Determining the optimal spectral sampling frequency and uncertainty thresholds for hyperspectral remote sensing of ocean color

RYAN A. VANDERMEULEN,[1,*] **ANTONIO MANNINO,**[2] **AIMEE NEELEY,**[1] **JEREMY WERDELL,**[2] **AND ROBERT ARNONE**[3]

[1]*Science Systems and Applications, Inc., Ocean Ecology Laboratory, NASA Goddard Space Flight Center, 8800 Greenbelt Rd. Greenbelt, MD 20771, USA*
[2]*National Aeronautics and Space Administration, Ocean Ecology Laboratory, NASA Goddard Space Flight Center, 8800 Greenbelt Rd. Greenbelt, MD 20771, USA*
[3]*University of Southern Mississippi, Department of Marine Science, 1020 Balch Blvd. Stennis Space Center, MS 39532, USA*
[*]ryan.a.vandermeulen@nasa.gov

Abstract: Using a modified geostatistical technique, empirical variograms were constructed from the first derivative of several diverse Remote Sensing Reflectance and Phytoplankton Absorbance spectra to describe how data points are correlated with "distance" across the spectra. The maximum rate of information gain is measured as a function of the kurtosis associated with the Gaussian structure of the output, and is determined for discrete segments of spectra obtained from a variety of water types (turbid river filaments, coastal waters, shelf waters, a dense *Microcystis* bloom, and oligotrophic waters), as well as individual and mixed phytoplankton functional types (PFTs; diatoms, eustigmatophytes, cyanobacteria, coccolithophores). Results show that a continuous spectrum of 5 to 7 nm spectral resolution is optimal to resolve the variability across mixed reflectance and absorbance spectra. In addition, the impact of uncertainty on subsequent derivative analysis is assessed, showing that a 3% Gaussian noise (SNR ~66) addition compromises data quality without smoothing the spectrum, and a 13% noise (SNR ~15) addition compromises data with smoothing.

References and links

1. E. J. Hochberg and M. J. Atkinson, "Capabilities of remote sensors to classify coral, algae, and sand as pure and mixed spectra," Remote Sens. Environ. **85**(2), 174–189 (2003).
2. D. R. Thompson, B. C. Gao, R. O. Green, D. A. Roberts, P. E. Dennison, and S. R. Lundeen, "Atmospheric correction for global mapping spectroscopy: ATREM advances for the HyspIRI preparatory campaign," Remote Sens. Environ. **167**, 64–77 (2015).
3. E. Devred, K. R. Turpie, W. Moses, V. V. Klemas, T. Moisan, M. Babin, M. G. Toro-Farmer, M. Forget, and Y. H. Jo, "Future retrievals of water column bio-optical properties using the Hyperspectral Infrared Imager (HyspIRI)," Remote Sens. **5**(12), 6812–6837 (2013).
4. E. L. Hestir, V. E. Brando, M. Bresciani, C. Giardino, E. Matta, P. Villa, and A. G. Dekker, "Measuring freshwater aquatic ecosystems: the need for a hyperspectral global mapping satellite mission," Remote Sens. Environ. **167**, 181–195 (2015).
5. S. E. Craig, S. E. Lohrenz, Z. Lee, K. L. Mahoney, G. J. Kirkpatrick, O. M. Schofield, and R. G. Steward, "Use of hyperspectral remote sensing reflectance for detection and assessment of the harmful alga, Karenia brevis," Appl. Opt. **45**(21), 5414–5425 (2006).
6. E. Torrecilla, D. Stramski, R. A. Reynolds, E. Millán-Núñez, and J. Piera, "Cluster analysis of hyperspectral optical data for discriminating phytoplankton pigment assemblages in the open ocean," Remote Sens. Environ. **115**(10), 2578–2593 (2011).
7. A. Bracher, M. Vountas, T. Dinter, J. P. Burrows, R. Röttgers, and I. Peeken, "Quantitative observation of cyanobacteria and diatoms from space using PhytoDOAS on SCIAMACHY data," Biogeosciences **6**(5), 751–764 (2009).

