LIMNOLOGY and OCEANOGRAPHY

ASLO

Limnol. Oceanogr. 69, 2024, 1547–1561

doi: 10.1002/lno.12606

Holistic optical water type classification for ocean, coastal, and inland waters

Shun Bi ,* **Martin Hieronymi**
Institute of Carbon Cycles, Helmholtz-Zentrum Hereon, Geesthacht, Germany

Abstract

Water constituents exhibit diverse optical properties across ocean, coastal, and inland waters, which alter their remote-sensing reflectance obtained via satellites. Optical water type (OWT) classifications utilized in satellite data processing aim to mitigate optical complexity by identifying fitting ocean color algorithms tailored to each water type. This facilitates comprehension of biogeochemical cycles ranging from local to global scales. Previous OWT frameworks have focused narrowly on either oceanic or inland waters and have relied too heavily on specific data collections. We propose a novel holistic OWT framework applicable to all natural waters, based on state-of-the-art bio-geo-optical modeling and radiative transfer simulations that encompass different phytoplankton groups. This framework employs a "knowledge-driven" paradigm, combining domain knowledge and insights from previous studies to simulate the reflectance spectrum from water constituent concentrations and inherent optical properties. Our method extracts optical variables to represent the full spectrum of reflectance, consolidating both spectral shape and magnitude. We apply the framework utilizing diverse in situ, synthetic, and satellite data (Sentinel-3 OLCI) and demonstrate its better classifiability than other frameworks. This framework lays the foundation for comprehensive global monitoring of natural waters.

Water color remote sensing plays a pivotal role in limnology and oceanography by enabling the deduction of water body compositions and substance concentrations through color information, a process often referred to as retrieval (Blondeau-Patissier et al. 2014). This method stands as the only approach for macroscopically and systematically observing aquatic ecosystems across oceans, coasts, and inland waters. Despite its importance, the endeavor faces challenges due to the significant spatial and temporal variability in the bio-geo-optical properties of these waters, complicating the universal application of retrieval models (Zheng and DiGiacomo 2017). Classification has proven to be a valuable strategy for overcoming these complexities by offering a structured approach to organizing and interpreting data, thus significantly enhancing our understanding of these intricate systems (Arnone et al. 2004).

*Correspondence: Shun.Bi@hereon.de

Additional Supporting Information may be found in the online version of this article.

Author Contribution Statement: SB and MH jointly developed the research idea and designed the framework. SB performed the data simulation and analysis. Both authors contributed to the result interpretation and the writing of the manuscript.

Historically, the classification of water colors dates back to early efforts such as the FU index (Forel 1890; Ule 1892), which visually distinguished 21 water color scales. Jerlov (1968) classified global waters into 10 types based on in situ measurements of the diffuse attenuation coefficient spectra laid foundational work in optical oceanography (Solonenko and Mobley 2015). Such classifications have been instrumental in elucidating the biological and chemical processes within water bodies, paralleling how temperature and salinity classifications have revolutionized physical oceanography (Arnone et al. 2004). The concept of optical water types (OWTs) has been fundamental, providing a common language for describing and understanding aquatic ecosystems.

In the satellite era, the classification into Case-1 and Case-2 types by Morel and Prieur (1977) has been particularly influential. Case-1 waters are characterized by optically active constituents directly related to phytoplankton concentration, whereas Case-2 waters are influenced by additional factors such as colored dissolved organic matter and suspended particulates. This binary classification simplified early bio-optical models and supported the initial focus of satellite missions on ocean color. However, the complexity of Case-2 waters has exposed limitations in this binary approach, leading to calls for more nuanced optical modeling for any water (Mobley et al. 2004).

Recent advancements have seen a focus on classifying water types, particularly based on remote-sensing reflectance,

R_{rs}, due to their accessibility from in situ and satellite observations. These efforts, described as "data-driven," involve collecting and analyzing spectral data through various clustering techniques, depending on the specific application. For instance, water type classification has been utilized in optimizing the retrieval of Chlorophyll *a* (Chl *a*) concentration and its uncertainty estimation, demonstrating superior performance over single-algorithm approaches in optically complex waters (Moore et al. 2014; Jackson et al. 2017; Neil et al. 2019). Spectrum-based classification has further aided in identifying marine phytoplankton pigment compositions (Torrecilla et al. 2011; Uitz et al. 2015) and studying the dynamic properties of the upper ocean (Martin Traykovski and Sosik 2003).

With a growing emphasis on optically complex waters, the normalization of water spectra to highlight spectral shape over magnitude has become common practice, reflecting the diverse influences on water color (Vantrepotte et al. 2012; Mélin and Vantrepotte 2015; Bi et al. 2019). Typological studies by Spyrakos et al. (2018) advocate for integrating inland and coastal water type frameworks, demonstrating their potential in exploring optical diversity. The statistical means of water types have also been recognized as a reliable tool for quality assurance in measurements (Wei et al. 2016, 2022) and for developing machine learning strategies (Hieronymi et al. 2017; Pahlevan et al. 2020). However, the specificity of emerging OWT frameworks to particular systems, based on data sets with limited bio-geo-optical variability, somewhat contradicts the goal of classification to facilitate a boarder understanding of aquatic ecosystems. This highlights the need for methodological advancements toward a more holistic OWT framework.

Several challenges currently limit the development of a holistic framework. These include methodological limitations in in situ data collection, which are not equally suitable for all water types and may include significant spectral uncertainties. Moreover, the measured variability is relatively limited, optically extreme waters are often less well sampled which limits the classifiable spectral features and lowers the resolution of OWTs (Campbell 1995; Lee and Hu 2006; Spyrakos et al. 2018). Additionally, the use of unsupervised clustering methods in water type classification, without strict type labels like land remote sensing, results in post hoc qualitative descriptions that depend heavily on co-measured bio-optical parameters and researcher experience. Variability in classification outcomes based on initial data conditions and clustering hyperparameters further complicates this issue (Bi et al. 2019, 2021). The high dimensionality and autocorrelation of hyperspectral R_{rs} data also represent challenges, making clustering analysis difficult due to data sparsity and feature redundancy (Bouveyron et al. 2007; Cael et al. 2023). Some studies address this by resampling spectra to configurations of satellite sensors or other custom multiband for dimensionality reduction, although this can hinder comparability across OWT frameworks due to the trade-off between high and low dimensions. Moreover, the incompatible focus on either spectral shape or magnitude in previous studies complicates the analysis of classification results, which calls for better high-level abstraction for spectrum data (Eleveld et al. 2017).

Forward modeling of remote sensing data offers a promising approach to overcome these data limitations. For instance, Feng et al. (2005) and Botha et al. (2020) utilized forward-simulated R_{rs} to define OWTs and optimize algorithms for each type. Unlike apparent optical properties, inherent optical properties, essential for forward simulation, are unaffected by the light fields and observation conditions (Bi et al. 2023*b*). However, the utility of simulated R_{rs} is still constrained by the availability of accurate inherent optical properties measurements. Our previous work (Bi et al. 2023*a*) introduced a model for calculating inherent optical properties from water constituent concentrations, demonstrating its potential to capture water optical variability more effectively. Yet, existing OWT frameworks did not fully classify the model-simulated R_{rs}, indicating the need for a new paradigm in OWT framework development.

This manuscript aims to establish a holistic optical water type classification framework for oceanic, coastal, and inland waters. We suggest a "knowledge-driven" approach, informed by domain knowledge and insights from prior "data-driven" studies, to enhance interpretability and applicability across diverse aquatic ecosystems (Arvor et al. 2019). It defines OWTs, employs specific data simulations based on forward models, and introduces a new classification scheme that accounts for both the shape and magnitude of reflectance spectra. By applying this framework to collected in situ spectral data and satellite remote sensing data, we evaluate its effectiveness in various water systems and its potential contributions to aquatic ecosystem studies. We anticipate that the new OWT framework will be broadly applicable, offering insights into the bio-optical states of different aquatic systems and supporting the development of ocean color algorithms for optically complex waters.

Data and methods

Forward modeling of remote-sensing reflectance

The model by Bi et al. (2023*a*) was adopted to simulate inherent optical properties (absorption and scattering). Further specifics of the model are detailed in their reference. Here, we summarize the key aspects for clarity and conciseness. Component optical properties of water are treated separately in the model, including those of pure water, phytoplankton, detritus, and colored dissolved organic matter (CDOM). Note the model uses "detritus" in a broad sense to encompass various particulate sources, beyond the traditional biological definition in marine optics, covering both biogenic and minerogenic materials across different water bodies. Parameters for each component are as follows: pure water is modeled based on temperature and salinity; phytoplankton is defined

by Chlorophyll *a* concentration, [Chl], and specific phytoplankton group fractions; detritus incorporates [Chl] and inorganic suspended matter concentration, [ISM]; and CDOM is quantified through its absorption at 440 nm, $a_g(440)$ and its spectral shape. The model also incorporates secondary parameters like the power law exponent of attenuation coefficient and single-scattering albedo to enhance the diversity of simulated optical data. It operates in two modes: the comprehensive "Four-term," which uses all four components, and the streamlined "Two-term," focusing on pure water and phytoplankton for efficiency in Case-1 waters, where other components are presumed to correlate with phytoplankton.

