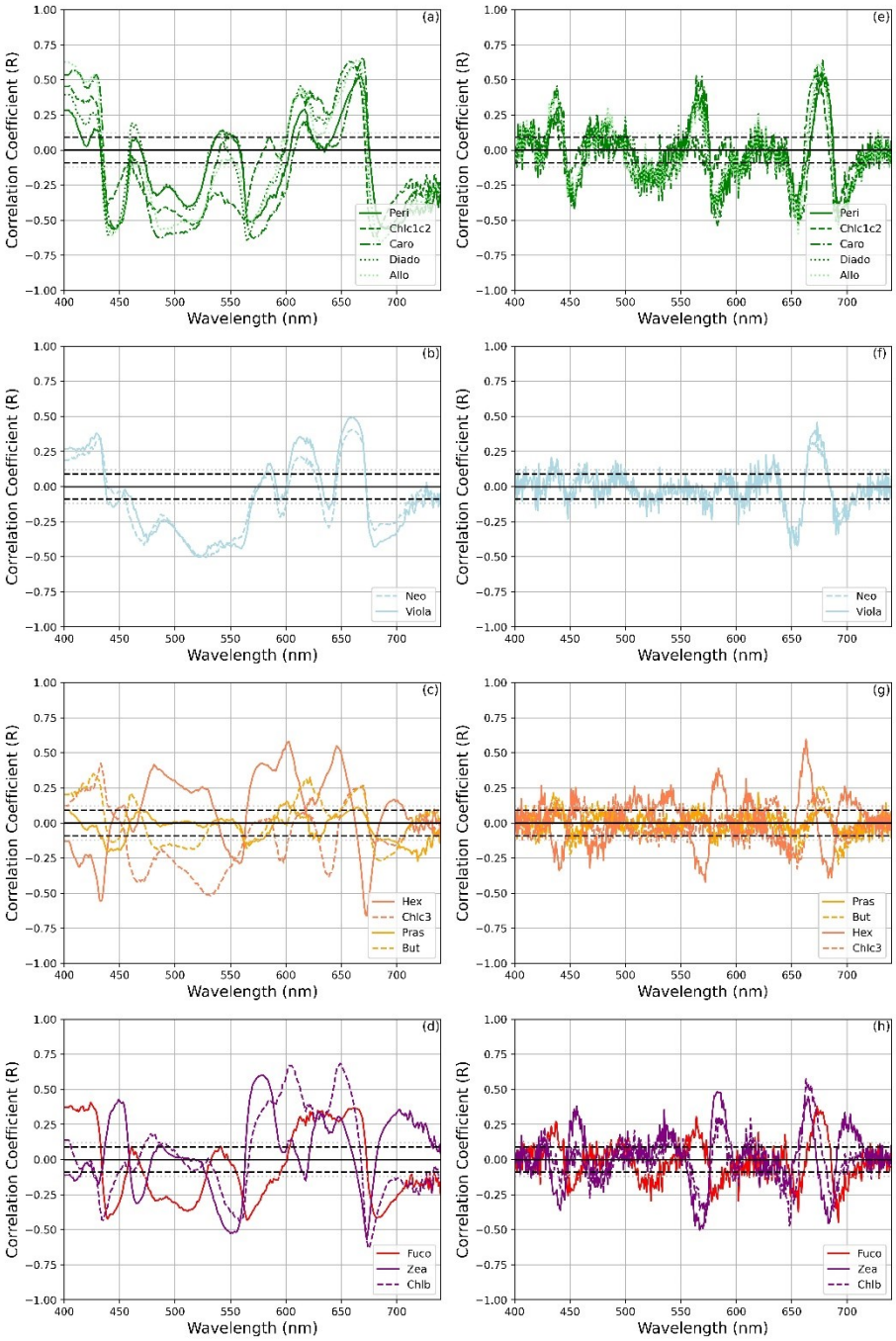
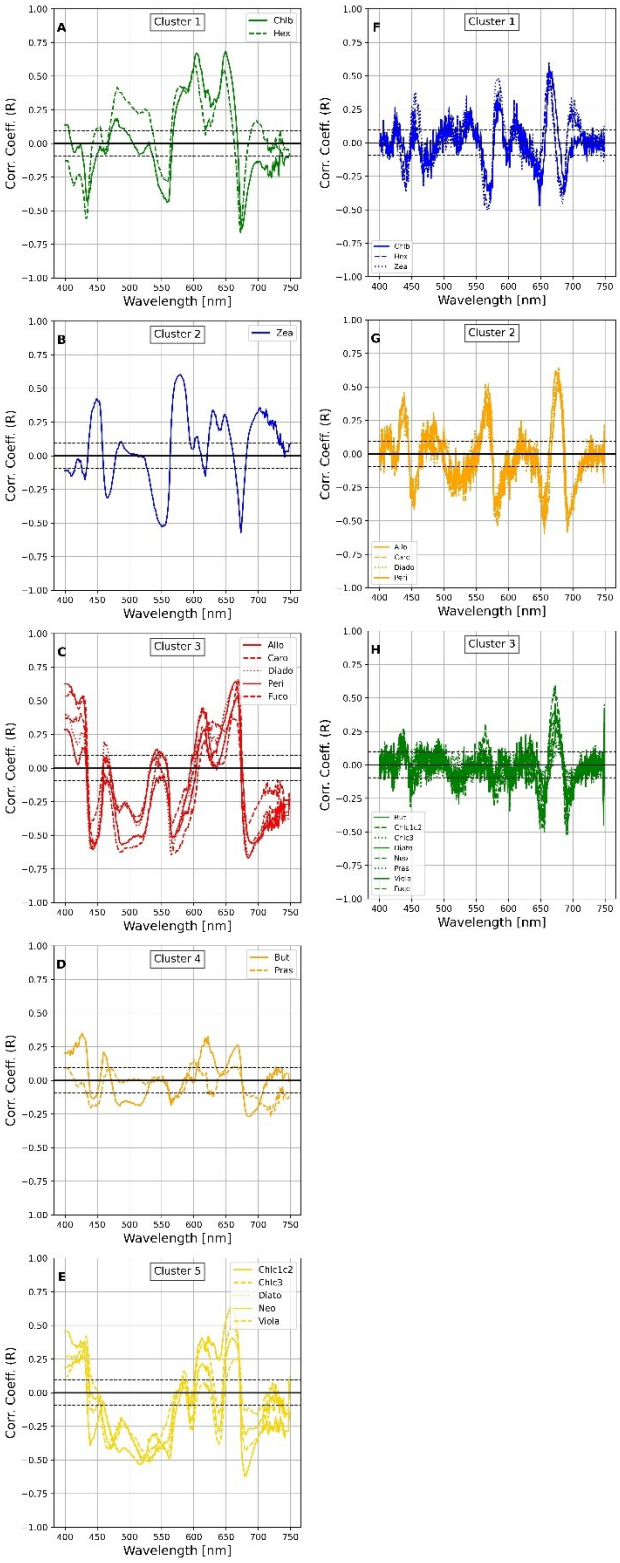


