11-7-2022

Optical Classification of an Urbanized Estuary Using Hyperspectral Remote Sensing Reflectance

Kyle J. Turner
City College of New York

Maria Tzortziou
City College of New York

Brice K. Grunert
Cleveland State University, b.grunert@csuohio.edu

Joaquim Goes
Columbia University

Jonathan Sherman
City College of New York

Research Article | Vol. 30, No. 23 / 7 Nov 2022 / Optics Express 41590
Optics EXPRESS

Optical classification of an urbanized estuary using hyperspectral remote sensing reflectance

KYLE J. TURNER,[1] MARIA TZORTZIOU,[1,*] BRICE K. GRUNERT,[2] JOAQUIM GOES,[3] AND JONATHAN SHERMAN[1]

[1]The City College of New York, CUNY, New York, NY 10031, USA
[2]Cleveland State University, Cleveland, OH 44115, USA
[3]Lamont-Doherty Earth Observatory, Columbia University, Palisades, NY 10964, USA
[*]mtzortziou@ccny.cuny.edu

Abstract: Optical water classification based on remote sensing reflectance ($R_{rs}(\lambda)$) data can provide insight into water components driving optical variability and inform the development and application of bio-optical algorithms in complex aquatic systems. In this study, we use an *in situ* dataset consisting of hyperspectral $R_{rs}(\lambda)$ and other biogeochemical and optical parameters collected over nearly five years across a heavily urbanized estuary, the Long Island Sound (LIS), east of New York City, USA, to optically classify LIS waters based on $R_{rs}(\lambda)$ spectral shape. We investigate the similarities and differences of discrete groupings (k-means clustering) and continuous spectral indexing using the Apparent Visible Wavelength (AVW) in relation to system biogeochemistry and water properties. Our $R_{rs}(\lambda)$ dataset in LIS was best described by three spectral clusters, the first two accounting for the majority (89%) of $R_{rs}(\lambda)$ observations and primarily driven by phytoplankton dynamics, with the third confined to measurements in river and river plume waters. We found AVW effective at tracking subtle changes in $R_{rs}(\lambda)$ spectral shape and fine-scale water quality features along river-to-ocean gradients. The recently developed Quality Water Index Polynomial (QWIP) was applied to evaluate three different atmospheric correction approaches for satellite-derived $R_{rs}(\lambda)$ from the Sentinel-3 Ocean and Land Colour Instrument (OLCI) sensor in LIS, finding Polymer to be the preferred approach. Our results suggest that integrative, continuous indices such as AVW can be effective indicators to assess nearshore biogeochemical variability and evaluate the quality of both *in situ* and satellite bio-optical datasets, as needed for improved ecosystem and water resource management in LIS and similar regions.

1. Introduction

Coastal and estuarine waters are highly productive and dynamic ecosystems and play a critical role in biogeochemical fluxes between the land, ocean, and atmosphere. Effective monitoring of coastal ocean regions is essential to assess spatiotemporal changes in water quality (e.g., eutrophication, hypoxia), identify harmful algal blooms (HABs), and understand the local-to-global scale effects of climate change and other anthropogenic pressures (e.g., urbanization, pollution, sea level rise). Satellite remote sensing of ocean color provides a particularly useful tool for recurrent, synoptic observation of coastal ocean dynamics which is not achievable with *in situ* point sampling alone. Yet, accurate estimation and interpretation of ocean color products in coastal waters remains challenging due to significant bio-optical complexity arising from unique physical and biogeochemical processes and interactions across land-water interfaces (e.g., river inputs, adjacency effects, meteorological phenomena such as sea breezes), tidal variability, shallow water influences (e.g., bottom reflectance, sediment re-suspension), and urban air pollution [1–3]. Thus, there is an ongoing need to better characterize the optical variability and underlying drivers and processes impacting the ocean color signal across diverse coastal,

#472765 https://doi.org/10.1364/OE.472765
Journal © 2022 Received 17 Aug 2022; revised 29 Sep 2022; accepted 11 Oct 2022; published 27 Oct 2022

Corrected 28 October 2022

Optics EXPRESS

estuarine, and nearshore environments, particularly in regions heavily influenced by human activity.

The spectral remote sensing reflectance ($R_{rs}(\lambda)$; λ denotes wavelength of light) is the fundamental radiometric variable used to quantify ocean color [4]. The shape and amplitude of $R_{rs}(\lambda)$ over the ultraviolet (UV), visible (Vis) and near-infrared (NIR) portions of the electromagnetic spectrum vary according to the concentration, composition, and size of optically active in-water components, including phytoplankton, non-algal particles (NAP), and colored dissolved organic matter (CDOM). The distinct contributions of these components to the total spectral absorption ($a(\lambda)$) and backscattering ($b_b(\lambda)$) of light constitute the inherent optical properties (IOPs) of the water. A variety of empirical and semi-analytical approaches, including band-ratio [5], non-linear optimization [6], multiple linear regression [7,8], and neural network models [9], have been developed to derive IOPs and other geophysical parameters (e.g., chlorophyll-a concentration (*Chl-a*) or CDOM absorption) from satellite-derived $R_{rs}(\lambda)$ data. Other studies have focused on using $R_{rs}(\lambda)$ data to classify discrete optical water types (OWTs), waters with similar optical and biogeochemical properties, based on $R_{rs}(\lambda)$ spectral shape and/or magnitude [10–18]. OWTs have been defined globally and regionally using various statistical clustering methods (e.g., *k*-means) applied to multi- and hyperspectral $R_{rs}(\lambda)$ from both *in situ* and satellite datasets. Useful applications of OWT frameworks have included the tuning and blending of bio-optical algorithms to improve retrievals and uncertainty assessment in coastal and inland waters [15,16,19], the delineation of episodic and spatially dynamic phenomena (e.g., river plumes, algal blooms) [17,20,21], and quality assurance of $R_{rs}(\lambda)$ datasets [22].

Vandermeulen et al. [23] introduced a novel metric, the Apparent Visible Wavelength (AVW), to quantitatively describe $R_{rs}(\lambda)$ spectral shape along a continuum of wavelength values, akin to other methods which relate measured $R_{rs}(\lambda)$ to the "true" color as perceived by the human eye [24–30]. Calculated as the weighted harmonic mean of $R_{rs}(\lambda)$ wavelengths, AVW provides a gradational, one-dimensional index that can be applied to *in situ* or satellite-derived $R_{rs}(\lambda)$ datasets of varying spectral resolution. As opposed to generating discrete classes, the continuous output of AVW enables the analysis of more gradual or subtle spatiotemporal shifts and trends in ocean color and associated variables. The AVW technique is also advantageous in that it is relatively simple and intuitive, computationally cost-effective, requires no *a priori* knowledge or training for application/interpretation, and produces consistent, comparable results across datasets. Dierssen at el. [31] extended the use of AVW to develop a quality control method for hyper- and multispectral $R_{rs}(\lambda)$ data using a robust polynomial relationship between AVW and a Normalized Difference Index (NDI) of blue-green and red wavelengths derived from a large global *in situ* hyperspectral $R_{rs}(\lambda)$ dataset. The Quality Water Index Polynomial (QWIP) score, representing the difference between a measured spectrum's AVW and NDI and the predicted NDI based on the polynomial, can be an effective tool for evaluating spectral data quality of large field or satellite datasets, optimizing radiometric data processing routines, or as shown for the first time in this study, assessing the performance of atmospheric correction (AC) approaches.

The Long Island Sound (LIS), east of New York City, USA, is one of the most highly populated, heavily urbanized estuaries in the world. Although previous studies have focused on characterizing bio-optical properties in specific regions across LIS [32–36], relatively few studies have explored the bio-optical complexity and $R_{rs}(\lambda)$ variability over the full spatial extent of LIS. Using detailed measurements of $R_{rs}(\lambda)$, IOPs, *Chl-a*, total suspended matter (TSM), and particle size across major subregions of the Sound, Aurin et al. [32] found a high degree of seasonal variability in $R_{rs}(\lambda)$. They identified two predominant $R_{rs}(\lambda)$ "types" based on spectral shape: (1) "algal-driven" spectra, typical in the western LIS and during summer months, with highly peaked green reflectance and low reflectance in the blue due to a combination of strong phytoplankton, CDOM, and non-algal particle absorption, and (2) "sediment-driven" spectra in the central and eastern LIS with broadened reflectance peaks due to relatively weaker phytoplankton absorption

Optics EXPRESS

and enhanced suspended sediment backscattering. Hyperspectral above-water radiometric measurements made at the Long Island Sound Coastal Observatory (LISCO) platform [37] in the southwestern LIS over the period 2011-2014 display similar seasonal changes in $R_{rs}(\lambda)$ spectral shape, with more peaked average reflectance during summer months (i.e., August, September) and broader blue-green reflectance during late fall/early winter (i.e., November to January) [38].

Here, we build on this previous work by extending measurements to nearshore LIS waters – where some of the strongest gradients in the amount and quality of biogeochemical inputs typically occur – using a spatiotemporally rich *in situ* dataset of hyperspectral $R_{rs}(\lambda)$ and surface biogeochemistry collected across the LIS and major tributaries. The dataset captures the optical and biogeochemical complexity of this dynamic system over a period of almost five years (2017-2022) across different seasons, environmental conditions, and extreme events. We use both statistical clustering and AVW to classify LIS waters based on $R_{rs}(\lambda)$ spectral shape and relate to IOPs and biogeochemistry to understand how these methods perform and compare in this complex, urbanized estuarine system. We further explore the utility of spectral classification for evaluating different AC approaches to retrieve satellite $R_{rs}(\lambda)$ in urban-influenced coastal waters and relate how this knowledge can improve our understanding of estuarine biogeochemistry, water quality, and ecology using satellite-observed ocean color data.

2. Data and methods

2.1. In situ measurements

An *in situ* dataset consisting of hyperspectral $R_{rs}(\lambda)$ and biogeochemical and physical measurements made across LIS from September 2017 to March 2022 was used in this study. Measurements in the mainstem of LIS were made largely in collaboration with the Connecticut Department of Energy and Environmental Protection (CTDEEP) on monthly to bi-monthly Water Quality and Hypoxia surveys aboard the *R/V* John Dempsey (https://portal.ct.gov/DEEP/Water/LIS-Monitoring/LIS-Water-Quality-and-Hypoxia-Monitoring-Program-Overview) (Fig. 1). Additional targeted sampling trips were made using chartered small fishing vessels to increase the density of sampling in LIS mainstem and collect measurements in nearshore waters, including along strong salinity gradients in the Housatonic and Connecticut Rivers – two of the primary freshwater inputs to LIS accounting for more than 40% of total nitrogen loads [39]. Together, the data represent all months of the year. The following sections describe in further detail the parameters used in this study and the methods for data collection and processing. Table 1 provides a summary of this information.

Table 1. Summary of *in situ* data in LIS (September 2017- March 2022). N denotes the number of observations where the number in parentheses is the number with coincident in situ $R_{rs}(\lambda)$.

Parameter	Units/Resolution	Method/Instrument	N
Remote sensing reflectance ($R_{rs}(\lambda)$)	sr^{-1}; 1 nm from 400-800 nm	Spectra Vista Corporation HR512-i; 3C model [47]	215
Colored dissolved organic matter absorption ($a_{CDOM}(\lambda)$)	m^{-1}; 1 nm from 240-750 nm	Agilent Cary 300 UV-Vis dual-beam spectrophotometer [51]	583 (213)
Dissolved organic carbon concentration (DOC)	mg L^{-1}	Shimadzu TOC-L analyzer	392 (131)
Chlorophyll-a concentration (*Chl-a*)	mg m^{-3}	Turner Trilogy fluorometer [52]	497 (157)
Salinity	unitless	EXO2 water quality sonde	270 (105)
Dissolved organic matter fluorescence (FDOM)	QSU	EXO2 water quality sonde	270 (105)
Turbidity	FNU	EXO2 water quality sonde	270 (105)

Optics EXPRESS

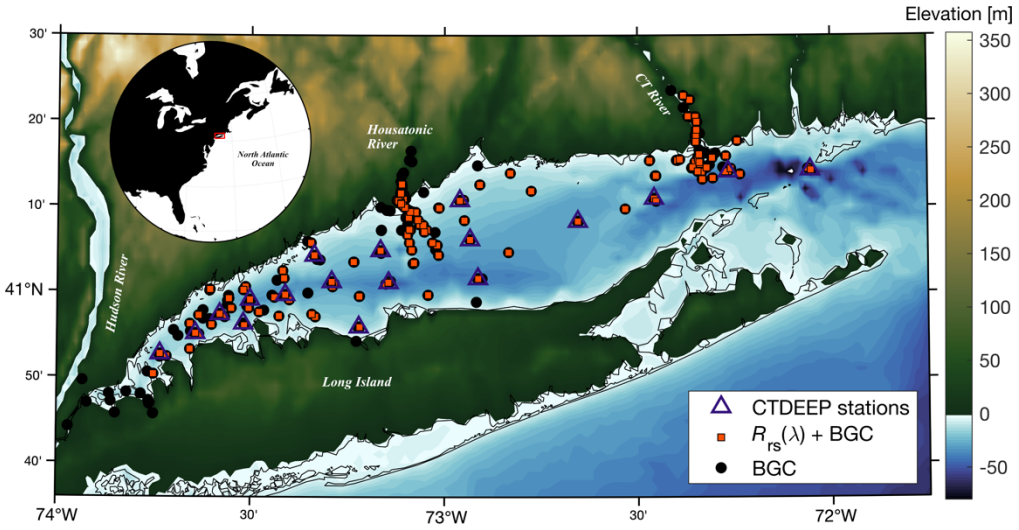

Fig. 1. Locations of *in situ* measurements in Long Island Sound (2017-2022). Orange squares indicate stations with concurrent biogeochemical (BGC) and hyperspectral above-water radiometric observations. Black circles indicate stations with only BGC measurements. Indigo triangles indicate sampling sites visited year-round by the Connecticut Department of Energy and Environmental Protection (CTDEEP).