To model remote-sensing reflectance from inherent optical properties, the formula proposed by Lee et al. (2011) was applied given as:

$$R_{\mathrm{rs}}(\lambda) = \left(G_0^w + G_1^w \frac{b_{\mathrm{bw}}(\lambda)}{k(\lambda)}\right)\frac{b_{\mathrm{bw}}(\lambda)}{k(\lambda)} + \left(G_0^p + G_1^p \frac{b_{\mathrm{bp}}(\lambda)}{k(\lambda)}\right)\frac{b_{\mathrm{bp}}(\lambda)}{k(\lambda)}, \quad (1)$$

where λ is the wavelength of light, $b_{\mathrm{bw}}(\lambda)$ and $b_{\mathrm{bp}}(\lambda)$ are the backscattering coefficients of pure water and total particulate matter, sum of which are total backscattering coefficient, $b_{\mathrm{bt}}(\lambda)$. $k(\lambda)$ is the sum of total absorption coefficient, $a_t(\lambda)$, and $b_{\mathrm{bt}}(\lambda)$. According to the lookup table presented by Lee et al. (2011), the coefficients denoted as G vary with respect to the solar zenith angle and the directions of observation. Specifically, when solar zenith angle is set at 30° and the observation is made from the nadir direction, the values of coefficients are as follows: $G_0^w = 0.05881474$, $G_1^w = 0.05062697$, $G_0^p = 0.03997009$, and $G_1^p = 0.1398902$. Note that inelastic scattering (e.g., Raman scattering) and fluorescence were not considered in the Lee et al. (2011) model but the effect is expected to be minimal for the training data used to build the framework (see Discussion and Supporting Information).

Definition of optical water types

We utilized domain expertise, empirical evidence, and a theoretical understanding of water color dynamics to define the OWTs, which range from clear to oligotrophic, eutrophic, and hyper-eutrophic. We discovered representative spectral features using previous "data-driven" OWT studies and accessible data sets. As to recap the relationship between absorption, scattering, and reflectance for defining water type, it is here helpful to simplify, albeit not exactly physically correct, that reflectance is influenced by the scattering-to-absorption ratio to better understand the intuitive changes in reflectance due to inherent optical properties.

We first consider four extreme scenarios as theoretical boundaries. Pure water, with its noticeable blue color due to lower short-wavelength absorption, is one extreme (Morel et al. 2007; Jackson et al. 2017; Casey et al. 2020). In contrast, waters dominated by CDOM appear darker and browner due to CDOM absorption, which contributes to negligible scattering (Kutser et al. 2016; Spyrakos et al. 2018). Detritus-rich waters are also brown, but they are much brighter due to significant scattering alongside absorption (Moore et al. 2009; Mélin and Vantrepotte 2015; Lavigne et al. 2022). Phytoplankton presence introduces the most optical variability, with two typical spectral reflectance peaks at around 560 and 709 nm that vary with pigment composition and biomass level (Spyrakos et al. 2018; Bi et al. 2021). High biomass conditions provide a distinct near-infrared reflectance plateau, whereas lower concentrations maintain green hues but with a lowered second peak, which transitions to turquoise in cleaner waters. This "knowledge-driven" paradigm also supports the definition of unique phytoplankton-related types, like *Coccolithophores*, which are known for enhanced scattering by coccoliths (Moore et al. 2012; Neukermans and Fournier 2018; Cazzaniga et al. 2021). Finally, we defined ten OWTs that span these extremes, ranging from indigo-blue oligotrophic waters (OWT 1) to dark brown, CDOM-rich waters (OWT 7), as shown in Table 1. The number of defined water types is subjective, but it is more than sufficient and manageable for all water conditions we have considered. OWT labels ending with "a" and "b" were intended for types with similar spectral shapes but differing magnitudes, with "b" usually representing a variant with higher brightness.

Simulation of training data

To numerically represent conceptual OWTs, we employed the forward model to generate R_{rs} data. This data was then assessed against specific criteria to ensure it met the expected characteristics for each water type. Only data meeting these criteria were retained in the respective training set. This selection process was repeated until the desired volume of data for each water type was achieved, with randomness incorporated into the forward modeling of R_{rs} by selecting values within the defined ranges for primary inputs, i.e., [Chl], [ISM], and $a_g(440)$. Additionally, we adjusted other model parameters, including the distribution of phytoplankton groups as outlined in Bi et al. (2023*a*). To streamline the R_{rs} selection criteria, we resampled the simulated hyperspectral data to match the Sentinel-3 OLCI wavebands.

The selection process was partly based on a set of R_{rs} inequalities to ensure that the spectral shape features appeared as expected. The peak position of reflectance, λ_p, was another criterion, indicating the spectral shape, with specific λ_p values corresponding to different waveband ranges to account for multimodal peaks. For example, the criteria of extremely clear blue water type (OWT 1) include $R_{\mathrm{rs}}(412) > R_{\mathrm{rs}}(443) > R_{\mathrm{rs}}(490)$ and $\lambda_p < 410$ nm, to ensure a monotonically decreasing spectral curve from shorter to longer wavebands. We also employed proportions of inherent optical properties to identify water types under extreme absorption or scattering conditions (OWTs 6 and 7).

The model of Bi et al. (2023*a*) incorporated seven phytoplankton groups, reflecting the optical variability and diversity across different aquatic ecosystems. The fraction of these

Table 1. Conceptual description of optical water types.

OWT	Description
1	Extremely clear and oligotrophic indigo-blue waters with high reflectance in the short visible wavelengths.
2	Blue waters with similar biomass level as OWT 1 but with slightly higher detritus and CDOM content.
3a	Turquoise waters with slightly higher phytoplankton, detritus, and CDOM compared to the first two types.
3b	A special case of OWT 3a with similar detritus and CDOM distribution but with strong scattering and little absorbing particles like in the case of Coccolithophore blooms. This type usually appears brighter and exhibits a remarkable ~490 nm reflectance peak.
4a	Greenish water found in coastal and inland environments, with higher biomass compared to the previous water types. Reflectance in short wavelengths is usually depressed by the absorption of particles and CDOM.
4b	A special case of OWT 4a, sharing similar detritus and CDOM distribution, exhibiting phytoplankton blooms with higher scattering coefficients, e.g., Coccolithophore bloom. The color of this type shows a very bright green.
5a	Green eutrophic water, with significantly higher phytoplankton biomass, exhibiting a bimodal reflectance shape with typical peaks at ~560 and ~709 nm.
5b	Green hyper-eutrophic water, with even higher biomass than that of OWT 5a (over several orders of magnitude), displaying a reflectance plateau in the Near Infrared Region, NIR (vegetation-like spectrum).
6	Bright brown water with high detritus concentrations, which has a high reflectance determined by scattering.
7	Dark brown to black water with very high CDOM concentration, which has low reflectance in the entire visible range and is dominated by absorption.

groups in the model was tailored to the characteristics of each water type, with adjustments based on biomass level, i.e., [Chl], and specific optical properties. For example, lower [Chl] values were associated with higher fractions of marine phytoplankton groups in oceanic water types (OWTs 1 and 2), while different distributions were applied for types with unique scattering and absorption properties (e.g., OWTs 3b and 4b for phytoplankton groups higher scattering).

For OWTs 1 and 2, simulations were divided evenly between the "Two-term" and "Four-term" models while the rest of the water types were simulated only by the "Four-term" model. In addition, we assigned a "global" set, taking up half of the simulations per type, with a common and broad set of ranges to avoid any excessive constraints. To ensure comprehensive and representative training data, we conducted 10,000 simulations for each water type, totaling 100,000 simulations. We refined the selection criteria through trial-and-error, initially setting broad concentration ranges and then adjusting based on the proportion of simulations that met the criteria. This iterative process, often exceeding five times, aimed to encompass a wide range of bio-geo-optical variations. Technical details were described in Supporting Information Text S1 and the finalized ranges and criteria were presented in Supporting Information Table S1.

Scheme of optical classification

Our classification scheme initially extracts three optical variables from the input R_{rs} spectrum: the apparent visible wavelength, AVW, the trapezoidal area at red, green, and blue (RGB) bands, A_{RGB}, and the normalized difference index at green and red, NDI. These features are selected to project hyperspectral or multispectral reflectance into a high-level abstraction space efficiently; thereby, the classification can leverage both the spectral shape and magnitude. For each water type, we compute the means and covariance matrix of these three optical variables based on the training data set, which are subsequently utilized to calculate membership values based on the Mahalanobis distance, as detailed by Moore et al. (2001).

The AVW, as proposed by Vandermeulen et al. (2020), is defined as the harmonic mean of the visible R_{rs} wavelengths ranging from 400 to 700 nm. This variable serves as an indicator of the spectral shape, where a lower AVW value suggests a higher reflectance in the blue regions. We have extended the range to 800 nm to capture more spectral variability in waters with high biomass content, such as OWTs 5a and 5b. AVW was calculated as

$$\mathrm{AVW} = \frac{\sum_{400}^{800} R_{rs}(\lambda)}{\sum_{400}^{800} \frac{R_{rs}(\lambda)}{\lambda}}. \quad (2)$$

For hyperspectral data with a 1-nm interval, AVW can be calculated directly. In the case of multispectral data, primarily from satellite observations, AVW values for multispectral data, $\mathrm{AVW}_{\mathrm{multi}}$ are obtained and then recalibrated to their hyperspectral equivalent, $\mathrm{AVW}_{\mathrm{eq}}$ through a polynomial equation, following the approach by Vandermeulen et al. (2020), as

$$\mathrm{AVW}_{\mathrm{eq}} = \sum_{i=0}^{5} c_i \, \mathrm{AVW}_{\mathrm{multi}}^{i}, \quad (3)$$

where specifically for Sentinel-3A OLCI band configuration, $c_0 = 5.924396 \times 10^{3}$, $c_1 = -5.805088 \times 10^{1}$, $c_2 = 2.304654 \times 10^{-1}$, $c_3 = -4.388190 \times 10^{-4}$, and $c_4 = 4.079200 \times 10^{-7}$, and $c_5 = -1.485240 \times 10^{-10}$. The used OLCI bands are 400, 412, 443, 490, 510, 560, 620, 665, 682, 709, 754, 779, and 866 nm. Coefficients for other satellite sensors were provided in Supporting Information Table S2.