Figure 11. Correlations of selected phytoplankton pigments with $a^*_{ph}{}'(\lambda)$ and $a^*_{ph}{}''(\lambda)$ group as the results of the $HCA_{spectra}$ [(A, B, C, D, E) for $a^*_{ph}{}'(\lambda)$)and (F, G, H) for $a^*_{ph}{}''(\lambda)$]. The dashed and dotted black lines show the magnitude of significant correlation coefficients at 95% and 99% confidence, respectively.



d. Proposed EOF-model to assess pigment concentrations from optical measurements

The purpose of this study was to develop a model for phytoplankton pigment concentrations based on spectra derived from the absorption coefficient, with the aim of assessing pigment levels by optical measurements using characteristic correlation patterns between pigments and $a_{ph}(\lambda)$. We tested $a_{ph}(\lambda)$, $a'_{ph}(\lambda)$, $a^*_{ph}(\lambda)$ and $a^*_{ph}{}'(\lambda)$ versus pigment concentrations and log-transformed concentrations. The pigments included in the analysis were all the DPA: Peri, Allo, Fuco, Zea, TChl *b,* TChl *c1c2* and Hex. These pigments were the bio-markers that were associated with specific phytoplankton groups and commonly used in phytoplankton pigments analysis. The performance metrics for MLR models predicting pigment concentrations using PC regression with the first 100 PCs were shown (Table 5). The optical metrics evaluated included $a_{ph}(\lambda)$, $a_{ph}'(\lambda)$ and $a^*_{ph}{}'(\lambda)$, with metrics comprising the coefficient of determination (R^2), RMSD, and RMSD coefficient of variation (RMSD CV%) from cross-validation with 500 permutations. Fuco models performed differently for different metrics. The $a_{ph}(\lambda)$ model achieved an R^2 of 0.645 with an RMSD mean of 0.3 and an RMSD CV% of 11, indicating moderate accuracy and consistent performance. Improved results were seen with the $a_{ph}'(\lambda)$ model, which had an R^2 of 0.739 with similar RMSD mean and CV% values. However, the $a^*_{ph}{}'(\lambda)$ model showed slightly lower performance with an R^2 of 0.47 and the highest RMSD CV%. For Zea, the $a_{ph}(\lambda)$)model showed an R^2 of 0.391, an RMSD mean of 1.1, and an RMSD CV% of 7.4, reflecting lower accuracy and variability. The $a_{ph}'(\lambda)$ model offered a modest improvement with an R^2 of 0.45, and similar RMSD mean and CV% values. The $a^*_{ph}{}'(\lambda)$ model achieved an R^2 of 0.45, showing modest performance. For Allo, the $a_{ph}(\lambda)$ model had an R^2 of 0.541 with an RMSD mean of 0.4 and a higher RMSD CV% of 25. The $a_{ph}'(\lambda)$ model improved to an R^2 of 0.56 but also had a higher RMSD CV%. In contrast, the $a^*_{ph}{}'(\lambda)$ model had a lower R^2 of 0.38, an RMSD mean of 0.23, and a significant RMSD CV% of 43.15, indicating considerable variability. The TChl *b* models showed good performance. The $a_{ph}(\lambda)$ model had an R^2 of 0.716, with an RMSD mean of 0.3 and an RMSD CV% of 20.5. The $a_{ph}'(\lambda)$ model showed similar efficacy with an R^2 of 0.714. The $a^*_{ph}{}'(\lambda)$ model performed best with an R^2 of 0.79, the lowest RMSD mean, and CV%, indicating superior predictive accuracy. For Hex, the $a_{ph}(\lambda)$ model achieved an R^2 of 0.624 with an RMSD mean of 0.9 and an RMSD CV% of 9.8. The $a_{ph}'(\lambda)$ model had a slightly lower R^2 of 0.603 with similar RMSD mean and CV% values. The $a^*_{ph}{}'(\lambda)$ model showed improved performance with an R^2 of 0.67, better RMSD mean, and CV%, reflecting improved accuracy. The Peri models varied, with $a_{ph}(\lambda)$ yielding an R^2 of 0.44, an RMSD mean of 1, and an RMSD CV% of 8.3. The $a_{ph}'(\lambda)$ model had similar R^2 and RMSD mean, with a slightly higher RMSD CV%. The $a^*_{ph}{}'(\lambda)$ model had an R^2 of 0.47, a lower RMSD mean but a higher CV%, indicating variable performance. The TChl *c1c2* models were particularly effective. The $a_{ph}(\lambda)$ model had an R^2 of 0.58, an RMSD mean of 0.4, and a higher RMSD CV% of 31.5. The $a_{ph}'(\lambda)$ model performed

similarly performance with a slightly better mean RMSD. The $a^*_{ph}'(\lambda)$ model stood out with an R^2 of 0.87, the lowest RMSD mean, and CV%, reflecting superior predictive ability. Finally, the $a_{ph}(\lambda)$ model for But showed lower performance with an R^2 of 0.36, an RMSD mean of 0.8, and an RMSD CV% of 8.1. The $a_{ph}'(\lambda)$ model had comparable R^2 and RMSD mean, with a slight increase in RMSD CV%. The $a^*_{ph}'(\lambda)$ model was less effective, with an R^2 of 0.27, a higher RMSD mean, and a low RMSD CV%. Overall, the models showed variable performance for different pigments and optical metrics. The highest accuracy and lowest variability were observed for TChl *c1+c2* using $a^*_{ph}'(\lambda)$, while pigments such as Zea and But showed lower prediction accuracy. These findings highlight the importance of selecting appropriate principal components and optical metrics to optimize model performance in predicting pigment concentrations.