8. A. Sadeghi, T. Dinter, M. Vountas, B. B. Taylor, M. Altenburg-Soppa, I. Peeken, and A. Bracher, "Improvements to the PhytoDOAS method for identification of coccolithophores using hyper-spectral satellite data," Ocean Sci. **8**(6), 1055–1070 (2012).
9. Z. Lee, C. Hu, R. Arnone, and Z. Liu, "Impact of sub-pixel variations on ocean color remote sensing products," Opt. Express **20**(19), 20844–20854 (2012).
10. G. Meister, C. R. McClain, Z. Ahmad, S. Bailey, R. A. Barnes, S. Brown, R. E. Eplee, B. Franz, A. Holmes, W. B. Monosmith, F. S. Patt, R. P. Stumpf, K. R. Turpie, and P. J. Werdell, "Requirements for an advanced ocean radiometer," NASA Report NASA/TM—2011–215883 (2011).
11. N. Hoepffner and S. Sathyendranath, "Effect of pigment composition on absorption properties of phytoplankton," Mar. Ecol. Prog. Ser. **73**, 11–23 (1991).
12. Z. Lee and K. L. Carder, "Effect of spectral band numbers on the retrieval of water column and bottom properties from ocean color data," Appl. Opt. **41**(12), 2191–2201 (2002).
13. Z. Lee, K. Carder, R. Arnone, and M. He, "Determination of primary spectral bands for remote sensing of aquatic environments," Sensors (Basel) **7**(12), 3428–3441 (2007).
14. T. Isada, T. Hirawake, T. Kobayashi, Y. Nosaka, M. Natsuike, I. Imai, K. Suzuki, and S. Saitoh, "Hyperspectral optical discrimination of phytoplankton community structure in Funka Bay and its implications for ocean color remote sensing of diatoms," Remote Sens. Environ. **159**, 134–151 (2015).
15. D. L. Roelke, C. D. Kennedy, and A. D. Weidemann, "Use of discriminant and fourth-derivative analyses with high resolution absorption spectra for phytoplankton research: limitations at varied signal-to-noise ratio and spectral resolution," Gulf Mex. Sci. **2**, 75–86 (1999).
16. A. Wolanin, M. A. Soppa, and A. Bracher, "Investigation of spectral band requirements for improving retrievals of phytoplankton functional types," Remote Sens. **8**(10), 871 (2016).
17. C. D. Mobley, "Estimation of the remote-sensing reflectance from above-surface measurements," Appl. Opt. **38**(36), 7442–7455 (1999).
18. G. S. Fargion and J. L. Mueller, "Ocean optics protocols for satellite ocean color sensor validation, revision 2," NASA, Goddard Space Flight Center (2000).
19. A. R. Neeley, S. A. Freeman, and L. A. Harris, "Multi-method approach to quantify uncertainties in the measurements of light absorption by particles," Opt. Express **23**(24), 31043–31058 (2015).
20. M. N. Kishino, N. Takahashi, N. Okami, and S. Ichimura, "Estimation of the spectral absorption coefficients of phytoplankton in the sea," Bull. Mar. Sci. **37**, 634–642 (1985).
21. D. Stramski, R. A. Reynolds, S. Kaczmarek, J. Uitz, and G. Zheng, "Correction of pathlength amplification in the filter-pad technique for measurements of particulate absorption coefficient in the visible spectral region," Appl. Opt. **54**(22), 6763–6782 (2015).
22. W. P. Bissett, R. A. Arnone, C. O. Davis, T. D. Dickey, D. Dye, D. D. Kohler, and R. W. Gould, "From meters to kilometers," Oceanography (Wash. D.C.) **17**(2), 32–43 (2004).
23. C. O. Davis, M. Kavanaugh, R. Letelier, W. P. Bissett, and D. Kohler, "Spatial and spectral resolution considerations for imaging coastal waters," Proc. SPIE **6680**, 66800P (2007).
24. D. Aurin, A. Mannino, and B. Franz, "Spatially resolving ocean color and sediment dispersion in river plumes, coastal systems, and continental shelf waters," Remote Sens. Environ. **137**, 212–225 (2013).
25. G. Matheron, "Principles of geostatistics," Econ. Geol. **58**(8), 1246–1266 (1963).
26. A. G. Journel and C. J. Huijbregts, *Mining geostatistics* (Academic, 1978).
27. H. Xi, M. Hieronymi, R. Rottgers, H. Krasemann, and Z. Qiu, "Hyperspectral differentiation of phytoplankton taxonomic groups: a comparison between using remote sensing reflectance and absorption spectra," Remote Sens. **7**(11), 14781–14805 (2015).
28. X.-G. Xing, D.-Z. Zhao, Y.-G. Liu, J.-H. Yang, P. Xiu, and L. Wang, "An overview of remote sensing of chlorophyll fluorescence," Ocean Sci. J. **42**(1), 49–59 (2007).
29. A. A. Gitelson, "The peak near 700 nm on radiance spectra of algae and water: Relationships of its magnitude and position with chlorophyll concentration," Int. J. Remote Sens. **13**(17), 3367–3373 (1992).
30. S. W. Wright, S. W. Jeffrey, and R. F. Mantoura, *Phytoplankton pigments in oceanography: guidelines to modern methods* (Unesco, 2005).
31. F. Tsai and W. Philpot, "Derivative analysis of hyperspectral data," Remote Sens. Environ. **66**(1), 41–51 (1998).
32. A. Wolanin, V. V. Rozanov, T. Dinter, S. Noël, M. Vountas, J. P. Burrows, and A. Bracher, "Global retrieval of marine and terrestrial chlorophyll fluorescence at its red peak using hyperspectral top of atmosphere radiance measurements: feasibility study and first results," Remote Sens. Environ. **166**, 243–261 (2015).
33. M. Tzortziou, J. R. Herman, Z. Ahmad, C. P. Loughner, N. Abuhassan, and A. Cede, "Atmospheric NO2 dynamics and impact on ocean color retrievals in urban nearshore regions," J. Geophys. Res. Oceans **119**(6), 3834–3854 (2014).