2.1.1. Above-water radiometry and calculation of $R_{rs}(\lambda)$

Above-water radiometric measurements were made using a Spectra Vista Corporation HR512i handheld hyperspectral spectroradiometer (SVC), which observes at a spectral range of 350-1050 nm (resolution ≤1.5 nm) and a 4° field-of-view. Sampling days were chosen to target ideal conditions (clear or fully overcast skies) but data were also collected under variable cloud cover when conditions were not limited by precipitation. A total of 215 stations in LIS included radiometric observations with the SVC. To ultimately derive $R_{rs}(\lambda)$, defined as the ratio of the water-leaving radiance ($L_w(\lambda)$) to the total downwelling plane irradiance ($E_d(\lambda)$)

$$R_{rs}(\lambda) = \frac{L_w(\lambda)}{E_d(\lambda)}, \tag{1}$$

consecutive scans of the water surface, sky, and a calibrated Lambertian reference plaque (5 scans each) were taken at each station. Scans were conducted using a consistent viewing geometry following IOCCG protocols [40], with a sensor viewing angle (θ) of 40° from nadir and zenith for water and sky, respectively, and a sun relative azimuth angle (ϕ) of 90°-135°. Care was taken to avoid any foam, bubbles, or floating debris on the water surface and any disturbance (i.e., shading or reflection) from the ship or small boat structure. The upwelling radiance from the water surface ($L_u(\lambda)$) measured directly by the above-water sensor is a combination of $L_w(\lambda)$ and the water surface-reflected radiance ($L_r(\lambda)$) such that Eq. (1) can be re-written as

$$R_{rs}(\lambda) = \frac{L_u(\lambda) - L_r(\lambda)}{E_d(\lambda)}. \tag{2}$$

The most commonly used above-water radiometric approach of Mobley [41] considers $L_r(\lambda)$ as the spectrally constant fraction, ρ, of sky radiance ($L_s(\lambda)$) measured in the specular direction of

Optics EXPRESS

$L_u(\lambda)$, such that

$$R_{rs}(\lambda) = \frac{L_u(\lambda) - \rho L_s(\lambda)}{E_d(\lambda)}, \quad (3)$$

with different values for ρ provided by Mobley [41] for a range of wind speeds (a proxy for surface roughness) and viewing geometries. However, several studies [42–46] have shown that $\rho L_s(\lambda)$ does not always fully represent $L_r(\lambda)$, especially when the sky radiance distribution is non-uniform, and highlight the importance of considering the spectral dependency of ρ to more effectively remove sun and sky glint signals and accurately estimate $L_w(\lambda)$. Taking this into account, $R_{rs}(\lambda)$ was computed using the 3C model of Groetsch et al. [47] which derives a spectrally resolved offset applied to Eq. (3) by means of a full-spectrum inversion that incorporates the analytical irradiance model of Gregg and Carder [48] and the bio-optical model of Albert and Mobley [49], developed for optically complex waters. The 3C processing was applied using the publicly available Python code package (https://gitlab.com/pgroetsch/rrs_model_3C). Observed IOP values were considered for select stations to further constrain initial conditions within the 3C model, with minimal impact on output spectra; thus, for simplicity, the default initial parameterization was used. Prior to processing, outlier $E_d(\lambda)$, $L_s(\lambda)$, or $L_u(\lambda)$ spectra were filtered using the MATLAB function *isoutlier.m*, where any spectra with >25 wavelengths exceeding three scaled median absolute deviations from the median between 400 and 600 nm were excluded. For input to 3C, the mean $E_d(\lambda)$ and $L_s(\lambda)$ spectra were used and $R_{rs}(\lambda)$ was calculated for each individual $L_u(\lambda)$ scan. The spectral average $R_{rs}(\lambda)$ was used as the final estimate of $R_{rs}(\lambda)$ for each station. The full workflow for this processing is available online (https://github.com/tzortziou-lab), including example code, custom MATLAB functions and the original 3C approach [50]. Output spectra were individually assessed for data quality. Spectra with particularly elevated or negative UV/NIR reflectance, a high level of noise, or highly irregular spectral shapes were excluded from the analysis.

2.1.2. CDOM absorption and spectral slope

Discrete water samples were collected within one meter of the water surface using a Niskin bottle attached to a CTD-rosette when sampling on CTDEEP surveys or by hand when sampling via small boat into 4 L Nalgene collection bottles that were triple rinsed with sample prior to collection. All samples were filtered on the research vessel (CTDEEP) or in the laboratory on the day of collection, first through 47 mm diameter 0.7 µm pore size GF/F glass fiber filters, then through 0.22 µm pore size polycarbonate membrane filters into pre-rinsed and combusted (450° C for six hours) borosilicate amber bottles and stored in the dark at 4° C until processing (typically within 2 days, all within 1 week). Once samples were equilibrated to room temperature, the spectral absorption of CDOM ($a_{CDOM}(\lambda)$) from 240-750 nm (1 nm resolution) was measured in triplicate using an Agilent Cary 300 UV-Vis dual-beam benchtop spectrophotometer with 5 cm optical pathlength quartz cuvettes following the procedure detailed in Tzortziou et al. [53]. The spectral slope of $a_{CDOM}(\lambda)$ over the wavelength range 275-295 nm ($S_{275:295}$) was estimated for all samples as an indicator of CDOM source, molecular weight and degradation [54]. For comparison with previous studies in LIS, we also calculated the spectral slope from 412-650 nm ($S_{412:650}$) [32].

2.1.3. Dissolved organic carbon (DOC) concentration

To determine DOC concentration, the 0.22 µm sample filtrate (filtration procedure described in Section 2.1.2) was collected in duplicate into pre-rinsed and combusted (450° C for six hours) borosilicate amber vials (30 or 40 mL) and acidified with a small amount of dilute (10% v/v) hydrochloric acid for preservation (final pH between 2 and 3). Samples were stored in the dark at 4° C and shipped to the Analytical Core Lab at the University of Wisconsin - Milwaukee School

Optics EXPRESS

of Freshwater Sciences for analysis (https://uwm.edu/freshwater/research/centers/analytical-core-lab/). Duplicate values were averaged to obtain final concentrations. If one value was >50% larger than the other, the sample was excluded from the analysis.

2.1.4. *Chl-a* concentration

For analysis of *Chl-a* concentration, water samples (collection procedure described in Section 2.1.2) were filtered in duplicate under gentle vacuum onto 25 mm diameter 0.7 µm pore size GF/F filters and immediately frozen (either in liquid nitrogen or −20° C chest freezer). The volumes filtered ranged from 50-200 mL, depending on the amount required to achieve a slight but noticeable coloration to the filter. *Chl-a* concentrations were determined fluorometrically [52] using a Turner Trilogy benchtop fluorometer. Duplicate values were averaged to obtain final concentrations. If one value was >50% larger than the other, the sample was excluded from the analysis.

2.1.5. Salinity, turbidity, and FDOM

Water salinity, turbidity, and dissolved organic matter fluorescence (FDOM) measurements were obtained using an EXO2 multiparameter water quality sonde. All sonde sensors were calibrated in the laboratory prior to each cruise using the recommended standards and protocols, as described in Menendez et al. [55]. The instrument was deployed to record data continuously (every two seconds) and manually profiled through the water column up to a depth of 10 m. Data were subsequently binned into 0.5 m depth increments and averaged. Here, we only consider the near-surface measurements (0.5 m) for comparison with $R_{\mathrm{rs}}(\lambda)$ and other discretely sampled water parameters.

2.2. Classification of $R_{rs}(\lambda)$ spectral shape

2.2.1. Apparent visible wavelength (AVW)

To characterize the spectral shape of the hyperspectral $R_{\mathrm{rs}}(\lambda)$ we applied the Apparent Visible Wavelength (AVW) approach of Vandermeulen et al. [23]. AVW provides a simple, one-dimensional metric of ocean color along a continuum of wavelength values that can be applied to hyperspectral or multispectral radiometric data from in situ or satellite-based sensors. The AVW is defined as the weighted harmonic mean of $R_{\mathrm{rs}}(\lambda)$ wavelengths constrained by the relative magnitude of reflectance at each wavelength and is expressed in units of nanometers following

$$AVW = \frac{\sum_{i=\lambda_1}^{\lambda_n} R_{rs}(\lambda_i)}{\sum_{i=\lambda_1}^{\lambda_n} \frac{R_{rs}(\lambda_i)}{\lambda_i}} = \left(\frac{\sum_{i=\lambda_1}^{\lambda_n} \lambda_i^{-1} R_{rs}(\lambda_i)}{\sum_{i=\lambda_1}^{\lambda_n} R_{rs}(\lambda_i)} \right)^{-1}. \quad (4)$$

AVW has been shown to be a highly effective index of spectral shape and has the advantage of being a continuous, quantitative descriptor that utilizes the full range of spectral information and can track gradational variations and trends in ocean color over space and time. In this study, AVW was calculated over the wavelength range 400-800 nm to minimize the impacts of noise/artifacts in the UV and NIR ends of the spectrum.

2.2.2. *k*-means clustering

As an alternative method for analyzing $R_{\mathrm{rs}}(\lambda)$ spectral shape, we applied the more traditional *k*-means clustering algorithm [56] to partition spectra into a discrete number of statistically similar groups. This method, along with the related fuzzy *c*-means clustering, have been applied extensively to $R_{\mathrm{rs}}(\lambda)$ datasets from diverse aquatic environments to define OWTs based on both spectral shape and magnitude [12,13,16–18,22,57]. As we were primarily interested in comparing to definitive clusters, and the clustering results were similar between the two approaches, we

Optics EXPRESS

opted to use *k*-means over fuzzy *c*-means for this study. To focus the clustering on spectral shape features rather than first order magnitude differences, each $R_{rs}(\lambda)$ spectrum was first normalized by the integral [10,11,23,58] from 400-800 nm following

$$nR_{rs}(\lambda) = \frac{R_{rs}(\lambda)}{\int_{400}^{800} R_{rs}(\lambda)d\lambda}, \tag{5}$$

where $nR_{rs}(\lambda)$ refers to the normalized reflectance spectrum. The integration was performed using the trapezoidal method (MATLAB function *trapz.m*). To determine the optimal number of clusters (k), we used two established cluster validation procedures: the Silhouette test [59] and the Davies-Bouldin index [60]. The tests were evaluated over a range of 2-10 clusters, with the results of both indicating $k = 3$ as the optimal k for our dataset. The *k*-means algorithm was then applied to the $nR_{rs}(\lambda)$ data using the MATLAB function *kmeans.m* with 20 replicates and the default Squared Euclidean distance metric to compute the cluster centroids.