In the Baltic Sea, the chlorophyll *c* is not representative of any phytoplankton community, while Allo, Fuco, Zea and TChl *b* are. Based on this consideration, the most promising outcomes were achieved when using $a_{ph}'(\lambda)$ and using log-transformed pigment concentration as the target (see Fig. 12). It is worth noting that the model showed better performance in the southern Baltic Sea, as demonstrated during validation with the isolation of a specific campaign (data not shown here). The model derived from $a_{ph}'(\lambda)$ performed better than from $a^*_{ph}'(\lambda)$ both for the Fuco (0.`74 against 0.47) and the Allo (0.56 against 0.38) and closely for Zea (0.45 in both models) and TChl b (0.71 and 0.79), while the model with $a^*_{ph}'(\lambda)$ had a better regression for the TChl c1c2 (0.87 vs 0.59 for the $a_{ph}'(\lambda)$ model).

While the Gaussian decomposition detected TChl *b* absorption primarily at 658 nm and 457 nm, our model identified additional significant correlations for TChl *b* at 440 nm, 600 nm, and 660 nm. These results suggest that our approach, using $a_{ph}'(\lambda)$ and its first derivatives, captures a broader spectral representation of TChl *b* absorption. Similarly, for TChl *c*, Gaussian decomposition highlighted peaks at 415 nm, 430 nm, 480 nm, and 665 nm, but our model found important correlations at additional wavelengths, particularly around 575-590 nm and 600 nm, broadening the understanding of TChl *c*'s spectral absorption.

Figure 12. Scatterplots of original data against the model in logarithmic scales for the MLR based on $a_{ph}'(\lambda)$ PCs for diagnostic pigments. The equations of regression lines, along with R-squared values, were shown.

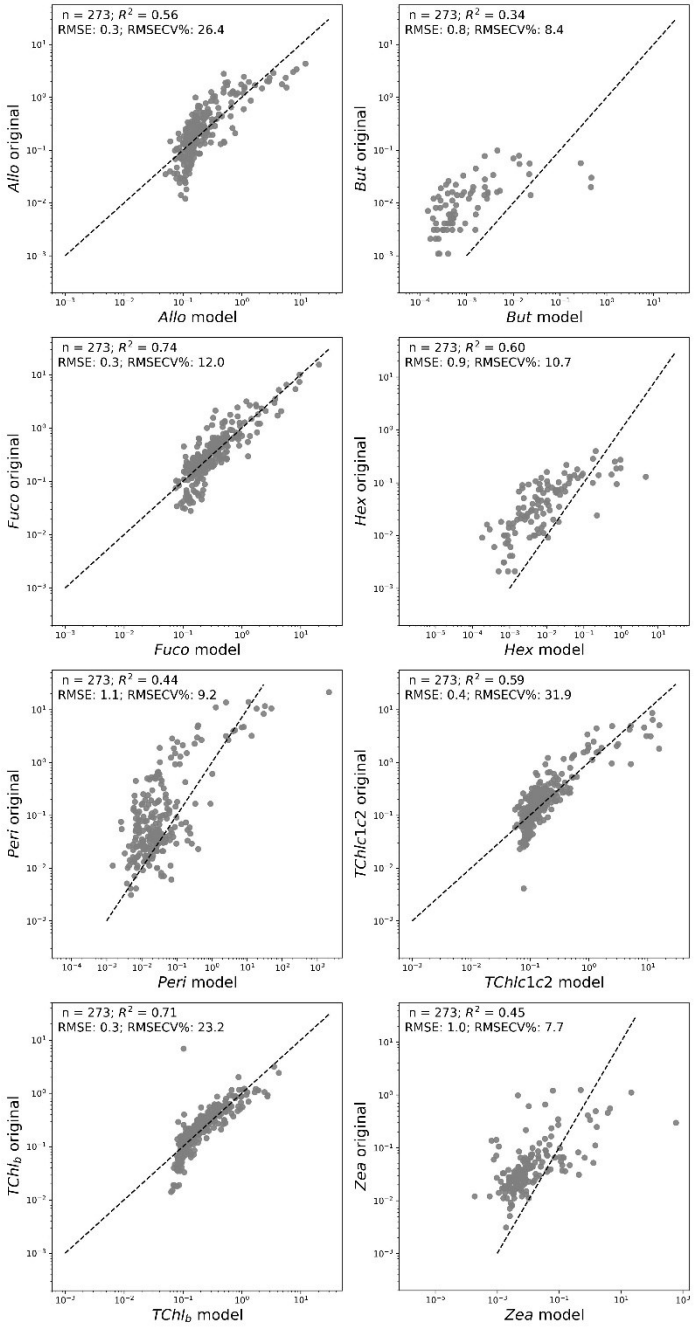


Table 5. Diagnostic pigments MLR model for PCs regression of $a_{ph}(\lambda)$, $a_{ph}'(\lambda)$ and $a^*_{ph}'(\lambda)$. For all the models were considered the first 100 PCs: coefficient of determination (R^2), root mean square difference (RMSD mean) and RMSD coefficient of variation from cross-validation (500 permutations) were shown.