We quantify spectral magnitude using the trapezoidal area between the spectrum and a zero baseline, a method also

applied in prior studies (Vantrepotte et al. 2012; Spyrakos et al. 2018). Unlike these studies, which calculate the area across the entire spectral range, we focus on the three RGB bands to enhance transferability across different sensor setups. A_{RGB} is calculated as

$$A_{\mathrm{RGB}} = \frac{1}{2}\sum_{i=2}^{N}(\lambda_i - \lambda_{i-1})(R_{\mathrm{rs}}(\lambda_i) - R_{\mathrm{rs}}(\lambda_{i-1})), \tag{4}$$

where $N = 3$, representing the blue, green, and red bands at 443, 560, and 665 nm, respectively. Despite a log10 transformation, A_{RGB} values showed a left skew, which was not conducive to classification. Thus, we applied a Box-Cox transformation (Box and Cox 1964), resulting in A_{BC} as

$$A_{\mathrm{BC}} = \frac{A_{\mathrm{RGB}}^{t_{\mathrm{BC}}} - 1}{t_{\mathrm{BC}}}, \tag{5}$$

with a transformation coefficient $t_{\mathrm{BC}} = 0.22675$, determined based on the training dataset.

The NDI is useful for distinguishing OWTs, especially when AVW values are high. It helps prevent the misclassification of extremely turbid and high biomass waters, which exhibit reduced red reflectance due to phytoplankton absorption compared to detritus. NDI is computed as

$$\mathrm{NDI} = \frac{R_{\mathrm{rs}}(560) - R_{\mathrm{rs}}(665)}{R_{\mathrm{rs}}(560) + R_{\mathrm{rs}}(665)}. \tag{6}$$

Upon calculating the three optical variables from the input R_{rs} spectrum, they are used to determine membership values as follows,

$$u_k = 1 - F\left(n; D_k^2\right), \tag{7}$$

where u_k is the membership value for the k-th water type, F denotes the cumulative χ^2 distribution function with n degrees of freedom ($n = 3$ for the three optical variables), and D_k is the Mahalanobis distance between the input variables and the mean of the k-th water type. The membership values are regarded as the classification results of the proposed framework but the dominant OWT, determined by the type with the maximum membership, can also be calculated for better visual inspection and understanding (Moore et al. 2001).

The squared Mahalanobis distance, D_k^2, is calculated as

$$D_k^2 = (x - \mu_k)^{\mathrm{T}}\Sigma_k^{-1}(x - \mu_k), \tag{8}$$

where x is a vector of three input variables [AVW, A_{BC}, NDI], μ_k is the corresponding mean vector for the k-th water type, and Σ_k^{-1} represents the inverse of the covariance matrix of the three variables. The symbol T denotes the matrix transpose operation.

The total membership of all types, u_{tot}, was computed by summing the individual memberships as

$$u_{\mathrm{tot}} = \sum_{k=1}^{C} u_k, \tag{9}$$

where $C = 10$ is the number of defined water types. This measure indicates the extent to which the input spectrum aligns within the established OWT framework (Moore et al. 2001; Jackson et al. 2017). An u_{tot} value exceeding one suggests redundant classification results, indicating overlaps among water types, whereas values less than one suggest that the classification result is underrepresented, deviating from the cluster centroids. Besides, conducting a statistical analysis on the classifiable rate, represented by the proportion $P(u_{\mathrm{tot}})$ of samples where u_{tot} exceeds a threshold of 0.1, provides critical insights into the capability of a framework for capturing optical variation (Mélin et al. 2011; Vantrepotte et al. 2012; Moore et al. 2014).

Other optical water type frameworks

We selected three existing "data-driven" OWT frameworks for comparison because they employ the same Mahalanobis distance-based method for calculating membership as this study. These frameworks are tailored for specific aquatic environment applications: Moore et al. (2014) for coastal and inland waters, Hieronymi et al. (2017) for oceanic and coastal waters, and Jackson et al. (2017) mainly for oceanic waters. Differences among these frameworks include the number of types and spectral processing techniques. The classification scheme of other frameworks aligns with that described in Eqs. 7–9, with the primary difference being the use of spectral centroids instead of the three optical variables specific to this study. For more details, see the original references or the summary in Hieronymi et al. (2023). Given the absence of an established method for comparing OWT frameworks, our analysis concentrates on their ability to classify spectra data, using the statistics of memberships, e.g., $P(u_{\mathrm{tot}})$, as a metric. This approach allows us to assess the performance of frameworks based on their handling of optical variability, streamlining the evaluation process.

Test data

In situ measurement and synthetic data

To evaluate the OWT framework, we compiled a comprehensive data set comprising both in situ measurements and synthetic (simulated) hyperspectral R_{rs} data, chosen for its diversity in water types and conditions. This diversity aims to test and compare both "knowledge-driven" and "data-driven" frameworks across various aquatic environments. The in situ measurement data ($N = 2411$) encompass a range from clean oceanic waters, "PACE" (Casey et al. 2020), to highly turbid coastal and inland waters, "HYPERMAQ" (Lavigne et al. 2022) and "GLORIA" (Lehmann et al. 2023), including eutrophic conditions, "C22" (Castagna et al. 2022). We also analyzed two synthetic data sets ($N = 16{,}000$), each focusing on Case-1,

"IOCCG5" (IOCCG 2006) and Case-2, "CCRR" (Nechad et al. 2015) waters.

Satellite data

Despite challenges with atmospheric correction in ocean color algorithms and optical water classification studies (Moore et al. 2014; Spyrakos et al. 2018), analyzing OWT distribution from space effectively tests a framework in capturing distinctive optical signatures across varied aquatic environments. Hieronymi et al. (2023) emphasize the utility of their new atmospheric correction method, A4O, for improved detection of optical shape features. The justification for using A4O was described in Supporting Information Text S2. We chose R_{rs} data from their Sentinel-3 OLCI test scenes to evaluate the performance of our framework. For better map visualization, five subsets were selected to represent regions with diverse optical complexities: the southern North Sea and Hangzhou Bay for their high turbidity; the Baltic Sea for its eutrophic conditions; the Barents Sea for its unique arctic optical properties; the Black Sea, Marmara Sea, and Aegean Sea for their role as transitional waters linking distinct marine ecosystems; various turbid and eutrophic inland rivers and lakes. The selected subsets also contained areas experiencing unique water events such as *Coccolithophores* and *Cyanobacteria* blooms.

Results

The optical water type framework

Figure 1 depicts the average R_{rs} spectra of simulated training data sets for each defined OWT on raw and normalized scales. Normalization was performed by dividing the original R_{rs} values by the A_{RGB} to improve the visibility of spectral shapes. Figure 2 adds to this by showing the distribution of water constituent concentrations (forward modeling inputs), optical variables (classification inputs), and absorption proportions of four water components. Significant variations in spectral shapes and magnitudes are evident.

From OWT 1 to 2 and 3a, the reflectance peak shifts from blue to green bands, indicating an increase in CDOM and [ISM] and a relative decrease in pure water absorption. CDOM has a significant effect on the spectral shape in the blue band. OWTs 3a and 3b have similar spectral shapes, but OWT 3b has significantly higher magnitudes than OWT 3a. This pattern of differentiation can also be seen between OWTs 4a and 4b, and to a lesser extent between OWT 5a and 5b, with the latter exhibiting higher near-infrared reflectance and variability. Although OWTs 6 and 7 have different magnitudes in Fig. 1a, their spectral shapes are somewhat similar in Fig. 1b, highlighting the possibility of misclassification based solely on spectral shape. This variation in R_{rs} spectra within one water type is more pronounced in magnitude than in shape (Supporting Information Figs. S1 and S2). The corresponding mean spectral inherent optical property coefficients also show substantial differences in magnitude and shape (Supporting Information Fig. S3).

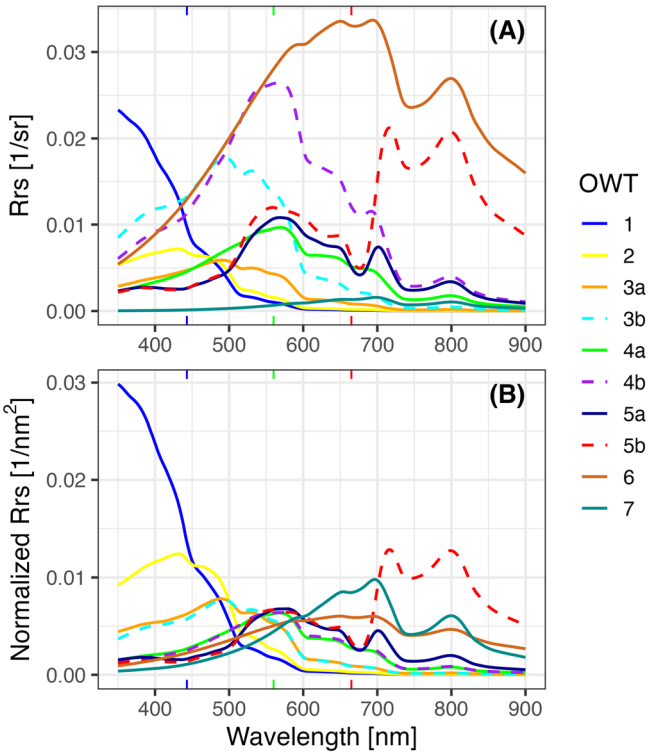

Fig. 1. Mean spectrum of simulated spectra for optical water types. Panel (**a**) displays the raw remote-sensing reflectance (unscaled), while Panel (**b**) shows the spectral normalized by trapezoidal-area. The positions of RGB bands are marked on the x-axis.