2.3. Satellite data and matchups

To explore the application of AVW and the QWIP technique [31] for assessing spectral data quality and the performance of different AC algorithms in LIS, data from the Ocean and Land Colour Instrument (OLCI) on the European Space Agency's (ESA's) Sentinel-3A and Sentinel-3B satellites were acquired from NASA's OceanColor Web (https://oceancolor.gsfc.nasa.gov/). Level-1B, full resolution (EFR) Top of Atmosphere (TOA) upwelling radiances covering the time-frame of our *in situ* measurements (2017-2021) were processed to $R_{rs}(\lambda)$ using three commonly applied AC methods: the standard NASA SeaDAS algorithm (L2gen), ACOLITE [61], and Polymer [62]. Standard quality control flags, including land and cloud flags, were applied for each AC approach, and flagged pixels were set to NaN. The satellite data were matched with the *in situ* hyperspectral $R_{rs}(\lambda)$ using the mean of a 3×3 pixel box centered on the sampling location within a ± 3-hour temporal window. Only matchups with at least 5 valid (non-NaN) pixels and a coefficient of variation (CV) between pixels of <0.15 for the 443 nm, 560 nm, and 665 nm bands were retained for analysis [63,64]. If a matchup meeting these criteria was available from both Sentinel-3A and Sentinel-3B, the average value was used.

2.4. Quality water index polymonial (QWIP) application

Based on a robust global relationship between AVW and a Normalized Difference Index (NDI) of blue-green and red wavelengths, the QWIP method has been recently introduced as a simple metric to assess the spectral quality of *in situ* or satellite-derived $R_{rs}(\lambda)$ [31]. To understand how this global relationship holds in the urban estuarine waters of LIS, we applied the QWIP to the *in situ* hyperspectral $R_{rs}(\lambda)$ measurements. We then applied the same relationship to the OLCI-derived multispectral $R_{rs}(\lambda)$ from the three AC algorithms described in Section 2.3, to assess the use of the QWIP for evaluating AC performance in coastal and nearshore systems. Following Dierssen et al. [31], AVW was calculated using Eq. (4) over the visible range (400-700 nm), as opposed to 400-800 nm, for coherence between the *in situ* and satellite data. For the OLCI AVW calculation, we applied sensor-specific offset coefficients to convert to hyperspectral-equivalent values, as described in Vandermeulen at al. [23], with updated coefficients provided in Vandermeulen [65]. The NDI at blue-green and red bands was calculated following

$$NDI = \frac{(R_{rs}(\lambda_2) - R_{rs}(\lambda_1))}{(R_{rs}(\lambda_2) + R_{rs}(\lambda_1))}, \tag{6}$$

where $\lambda_1 = 492$ nm and $\lambda_2 = 665$ nm for the hyperspectral $R_{rs}(\lambda)$ and $\lambda_1 = 490$ nm and $\lambda_2 = 665$ nm for the OLCI $R_{rs}(\lambda)$. The global QWIP relationship was calculated following

$$QWIP = p_1 AVW^4 + p_2 AVW^3 + p_3 AVW^2 + p_4 AVW + p_5, \tag{7}$$

Optics EXPRESS

where AVW is the AVW calculated from 400-700 nm and p is the set of five coefficients provided in Eq. (4) of Dierssen et al. [31]. The QWIP score was then determined as the difference between the calculated NDI and the QWIP.

3. Results and discussion

3.1. Biogeochemical variability in LIS waters

Table 2 provides a basic statistical summary of select biogeochemical and optical parameters from the full LIS dataset, collected over every season and including river/river plume observations. Where available, matching parameters from the bio-optical characterization of Aurin et al. [32] are included for comparison.

Table 2. General statistics of biogeochemical and optical parameters from the full LIS dataset (2017-2022). N denotes the number of observations. Values for matching parameters from Table 3 in Aurin et al. [32] are included for comparison.

Parameter	This study			Aurin et al. [32]		
	N	Median	Range	N	Median	Range
Chl-a [mg m^{-3}]	497	4.42	0.55-79	121	7.1	0.7-80.6
DOC [mg L^{-1}]	392	1.76	0.92-6.87	n/a	n/a	n/a
$a_{CDOM}(300)$ [m^{-1}]	583	3.55	0.79-37.35	n/a	n/a	n/a
$a_{CDOM}(440)$ [m^{-1}]	583	0.32	0.08-4.08	110	0.3	0.12-0.75
$S_{275:295}$ [nm^{-1}]	583	0.021	0.013-0.024	n/a	n/a	n/a
$S_{412:650}$ [nm^{-1}]	583	0.015	0.009-0.021	110	0.015	0.011-0.02
Turbidity [FNU]	74	2.08	0.22-8.33	n/a	n/a	n/a

The concentration of *Chl-a* in our dataset ranged from 0.55 mg m^{-3} to 79 mg m^{-3}, very close to the range reported by Aurin et al. [32], although with a slightly lower median value of 4.42 mg m^{-3}. Concentrations of DOC were also highly variable, ranging from 0.92 mg L^{-1} to 6.87 mg L^{-1}, with a median value of 1.76 mg L^{-1}, which is similar to the average values reported by Vlahos and Whitney [66] for multi-year, near-surface measurements by CTDEEP in the mainstem of LIS. CDOM absorption, spectral slope, and turbidity measurements likewise covered a very wide range of values, attesting to the contrasting water environments sampled, from the marine-influenced eastern LIS to terrestrially and anthropogenically influenced freshwater sources into the Sound. When comparing CDOM absorption at 440 nm ($a_{CDOM}(440)$) and spectral slope from 412 to 650 nm ($S_{412:650}$) to those reported by Aurin et al. [32], the median values are quite similar, however, a significantly larger range is observed in our dataset with CDOM absorption reaching more than five times higher than previously reported [32], typically associated with export of strongly absorbing, humic, high molecular weight and bioavailable DOM from the Housatonic and Connecticut Rivers [67].

3.2. AVW and relationship with other variables

Driven by the variability in water biogeochemical and inherent optical properties, measured hyperspectral $R_{rs}(\lambda)$ in LIS exhibited high variability in spectral shape and amplitude, with absolute reflectance values varying by over an order of magnitude (Fig. 2). The AVW calculated from 400 to 800 nm ranged from 524 to 597 nm, with a median value of 554 nm. For comparison, Stramski et al. [68] reported a median AVW of 460 nm for a global ocean dataset, while Valerio et al. [12] reported AVW values from 543 to 612 nm for the Lower Amazon region. It is important to note that AVW values shift depending on the wavelength range used. For example, when calculated only over the visible range, AVW for the LIS dataset ranged from 520 to 577 nm, with a median of 546 nm. The AVW values fell along a smooth gradient of spectral shape, as

Optics EXPRESS

can be seen from the $nR_{rs}(\lambda)$ plot in Fig. 2. Generally, lower AVW spectra were associated with elevated reflectance in the blue-to-green (400-550 nm) and lower reflectance in the orange-to-NIR (600-800 nm) portions of the spectrum. Spectra with intermediate AVW values showed the most pronounced green reflectance peaks, while the highest AVW spectra overall displayed relatively lower reflectance in the green with more elevated reflectance in the orange-to-NIR.

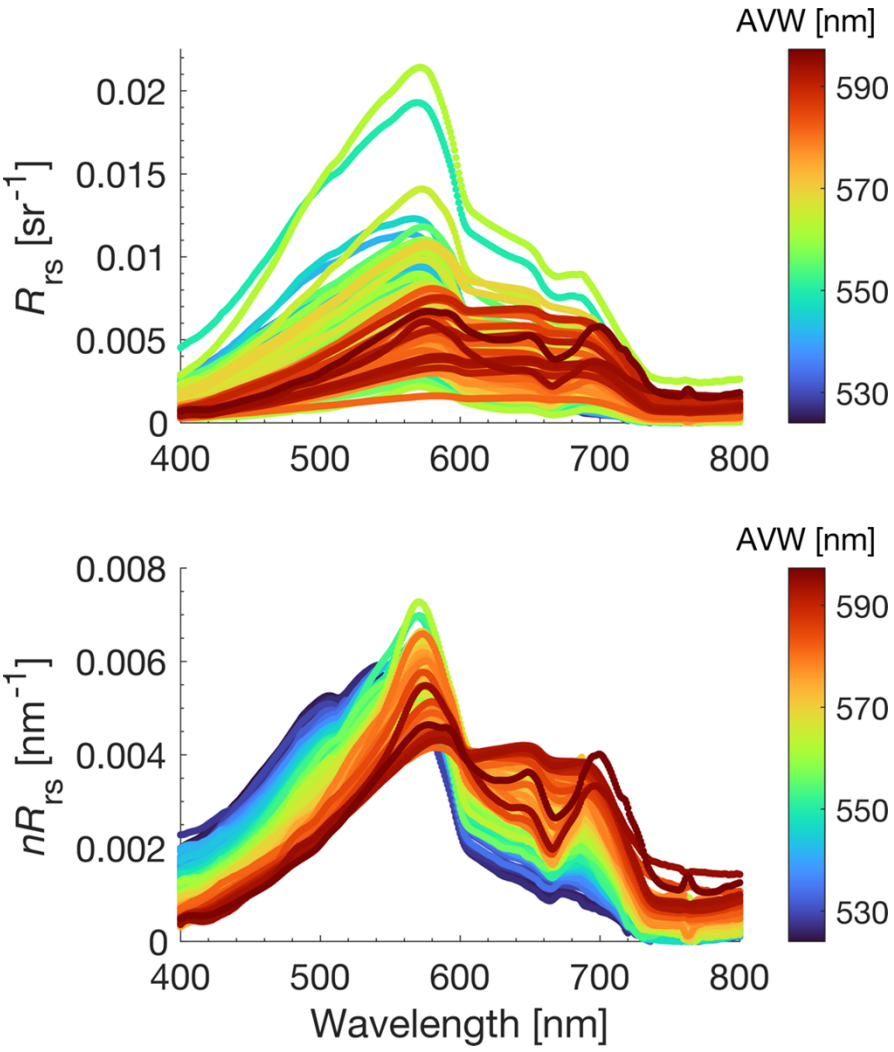

Fig. 2. *In situ* hyperspectral remote sensing reflectance ($R_{rs}(\lambda)$) (top) and integral normalized $R_{rs}(\lambda)$ (bottom) colored by Apparent Visible Wavelength (AVW).

Regression of AVW and $a_{CDOM}(300)$ revealed a moderately strong positive correlation ($r = 0.63$) between the two parameters over the full dataset (Fig. 3). This is not surprising, given that $a_{CDOM}(\lambda)$ increases exponentially toward the blue-UV and is negligible in the red, thus serving to reduce the $R_{rs}(\lambda)$ signal significantly in the blue with less impact at longer wavelengths (i.e., shifting the weighted mean to a higher value). However, a stark divergence in the relationship between AVW and $a_{CDOM}(300)$ suggested change in underlying processes driving the increase in AVW (i.e., change in spectral shape), which was further revealed by incorporating the *Chl-a* observations (Fig. 3, right). In the most basic sense, two contrasting

Optics EXPRESS

water "types" are shown to exhibit similar AVW values: (1) high CDOM, generally low *Chl-a* waters where AVW is driven by CDOM and/or co-varying non-algal particles and (2) relatively low CDOM, high *Chl-a* waters where phytoplankton have a greater impact on $R_{rs}(\lambda)$ spectral shape. When considering only points with *Chl-a* < 5 mg m^{-3}, the correlation between AVW and $a_{CDOM}(300)$ was greatly enhanced ($r = 0.89$), whereas for points with *Chl-a* > 5 mg m^{-3}, only a weak correlation ($r = 0.26$) remained (Fig. 3, right). A similar relationship was seen between AVW and *Chl-a*, which showed a weak positive correlation over the whole dataset ($r = 0.25$), but a much stronger positive correlation ($r = 0.68$) for points with *Chl-a* exceeding 5 mg m^{-3} (not shown). At these higher concentrations, increased phytoplankton absorption in the blue, increased backscattering in the green, and more pronounced chlorophyll fluorescence peaks in the red become dominant spectral shape features driving the increase in AVW.