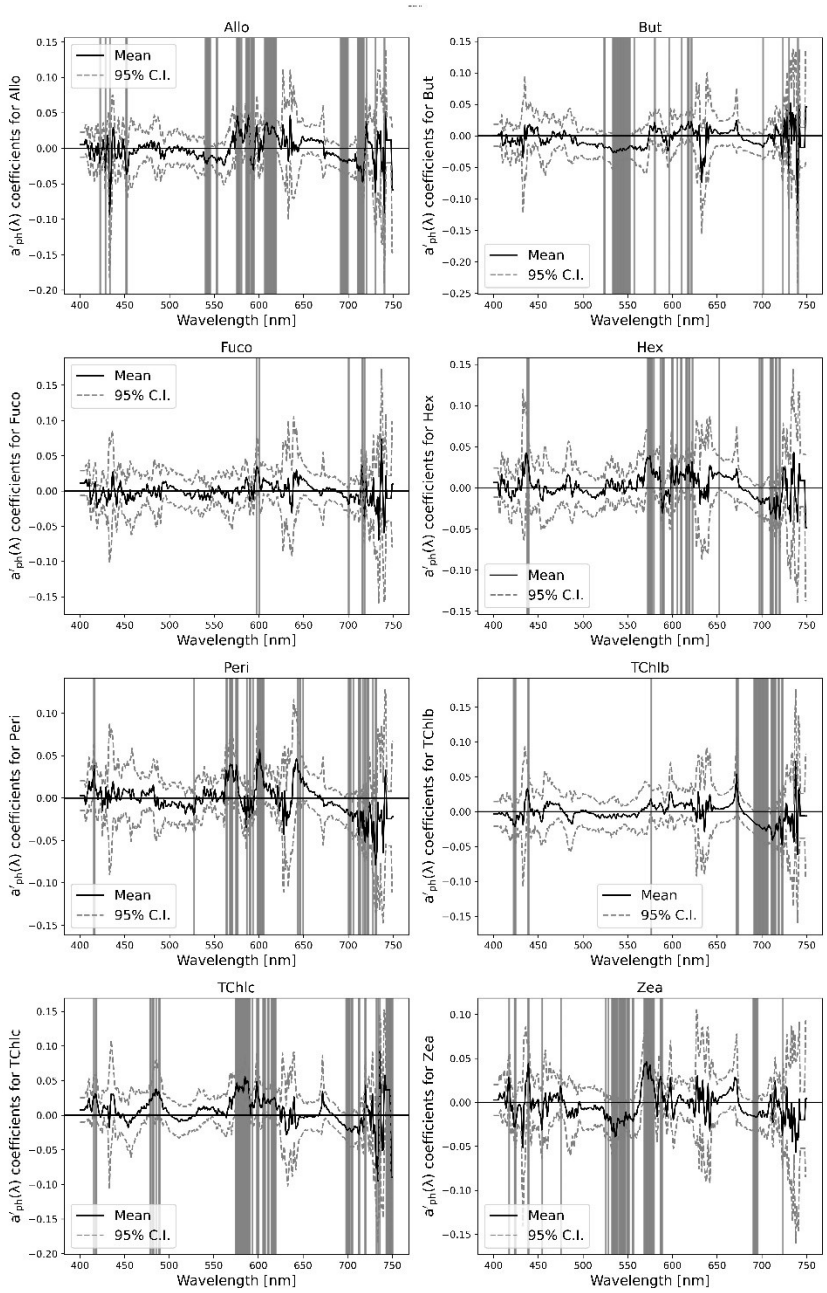
	$a_{ph}(\lambda)$			$a_{ph}'(\lambda)$			$a^*_{ph}'(\lambda)$		
	R^2	RMSD mean	RMSD (CV%)	R^2	RMSD mean	RMSD (CV%)	R^2	RMSD mean	RMSD (CV%)
Fuco	0.645	0.3	11	0.739	0.3	11.4	0.47	0.2	11.51
Zea	0.391	1.1	7.4	0.455	1	7.4	0.45	0.78	7.45
Allo	0.541	0.4	25	0.56	0.3	26.8	0.38	0.23	43.15
TChl b	0.716	0.3	20.5	0.714	0.3	21.6	0.79	0.2	12.8
Hex	0.624	0.9	9.8	0.603	0.9	11.2	0.67	0.6	9.6
Peri	0.44	1	8.3	0.442	1.1	9.3	0.47	0.79	10.2
TChl c1c2	0.58	0.4	31.5	0.592	0.4	31.9	0.87	0.1	32.1
But	0.36	0.8	8.1	0.361	0.8	8.4	0.27	0.6	0.7

A variability for Allo demonstrated correlation coefficients across the visible spectrum (Fig. 13). Significant correlations were seen in the 450-500 nm and 600-650 nm wavelength ranges, as indicated by the narrowing of the 95% confidence intervals (CI). These regions likely corresponded to absorption peaks associated with the the carotenoid structure of Allo suggesting that these wavelengths are essential for accurate prediction Allo concentrations. The narrow confidence intervals in these areas reinforced the model's consistency in capturing the relevant spectral signals. In the case of But, correlation coefficients hovered close to zero across most of the spectrum, with minor dips in the 450-550 nm range. The wide C.I. reflected significant uncertainty in these predictions, indicating that the absorption features of But were either weakly represented in the $a_{ph}'(\lambda)$ -based model or confounded by signals from other pigments. This result showed the poor performance for But (see Fig. 12), with an R^2 of 0.34.

For Fuco, stronger correlations emerged, especially in the 500-550 nm and 620-680 nm ranges, which matched the known absorption features of Fuco. The tight C.I. around these wavelengths suggested high model confidence in predicting Fuco concentrations. The strong response at around 540 nm, a significant Fuco absorption peak, underlined the significant influence of pigment on the high performance of the model (R^2 = 0.74). Hex showed similar trend to But, with minimal correlation across the spectrum, small peaks in the 400-450 nm range and slight variations around 650-700 nm suggesting some wavelength influence. However, the wide C.I. and relatively flat response indicate that the model struggles to predict Hex accurately, consistent with the lower R^2 (0.60). This suggested that the absorption signature of Hex was not well captured by the $a_{ph}'(\lambda)$-based model. For Peri, notable correlation peaks appeared in the 450-500 nm and 650-700 nm regions, corresponding to Peri's characteristic carotenoid absorption in the blue-

green spectrum. These peaks were accompanied by narrower confidence intervals, indicating a reliable model fit in these bands. However, the model's overall predictive accuracy was moderate (R^2 = 0.44), and broader variability suggested some difficulty in isolating Peri's signal from others, leading to a higher RMSD (1.1). TChl *b* exhibited strong correlations in the red region (600-700 nm), particularly around 650-675 nm,

Figure 13. Correlation coefficient between the model based on the first derivative of the absorption coefficient $a_{ph}'(\lambda)$ and individual pigment concentrations across different wavelengths.



where chlorophyll *b* had its key absorption peak. The model's high performance for TChl *b* (R^2 = 0.71) was reflected in the narrow confidence intervals at these wavelengths, indicating their importance in accurately predicting TChl *b* concentrations. Minor peaks also appeared in the 450-500 nm region, reflecting additional absorption by TChl *b* in the blue range. For TChl *c*, the correlation plot showed scattered peaks, particularly around 450-500 nm and 650-700 nm, which matched with the absorption features of TChl *c*. However, the broader, more variable confidence intervals suggested a lower level of confidence in these predictions, probably due to overlapping signals from other pigments. This was consistent with the moderate performance for TChl *c* (R^2 = 0.59). Finally, Zea showed a flatter correlation pattern, with weak but noticeable peaks in the 450-500 nm range and a less distinct response in the 600-650 nm region. The broad C.I. suggested high uncertainty, consistent with the low R^2 value (0.45) in Fig. 12. This indicates that Zea's absorption characteristics were difficult to distinguish, contributing to the challenges in modeling Zea concentrations accurately.