Figure 2a–c shows that some simulated samples fall outside the defined ranges of water constituent concertation in forward modeling. This is because the "global" set samples, which account for half of each type, increase optical variability even when using the same selection criteria, which is beneficial to the OWT framework. AVW is an effective optical indicator for the similar trend in [Chl] from OWT 1 to 5a, as shown in Fig. 2a,d. However, for some water types, such as OWTs 3a, 3b, 4a, 4b, 6, and 7, AVW alone is insufficient to distinguish them, whereas differences in A_{BC} and NDI can aid in classification. Figure 2g–i shows the mean statistic of proportions of component absorption at RGB bands, which supports the expected classification of optical variables. Pure water absorption is dominant in OWTs 1–3b, whereas all components are mixed in OWTs 4a and 4b. a_{ph} dominates OWTs 5a and 5b, even in the red band, whereas a_d and a_g dominate OWTs 6 and 7, respectively. Supporting Information Fig. S4 complements the proportional distribution over the entire visible spectrum.

Figure 3a,b present the pairwise scatter plots for the three optical variables of the training data, also showing the data space of our classification framework. When AVW values are <460 nm, the A_{BC} variation is constrained due to little particle scattering, particularly from detritus. The observed decrease in A_{BC} as AVW decreases in OWT 1 suggests a shift in the

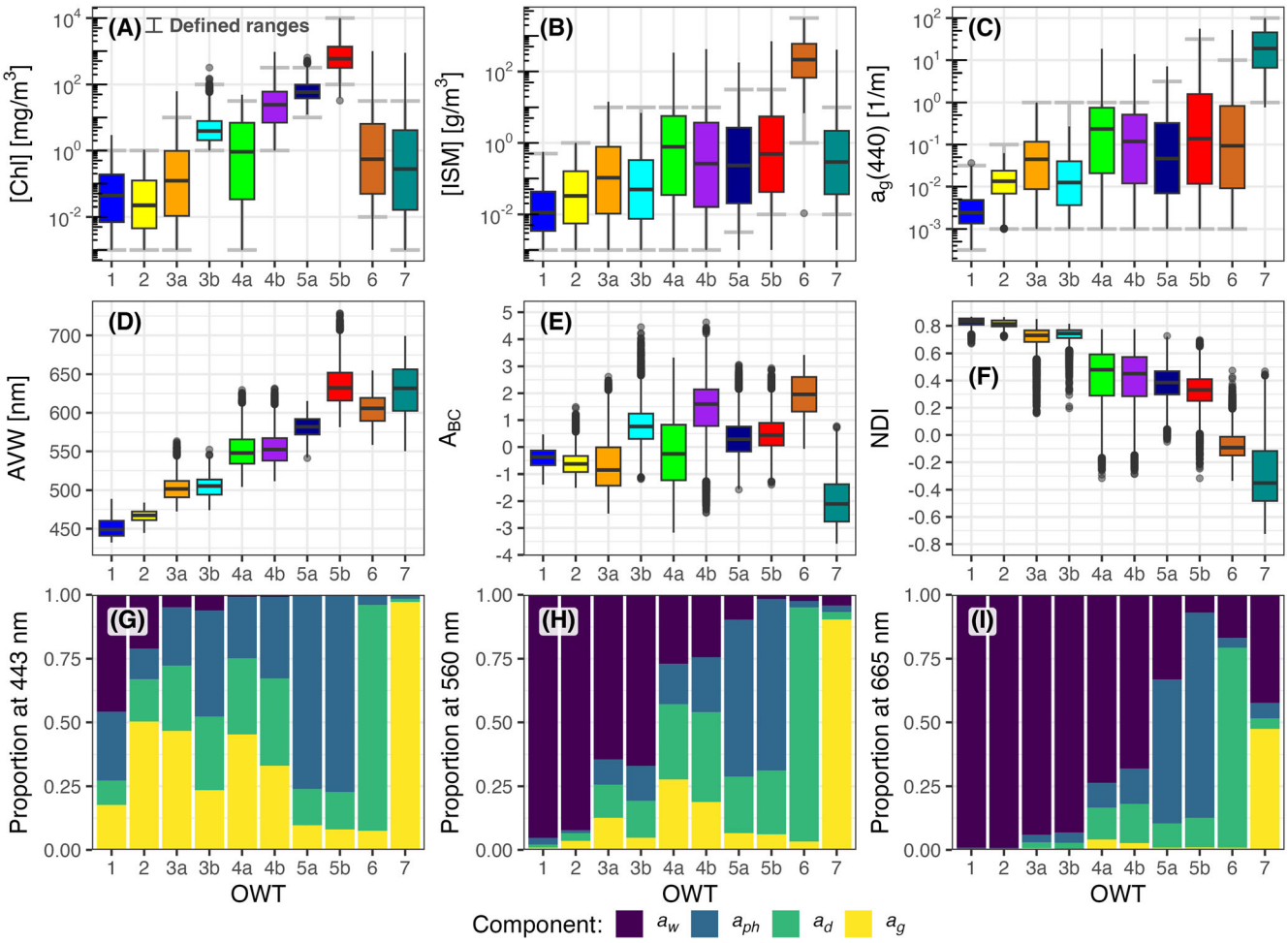

Fig. 2. Distribution of water constituent concentration, optical variables, and component proportions in total absorption for each optical water type. Panels (**a**–**c**) show predefined concentration ranges from Table 1 as gray crossbars, with colored boxplots depicting simulated samples that meet each type's criteria. The panels detail the following: (a) Chlorophyll *a* concentration, [Chl], (b) inorganic suspended matter concentration, [ISM], (c) CDOM absorption at 440 nm, $a_g(440)$, (d) Box-Cox transformed trapezoidal area at RGB bands, A_{BC}, (e) apparent visible wavelength, AVW, and (f) normalized difference index at 560 and 665 nm, NDI, and (g–i) component absorption proportions—pure water, a_w, phytoplankton, a_{ph}, detritus, a_d, and CDOM, a_g—to total absorption at blue (443 nm), green (560 nm), and red (665 nm) bands.

reflectance peak to longer wavelengths as absorption in the blue bands becomes stronger. Similar patterns can be discerned in some lower margins of other types. In contrast, a rise in A_{BC} with AVW implies the impact of particle scattering, like phytoplankton scattering in OWTs 3b and 4b, and detritus scattering in OWT 6, leading to higher reflectance (increased A_{BC} values). Differences arise in OWTs 5a and 5b from other OWTs due to the application of a phytoplankton absorption wavelength (665 nm). In certain intersection regions among water types, it is important to note that there are points overlapped. We decided to retain these points to calculate the centroids of OWTs, as these overlaps provide the necessary variance for each type rather than isolating them.

Framework efficacy: capturing in situ and synthetic data

Figure 3 illustrates the capability of our framework to accurately identify the expected water types within both in situ and synthetic data sets. In the "C22" data, collected from eutrophicated waters, OWT 5b is notably present. The "HYPERMAQ" data, characterized by its turbid water samples, features OWT 6, while the "PACE" data reveals the presence of cleaner oceanic waters, OWTs 1 and 2. The "GLORIA" data stands out for its diversity, covering nearly all water types except OWTs 1 and 2, with more than half of the samples classified as greenish OWTs 4a and 4b. It uniquely captures instances of OWT 7, specifically from black lakes in Estonia. Two synthetic data sets "CCRR" and "IOCCG5" are also well captured under the new framework, while Fig. 3e,f highlight their modeling limitation on specific water types such as OWT 3b and OWT 5b, which are characterized by higher reflectance magnitudes, and OWT 7, noted for its distinct dark brown color due to high CDOM absorption. Our framework achieves a