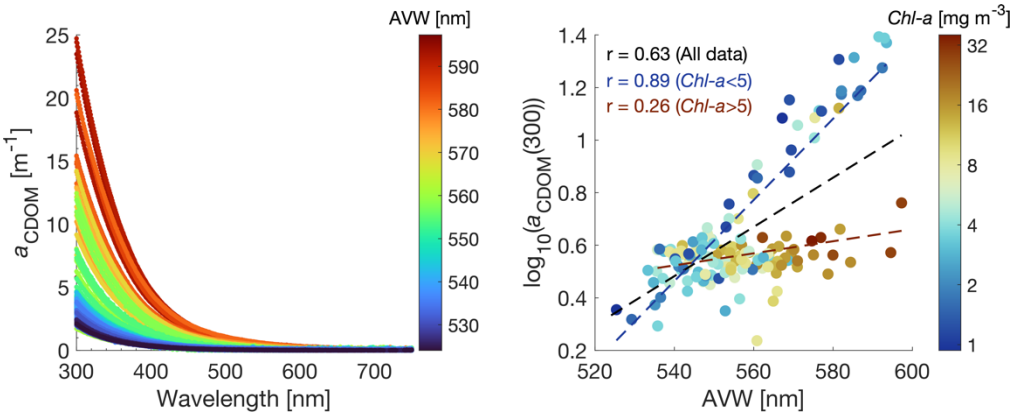

Fig. 3. CDOM absorption spectra colored by AVW (left) and the relationship between AVW and $\log_{10}$-transformed $a_{CDOM}(300)$ with points colored by *Chl-a* (right, color shown on a log scale). Dashed lines indicate Type-II major axis linear regressions and r is the Pearson correlation coefficient of each regression.

Turbidity and FDOM likewise showed positive trends with AVW (Fig. 4). In addition to AVW being impacted by the amount of DOM, it was also strongly connected to the quality of DOM, with AVW decreasing with increasing $S_{275:295}$ (Fig. 4, right). Low salinity river and river plume waters had the highest AVW values, were exceptionally high in FDOM, typically (but not always) more turbid, and had significantly lower $S_{275:295}$ than waters further into the mainstem of LIS. FDOM and $S_{275:295}$ are both highly correlated with the magnitude of CDOM absorption at blue wavelengths, while higher turbidity is associated with increased suspended particulate backscattering in the orange-to-NIR [69,70].

3.3. *k*-means clustering results

The *k*-means algorithm was applied to the n$R_{rs}(\lambda)$ spectra using $k = 3$ as the optimal number of clusters, as described in Section 2.2.2. The number of clusters reported in previous studies applying the *k*-means approach to normalized $R_{rs}(\lambda)$ varies widely. For example, Xue et al. [57] found four OWTs using a large dataset comprising multiple inland lakes in China, while Wei et al. [22] produced 23 optical clusters based on an extensive global dataset covering a range of marine environments. The resulting three optical clusters showed distinctive spectral shape features, as illustrated by their spectral cluster centers (Fig. 5, left). Cluster 1 ($N = 97$) was characterized by relatively broader and elevated blue-to-green reflectance, with lower reflectance in the orange-to-NIR spectral range. Cluster 2 ($N = 95$) had "sharper" green reflectance (~575 nm) and secondary red reflectance (~685 nm) peaks, with lower blue-to-green reflectance and

Optics EXPRESS

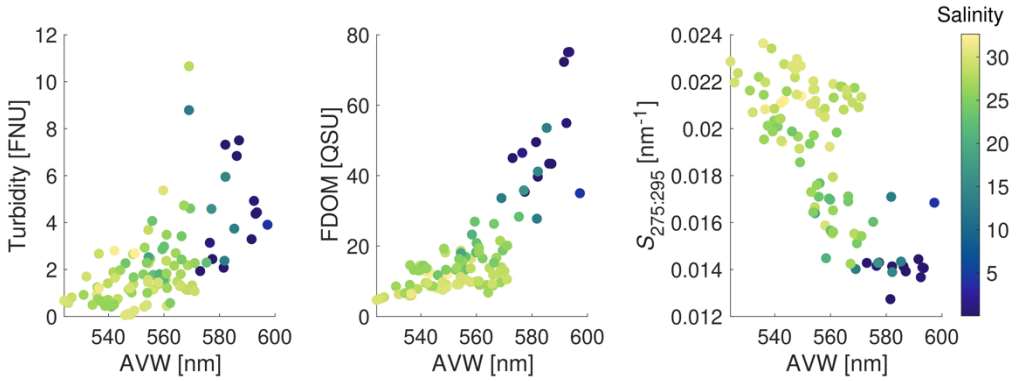

Fig. 4. Turbidity, FDOM, and $S_{275:295}$ plotted against AVW with points colored by salinity.

higher orange-to-NIR reflectance compared with Cluster 1. Cluster 3 ($N = 23$) had the lowest blue-to-green reflectance of the three clusters with significantly elevated reflectance in the orange-to-NIR.

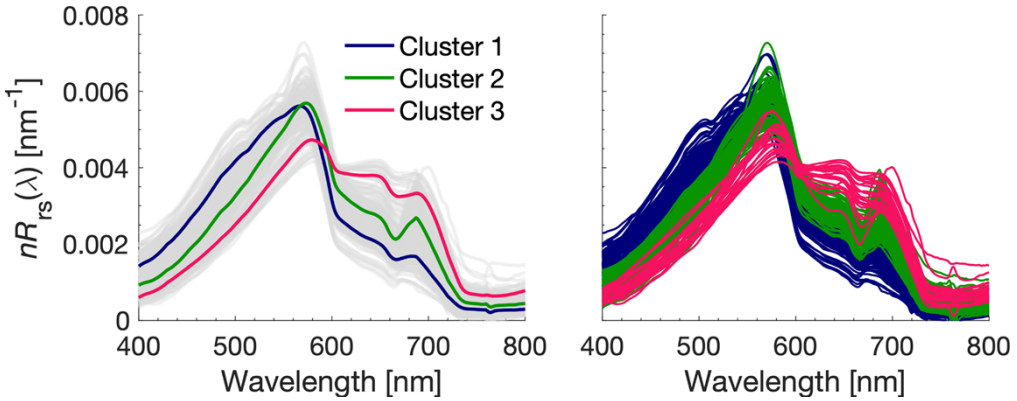

Fig. 5. Integral normalized $R_{rs}(\lambda)$ spectra with the three k-means cluster centroids overlain (left) and all spectra colored according to their assigned cluster (right).

Further description of the spectral shape and biogeochemical properties of the three n$R_{rs}(\lambda)$ clusters is provided in Fig. 6 and Table 3. The three clusters were well separated along the continuum of AVW values, highlighting the effectiveness of AVW as an index of spectral shape (Fig. 6, top left). Cluster 1 had the lowest median AVW (542 nm), and likewise the highest median blue-green/red (492/665 nm) reflectance ratio (2.40). Cluster 2 had a higher median AVW (561 nm) and lower median blue-green/red reflectance ratio (1.40) than Cluster 1. Cluster 3 had the highest median AVW (582 nm) and the lowest median blue-green/red ratio (0.68) due to the elevated red reflectance of this cluster.

All three clusters exhibited a large range of *Chl-a*; however, Cluster 2 was characterized by the highest median *Chl-a* (9.58 mg m^{-3}), more than twice that of Cluster 1 (4.03 mg m^{-3}) and three times that of Cluster 3 (2.80 mg m^{-3}). In fact, *Chl-a* (phytoplankton) appears to be the primary variable driving the shape differences between Cluster 1 and Cluster 2, as these two clusters were fairly similar in terms of CDOM and turbidity, although the median and maximum values of DOC, $a_{CDOM}(300)$, FDOM, and turbidity were consistently higher for Cluster 2 than Cluster 1. Cluster 3 is clearly distinct as riverine and river-influenced waters which are much

Optics EXPRESS

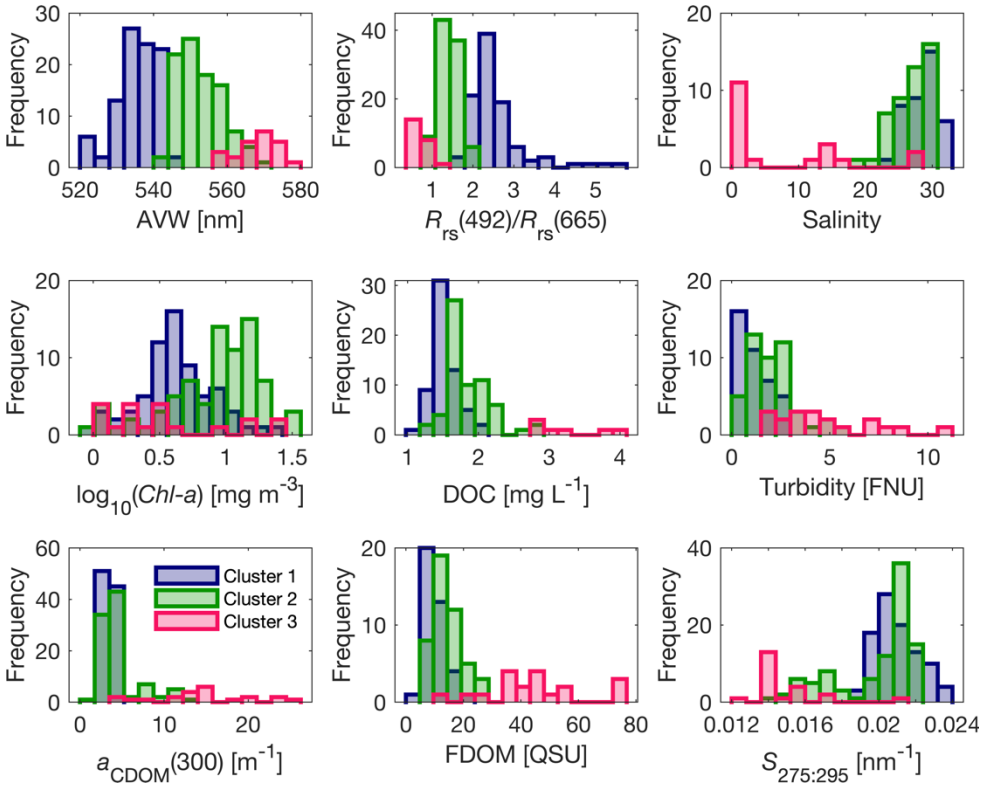

Fig. 6. Histograms of optical and biogeochemical parameters for the three *k*-means clusters.

Table 3. Statistics of the parameters shown in Fig. 6 for the three k-means clusters.

Parameter	Cluster 1 (N = 97)		Cluster 2 (N = 95)		Cluster 3 (N = 23)	
	Median	Range	Median	Range	Median	Range
AVW [nm]	542.3	521.1-556.4	561.1	551.8-584.4	582.1	568.5-597.3
$R_{rs}(492)/R_{rs}(665)$	2.4	1.73-5.5	1.4	0.95-2.04	0.68	0.51-1.08
Salinity	28.7	23.8-32.66	26.8	18.87-30.38	1.11	0.03-28.27
Chl-a [mg m^{-3}]	4.03	0.94-26.42	9.58	0.96-35.99	2.8	1.19-28.25
DOC [mg L^{-1}]	1.49	1.12-2	1.74	1.18-2.88	2.97	2.81-3.91
Turbidity [FNU]	1.07	0.06-2.8	2.02	0.41-5.37	4.38	1.94-10.66
$a_{CDOM}(300)$ [m^{-1}]	3.37	1.96-4.64	3.66	1.73-12.45	14.18	3.73-24.72
FDOM [QSU]	9.48	4.68-19.27	13.67	6.77-28.34	43.37	14.02-75.13
$S_{275:295}$ [nm^{-1}]	0.021	0.019-0.024	0.021	0.014-0.022	0.014	0.013-0.021

Optics EXPRESS

fresher, more turbid, with significantly higher DOC, $a_{\mathrm{CDOM}}(300)$ and FDOM. This cluster was also characterized by lower $S_{275:295}$, indicative of higher molecular weight, less photodegraded, terrestrially derived CDOM.