3. **DISCUSSION**

The present study examined the surface phytoplankton community distribution in the Baltic Sea, derived from IOPs and HPLC datasets covering different seasons and different areas of the Baltic Sea.

The high variations in these results were in agreement with the findings on previous pigments and absorption dataset analysis representative of the Southern Baltic (Woźniak et al., 2020; Meler et al., 2018, 2020). The variability expressed as coefficients of variation (CV%, defined as the ratio of the standard deviation to the average value) in Meler et al. (2020) was of 163 % for the TChl *a* (here 159%) and of 137% for a_{ph} at 440 nm (here of 128.8% for a_{ph} at 443 nm). Notably, the yellow substance exhibits high variation throughout all the dataset, in agreement with the finding of Kratzer (2019) and Harvey (2015). Meler et al. (2023) in a study dedicated to the Southern Baltic and the Gulf of Gandsk found that the average contribution of a_{ph} was 29 % ± 14 %, while for the detritus it was 19 % ± 9 % and, in agreement with our findings, the greatest contribution to the total light absorption was made by CDOM: 52 % ± 20 %.

Our data-driven statistical analyses on the absorption coefficient dataset identified distinct taxonomically defined phytoplankton communities in the Baltic Sea, characterized by five biomarker pigments: diatoms (Fuco), dinoflagellates (Peri), cryptophytes (Allo), green algae (TChl *b*) and cyanobacteria-pico-plankton (Zea). As already addressed in the introduction to this study, the use of biomarker pigments as representative of single taxa (i.e., Fuco for the Diatoms, while Fuco contributes to different taxa) is a simplification.

The unsupervised statistical techniques used in the data analysis were applied to both the HPLC pigment dataset and the absorption coefficient dataset. The analysis of the CDOM coefficient showed that the $a_g(\lambda)$ attributed to non-living particles was found to be consistently more pronounced when microplankton groups dominated the phytoplankton community, as observed in Modes 1 and 4. This finding was consistent with previous observations (Barrón et al. 2014). This can be explained by the higher production of detrital organic matter by larger microplankton, such as diatoms and dinoflagellates, which tend to have shorter life cycles and faster sinking rates, leading to an accumulation of non-living organic particles in the water column. These particles contributed to the CDOM pool and enhance the $a_g(\lambda)$ signal. For Modes 2 and 3 corresponding to the phytoplankton community composition with a predominant nanoplankton fraction, there were no significant correlations observed at any wavelength. This suggested that there were no significant relationships between the non-living portion of the absorption spectrum during periods when nanoplankton were dominant. Smaller phytoplankton, such as cyanobacteria and haptophytes, generally have slower sinking rates and lower production of particulate organic matter, resulting in a less pronounced CDOM signal. Similarly, the $a_{ph}(\lambda)$ was correlated with the 4 PCA modes. A remarkable correlation between $a_{ph}(\lambda)$ and Mode 1 was observed in the spectral ranges of 400–500 nm and 660–685 nm, which coincided with the regions of highest TChl *a* absorption. However, the correlation with other modes was less significant. The $a^*_{ph}(\lambda)$ on the other hand, showed a strong negative correlation with the first and second Modes: the trend was more pronounced with respect to $a_g(\lambda)$ and $a_{ph}(\lambda)$.

Furthermore, based on Catlett & Siegel previous work (2018), we explored the relationships between pigments and absorption spectra and their derivatives, extending the approach to $a^*_{ph}(\lambda)$. This modeling approach was found to be effective, as it exploited the covariance among DPA pigments, as well as their correlations with more common pigments and their corresponding absorption patterns. This approach differs from conventional methods, which typically assume that the abundances of distinct taxa are not correlated and that pigment ratios remain constant within a given dataset. The application of this modeling approach can be used in conjunction with the discrete approach applied to the HPLC database. The correlation with pigments $a_{ph}'(\lambda)$, $a_{ph}''(\lambda)$, $a^*_{ph}'(\lambda)$, and $a^*_{ph}''(\lambda)$, were analysed in relation to the results obtained by HCA_{pig} (Catlett & Siegel, 2017, Sun et al., 2022) and $HCA_{spectra}$. The analysis of the obtained correlation suggested that $a_{ph}'(\lambda)$ were $a^*_{ph}'(\lambda)$ of significance in the identifying consistently the biomarker pigments distribution in the Baltic and were taken in consideration when developing the EOF model. This analysis was preliminary to the application development of an EOF model, adapted from EOF of Bracher et al. (2015), to recunstruct biomarker pigments composition from hyperspectral optical observations. Among the analysis conducted on the spectral dataset, the Gaussian decomposition was also included to

compare the results of the EOF model in predicting the different phytoplankton pigments considered in the present study.

Among the model compared in the present stuty, the EOF model emerged as best performer in predicting the biomarker pigments distribution in the Baltic, by incorporating first derivative spectral absorption. Results from the EOF model performance scatterplots showed varying degrees of success in predicting pigment concentrations based on the first derivative of absorption coefficient, $a_{ph}'(\lambda)$. Pigments like Fuco and TCh *b* showed strong predictability indicating a strong linear relationship between the observed and predicted values. On the other hand, pigments like But and Zea shown weaker correlations suggesting more uncertainty in their predictions. The RMSD and cross-validation error (RMSD CV%) showed some pigments (e.g., Fuco) with low error rates, while others (e.g., Peri) had higher RMSE values, indicating potential limitations in the model performance for certain pigments. Our model found correlations at the 440, 600, and 660 nm wavelengths with TChl *b* and the first derivatives $a_{ph}'(\lambda)$, whereas the Gaussian decomposition considers the main TChl *b* contributions at 658 nm and 457 nm. For TChl *c*, our model finds the main correlations at 415, 430, 480, 575-590, 600, and 665 nm, which was partially consistent with the results of the Gaussian decomposition. The main contribution of EOF model was to find correlations at specific wavelengths for other pigments such as Allo, Fuco, Zea, and Peri, which were missed by the Gaussian decomposition as it did not consider carotenoids as individual pigments but only as groups (i.e., PPC, PSC). By transforming the original data into a set of orthogonal principal components, EOF captures the most significant variance while integrating information from multiple wavelengths while correlation-based methods (such as The Gaussian decomposition) relied on specific wavelengths and show variable performance. At least, is notewarty to add that the EOF model presented here used a more extensive dataset than the Bracher et al. (2015), that captured spatial and seasonal variability over multiple years. The spatial variability and the optical complexity of the Baltic Sea required a dataset that could cover different campaigns, multiple years and seasons. The present dataset used for the development of the method fullfilled this requirement. This, associated with the analysis of the model robustness, made the proposed model suitable for application in this complex basin of Baltic Sea.