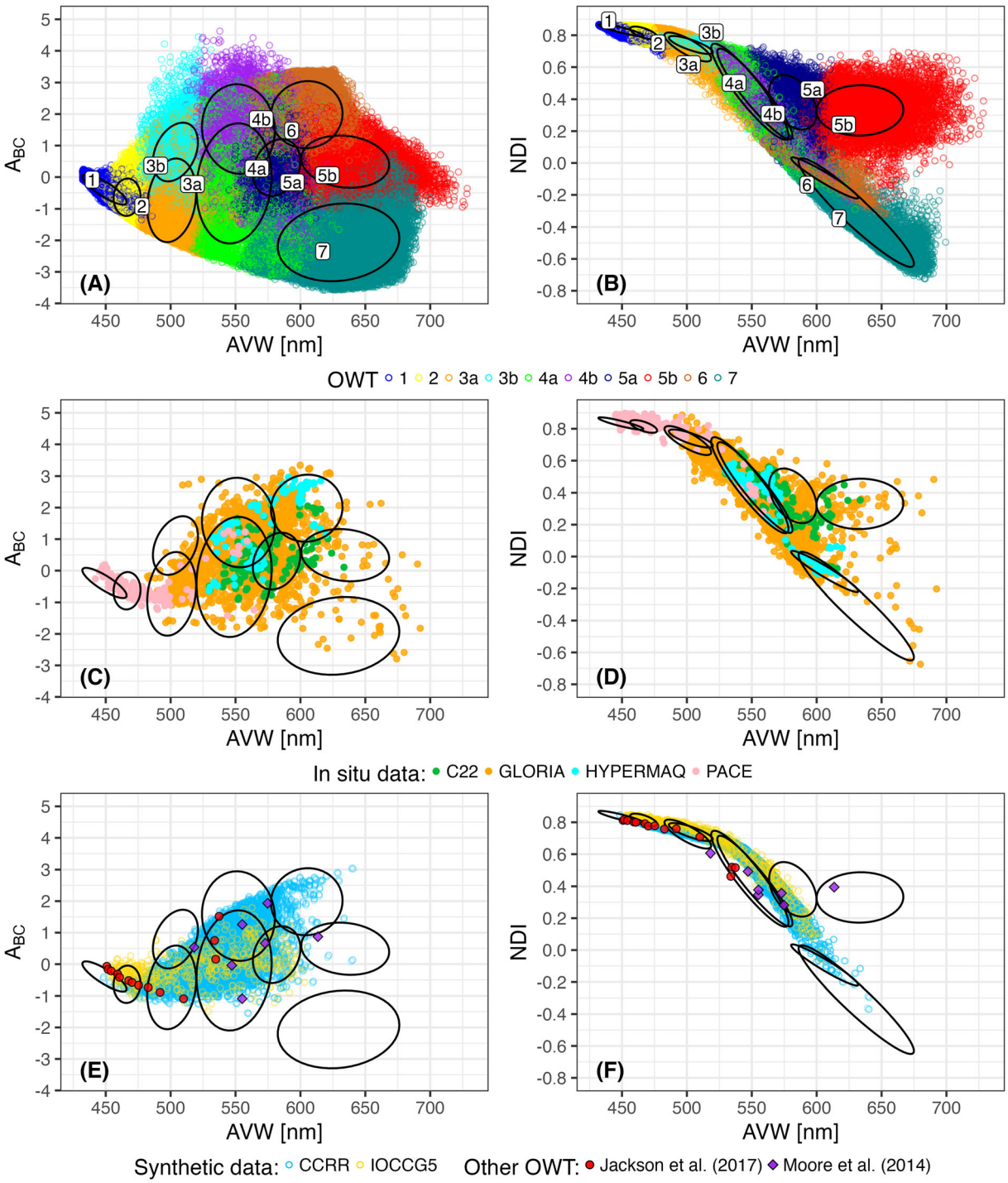

Fig. 3. Distribution of optical variables (apparent visible wavelength, AVW, Box-Cox transformed trapezoidal area at RGB bands, A_{BC}, and normalized difference index at 560 and 665 nm, NDI) across simulated samples used to determine classification centroids, illustrated in Panels (**a**) and (**b**). Ellipses represent one standard deviation of the bivariate distribution, with their orientation indicating the covariance of the two variables. Panels (**c**) and (**d**) present test results for four in situ data sets, and Panels (**e**) and (**f**) for synthetic data sets, with overlay points indicating optical variables as determined by previous centroids from Jackson et al. (2017) and Moore et al. (2014).

classifiable rate $P(u_{tot} > 0.1)$ of approximately 99.3% for all test data, demonstrating that nearly all samples in the test data can be effectively classified. The spectral distribution across all test data is detailed in Supporting Information Fig. S5, further evidencing the remarkable efficacy of our framework.

Application to satellite imagery

Figure 4 displays five subsets of satellite images and the distribution of dominant OWTs based on maximum membership. The first subset (Fig. 4a) includes the Wadden Sea and southern North and Baltic Sea, with transitions from OWT 6 to 4a and 3a near rivers and coasts, to OWT 1 and 2 further away, with some pixels affected by land and submerged vegetation marked as OWT 5b. The southern Baltic displays two water masses of OWTs 3a and 4a, which may indicate the inflow of saline water from the North Sea to the Baltic (Mohrholz et al. 2015). Pixels indicating *Cyanobacteria* blooms are also observed as OWTs 5a and 5b with OWT 4 as the background. Figure 4b covers the Baltic Sea and lakes in Sweden, where extensive *Cyanobacteria* blooms are observed in the Baltic Sea. Among the Swedish lakes, Lake Dalbosjön and Lake Vänern (on the west) display higher biomass and turbidity (OWT 4a) compared to Lake Vättern (OWT 3a), as also found in Eleveld et al. (2017). The third subset (Fig. 4c) features high eutrophication and *Cyanobacteria* blooms in Lake Taihu, classified as OWTs 5a, 5b, and 6, as also supported by Bi et al. (2019). Along the coast, the Yangtze River and Hangzhou Bay transition from OWT 6 near the land to OWTs 4b, 4a, 3a, and 1 toward the ocean. The last subset (Fig. 4e) shows three connected aquatic systems with distinct water types including the Black and Marmara Seas (OWT 3a) and the Aegean Sea (OWT 2). Three Turkish lakes are classified as OWTs 4a, 5a, and 5b, difference of which are also obvious from their RGB image. Satellite spectra extracted from pixels match well with mean OWT spectra, especially in shape, as shown in Supporting Information Fig. S6. However, the membership values of OWT 7 are generally lower, possibly due to atmospheric correction challenges with very dark waters (Hieronymi et al. 2023). Overall, the satellite results showcase that our framework can identify reasonable water type transitions across aquatic ecosystems and spot water events (e.g., phytoplankton blooms and turbid plumes).

Comparison with other frameworks

We converted the mean OWT spectra from Jackson et al. (2017) and Moore et al. (2014) into R_{rs} values, which were then used to calculate AVW, A_{BC}, and NDI for comparison against our framework. The centroids by Hieronymi et al. (2017) was excluded from this round due to the inability to convert their spectra to R_{rs}. The results, presented in Fig. 3e,f, reveal that the centroids from Jackson et al. (2017) align closely with our classifications for OWT 1, 2 to 3a, and the three turbid types derived from Moore et al. (2009), which correspond to our OWTs 4a and 4b. Moreover, the centroids from Moore et al. (2014) showed relatively better variability, encompassing the typical green OWTs 4a and 4b, the very turbid OWT 6, and the high biomass OWT 5b, but did not include clear (OWTs 1, 2, and 3a) and dark waters (OWT 7). Figure 5 shows the classification results by directly applying four OWT frameworks to both training and testing data. Our framework demonstrates a significantly higher rate of classifiability, $P(u_{tot} > 0.1)$, compared to other frameworks.

Discussion and conclusion

A number of studies on OWT classification have been proposed for various applications since the study by Moore et al. (2001), which applied fuzzy clustering technology to water reflectance spectra. Such emergence is mainly attributed to increasing data collection efforts (Spyrakos et al. 2018; Lehmann et al. 2023). However, data limitations have always been an obstacle to the further development of a "data-driven" framework. Even with more data collected, new clustering is required, which usually leads to different centroids due to their unsupervised features (Bi et al. 2019, 2021). The challenge of expanding the established "data-driven" framework has been demonstrated by Moore et al. (2012), where significant efforts were made in data extraction from satellites and in more complex clustering to accommodate just one additional type, let alone the need for more. The inception of our new framework was inspired by satellite data evaluations in Hieronymi et al. (2023), revealing that existing OWT frameworks are overly specific to certain water ecosystems. This necessitates a new, comprehensive framework built upon an advanced paradigm. However, the concept of a "knowledge-driven" framework relies on a sturdy forward model capable of identifying various water types, which has been achieved in our previous study (Bi et al. 2023*a*).

Comparing different OWT frameworks remains challenging due to the diversity of waveband configurations (hyperspectral or multispectral, Moore et al. 2014; Jackson et al. 2017), spectra processing methods (normalization techniques, Wei et al. 2016; Eleveld et al. 2017), and clustering approaches (fuzzy vs. hard classification, Bi et al. 2021) applied to original R_{rs} data. Our statistical analysis emphasizes total membership as an indicator of classifiability across varied spectral datasets, which is affected by the data covariance and number of dimensions (Mélin et al. 2011; Vantrepotte et al. 2012; Moore et al. 2014). The higher classification rate observed in Jackson et al. (2017) may be due to the utilization of fewer wavebands. However, the statistics do not necessarily imply that the framework of Jackson et al. (2017) is superior in classifying a greater variety of types compared to Moore et al. (2014) and Hieronymi et al. (2017). It is worth noting that while simplifying the framework, using fewer wavebands might limit its ability to effectively distinguish between complex water types, especially when spectral information beyond 670 nm is not included. We explored thresholds beyond 0.1, finding our framework consistently outperforms others in classification rate. Nonetheless, we still foresee a potential knowledge gap in the further when we will have new understandings of water optics. However, leveraging a "knowledge-driven" paradigm allows us to add new types to our existing framework without needing to reconstruct the whole work—an approach not