3.4. Spatial and seasonal variability in $R_{rs}(\lambda)$ spectral shape

The spatial distribution of $R_{\mathrm{rs}}(\lambda)$ spectral shape for the full dataset, indexed continuously by AVW and discretely by the three *k*-means clusters, is shown in Fig. 7. As described previously, the highest AVW was observed very nearshore within the Housatonic and Connecticut Rivers and river plume regions, coinciding with the location of *k*-means Cluster 3. AVW was also higher in the western LIS, where *k*-means Cluster 2 tended to be more dominant. The lowest AVW was observed most consistently in the deeper, more marine-influenced waters of the central and eastern LIS where *k*-means Cluster 1 was most prevalent, consistent with the higher salinity and lower FDOM and turbidity associated with this cluster. There are, however, several point locations where multiple different clusters were observed, highlighting the importance of temporal/seasonal dynamics in the processes driving spectral shape at a given location (e.g., river discharge, storm events, phytoplankton blooms). Indeed, measurements in the western LIS (west of 73° W longitude), where the dataset had the fullest temporal representation over the year, showed a strong seasonal shift in spectral shape between late fall/early winter (Nov-Jan) and summer months (Jul-Sep) (Fig. 8). In summer, one of the typical phytoplankton bloom periods in LIS [71], spectra have a more peaked shape, with lower reflectance in the blue and more prominent chlorophyll fluorescence peaks around 685 nm, consistent with previous studies [32,38,71]. This is accompanied by a positive shift in AVW, with the median AVW increasing from 545.4 nm to 564.7 nm (shifting from *k*-means Cluster 1 to Cluster 2), and a doubling of the median *Chl-a* from 3.99 to 8.03 mg m^{-3}. While *Chl-a* significantly increased in the summer, a slight overall decrease in $a_{\mathrm{CDOM}}(300)$ was observed (data not shown), confirming phytoplankton as the primary seasonal control on $R_{\mathrm{rs}}(\lambda)$ spectral shape in the western region of the Sound. It is important to note that, in addition to the increase in total algal biomass (*Chl-a*), phytoplankton community composition and size structure are also highly variable spatially and seasonally in LIS, which can have secondary impacts on phytoplankton scattering and absorption characteristics and resulting $R_{\mathrm{rs}}(\lambda)$ spectral shape [32,33,72].

3.5. Closer look at river-to-sound gradients

An open question in nearshore biogeochemistry is to what extent rivers and their plumes act as transformers, rather than transporters, of natural and anthropogenic terrigenous materials to coastal waters. To better understand the extent of optical variability within these localized zones in LIS and capture associated spatial gradients in water properties – both relevant to amount and quality of organic carbon - we conducted multiple sampling trips in the Housatonic and Connecticut Rivers and their outflows into LIS. Here, we focus on two of these trips as case studies of $R_{\mathrm{rs}}(\lambda)$ spectral variability along the river-to-Sound aquatic continuum. Figure 9 shows the n$R_{\mathrm{rs}}(\lambda)$ spectra and ancillary measurements from a transect in the Housatonic River on July 16, 2021. From upriver to the river mouth (stations 1-3; ~4.5 km distance), spectral shape and water properties were quite similar, illustrating a typical river freshwater endmember with elevated reflectance in the red and diminished reflectance in the blue resultant from higher turbidity and $a_{\mathrm{CDOM}}(300)$ with relatively low values of $S_{275:295}$ (0.014 nm^{-1}) These IOPs result in higher AVW values and the "tri-peaked" spectral shape characterized by *k*-means Cluster 3. Interestingly, a slight increase in turbidity was observed towards the river mouth accompanied by a slight increase in AVW, potentially a product of greater turbulence, vertical mixing, and flocculation in this area [73]. At station 4 (1.7 km downstream of station 3), a shift in spectral shape occurs with a slightly more pronounced green peak along with a decrease in AVW, turbidity and $a_{\mathrm{CDOM}}(300)$. From station 5 to station 7, ultimately extending ~6 km southeast of the river mouth, salinity,

Optics EXPRESS

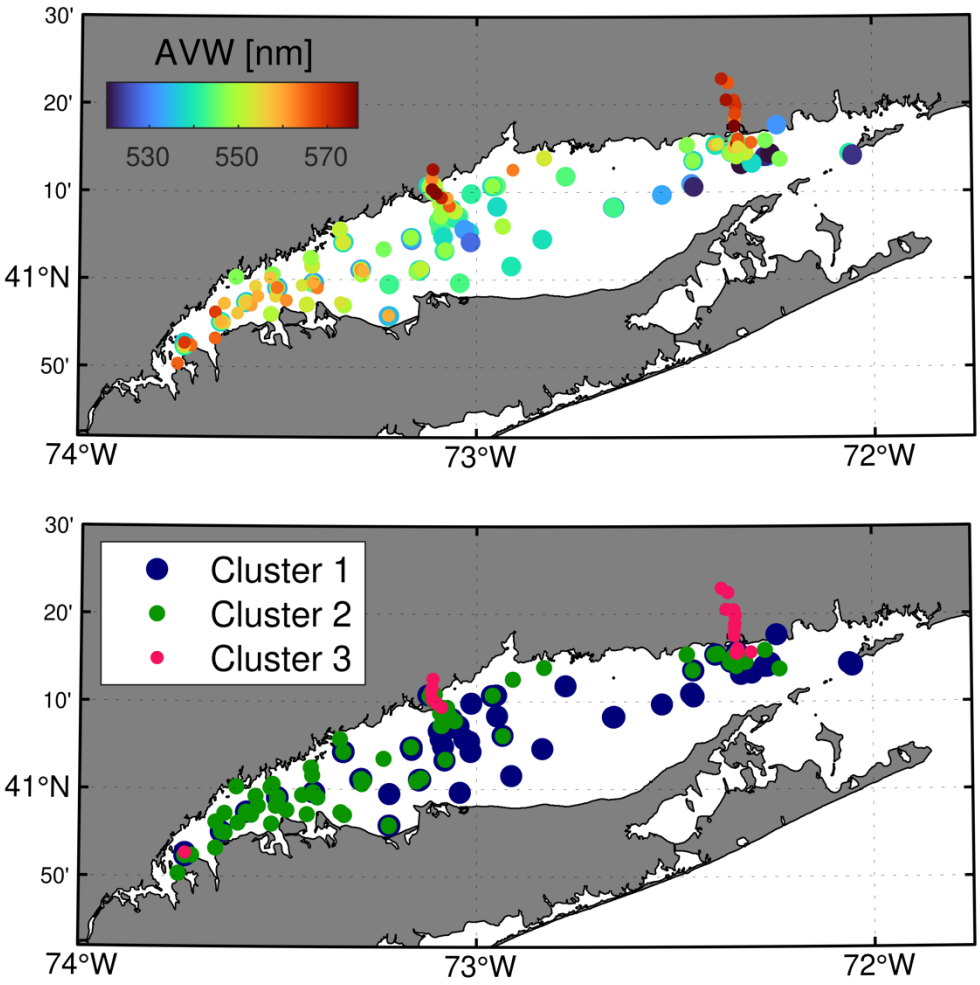

Fig. 7. Maps showing the spatial distribution of AVW (top) and the three *k*-means clusters (bottom). Marker size in the AVW map is scaled discretely by thirds of the AVW range for visual comparison (smaller marker = higher AVW).

Optics EXPRESS

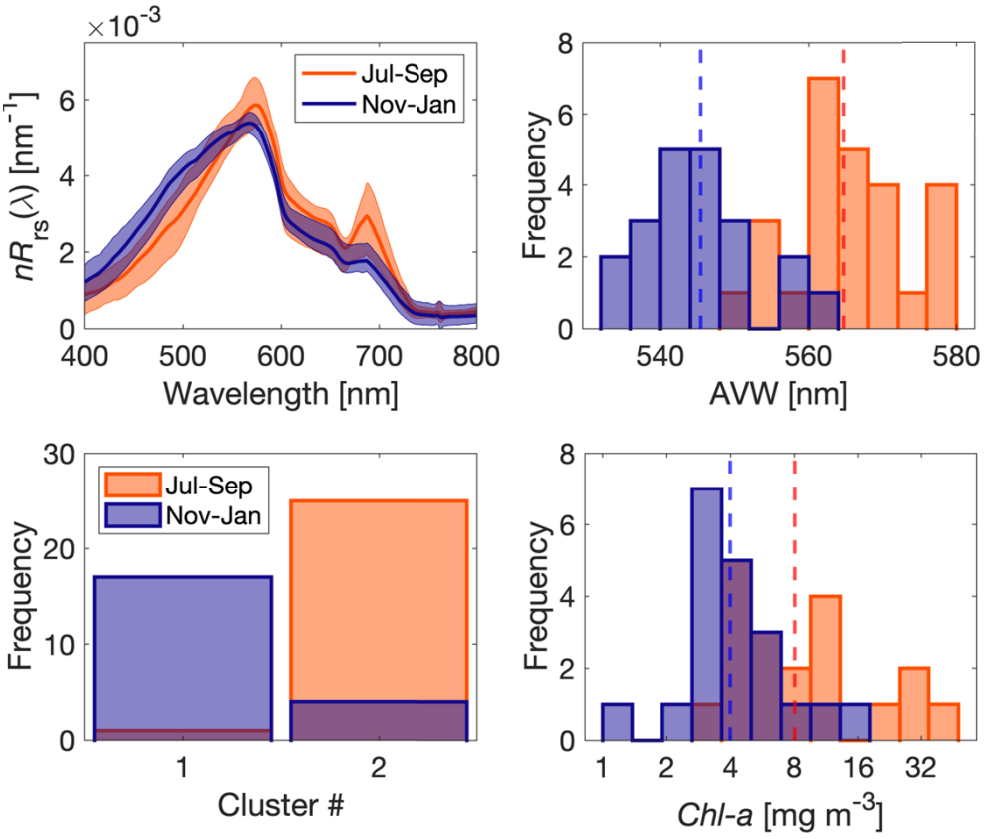

Fig. 8. Seasonal shift in $R_{rs}(\lambda)$ spectral shape, AVW, *k*-means cluster number, and *Chl-a* in the western LIS between late fall/early winter (Nov-Jan) and summer (Jul-Sep). In the spectral plot, the filled region shows the range of all spectra within each seasonal grouping and the solid line is the spectral mean. Dashed lines in the histogram plots indicate median values.

Chl-a and $S_{275:295}$ increase to their maximum values of 26.45, 4.96 mg m^{-3}, and 0.017 nm^{-1} respectively, while turbidity and $a_{CDOM}(300)$ decrease to their minimum values of 1.83 FNU and 8.06 m^{-1}. These biogeochemical changes led to even more pronounced green reflectance peaks, lower reflectance in the yellow-to-red, and lower AVW values. While stations 1-4 are all grouped into *k*-means Cluster 3, and stations 5-7 into Cluster 2, AVW is shown here to be highly effective at tracking more subtle changes in spectral shape within each cluster, integrating the impact of several water quality parameters and demonstrating an advantage of using a continuous metric over discrete classification schemes.

Figure 10 shows the same as Fig. 9 but for a sampling trip in the Connecticut River plume on May 6, 2021. Compared to the Housatonic River, the Connecticut River endmember (i.e., salinity < 1) had a very similar $nR_{rs}(\lambda)$ spectrum, with higher turbidity and lower $a_{CDOM}(300)$ (4.92 FNU and 18.83 m^{-1} vs. 4.44 FNU and 24.72 m^{-1}) at the time of observation (two months earlier in the year). Stations 1-4, representing a linear transect from the river into the Sound, displayed a high degree of spectral variability within a relatively small spatial extent (~7.5 km), including spectra from all three *k*-means clusters. From station 1 to station 2 (3.6 km downstream), both assigned to *k*-means Cluster 3, AVW and $a_{CDOM}(300)$ decreased while turbidity, salinity, $S_{275:295}$ and *Chl-a* increased. From station 2 to station 3, AVW and $a_{CDOM}(300)$ decreased