4. **CONCLUSION**

This study investigates the distribution of phytoplankton communities in the Baltic Sea, using in situ datasets of IOPs and HPLC pigments. Using data-driven statistical analyses, the study identifyes five key biomarker pigments representing distinct phytoplankton communities: diatoms (Fuco), dinoflagellates (Peri), cryptophytes (Allo), green algae (TChl *b*), and cyanobacteria-picoplankton (Zea). Although using

these biomarker pigments as proxies for individual taxa is a simplification, they provide valuable insights into phytoplankton community structure.

Unsupervised statistical methods were applied to both the HPLC pigment and absorption coefficient datasets, suggesting that non-living particles, $a_g(\lambda)$, were more pronounced when microplankton dominated the community, while this correlation was absent when nanoplankton were dominant. Moreover, the absorption coefficient, $a_{ph}(\lambda)$, was strongly correlated with specific modes of the phytoplankton biomarker pigments, particularly in the spectral regions of 400–500 nm and 660–685 nm, highlighting their relevance for understanding TChl *a* absorption. The correlation between the spectral datasets and its derivatives with the pigment datasets was also investigated, to evaluate whether the development of a model should incorporate the derivative to better represent the phytoplankton biomarkers. A key contribution of this study is the development of a model to reconstruct phytoplankton biomarker pigments composition from hyperspectral optical data, extending the EOF model of Bracher et al. (2015) with first derivative analysis. The EOF model provided robust correlations at specific wavelengths for pigments such as TChl *b*, TChl *c*, Allo, Fuco, Zea, and Peri, allowing higher prediction accuracy, especially for Fuco and TChl *b* representative of diatoms and green algae communities respectively. In our study we found the EOF model outperforming the correlation-based methods in predicting pigment concentrations, particularly when using $a_{ph}'(\lambda)$ and added information on pigment not resolved by the Gaussian decomposition approach, such as Fuco.

In conclusion, this study presents a comprehensive framework for modeling phytoplankton biomarker pigments composition in the Baltic Sea using IOPs data in conjunction with an HPLC matching dataset. The results highlighted the value of advanced statistical and optical modeling approaches, such as EOF model with $a_{ph}'(\lambda)$, in describing the phytoplankton dynamics and providing more accurate predictions of pigment concentrations. This method shows promise for further application in other optically complex basins and contributes to a deeper understanding of phytoplankton dynamics in response to environmental change.

REFERENCES

Anderson, C. R., Siegel, D. A., Brzeniski, M. A., Guillocheau, N. (2008) Controls on temporal patterns in phytoplankton community structure in the Santa Barbara Channel, California. *J. Geophys. Res.*

Barrón, R. K., Siegel, D. A., Guillocheau, N. (2014) Evaluating the importance of phytoplankton community structure to the optical properties of the Santa Barbara Channel, California, *Limn. Oceanogr.,* 59, doi: 10.4319/lo.2014.59.3.0927.

Blanz, T., Emeis, K., Siegel, H. (2005) Controls on alkenone unsaturation ratios along the salinity gradient between the open ocean and the Baltic Sea. *Geochim. Cosmochim. Acta*. 69. 3589-3600. 10.1016/j.gca.2005.02.026.

Bracher, A., Taylor, M. H., Taylor, B., Dinter, T., Rottgers, R., and Steinmetz, F. (2015) Using empirical orthogonal functions derived from remote-sensing reflectance for the prediction of phytoplankton pigment concentrations. *Oc. Sci*., 11(1), 139–158. https://doi.org/10.5194/os-11-139-201

Brewin, R.J.W., Sathyendranath, S., Hirata, T. Lavender, S.J., Barciela, R.M., Hardman-Mountford, N.J. (2010) A three-component model of phytoplankton size class for the Atlantic. *Oc. Ecol. Model*., 221 pp. 1472-1483

Bricaud, A., Claustre, H., Ras, J., Oubelkheir, K. (2004) Natural variability of phytoplanktonic absorption in oceanic waters: Influence of the size structure of algal populations. *J. Geophys. Res.* 109:C11010, doi:10.1029/98JC02712

Bricaud, A., Mejia, C., Biondeau-Patissier, D., Claustre, H., Crepon, M., Thiria, S. (2007) Retrieval of pigment concentrations and size structure of algal populations from their absorption spectra using multilayered perceptrons. *App. Opt.*, 46(8), 1251–1260. https://doi.org/10.1364/ao.46.001251

Candela, L., Formaro, R., Guarini, R., Loizzo, R., Longo F., Varacalli, G. 2016. The PRISMA mission. (2016) IEEE International Geoscience and Remote Sensing Symposium (IGARSS), Beijing, China, pp. 253-256, doi: 10.1109/IGARSS.2016.7729057.

Canuti, E., Ras, J., Grung, M., Röttgers, R., Costa Goela, P., Artuso, F., Cataldi, D. (2016) HPLC/DAD Intercomparison on Phytoplankton Pigments (HIP-1, HIP-2, HIP-3 and HIP-4), *EUR 28382 EN*, Publications Office of the European Union, Luxembourg

Canuti, E. (2023) Phytoplankton pigment *in situ* measurements uncertainty evaluation: an HPLC interlaboratory comparison with a European-scale dataset. *Front. Mar. Sci.* 10:1197311. doi: 10.3389/fmars.2023.1197311

Canuti, E., Penna, A. (2024) Dynamics of Phytoplankton Communities in the Baltic Sea: Insights from a Multi-dimensional Analysis of Pigment and Spectral Data: Part I, Pigment Dataset, *Front. Mar. Sci. Sec. Oc. Obs.,* Vol. 11 - doi: 10.3389/fmars.2024.1425347

Chase, A., Boss, E., Zaneveld, R., Bricaud, A., Claustre, H., Ras, J., Westberry, T. K. (2013) Decomposition of in situ particulate absorption spectra. *Meth. Oceanogr.*, 7, 110–124. https://doi.org/10.1016/j.mio.2014.02.002

Chase, A. P., Boss E., Cetinić, I., Slade, W. (2017) Estimation of phytoplankton accessory pigments from hyperspectral reflectance spectra:toward a global algorithm. *Journal of Geophysical Research:Oceans*, **122**(12): 9725-9743. DOI:10.1002/2017JC012859

Catlett, D., Siegel, D. A. (2018) Phytoplankton pigment communities can be modeled using unique relationships with spectral absorption signatures in a dynamic coastal environment. *J. Geophys. Res.: Oc.*, 123, 246– 264. https://doi.org/10.1002/2017JC013195.