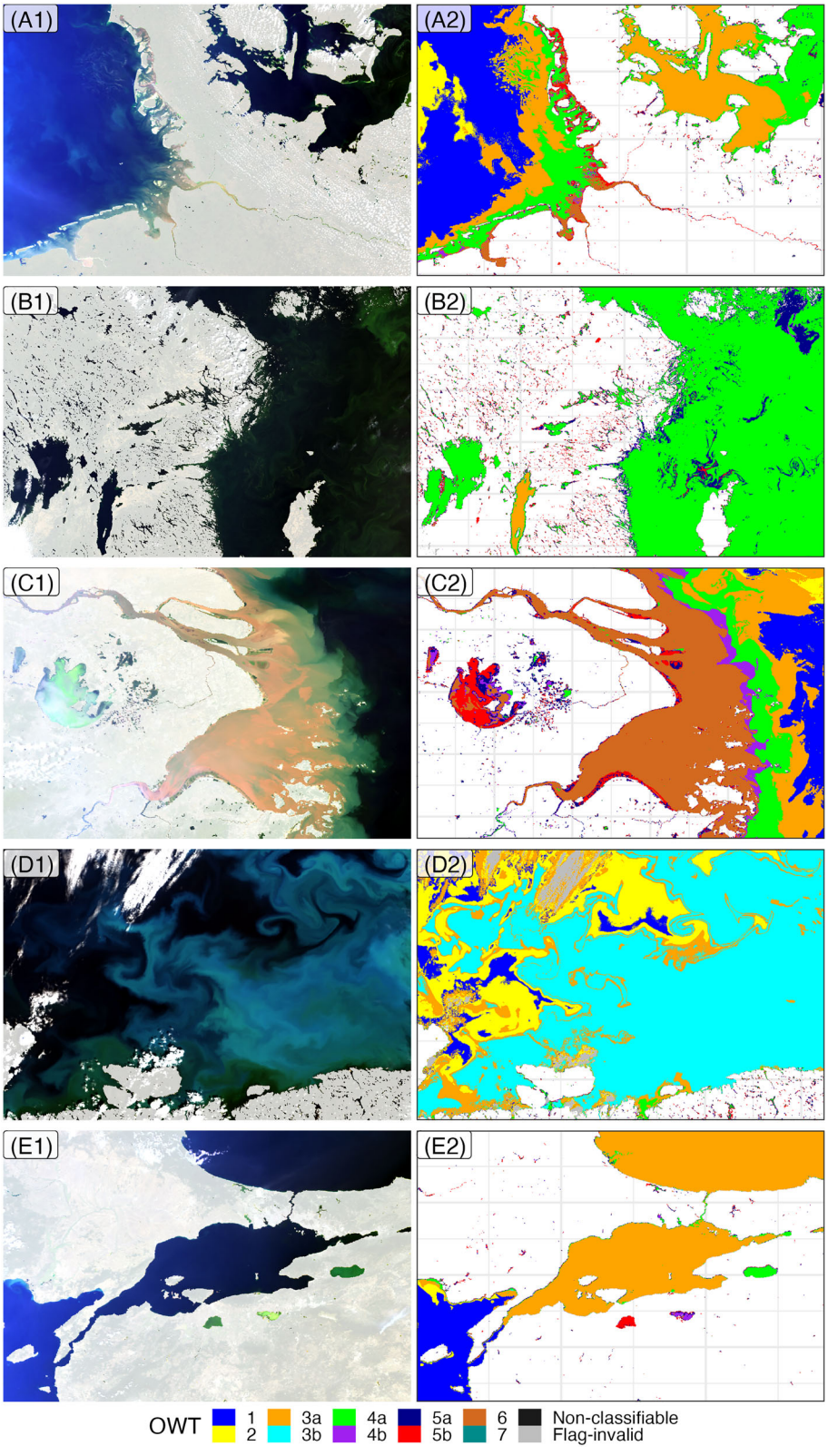

Fig. 4. Sentinel-3 OLCI satellite image subsets and their classification results. The first column displays RGB compositions of top-of-atmosphere reflectance, masking non-water pixels with reduced transparency. The second column shows classification results based on A4O-derived remote-sensing reflectance data. Panels (**a**1) and (**a**2) feature the German Bight, North Sea, and western Baltic Sea on July 20, 2016; (**b**1) and (**b**2) Swedish lakes and northern Baltic Sea on July 7, 2018; (**c**1) and (**c**2) Lake Taihu, Yangtze River, and Hangzhou Bay on May 27, 2017; (**d**1) and (**d**2) southern Barents Sea on July 7, 2016; (**e**1) and (**e**2) Marmara Sea, southern Black Sea, northern Aegean Sea, and Turkish lakes on November 13, 2017.

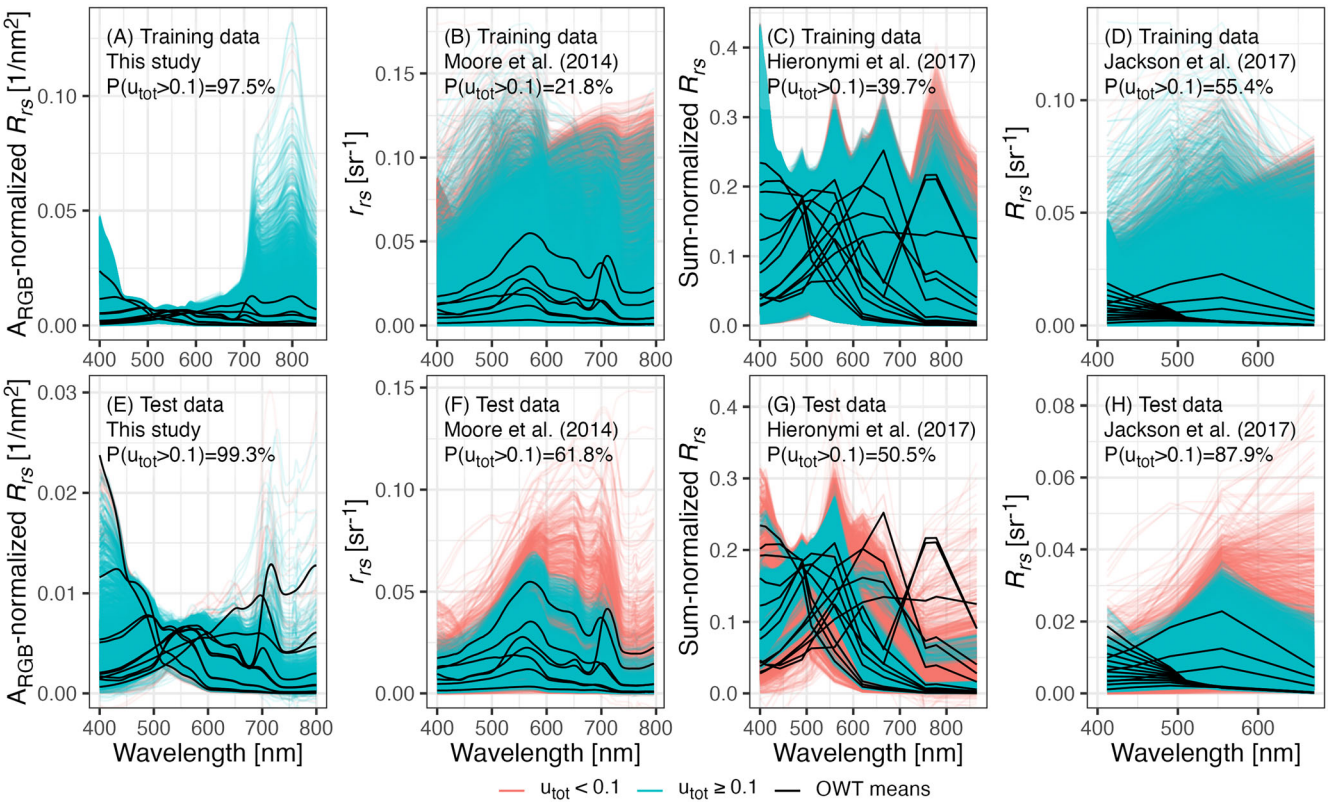

Fig. 5. Classification results of our framework and three published optical water type frameworks including Moore et al. (2014), Hieronymi et al. (2017), and Jackson et al. (2017). Spectra in panels (**a**–**d**) and panels (**e**–**h**) are the training and test data, respectively. The statistical analysis on the classifiable rate $P(u_{tot} > 0.1)$ is on the top-left corner of each panel. Black lines are the mean spectra of optical water types for each framework. The colored lines denote the classified spectra with two different u_{tot} levels.

viable with "data-driven" approaches. This paradigm will facilitate easier collaboration among researchers working on different aquatic systems and ensure the sustainable development of the current framework.

An important element of the "knowledge-driven" approach comprises an inclusive forward bio-geo-optical model that aims to encompass a wide range of variations. To convert inherent optical properties to R_{rs}, we applied the Lee et al. (2011) model due to its higher calculation efficiency and compared it to HydroLight (Mobley 1994), a fully radiative transfer model, detecting a substantial resemblance between the two (Supporting Information Fig. S7). Note that some factors were not taken into account while simulating R_{rs}, such as light-viewing geometries, inelastic scattering, volume scattering function, and heterogeneous water profiles, including surface scum (Bi et al. 2023*a*). We believe that our forward simulations offer more consistent observing conditions compared to in situ measurement data collections, which can be affected by varying conditions and measurement protocols (Lehmann et al. 2023). Moreover, the current framework has limitations in certain scenarios, such as optically shallow waters with varying bottom reflectance (Dierssen et al. 2015; Qi et al. 2020) and unique floating aggregations (e.g., Noctiluca; Luang-on et al. 2023; Kordubel et al. 2024). Addressing these scenarios will require additional knowledge from in situ observations and radiative transfer modeling in future research.

The three optical variables, AVW, A_{BC}, and NDI, serve as a method of reducing dimensionality from spectral data, effectively decreasing uncertainties in forward modeling. This approach, which conservatively combines spectral shape and magnitude, offers advantages over unsupervised dimensionality reduction techniques like principal component analysis (Uitz et al. 2015). For instance, when we extracted spectra from pixels in the German Bight and from the *Coccolithophore* blooms in the Barents Sea, we noticed that, although these spectra are tangled on a normalized scale, they significantly differ in magnitude (Supporting Information Fig. S8). These optical variables not only provide a more coherent multivariate normal distribution, enhancing the accuracy of fuzzy classification results (Campbell 1995; D'Alimonte et al. 2003; Schiller et al. 2007) but also simplify the visualization process for developing water algorithms. This is illustrated through scatter plots in Fig. 3c,d, which clarify the data coverage, aiding in the distribution analysis for training data in retrieval

algorithm development. The mapping of optical variables across satellite subsets shows how AVW effectively differentiates water types across coastal to open ocean regions for AVW < 560 nm (Supporting Information Fig. S9). However, for certain water types, such as OWTs 3a and 3b, AVW alone is insufficient for reliable differentiation. A detailed examination at AVW > 600 nm unveils a complex distribution, with OWTs 6 and 7 sharing similar AVW ranges but differing significantly in magnitude (see Fig. 3). The introduction of NDI as a third variable allows for the clear differentiation of closely related OWTs, such as 5b and 6, which share bright spectra and similar AVW ranges.