Optics EXPRESS

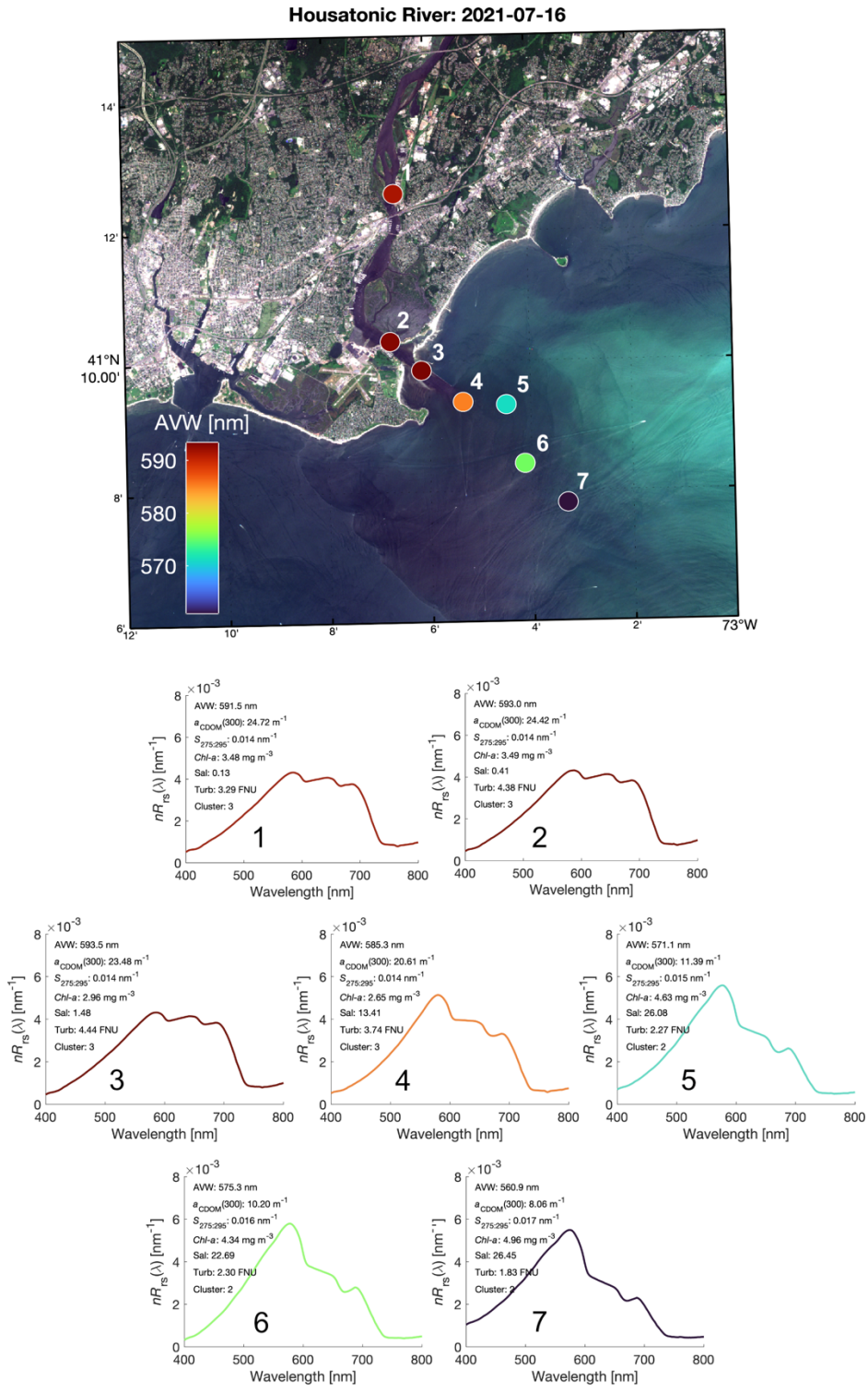

Fig. 9. Gradients in $R_{rs}(\lambda)$ spectral shape and ancillary measurements from a transect in the Housatonic River plume on July 16, 2021. Background is a Sentinel-2 MSI RGB composite (20 m resolution) from the same day, downloaded from https://scihub.copernicus.eu/.

Optics EXPRESS

substantially from 582.1 nm to 560.9 nm and 15.39 m^{-1} to 7.17 m^{-1}, respectively, along with a decrease in turbidity and *Chl-a* and a shift from *k*-means Cluster 3 to Cluster 2. From station 3 to station 4, $a_{CDOM}(300)$, turbidity, and AVW continued to decrease and *Chl-a* remained low, evident in lower reflectance across the orange-to-NIR and more domed reflectance in the blue-to-green, as characterized by *k*-means Cluster 1. The remaining stations showed varying levels of AVW, turbidity and $a_{CDOM}(300)$ depending on proximity to the plume waters. Interestingly, station 6, located within a very nearshore cove to the east of the Connecticut River mouth, had the lowest AVW, $a_{CDOM}(300)$, turbidity (536.5 nm, 2.52 m^{-1}, and 1.21 FNU, respectively), and highest salinity (31.33) and $S_{275:295}$ (0.021 nm^{-1}) of all the stations sampled. This more marine signature (i.e., Cluster 1) in such a nearshore location exemplifies the complex water circulation of the Connecticut River plume, which varies in response to the combined influence of freshwater discharge, tidal stage, and wind forcing [74].

3.6. Application of AVW and the Quality Water Index Polynomial (QWIP) to field and satellite $R_{rs}(\lambda)$ data in LIS

Figure 11 shows how the *in situ* and OLCI-derived $R_{rs}(\lambda)$ in LIS fall along the global QWIP relationship, along with histograms of the QWIP scores, defined as the difference between the measured spectrum's NDI and the QWIP. The *in situ* hyperspectral observations generally followed the QWIP, with the majority of points falling along or slightly below the polynomial. This is reflected in nearly all spectra (96%) having QWIP scores within the ±0.2 threshold used in Dierssen et al. [31] to determine "passing" spectra. The three optical clusters were well-separated along the QWIP and show similar QWIP score distributions, although Cluster 2 and Cluster 3 appeared to fall outside of the ±0.2 threshold slightly more often than Cluster 1. For the OLCI-derived $R_{rs}(\lambda)$, differences between the three AC methods were apparent. First, the number of "valid" matchups varied between the three algorithms due to differences in flagging criteria and thresholds ($N = 30$, $N = 76$, and $N = 52$ for SeaDAS, Polymer, and ACOLITE, respectively). In addition to having the highest number of matchups, the Polymer-derived data most closely followed the QWIP and generally fell more within the range of the *in situ* data than those derived from SeaDAS and ACOLITE. When viewed in terms of the QWIP scores for each method, 100% of the Polymer data had "passing" QWIP scores within ±0.2, while only 70% and 44% of the spectra from SeaDAS and ACOLITE, respectively, had scores within the ±0.2 threshold. SeaDAS showed a large spread in QWIP scores, exceeding both positive and negative 0.2, while ACOLITE had a strong positive bias due to frequent overestimation of $R_{rs}(\lambda)$ at the blue end of the spectrum. This is demonstrated by the inset comparison of $R_{rs}(\lambda)$ spectra from the three methods with the *in situ* $R_{rs}(\lambda)$ weighted to OLCI wavebands in Fig. 11. While the QWIP technique provides an assessment of $R_{rs}(\lambda)$ spectral quality based on spectral shape, validation of satellite retrievals with *in situ* data is still essential to evaluate the accuracy of reflectance magnitude. The results presented here are in agreement with Sherman et al. (*submitted*), who, using different metrics, also found that the Polymer AC approach resulted in more valid matchups and better agreement between OLCI and *in situ* $R_{rs}(\lambda)$ in LIS, compared to SeaDAS and ACOLITE.

Optics EXPRESS

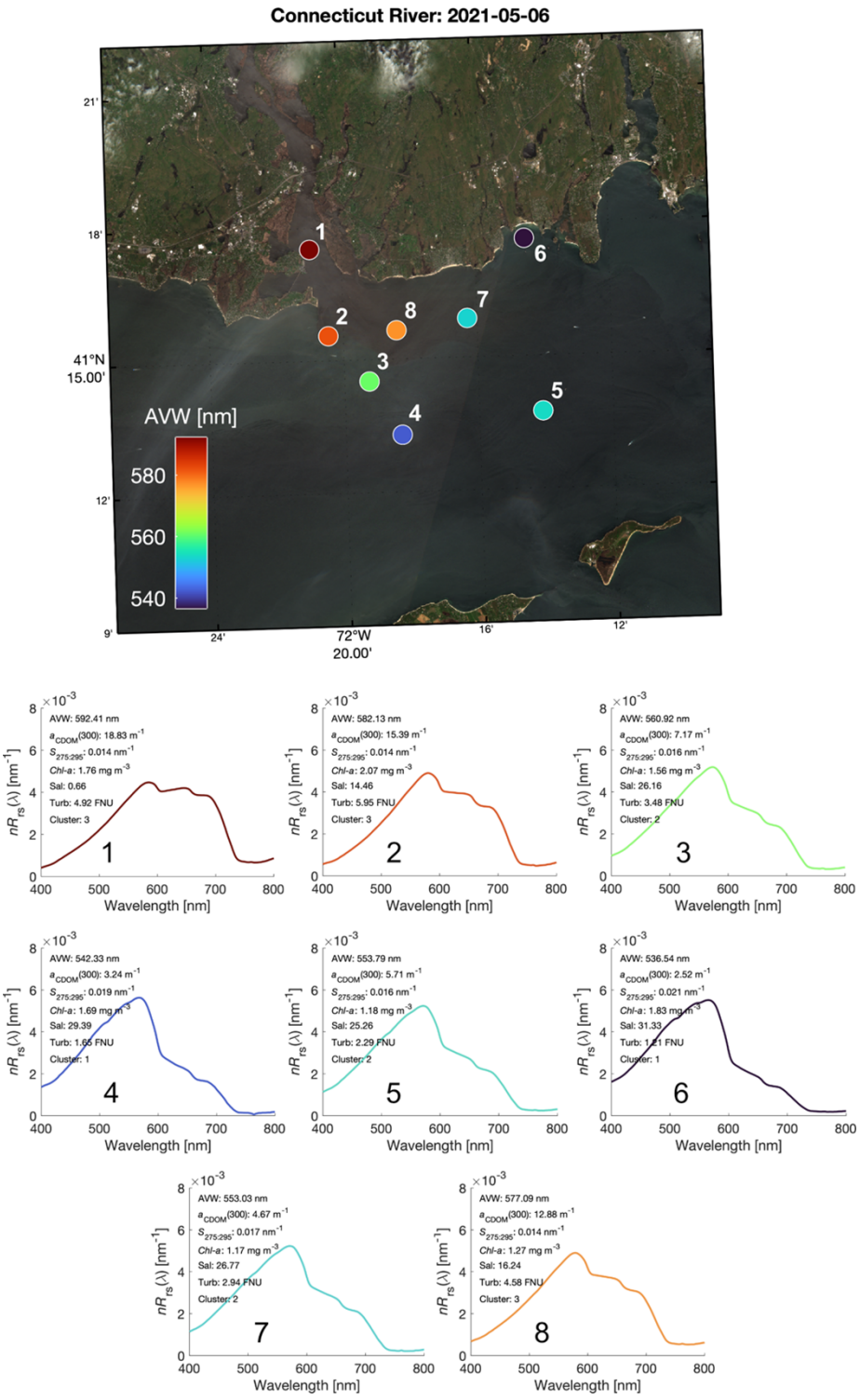

Fig. 10. Gradients in $R_{rs}(\lambda)$ spectral shape and ancillary measurements from a transect in the Connecticut River plume on May 6, 2021. Background is a Sentinel-2 MSI RGB composite (20 m resolution) from the following day, downloaded from https://scihub.copernicus.eu/.

Optics EXPRESS

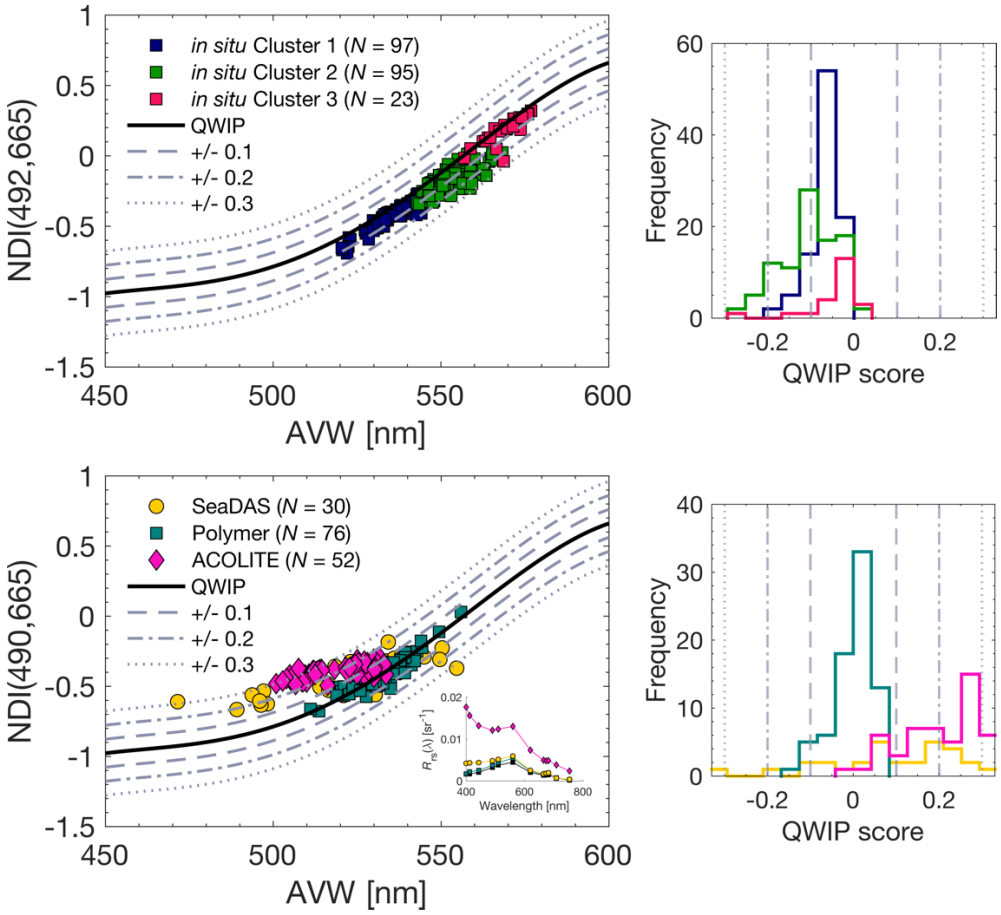

Fig. 11. The QWIP relationship between AVW and the NDI at blue-green and red bands, as described in Dierssen et al. [31], with the *in situ* hyperspectral $R_{rs}(\lambda)$ from the LIS dataset colored by *k*-means cluster membership (top left) and the satellite matchup $R_{rs}(\lambda)$ from Sentinel-3A/B OLCI (2018-2021) processed using three different AC algorithms (bottom left): SeaDAS L2gen (yellow circles), Polymer (teal squares), and ACOLITE (pink diamonds). The distribution of QWIP scores are shown with corresponding colors in the right panels. The inlaid figure shows an example of satellite-derived $R_{rs}(\lambda)$ spectra from the three AC approaches compared with the *in situ* $R_{rs}(\lambda)$ weighted to OLCI wavebands (shown in black).