Ciotti, A. M., Lewis, M. R., Cullen. J. J. (2002) Assessment of the relationships between dominant cell size in naturalphytoplankton communities and the spectra shape of the absorption coefficient. *Limnol. Oceanogr*. 47:404–417,doi:10.4319/lo.2002.47.2.0404

Dugué, N., Anthony Perez, A.(2015) Directed Louvain: maximizing modularity in directed networks. Université d'Orléans. hal-01231784. https://hal.archives-ouvertes.fr/hal-01231784

Ficek, D., Kaczmarek, S., Stoń-Egiert, J., Woźniak, B., Majchrowski, R., Dera, J. (2004) Spectra of light absorption by phytoplankton pigments in the Baltic; conclusions to be drawn from a Gaussian analysis of empirical data. *Oceanologia*, **46**: 533-555.

Guanter, L., Kaufmann, H., Segl, K., Foerster, S., Rogass, C., Chabrillat, S., et al. (2015) The EnMAP Spaceborne Imaging Spectroscopy Mission for Earth Observation. Remote Sens. 7, 8830-8857. https://doi.org/10.3390/rs70708830

Harvey, E.T., Kratzer, S., Andersson, A. (2015) Relationships between colored dissolved organic matter and dissolved organic carbon in different coastal gradients of the Baltic Sea. AMBIO 44 (Suppl 3), 392–401 (2015). https://doi.org/10.1007/s13280-015-0658-4

Hirata, T., Hardman-Mountford, N.J., Brewin, R.J.W., Aiken, J., Barlow, R.G., Suzuki, K., Isada, T., Howell, E., Hashioka, T., Noguci-Aita, M., Yamanaka, Y. (2011) Synoptic relationships between surface chlorophyll-a and diagnostic pigments specific to phytoplankton functional types. *Biogeosc.*, 8 pp. 311-327

Hooker, S.B., Thomas, C.S., Van Heukelem, L., Schluter, L., Ras, J., Claustre, H.,Russ, M.E., Clementson, L., Canuti, E., Berthon, J-F., Perl, J., Normandeau, C., Cullen, J., Kienast, M., Pickney J.L. (2010) The Forth SeaWiFS HPLC Analysis Round-Robin Experiment (SeaHARRE-4) *NASA Tech. Memo. 2010-215857*, NASA Goddard Space Flight Center, Greenbelt, Maryland, 74 pp.

Hoepffner, N., Sathyendranath, S. (1991) Effect of pigment composition on absorption properties of phytoplankton *Mar. Ecol. Prog. Ser*., 73 , pp. 11-23

Hoepffner, N., Sathyendranath, S. (1993) Determination of the major groups of phytoplankton pigments from the absorption spectra of total particulate matter *J. Geophys. Res*., 98 (C12), pp. 22,789-22,803

Huping, Y.E., ZHANG, B., LIAO, X., LI, T., SHEN, Q., ZHANG, F., ZHU, J., LI, J. (2019) Gaussian decomposition and component pigment spectral analysis of phytoplankton absorption spectra. *J. Oc. Limn.,* 37(5): 1542-1554.

IOCCG (2014). Phytoplankton Functional Types from Space. Sathyendranath, S. (ed.), Reports of the International Ocean-Colour Coordinating Group, No. 15*, IOCCG*, Dartmouth, Canada.

Jeffrey, S.W., R.F.C. Mantoura, Wright, S.W. (1997) Phytoplankton Pigments in Oceanography, Monographs on Oceanographic Methodology, UNESCO.

Kowalczuk, P., Olszewski, J., Darecki, M., and Kaczmarek, S. (2005) Empirical relationships between coloured dissolved organicmatter (CDOM) absorption and apparent optical properties in Baltic Sea waters. Int . *J. Remote Sens.* 26 (2), 345–370

Mackey, M., D. Mackey, H. W. Higgins, Wright, S. W. (1996) CHEMTAX - a program for estimating class abundances from chemical markers: Application to HPLC measurements of phytoplankton. *Mar. Ecol. Prog. Ser.* 144: 265– 283.

Mannino, A., et al., (2019) Measurement protocol of absorption by chromophoric dissolved organic matter (CDOM) and other dissolved materials, In Inherent Optical Property Measurements and Protocols: Absorption Coefficient, Mannino, A. and Novak, M. G. (eds.), IOCCG Ocean Optics and Biogeochemistry

Protocols for Satellite Ocean Colour Sensor Validation, Volume 5.0, IOCCG, Dartmouth, NS, Canada (DRAFT).

Meister, G. Knuble, J. J. Gliese, U. Bousquet, R. Chemerys, L. H. Choi, H. Eplee, R. E. Estep, R. H. Gorman, E. T. Kitchen-McKinley, S. Kubalak, D. Lee, S. McClain, C. McIntire, J. W. Patt, F. S. Rhodes, Z. Werdell, P. J. (2024) The Ocean Color Instrument (OCI) on the Plankton, Aerosol, Cloud, ocean Ecosystem (PACE) Mission: System Design and Prelaunch Radiometric Performance. *Transactions on Geoscience and Remote Sensing*, vol. 62, pp. 1-18, 2024, Art no. 5517418, doi: 10.1109/TGRS.2024.3383812.