The proposed classification scheme has another key advantage for its application across various satellite missions, each with its unique waveband configurations. In previous studies, aligning common wavebands might require compromises between high and low dimensions, potentially leading to loss of useful spectral information. Otherwise, hyperspectral centroids can be proposed and resampled to fit these configurations based on spectral responding functions. However, the covariance matrix cannot be resampled directly and needs original data, and such transformation will offer variable degrees of classification freedom and impact the membership matrix outcomes (Moore et al. 2014; Jia et al. 2021; Guo et al. 2022). Our tests comparing membership matrices derived from optical variables versus spectral centroids across different satellite missions (Supporting Information Table S3) indicate that optical variables provide a more consistent membership distribution compared to traditional spectral centroid-based classification results. This consistency is crucial for applications like fuzzy clustering, for instance in algorithm blending (Jackson et al. 2017), where maintaining a stable membership distribution is more important than the type labels derived from maximum membership (Bi et al. 2019).

The application of the new OWT framework offers several key advantages. First, it can aid in testing and developing atmospheric correction processors by analyzing the classification of corrected spectral data, where unrealistic classification distributions can indicate processor underperformance for specific water types or regions (Liu et al. 2022; Hieronymi et al. 2023). This framework provides a comprehensive overview of various water types, extending beyond mere statistical analysis from limited validation points (Supporting Information Fig. S10). Additionally, the holistic design of our framework allows for better inclusion of diverse water types, suggesting that unclassified spectra from satellite spectra are presumably due to atmospheric correction errors, as evidenced in Supporting Information Fig. S11. Second, it identifies potential data gaps in measurement collections, highlighting that while the "GLORIA" dataset in our study offers extensive coverage, there is a need for more samples in oceanic waters (OWTs 1 and 2), high biomass waters (OWT 5b), and black lakes (OWT 7). Knowing the data distribution is crucial for the development of ocean color algorithms, especially neural network approaches (Hieronymi et al. 2017; Pahlevan et al. 2020). Third, the framework provides consistent type information across different aquatic systems. For example, as shown in Fig. 4, high biomass waters were classified in both oceanic and inland waters (the Baltic Sea and Lake Taihu), while and highly turbid type were identified in two different estuarine systems (the German Bight and Hangzhou Bay). This enables universal categorization and examination of the spatial and temporal dynamics of ecological and biogeochemical properties across different water bodies.

In summary, this study presents a novel OWT classification framework that transitions from traditional "data-driven" approaches to a "knowledge-driven" paradigm. By integrating forward remote sensing modeling, qualitative description, and data simulation, this framework constructs a classification encompassing ten distinct water types across diverse aquatic environments. The utilization of three significant optical variables—AVW, A_{BC}, and NDI—enables a nuanced and effective classification of water bodies, taking into account both spectral shapes and magnitudes. Demonstrated through its application to in situ, synthetic, and satellite data, our framework surpasses previous frameworks with better classifiability. By combining scientific knowledge with advanced classification techniques, this methodology provides a reliable tool for comprehending aquatic systems, identifying distinct characteristics, and guiding future research and monitoring efforts. The new framework has the capability to identify different water types based on a more holistic understanding, and its compatibility with satellite imagery allows for enhanced interpretation of intricate water systems from ocean to coast and inland. This feature makes it possible to apply the framework in practical settings such as ocean color algorithm development, water quality assessment, and ecosystem monitoring.

Data availability statement

The training data of the proposed OWT framework was generated using our previous open-source R package, which can be found on GitHub (https://github.com/bishun945/IOPmodel). The training data is available upon request from the authors. The in situ measured R_{rs} data sources are available through their respective references. The test satellite data can be accessed on Zenodo (https://zenodo.org/records/7567534). Additionally, the source code and related parameters for the proposed OWT framework are available on GitHub (https://github.com/bishun945/pyOWT).

References

Arnone, R., M. Wood, and R. Gould. 2004. The Evolution of Optical Water Mass Classification. Oceanography **17**: 14–15. doi:10.5670/oceanog.2004.42

Arvor, D., M. Belgiu, Z. Falomir, I. Mougenot, and L. Durieux. 2019. Ontologies to interpret remote sensing images: why

do we need them? GISci Remote Sens. **56**: 911–939. doi:10.1080/15481603.2019.1587890

Bi, S., and others. 2019. Optical classification of inland waters based on an improved Fuzzy C-Means method. Opt. Express **27**: 34838–34856. doi:10.1364/OE.27.034838

Bi, S., and others. 2021. Assessment of algorithms for estimating chlorophyll-a concentration in inland waters: A Round-Robin scoring method based on the optically fuzzy clustering. IEEE Trans. Geosci. Remote Sens. **60**: 1–17. doi:10.1109/TGRS.2021.3058556

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

Bi, S., R. Röttgers, and M. Hieronymi. 2023*b*. A transfer model to determine the above-water remote-sensing reflectance from the underwater remote-sensing ratio. Opt. Express. **31**: 10512–10524. doi:10.1364/OE.482395

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

Botha, E. J., J. M. Anstee, S. Sagar, E. Lehmann, and T. A. G. Medeiros. 2020. Classification of Australian waterbodies across a wide range of optical water types. Remote Sens. **12**: 3018. doi:10.3390/rs12183018

Bouveyron, C., S. Girard, and C. Schmid. 2007. High-dimensional data clustering. Comput. Stat. Data Anal. **52**: 502–519. doi:10.1016/j.csda.2007.02.009

Box, G. E., and D. R. Cox. 1964. An analysis of transformations. J. R. Stat. Soc. Ser. B Stat. Methodol. **26**: 211–243.

Cael, B. B., K. Bisson, E. Boss, and Z. K. Erickson. 2023. How many independent quantities can be extracted from ocean color? Limnol. Oceanogr. Lett. **8**: 603–610. doi:10.1002/lol2.10319

Campbell, J. W. 1995. The lognormal distribution as a model for bio-optical variability in the sea. J. Geophys. Res. **100**: 13237. doi:10.1029/95JC00458

Casey, K. A., and others. 2020. A global compilation of in situ aquatic high spectral resolution inherent and apparent optical property data for remote sensing applications. Earth Syst. Sci. Data **12**: 1123–1139. doi:10.5194/essd-12-1123-2020

Castagna, A., and others. 2022. Optical and biogeochemical properties of diverse Belgian inland and coastal waters. Earth Syst. Sci. Data **14**: 2697–2719. doi:10.5194/essd-14-2697-2022

Cazzaniga, I., G. Zibordi, and F. Mélin. 2021. Spectral variations of the remote sensing reflectance during coccolithophore blooms in the Western Black Sea. Remote Sens. Environ. **264**: 112607. doi:10.1016/j.rse.2021.112607

D'Alimonte, D., F. Melin, G. Zibordi, and J.-F. Berthon. 2003. Use of the novelty detection technique to identify the range of applicability of empirical ocean color algorithms. IEEE Trans. Geosci. Remote Sens. **41**: 2833–2843. doi:10.1109/TGRS.2003.818020

Dierssen, H. M., A. Chlus, and B. Russell. 2015. Hyperspectral discrimination of floating mats of seagrass wrack and the macroalgae Sargassum in coastal waters of Greater Florida Bay using airborne remote sensing. Remote Sens. Environ. **167**: 247–258. doi:10.1016/j.rse.2015.01.027

Eleveld, M., A. Ruescas, A. Hommersom, T. Moore, S. Peters, and C. Brockmann. 2017. An optical classification tool for global lake waters. Remote Sens. **9**: 420. doi:10.3390/rs9050420

Feng, H., J. Campbell, M. Dowell, and T. Moore. 2005. Modeling spectral reflectance of optically complex waters using bio-optical measurements from Tokyo Bay. Remote Sens. Environ. **99**: 232–243. doi:10.1016/j.rse.2005.08.015

Forel, F. 1890. Une nouvelle forme de la gamme de couleur pour létude de leau des lacs. Arch. Sci. Phys. Nat. Phys. D'histoire Nat. **6**.

Guo, Y., and others. 2022. Hyperspectral reconstruction method for optically complex inland waters based on bio-optical model and sparse representing. Remote Sens. Environ. **276**: 113045. doi:10.1016/j.rse.2022.113045

Hieronymi, M., D. Müller, and R. Doerffer. 2017. The OLCI neural network swarm (ONNS): A bio-geo-optical algorithm for open ocean and coastal waters. Front. Mar. Sci. **4**: 140. doi:10.3389/fmars.2017.00140

Hieronymi, M., and others. 2023. Ocean color atmospheric correction methods in view of usability for different optical water types. Front. Mar. Sci. **10**: 1129876. doi:10.3389/fmars.2023.1129876

IOCCG. 2006. Remote sensing of inherent optical properties: fundamentals, tests of algorithms, and applications. International Ocean Colour Coordinating Group (IOCCG).