1. Introduction

Near-term planning for the future launch of space-borne hyperspectral ocean color sensing systems is currently underway, including the design specifications for the National Aeronautics and Space Administration (NASA) Plankton, Aerosol, Cloud, ocean Ecosystem (PACE) mission, NASA Geostationary Coastal and Air Pollution Events (GEO-CAPE)

mission, NASA Hyperspectral Infrared Imager (HyspIRI), and the German Aerospace Center (DLR) Environmental Mapping and Analysis Program (EnMAP). Global hyperspectral measurements will have a distinct advantage over multispectral measurements, as the synoptic mapping of the fine-scale spectral features of the Earth will enable an unprecedented potential to resolve benthic substrate types [1], improve atmospheric correction [2], enhance bio-optical retrievals [3], improve monitoring of inland water quality [4], detect harmful algal species [5], distinguish specific phytoplankton pigments [6], and enable the simultaneous quantification of multiple phytoplankton functional types (PFTs) through spectral inversions [7,8] on global scales, all of which ultimately contribute to a better characterization of the global carbon budget.

However, it is noteworthy that higher spectral resolution comes at a cost of reducing the signal-to-noise ratio (SNR) of the instrument, as there is less signal (photons) reaching the detector(s) when shorter spectral intervals are sampled. This can be compensated for by modifying certain engineering parameters in the design of the sensor, but can subsequently have an adverse impact on the quality of satellite retrievals, as well as the cost and complexity of a mission. For example, increased SNR in hyperspectral measurements can be obtained by reducing the spatial resolution (i.e. larger ground sampling footprint) to allow more photons into the sensor and increase the signal, but this can also introduce additional uncertainty as a result of subpixel variation in bio-optical properties [9]. Alternatively, reducing the dynamic range of the detector(s) optimizes the sensor's ability to detect a given range of photon densities, but can lead to saturation or non-detection of radiance signals if this range is too narrow and may require multiple gain settings [10]. Therefore, it is necessary to quantify the optimal spectral sampling frequency in order to maximize the quality and efficacy of data retrievals.

Previous studies analyzing the spectral requirements for ocean color remote sensing [11–14] presented in the context of retrieving primary ocean color variables (e.g. chlorophyll-*a*, absorption, backscattering) generally conclude that a 10 nm continuous spectrum or 13-17 discrete bands are sufficient for resolving most ocean color features. However, when examining these requirements in the context of detecting subtle spectral features that require derivative analysis to extract, there can be significant improvements to PFT distinction, for example, by using a 4 nm [15] or 5 nm [16] continuous spectrum. Varying methodologies as well as data set origins yield some discrepancy among recommendations for spectral band requirements [14], and can be dependent on specific applications (e.g. detection of specific products or PFTs). This study intends to add to the existing body of work by testing a new methodology to quantify the optimal spectral resolution and noise thresholds required to discern subtle spectral features that may be important for PFT algorithm development or other applications over a variety of water-types.

2. Methods

2.1 Data collection

Hyperspectral data were collected from multiple sources to simulate a highly idealized environment (phytoplankton absorbance from filter pad-based optical density, OD_f, obtained on various cultures), as well as realistic environments with complex optical features (phytoplankton absorption, a_{ph}, and above water remote sensing reflectance, R_{RS}) from field measurements. These data were the basis for subsequent statistical analysis to help address the optimal spectral resolution at which spectral peaks and valleys can be resolved, as described in section 2.2. In order to ensure that the complete characterization of the spectral shape of peaks and valleys, data were collected at the highest available spectral resolutions.