Meler, J., Ostrowska, M., Ficek, D., Zdun, A., (2017) Light absorption by phytoplankton in the southern Baltic and Pomeranian lakes: mathematical expressions for remote sensing applications. Oceanologia 59 (2017) 195-212, doi:10.1016/j.oceano.2017.03.010

Meler, J., Woźniak, S., Stoń-Egiert, J., Woźniak, B. (2018) Parameterization of the spectral light absorption coefficient of phytoplankton in the Baltic Sea: general, monthly and two-component variants of approximation formulas. Ocean Science Discussions. 1-39. 10.5194/os-2018-69.

Meler, J., Wo´zniak, S.B., Ston-Egiert, J. (2020) Comparison of methods for indirectly estimating the phytoplankton population size structure and their preliminary modifications adapted to the specific conditions of the Baltic Sea. *J. Mar. Syst.* 212, 103446. https://doi.org/10.1016/j.jmarsys.2020.103446

Meler, J., Litwicka, D., Zabłocka, M. (2023) Variability of light absorption coefficients by different size fractions of suspensions in the southern Baltic Sea. Biogeosciences. 20. 2525-2551. 10.5194/bg-20-2525-2023.

Mitchell, B.G., 1990. Algorithm for determining the absorption coefficient of aquatic particulates using the quantitative filter technique. Proceedings of SPIE. 1302:137-148

Mouw, C.B., Hardman-Mountford, N.J., Alvain, S., Bracher, A. , Brewin, R.J.W. , Bricaud, A., Ciotti, A.M., Devred, E. Fujiwara, A., Hirata, T. Hirawake, T. , Kostadinov, T.S., Roy, S., Uitz, J. (2017) A consumer's guide to satellite remote sensing of multiple phytoplankton groups in the Global Ocean. *Front. Mar. Sci.,* 4

Nelson, N.B. and D. A. Siegel. (2013) Global distribution and dynamics of chromophoric dissolved organic matter. *Annu. Rev. Mar. Sci.* 5:447–476, doi:10.1146/annurev-marine-120710-100751

Olli, K., Klais-Peets, R., Tamminen, T., Ptacnik, R., Andersen, T. (2011) Long term changes in the Baltic Sea phytoplankton community. *Boreal Env. Res.* 16 (SUPPL A). 3-14.

Roy, S., Llewellyn, C.A., Egeland, E.S., Johnsen, G., (2011) Phytoplankton Pigments, Characterization, Chemotaxonomy and Applications in Oceanography. Cambridge Univ. Press, p. 845.

Schlüter, L., Møhlenberg, F., Havskum, H., and Larsen, S. (2000) The use of phytoplankton pigments for identifying and quantifying phytoplankton groups in coastal areas: testing the influence of light and nutrients on pigment/chlorophyll a ratios. *Mar. Ecol. Prog. Ser.* 192, 49–63. doi: 10.3354/meps192049.

Seppälä, J. 2009. Fluorescence properties of Baltic Sea phytoplankton*, Mono. Boreal Env. Res.* vol. 34

Stoń-Egiert, J., Ostrowska, M., (2022) Long-term changes in phytoplankton pigment contents in the Baltic Sea: Trends and spatial variability during 20 years of investigations, *Cont. Shelf Res.,* Vol. 236, https://doi.org/10.1016/j.csr.2022.104666.

Tassan, S., Ferrari, G.M. (1995) An alternative approach to absorption measurements of aquatic particles retained on filters *Limnol. Oceanogr.,* 40 (8) pp. 1358-1368

Teng, J., Zhang, T., Sun, K., Gao, H. (2022) Retrieving Pigment Concentrations Based on Hyperspectral Measurements of the Phytoplankton Absorption Coefficient in Global Oceans. *Remote Sens.,* 14, 3516. https://doi.org/ 10.3390/rs14153516

Uitz, J. Claustre, H., Morel, A., Hooker, S.B. (2006) Vertical distribution of phytoplankton communities in open ocean: an assessment based on surface chlorophyll *J. Geophys. Res.,* 111 p. C08005

Van Heukelem, L., Thomas, C. S. (2001) Computer-assisted high-performance liquid chromatography method development with applications to the isolation and analysis of phytoplankton pigments. *J. Chromatogr. A* 910, 31–49. doi: 10.1016/S0378-4347100000603-00604

Vidussi, F. , Claustre, H., Manca, B.B., Luchetta, A., Marty, J.C. (2001) Phytoplankton pigment distribution in relation to upper thermocline circulation in the eastern Mediterranean Sea during winter *J. Geophys. Res*., 106 (C9) pp. 19939-19956

Sun, X., Shen, F., Brewin, R.J.W., Li, M. and Zhu, Q. (2022) Light absorption spectra of naturally mixed phytoplankton assemblages for retrieval of phytoplankton group composition in coastal oceans. *Limnol Oceanogr*., 67: 946-961. https://doi.org/10.1002/lno.12047

Wasmund, N., Nausch G., Matthäus, W. (1998) Phytoplankton spring blooms in the southern Baltic Sea — spatio temporal development and long-term trends. *J. Plankton Res*. 20: 1099–1117.

Wasmund, N., Uhlig S. (2003) Phytoplankton trends in the Baltic Sea. ICES *J. Mar. Sci.* 60: 177–186.

Wasmund N., Göbel J., von Bodungen B. (2008) 100-years-change in the phytoplantkon community of Kiel Bight (Baltic Sea). *J. Mar. Syst.* 73: 300–322

Woźniak, S., Meler, J., Stoń-Egiert, J. 2022. Inherent optical properties of suspended particulate matter in the southern Baltic Sea in relation to the concentration, composition and characteristics of the particle size distribution; new forms of multicomponent parameterizations of optical properties*, J. Mar. Sys.*, Vol. 229, https://doi.org/10.1016/j.jmarsys.2022.103720

Acknowledgments

The author would like to acknowledge Juha Flinkman, Seppo Kaitala and Jukka Seppälä from the Finnish Environment Institute (former Finnish Institute of Marine Research) for the opportunity to participate in the Oceanographic campaign on-board to the R/V “Aranda” and to the crew of R/V “Aranda” for their support during all the cruises. Acknowledgments are due Dirk Van der Linde from JRC for water sampling and laboratory analyses.

Conflict of Interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Author Contributions

First authorship: Elisabetta Canuti.

EC contributed to this paper by conceptualizing and designing the study, performing analysis and calculations and original draft. AP contributed to the paper by reviewing and editing the content.

Funding

The study has been supported by the European Commission Directorate General Join Research Centre (JRC) and the Copernicus Program.