4. Conclusions

The interplay of biogeochemical, ecological, and physical processes – intertwined with human influence – shapes the bio-optical complexity of coastal waters, also driving variability in $R_{rs}(\lambda)$, the fundamental quantity used in ocean color remote sensing. Here, we investigated the bio-optical characteristics of nearshore waters across a heavily urbanized and highly dynamic estuarine ecosystem and their connection to the shape of hyperspectral $R_{rs}(\lambda)$ through two optical classification approaches. Continuous spectral indexing using the AVW technique and classification using *k*-means clustering revealed comparable patterns in $R_{rs}(\lambda)$ spectral shape and underlying biogeochemistry. Our $R_{rs}(\lambda)$ dataset in LIS, spanning every season over multiple years and covering a wide range of water properties, was best described by three optical clusters,

Optics EXPRESS

which were well-separated along the continuum of AVW values and displayed notable differences in physical, biogeochemical, and optical parameters. The first two clusters described 89% of the $R_{rs}(\lambda)$ observations, the first characterized by lower AVW, *Chl-a*, turbidity, and DOC/DOM, and higher salinity and $S_{275:295}$. Phytoplankton dynamics was the strongest driver of spectral shape variability between these two clusters, with CDOM and NAP playing a secondary role. Spatially, Cluster 2, characterized by more peaked green reflectance and cholorophyll fluorescence in the red, was most prevalent in the western LIS, a region of high productivity and terrestrial influence. Cluster 1 was predominant in the more marine-influenced central and eastern LIS basins. However, strong seasonality in $R_{rs}(\lambda)$ spectral shape was observed in the western LIS, with higher *Chl-a*, AVW, and more peaked spectra (i.e., Cluster 2) during the summer bloom period (Jul-Sep) compared with less productive fall and winter months (Nov-Jan). Cluster 3 was confined to waters with a strong riverine influence and exhibited the highest AVW, a combined effect of higher orange-to-NIR and lower blue reflectance resulting from increased sediment backscattering (turbidity) and CDOM absorption, respectively.

In the river and river plume regions of the Housatonic and Connecticut Rivers, strong gradients in $R_{rs}(\lambda)$ spectral shape were observed, corresponding with changes in AVW, salinity, CDOM absorption and spectral slope, turbidity, and *Chl-a*. While all three of the optical clusters were observed in these highly dynamic, localized zones, more subtle shifts in spectral shape and water properties within each cluster were effectively captured using AVW. Compared to *k*-means clustering, AVW enabled the detection of finer-scale spatial, temporal, and spectral variability without being limited by the need for training or *a priori* determination of an optimal number of clusters. The continuous output of AVW can potentially be applied to bio-optical algorithm development, refinement, and blending in coastal and inland waters, where efforts have previously relied on OWT frameworks from traditional or "fuzzy" clustering techniques. We also demonstrated the application of AVW and the QWIP method for evaluating the quality of *in situ* hyperspectral $R_{rs}(\lambda)$ data and the performance of AC approaches for retrieving satellite $R_{rs}(\lambda)$ in urban coastal environments. From our assessment of three different AC methods applied to Sentinel-3 OLCI, we determined Polymer to be the most effective for the LIS region.

Utilizing the full range of available spectral information, AVW can be applied to the constellation of existing multispectral sensors, but is particularly well suited for hyperspectral measurements from airborne and space-based platforms including NASA's global-scale Plankton, Aerosol, Cloud, ocean Ecosystems (PACE) mission, the Geosynchronous Littoral Imaging and Monitoring Radiometer (GLIMR), and the Surface Biology and Geology (SBG) Designated Observable. Such simple but scalable and integrative indicators of water quality state would be a valuable complement to a suite of *in situ* and satellite-estimated water quality products across the continuum of inland, estuarine, and coastal waters in LIS and beyond.

Funding. National Aeronautics and Space Administration (80NSSC17K0258, 80NSSC20K1287); U.S. Environmental Protection Agency (LI96261317-0, R/CMC-17 NYCT).

Acknowledgments. The authors would like to thank CTDEEP, specifically Matt Lyman, Katie O'Brien-Clayton, and Captain Tommy Seda of the R/V John Dempsey, for their ongoing collaboration and work to collect valuable long-term data in Long Island Sound. We also sincerely thank Philipp Groetsch for his generous help with the application of the 3C glint model, Ryan Vandermeulen at NASA Goddard Space Flight Center for helpful discussions about the Apparent Visible Wavelength and Quality Water Index Polynomial methods, and Tim Wahl at the University of Wisconsin - Milwaukee School of Freshwater Sciences for analysis of dissolved organic carbon samples. Finally, the authors are grateful to the Associate Editor and anonymous reviewers who provided feedback and suggestions to significantly improve this manuscript.

Disclosures. The authors declare no conflicts of interest.

Data Availability. The CDOM absorption, DOC, and *Chl-a* data collected from 2017-2021 are publicly available on NASA's SeaBASS archive. Additional parameters, including $R_{rs}(\lambda)$ and EXO2 sonde measurements, and data collected in 2022 are in preparation for submission to SeaBASS and can be made available upon request.

Optics EXPRESS

References

1. C. B. Mouw, S. Greb, D. Aurin, P. M. DiGiacomo, Z. Lee, M. Twardowski, C. Binding, C. Hu, R. Ma, T. Moore, W. Moses, and S. E. Craig, "Aquatic color radiometry remote sensing of coastal and inland waters: Challenges and recommendations for future satellite missions," Remote Sens. Environ. **160**, 15–30 (2015).
2. M. Tzortziou, C. F. Kwong, D. Goldberg, L. Schiferl, R. Commane, N. Abuhassan, J. J. Szykman, and L. C. Valin, "Declines and peaks in NO2 pollution during the multiple waves of the COVID-19 pandemic in the New York metropolitan area," Atmos. Chem. Phys. **22**(4), 2399–2417 (2022).
3. C. P. Loughner, M. Tzortziou, S. Shroder, and K. E. Pickering, "Enhanced dry deposition of nitrogen pollution near coastlines: A case study covering the Chesapeake Bay estuary and Atlantic Ocean coastline," J. Geophys. Res. Atmos. **121**(23), 14221–14238 (2016).
4. J. T. O. Kirk, *Light and Photosynthesis in Aquatic Ecosystems* (Cambridge University Press, 1994).
5. J. E. O'Reilly and P. J. Werdell, "Chlorophyll algorithms for ocean color sensors - OC4, OC5 & OC6," Remote Sens. Environ. **229**, 32–47 (2019).
6. P. J. Werdell, L. I. W. McKinna, E. Boss, S. G. Ackleson, S. E. Craig, W. W. Gregg, Z. Lee, S. Maritorena, C. S. Roesler, C. S. Rousseaux, D. Stramski, J. M. Sullivan, M. S. Twardowski, M. Tzortziou, and X. Zhang, "An overview of approaches and challenges for retrieving marine inherent optical properties from ocean color remote sensing," Prog. Oceanogr. **160**, 186–212 (2018).
7. F. Cao and M. Tzortziou, "Capturing dissolved organic carbon dynamics with Landsat-8 and Sentinel-2 in tidally influenced wetland–estuarine systems," Sci. Tot. Environ. **777**, 145910 (2021).
8. F. Cao, M. Tzortziou, C. Hu, A. Mannino, C. G. Fichot, R. Del Vecchio, R. G. Najjar, and M. Novak, "Remote sensing retrievals of colored dissolved organic matter and dissolved organic carbon dynamics in North American estuaries and their margins," Remote Sens. Environ. **205**, 151–165 (2018).
9. N. Pahlevan, B. Smith, K. Alikas, J. Anstee, C. Barbosa, C. Binding, M. Bresciani, B. Cremella, C. Giardino, D. Gurlin, V. Fernandez, C. Jamet, K. Kangro, M. K. Lehmann, H. Loisel, B. Matsushita, N. Hà, L. Olmanson, G. Potvin, S. G. H. Simis, A. VanderWoude, V. Vantrepotte, and A. Ruiz-Verdù, "Simultaneous retrieval of selected optical water quality indicators from Landsat-8, Sentinel-2, and Sentinel-3," Remote Sens. Environ. **270**, 112860 (2022).
10. F. Mélin and V. Vantrepotte, "How optically diverse is the coastal ocean?" Remote Sens. Environ. **160**, 235–251 (2015).
11. V. Vantrepotte, H. Loisel, D. Dessailly, and X. Mériaux, "Optical classification of contrasted coastal waters," Remote Sens. Environ. **123**, 306–323 (2012).
12. A. de M. Valerio, M. Kampel, V. Vantrepotte, N. D. Ward, and J. E. Richey, "Optical Classification of Lower Amazon Waters Based on In Situ Data and Sentinel-3 Ocean and Land Color Instrument Imagery," Remote Sens. **13**(16), 3057 (2021).
13. E. F. F. da Silva, E. M. L. de M. Novo, F. de L. Lobo, C. C. F. Barbosa, M. A. Noernberg, L. H. da S. Rotta, C. T. Cairo, D. A. Maciel, and R. Flores Júnior, "Optical water types found in Brazilian waters," Limnology **22**(1), 57–68 (2021).
14. S. Monolisha, T. Platt, S. Sathyendranath, J. Jayasankar, G. George, and T. Jackson, "Optical Classification of the Coastal Waters of the Northern Indian Ocean," Front. Mar. Sci. **5**, 87 (2018).
15. T. Jackson, S. Sathyendranath, and F. Mélin, "An improved optical classification scheme for the Ocean Colour Essential Climate Variable and its applications," Remote Sens. Environ. **203**, 152–161 (2017).
16. T. S. Moore, M. D. Dowell, S. Bradt, and A. R. Verdu, "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 (2014).
17. J. T. Trochta, C. B. Mouw, and T. S. Moore, "Remote sensing of physical cycles in Lake Superior using a spatio-temporal analysis of optical water typologies," Remote Sens. Environ. **171**, 149–161 (2015).
18. E. Spyrakos, R. O'Donnell, P. D. Hunter, C. Miller, M. Scott, S. G. H. Simis, C. Neil, C. C. F. Barbosa, C. E. Binding, S. Bradt, M. Bresciani, G. Dall'Olmo, C. Giardino, A. A. Gitelson, T. Kutser, L. Li, B. Matsushita, V. Martinez-Vicente, M. W. Matthews, I. Ogashawara, A. Ruiz-Verdú, J. F. Schalles, E. Tebbs, Y. Zhang, and A. N. Tyler, "Optical types of inland and coastal waters," Limnol. Oceanogr. **63**(2), 846–870 (2018).
19. T. S. Moore, J. W. Campbell, and M. D. Dowell, "A class-based approach to characterizing and mapping the uncertainty of the MODIS ocean chlorophyll product," Remote Sens. Environ. **113**(11), 2424–2430 (2009).
20. D. Blondeau-Patissier, J. F. R. Gower, A. G. Dekker, S. R. Phinn, and V. E. Brando, "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 (2014).
21. A. W. Yu, "Assessing the spatial and temporal variability of the Detroit River and harmful algal blooms in western Lake Erie," Master of Science in Geology, Michigan Technological University (2019).
22. J. Wei, Z. Lee, and S. Shang, "A system to measure the data quality of spectral remote-sensing reflectance of aquatic environments," J. Geophys. Res. Oceans **121**(11), 8189–8207 (2016).
23. R. A. Vandermeulen, A. Mannino, S. E. Craig, and P. J. Werdell, "150 shades of green: Using the full spectrum of remote sensing reflectance to elucidate color shifts in the ocean," Remote Sens. Environ. **247**, 111900 (2020).
24. H. J. Van der Woerd and M. R. Wernand, "Hue-Angle Product for Low to Medium Spatial Resolution Optical Satellite Sensors," Remote Sens. **10**(2), 180 (2018).
25. H. J. Van der Woerd and M. R. Wernand, "True Colour Classification of Natural Waters with Medium-Spectral Resolution Satellites: SeaWiFS, MODIS, MERIS and OLCI," Sensors **15**(10), 25663–25680 (2015).