Above-water remote sensing reflectance measurements at < 3 nm spectral resolution (sampled at 1 nm) were taken using Analytical Spectral Device (ASD) FieldSpec™ Spectroradiometers with a 10° field of view foreoptic attached. These instruments enable the derivation of above-water R_{RS} using un-calibrated radiance of the water (S_{sfc}) and sky (S_{sky}),

corrected for Fresnel reflectance ($\rho = 0.028$) [17], relative to reflectance plaque measurements (S_g). The reflectance plaque is a 10% gray card with a known bi-directional reflectance function (R_g), and is assumed to be a semi-Lambertian surface. The optical sensor zenith angles for the water (θ_{sfc}), grey card (θ_g), and sky (θ_{sky}) measurements are 135°, 135°, and 45°, respectively. The relative azimuth angle of the sensor to the sun (φ) was 135°. Remote sensing reflectance is computed following protocols from Fargion and Mueller [18]:

$$R_{RS} = \frac{S_{sfc}(\lambda) - \rho S_{sky}(\lambda)}{\pi S_g(\lambda) / R_g(\lambda)} \qquad (1)$$

Data were collected over three years (2012–2015) from various locations [Fig. 1], including the turbid Mississippi River plume (4 spectra, average chlorophyll-a = 6.18 ± 2.58 μg L^{-1}), coastal waters of the Gulf of Mexico/U.S. East Coast (4 spectra, average chlorophyll-a = 1.40 ± 0.47 μg L^{-1}), shelf waters from the Gulf of Mexico/U.S. East Coast (4 spectra, average chlorophyll-a = 0.59 ± 0.09 μg L^{-1}), the Gulf stream (4 spectra, average chlorophyll-a = 0.27 ± 0.09 μg L^{-1}), the Bahamas (4 spectra, average chlorophyll-a = 0.19 ± 0.03 μg L^{-1}), and a *Microcystsis* bloom in Lake Erie (1 spectrum, chlorophyll-a = 124.23 μg L^{-1}). The R_{RS} values were smoothed in 1 nm increments with a 5 nm boxcar Gaussian-weighted smoothing filter to avoid noise artifacts, and then segregated according to spectral shape and normalized to the maximum value in each category before running the statistical analysis, to ensure that higher magnitude reflectance spectra did not bias the results. A Gaussian-weighed boxcar smoothing filter was chosen to help minimize the potential dampening of spectral patterns, as more weight is given to the center value of the 5nm window and less so to the direct neighboring values.
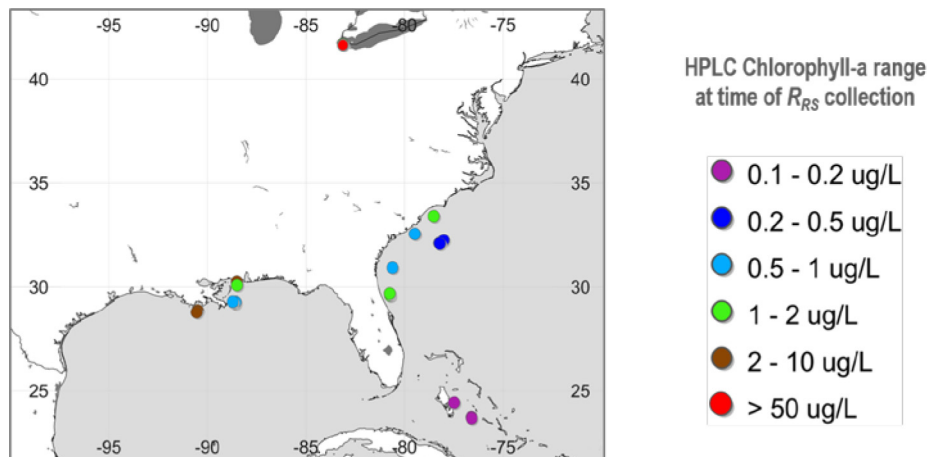

Fig. 1. Locations of above-water remote sensing reflectance measurements collected from an ASD radiometer. The corresponding range of HPLC-derived chlorophyll-a concentrations at each station are displayed in the legend, showing a diverse range of water types that were sampled.