Jackson, T., S. Sathyendranath, and F. Mélin. 2017. An improved optical classification scheme for the Ocean Colour Essential Climate Variable and its applications. Remote Sens. Environ. **203**: 152–161. doi:10.1016/j.rse.2017.03.036

Jerlov, N. G. 1968. Optical Oceanography. Elsevier.

Jia, T., Y. Zhang, and R. Dong. 2021. A universal fuzzy logic optical water type scheme for the global oceans. Remote Sens. **13**: 4018. doi:10.3390/rs13194018

Kordubel, K., B. Baschek, M. Hieronymi, Y. G. Voynova, and K. O. Möller. 2024. Improving the sampling of red *Noctiluca scintillans* to understand its impact on coastal ecosystem dynamics. J. Plankton Res. **46**: 251–271. doi:10.1093/plankt/fbae010

Kutser, T., B. Paavel, C. Verpoorter, M. Ligi, T. Soomets, K. Toming, and G. Casal. 2016. Remote sensing of black lakes and using 810 nm reflectance peak for retrieving water quality parameters of optically complex waters. Remote Sens. **8**: 497. doi:10.3390/rs8060497

Lavigne, H., and others. 2022. The HYPERMAQ dataset: bio-optical properties of moderately to extremely turbid waters.

Earth Syst. Sci. Data **14**: 4935–4947. doi:10.5194/essd-14-4935-2022

Lee, Z., and C. Hu. 2006. Global distribution of Case-1 waters: An analysis from SeaWiFS measurements. Remote Sens. Environ. **101**: 270–276. doi:10.1016/j.rse.2005.11.008

Lee, Z., K. Du, K. J. Voss, G. Zibordi, B. Lubac, R. Arnone, and A. Weidemann. 2011. An inherent-optical-property-centered approach to correct the angular effects in water-leaving radiance. Appl. Opt. **50**: 3155. doi:10.1364/AO.50.003155

Lehmann, M. K., and others. 2023. GLORIA–A globally representative hyperspectral in situ dataset for optical sensing of water quality. Sci. Data **10**: 100. doi:10.1038/s41597-023-01973-y

Liu, H., and others. 2022. Evaluation of ocean color atmospheric correction methods for sentinel-3 OLCI using global automatic *in situ* observations. IEEE Trans. Geosci. Remote Sens. **60**: 1–19. doi:10.1109/TGRS.2021.3136243

Luang-On, J., and others. 2023. MODIS-derived green Noctiluca blooms in the upper Gulf of Thailand: Algorithm development and seasonal variation mapping. Front. Mar. Sci. **10**: 1031901. doi:10.3389/fmars.2023.1031901

Martin Traykovski, L. V., and H. M. Sosik. 2003. Feature-based classification of optical water types in the Northwest Atlantic based on satellite ocean color data. J. Geophys. Res. **108**: 3150. doi:10.1029/2001JC001172

Mélin, F., V. Vantrepotte, M. Clerici, D. D'Alimonte, G. Zibordi, J.-F. Berthon, and E. Canuti. 2011. Multi-sensor satellite time series of optical properties and chlorophyll-a concentration in the Adriatic Sea. Prog. Oceanogr. **91**: 229–244.

Mélin, F., and V. Vantrepotte. 2015. How optically diverse is the coastal ocean? Remote Sens. Environ. **160**: 235–251.

Mobley, C. D. 1994. Light and water: radiative transfer in natural waters. Academic press.

Mobley, C., D. Stramski, P. Bissett, and E. Boss. 2004. Optical Modeling of Ocean Waters: Is the Case 1-Case 2 Classification Still Useful? Oceanography **17**: 60–67. doi:10.5670/oceanog.2004.48

Mohrholz, V., M. Naumann, G. Nausch, S. Krüger, and U. Gräwe. 2015. Fresh oxygen for the Baltic Sea—An exceptional saline inflow after a decade of stagnation. J. Mar. Syst. **148**: 152–166. doi:10.1016/j.jmarsys.2015.03.005

Moore, T. S., J. W. Campbell, and H. Feng. 2001. A fuzzy logic classification scheme for selecting and blending satellite ocean color algorithms. IEEE Trans. Geosci. Remote Sens. **39**: 1764–1776.

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

Moore, T. S., M. D. Dowell, and B. A. Franz. 2012. Detection of coccolithophore blooms in ocean color satellite imagery: A generalized approach for use with multiple sensors. Remote Sens. Environ. **117**: 249–263. doi:10.1016/j.rse.2011.10.001

Moore, T. S., M. D. Dowell, S. Bradt, and A. R. Verdu. 2014. An optical water type framework for selecting and blending retrievals from bio-optical algorithms in lakes and coastal waters. Remote Sens. Environ. **143**: 97–111.

Morel, A., and L. Prieur. 1977. Analysis of variations in ocean color: Ocean color analysis. Limnol. Oceanogr. **22**: 709–722. doi:10.4319/lo.1977.22.4.0709

Morel, A., H. Claustre, D. Antoine, and B. Gentili. 2007. Natural variability of bio-optical properties in Case 1 waters: attenuation and reflectance within the visible and near-UV spectral domains, as observed in South Pacific and Mediterranean waters. Biogeosciences **4**: 913–925. doi:10.5194/bg-4-913-2007

Nechad, B., and others. 2015. CoastColour Round Robin data sets: a database to evaluate the performance of algorithms for the retrieval of water quality parameters in coastal waters.

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

Neukermans, G., and G. Fournier. 2018. Optical modeling of spectral backscattering and remote sensing reflectance from emiliania huxleyi blooms. Front. Mar. Sci. **5**: 146. doi:10.3389/fmars.2018.00146

Pahlevan, N., and others. 2020. Seamless retrievals of chlorophyll-a from Sentinel-2 (MSI) and Sentinel-3 (OLCI) in inland and coastal waters: A machine-learning approach. Remote Sens. Environ. **240**: 111604. doi:10.1016/j.rse.2019.111604

Qi, L., and others. 2020. In search of floating algae and other organisms in global oceans and lakes. Remote Sens. Environ. **239**: 111659. doi:10.1016/j.rse.2020.111659

Schiller, H., W. Schönfeld, H. L. Krasemann, and K. Schiller. 2007. Novelty detection–recognition and evaluation of exceptional water reflectance spectra. Environ. Monit. Assess. **132**: 339–350. doi:10.1007/s10661-006-9538-5

Solonenko, M. G., and C. D. Mobley. 2015. Inherent optical properties of Jerlov water types. Appl. Opt. **54**: 5392. doi:10.1364/AO.54.005392

Spyrakos, E., and others. 2018. Optical types of inland and coastal waters. Limnol. Oceanogr. **63**: 846–870. doi:10.1002/lno.10674

Torrecilla, E., D. Stramski, R. A. Reynolds, E. Millán-Núñez, and J. Piera. 2011. Cluster analysis of hyperspectral optical data for discriminating phytoplankton pigment assemblages in the open ocean. Remote Sens. Environ. **115**: 2578–2593. doi:10.1016/j.rse.2011.05.014

Uitz, J., D. Stramski, R. A. Reynolds, and J. Dubranna. 2015. Assessing phytoplankton community composition from

hyperspectral measurements of phytoplankton absorption coefficient and remote-sensing reflectance in open-ocean environments. Remote Sens. Environ. **171**: 58–74. doi:10.1016/j.rse.2015.09.027

Ule, W. 1892. Die bestimmung der Wasserfarbe in den Seen. Kleinere Mitth. Dr Petermanns Mitth. Aus Justus Perthes Geogr. Anst.: 70–71.

Vandermeulen, R. A., A. Mannino, S. E. Craig, and P. J. Werdell. 2020. 150 shades of green: Using the full spectrum of remote sensing reflectance to elucidate color shifts in the ocean. Remote Sens. Environ. **247**: 111900. doi:10.1016/j.rse.2020.111900

Vantrepotte, V., H. Loisel, D. Dessailly, and X. Mériaux. 2012. Optical classification of contrasted coastal waters. Remote Sens. Environ. **123**: 306–323.

Wei, J., Z. Lee, and S. Shang. 2016. A system to measure the data quality of spectral remote sensing reflectance of aquatic environments. J. Geophys. Res. Oceans. **121**: 8189–8207. doi:10.1002/2016JC012126

Wei, J., and others. 2022. Global satellite water classification data products over oceanic, coastal, and inland waters. Remote Sens. Environ. **282**: 113233. doi:10.1016/j.rse.2022.113233

Zheng, G., and P. M. DiGiacomo. 2017. Uncertainties and applications of satellite-derived coastal water quality products. Prog. Oceanogr. **159**: 45–72. doi:10.1016/j.pocean.2017.08.007

Acknowledgment

The Helmholtz Association with the research program Earth and Environment (PoF IV) funded this study. Additional support was provided by the Hereon-I^{2}B project PhytoDive. The work is based on free and open satellite data from the European Union's Copernicus Programme provided by ESA and EUMETSAT. We would like to thank Rüdiger Röttgers, Hajo Krasemann, Kerstin Heymann, Henning Burmester, and Daniel Behr for data processing and discussion. We are also grateful to Alexandre Castagna, Moritz Lehmann, Héloïse Lavigne, Kimberly Casey, Bouchra Nechad, IOCCG, and their colleagues for providing their data sets. Our sincere thanks to Tim Moore and Thomas Jackson for sharing their OWT data. We also extend our sincere thanks to the two anonymous reviewers for their valuable comments and insights, and we are grateful to the editors for their constructive suggestions and guidance throughout the revision process. Open Access funding enabled and organized by Projekt DEAL.

Conflict of Interest

None declared.

Submitted 02 November 2023
Revised 25 March 2024
Accepted 27 May 2024

Associate editor: David Antoine