Optics EXPRESS

26. M. R. Wernand, A. Hommersom, and H. J. van der Woerd, "MERIS-based ocean colour classification with the discrete Forel–Ule scale," Ocean Sci. **9**(3), 477–487 (2013).
27. H. M. Dierssen, R. M. Kudela, J. P. Ryan, and R. C. Zimmerman, "Red and black tides: Quantitative analysis of water-leaving radiance and perceived color for phytoplankton, colored dissolved organic matter, and suspended sediments," Limnol. Oceanogr. **51**(6), 2646–2659 (2006).
28. J. Pitarch, H. J. van der Woerd, R. J. W. Brewin, and O. Zielinski, "Optical properties of Forel-Ule water types deduced from 15 years of global satellite ocean color observations," Remote Sens. Environ. **231**, 111249 (2019).
29. J. Pitarch, M. Bellacicco, S. Marullo, and H. J. van der Woerd, "Global maps of Forel–Ule index, hue angle and Secchi disk depth derived from 21 years of monthly ESA Ocean Colour Climate Change Initiative data," Earth Syst. Sci. Data **13**(2), 481–490 (2021).
30. Z. Lee, S. Shang, Y. Li, K. Luis, M. Dai, and Y. Wang, "Three-Dimensional Variation in Light Quality in the Upper Water Column Revealed With a Single Parameter," IEEE Trans. Geosci. Remote Sens. **60**, 1–10 (2022).
31. H. M. Dierssen, R. A. Vandermeulen, B. B. Barnes, A. Castagna, E. Knaeps, and Q. Vanhellemont, "QWIP: A Quantitative Metric for Quality Control of Aquatic Reflectance Spectral Shape Using the Apparent Visible Wavelength," Front. Remote Sens. **3**, 869611 (2022).
32. D. A. Aurin, H. M. Dierssen, M. S. Twardowski, and C. S. Roesler, "Optical complexity in Long Island Sound and implications for coastal ocean color remote sensing," J. Geophys. Res. **115**(C7), C07011 (2010).
33. D. A. Aurin and H. M. Dierssen, "Advantages and limitations of ocean color remote sensing in CDOM-dominated, mineral-rich coastal and estuarine waters," Remote Sens. of Environ. **125**, 181–197 (2012).
34. P. J. Werdell and C. S. Roesler, "Remote assessment of benthic substrate composition in shallow waters using multispectral reflectance," Limnol. Oceanogr **48**(1part2), 557–567 (2003).
35. A. B. Branco and J. Kremer, "The relative importance of chlorophyll and colored dissolved organic matter (CDOM) to the prediction of the diffuse attenuation coefficient in shallow estuaries," Estuaries **28**(5), 643–652 (2005).
36. S. G. Ackleson, "Optical determinations of suspended sediment dynamics in western Long Island Sound and the Connecticut River plume," J. Geophys. Res. **111**(C7), C07009 (2006).
37. T. Harmel, A. Gilerson, S. Hlaing, A. Tonizzo, T. Legbandt, A. Weidemann, R. Arnone, and S. Ahmed, "Long Island Sound Coastal Observatory: Assessment of above-water radiometric measurement uncertainties using collocated multi and hyperspectral systems," Appl. Opt. **50**(30), 5842 (2011).
38. P. M. M. Groetsch, R. Foster, and A. Gilerson, "Exploring the limits for sky and sun glint correction of hyperspectral above-surface reflectance observations," Appl. Opt. **59**(9), 2942 (2020).
39. J. Vaudrey, "*New York City's Impact on Long Island Sound Water Quality Technical Report*," (University of Connecticut, 2017).
40. IOCCG, *Ocean Optics and Biogeochemistry Protocols for Satellite Ocean Colour Sensor Validation, Volume 3.0: Protocols for Satellite Ocean Colour Data Validation: In Situ Optical Radiometry*. (International Ocean Colour Coordinating Group (IOCCG), 2019).
41. C. D. Mobley, "Estimation of the remote-sensing reflectance from above-surface measurements," Appl. Opt. **38**(36), 7442 (1999).
42. Z. Lee, Y. H. Ahn, C. Mobley, and R. Arnone, "Removal of surface-reflected light for the measurement of remote-sensing reflectance from an above-surface platform," Opt. Express **18**(25), 26313–26324 (2010).
43. T. W. Cui, Q. J. Song, J. W. Tang, and J. Zhang, "Spectral variability of sea surface skylight reflectance and its effect on ocean color," Opt. Express **21**(21), 24929–24941 (2013).
44. L. G. Sokoletsky and F. Shen, "Optical closure for remote-sensing reflectance based on accurate radiative transfer approximations: the case of the Changjiang (Yangtze) River Estuary and its adjacent coastal area, China," Int. J. Remote Sens. **35**(11-12), 4193–4224 (2014).
45. A. Gilerson, C. Carrizo, R. Foster, and T. Harmel, "Variability of the reflectance coefficient of skylight from the ocean surface and its implications to ocean color," Opt. Express **26**(8), 9615–9633 (2018).
46. X. Zhang, S. He, A. Shabani, P.-W. Zhai, and K. Du, "Spectral sea surface reflectance of skylight," Opt. Express **25**(4), A1–A13 (2017).
47. P. M. M. Groetsch, P. Gege, S. G. H. Simis, M. A. Eleveld, and S. W. M. Peters, "Validation of a spectral correction procedure for sun and sky reflections in above-water reflectance measurements," Opt. Express **25**(16), A742–A761 (2017).
48. W. W. Gregg and K. L. Carder, "A simple spectral solar irradiance model for cloudless maritime atmospheres," Limnol. Oceanogr. **35**(8), 1657–1675 (1990).
49. A. Albert and C. D. Mobley, "An analytical model for subsurface irradiance and remote sensing reflectance in deep and shallow case-2 waters," Opt. Express **11**(22), 2873–2890 (2003).
50. B. Grunert, K. Turner, and M. Tzortziou, "process_svc: v2.1," https://github.com/bricegrunert/process_svc (Github, 2022).
51. B. K. Grunert, M. Tzortziou, P. Neale, A. Menendez, and P. Hernes, "DOM degradation by light and microbes along the Yukon River-coastal ocean continuum," Sci. Rep. **11**(1), 10236 (2021).
52. J. Marra and C. J. Langdon, "*An evaluation of an in situ fluorometer for the estimation of chlorophyll a*," (Columbia University, 1993).
53. M. Tzortziou, P. J. Neale, C. L. Osburn, J. P. Megonigal, N. Maie, and R. JaffÉ, "Tidal marshes as a source of optically and chemically distinctive colored dissolved organic matter in the Chesapeake Bay," Limnol. Oceanogr. **53**(1), 148–159 (2008).

Optics EXPRESS

54. J. R. Helms, A. Stubbins, J. D. Ritchie, E. C. Minor, D. J. Kieber, and K. Mopper, "Absorption spectral slopes and slope ratios as indicators of molecular weight, source, and photobleaching of chromophoric dissolved organic matter," Limnol. Oceanogr. **53**(3), 955–969 (2008).
55. A. Menendez, M. Tzortziou, P. Neale, P. Megonigal, L. Powers, P. Schmitt-Kopplin, and M. Gonsior, "Strong dynamics in tidal marsh DOC export in response to natural cycles and episodic events from continuous monitoring," J. Geophys. Res.: Biogeosci. **127**(7), e2022JG006863 (2022).
56. J. MacQueen, "Some methods for classification and analysis of multivariate observations," Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability, Volume 1: Statistics 5.1, 281–298 (1967).
57. K. Xue, R. Ma, D. Wang, and M. Shen, "Optical Classification of the Remote Sensing Reflectance and Its Application in Deriving the Specific Phytoplankton Absorption in Optically Complex Lakes," Remote Sens. **11**(2), 184 (2019).
58. B. Lubac and H. Loisel, "Variability and classification of remote sensing reflectance spectra in the eastern English Channel and southern North Sea," Remote Sens. Environ. **110**(1), 45–58 (2007).
59. P. J. Rousseeuw, "Silhouettes: A graphical aid to the interpretation and validation of cluster analysis," J. Comput. Appl. Math. **20**, 53–65 (1987).
60. D. L. Davies and D. W. Bouldin, "A Cluster Separation Measure," IEEE Trans. Pattern Anal. Mach. Intell. **PAMI-1**(2), 224–227 (1979).
61. Q. Vanhellemont and K. Ruddick, "Atmospheric correction of Sentinel-3/OLCI data for mapping of suspended particulate matter and chlorophyll-a concentration in Belgian turbid coastal waters," Remote Sens. Environ. **256**, 112284 (2021).
62. F. Steinmetz, P. Y. Deschamps, and D. Ramon, "Atmospheric correction in presence of sun glint: application to MERIS," Opt. Express **19**(10), 9783–9800 (2011).
63. S. W. Bailey and P. J. Werdell, "A multi-sensor approach for the on-orbit validation of ocean color satellite data products," Remote Sens. Environ. **102**(1-2), 12–23 (2006).
64. F. Mélin, G. Zibordi, J. F. Berthon, S. Bailey, B. Franz, K. Voss, S. Flora, and M. Grant, "Assessment of MERIS reflectance data as processed with SeaDAS over the European seas," Opt. Express **19**(25), 25657–25671 (2011).
65. R. A. Vandermeulen, "*Apparent Visible Wavelength (AVW): NASA Algorithm Theoretical Basis Document*," (NASA, 2022).
66. P. Vlahos and M. M. Whitney, "Organic carbon patterns and budgets in the Long Island Sound estuary," Limnol. Oceanogr. **62**(S1), S46–S57 (2017).
67. J. Supino, "*Microbial Degradability of Terrigenous, Marsh-Exported, and Estuarine Colored Dissolved Organic Matter in Long Island Sound*," (City University of New York, 2020).
68. D. Stramski, I. Joshi, and R. A. Reynolds, "Ocean color algorithms to estimate the concentration of particulate organic carbon in surface waters of the global ocean in support of a long-term data record from multiple satellite missions," Remote Sens. Environ. **269**, 112776 (2022).
69. D. Doxaran, J. M. Froidefond, S. Lavender, and P. Castaing, "Spectral signature of highly turbid waters: Application with SPOT data to quantify suspended particulate matter concentrations," Remote Sens. Environ. **81**(1), 149–161 (2002).
70. Y. Luo, D. Doxaran, K. Ruddick, F. Shen, B. Gentili, L. Yan, and H. Huang, "Saturation of water reflectance in extremely turbid media based on field measurements, satellite data and bio-optical modelling," Opt. Express **26**(8), 10435 (2018).
71. F. H. Freitas and H. M. Dierssen, "Evaluating the seasonal and decadal performance of red band difference algorithms for chlorophyll in an optically complex estuary with winter and summer blooms," Remote Sens. Environ. **231**, 111228 (2019).
72. E. Suter, K. Lwiza, J. Rose, C. Gobler, and G. Taylor, "Phytoplankton assemblage changes during decadal decreases in nitrogen loadings to the urbanized Long Island Sound estuary, USA," Mar. Ecol. Prog. Ser. **497**, 51–67 (2014).
73. E. Asmala, D. G. Bowers, R. Autio, H. Kaartokallio, and D. N. Thomas, "Qualitative changes of riverine dissolved organic matter at low salinities due to flocculation," J. Geophys. Res. Biogeosci. **119**(10), 1919–1933 (2014).
74. Y. Jia and M. M. Whitney, "Summertime Connecticut River Water Pathways and Wind Impacts," J. Geophys. Res. Oceans **124**(3), 1897–1914 (2019).