For phytoplankton absorbance data, four cultures (*Nannochloropsis sp., Thalassiosira weissflogii, Emiliana huxleyi, Synechococcus sp.*) were grown under simulated sunlight conditions, and filtered onto GF/F filter pads. A mixed culture was obtained for spectral analysis by combining equal volumes of the dense cultures into a diluted seawater medium. The spectra were analyzed on an Agilent Cary 4000 spectrophotometer equipped with an integrating sphere (Labsphere DRA-CA-900). The filters were held in the center of the integrating sphere using jaw mount and a Plexiglas slide [19]. Data were gathered from a spectral range of 342–750 nm, at 0.3125, 0.625, 1.00, and 5.00 nm respective Slit Band Width

(SBW) and data intervals. For lab-based absorbance analysis, blank-corrected filter pad-based optical density OD_f values are utilized for statistical analysis of variability, since the corrections to obtain absolute absorption values would make minimal difference to the spectral shape. The OD_f values presented graphically are normalized to the maximum value for inter-comparisons between different PFTs, however, the maximum OD_f values were 0.245 m^{-1} *Nannochloropsis sp.*, 0.159 m^{-1} for *Thalassiosira weissflogii*, 0.194 m^{-1} for *Emiliana huxleyi*, 0.123 m^{-1} for *Synechococcus sp.*, and 0.180 m^{-1} for the mixed assemblage.

Based on findings from our field R_{RS} and laboratory OD_f experimental data, we chose a third independent data source to check the consistency of our results through a comparative analysis of phytoplankton absorption (a_{ph}) that was performed on a subset of natural seawater samples collected on a GF/F filter pad along the Yellow Sea/Korean Strait, from the GOCI validation field campaign (2013-09-27 to 2013-10-02). The extraction protocol was based on Kishino et al. [20] using two consecutive extractions of 95% methanol/5% ultrapure water. These samples were measured on an Agilent Cary 4000 spectrophotometer equipped with an integrating sphere (Labsphere DRA-CA-900). Scans were performed between 290 and 850 nm with a 2 nm SBW and 0.2 nm data interval. Particulate and detrital absorption (a_p and a_d) were computed using the pathlength amplification correction (β) from Stramski et al. [21], and a_{ph} was determined by the subtraction of a_d from a_p. These data were compiled from the NASA SeaBASS data archive (https://seabass.gsfc.nasa.gov/).

2.2. Statistical analysis

Variogram analysis is a geo-statistical technique that has been used to determine the optimal frequency of spatial sampling (i.e. spatial resolution) for resolving bio-optical processes in various water types [22–24]. In this study, the use of the empirical variogram is extended to analyze the minima and maxima of the (de-trended) first derivative of spectral data, essentially treating spectral data in the same manner as spatial data. The fundamental concept of the empirical variogram, *γ(h)*, is to quantify how data are related (correlated) with "distance":

$$\gamma(h) = \frac{1}{2|N(h)|} \sum_{N(h)} (z_i - z_j)^2 \tag{2}$$

where h is spectral distance between locations i and j (e.g. 1 nm), $N(h)$ is the number of all possible pairwise distances (e.g. all instances of 1 nm separation), z_i and z_j are the data values at locations i and j, respectively.

The empirical variogram is a measurement of the average squared difference between data separated by distance h, and each average(h) is a point on the variogram [Fig. 2(a)]. The calculation proceeds for all distances, h = 1 nm, 2 nm, 3 nm… 50 nm. At some point, the variance measured at different distances (spectral resolutions) reaches a maximum, indicating the data are no longer auto-correlated, and this point represents a measurement of the variance of the random field (the sill, σ_n^2, [Fig. 2(a)]. The range, a, describes the distance at which data are not auto-correlated, and the nugget (c_n, y-intercept) represents the uncertainty, or micro-scale variations in the data, or both [25]. The variogram can be mathematically represented by one of several models depending on the data structure. In this case, the data were best represented by a Gaussian curve:

$$g(h) = c_n + \sigma_n^2 \left(1 - \exp\left(\frac{-h^2}{a^2} \right) \right) \tag{3}$$
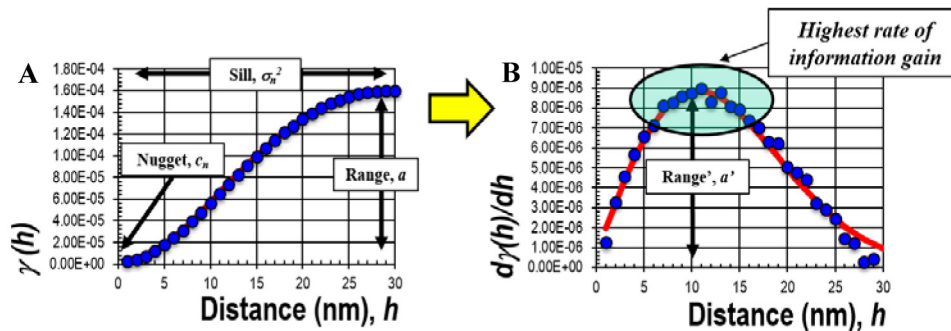
Fig. 2. (A) Empirical variogram performed on the first derivative of spectral phytoplankton absorbance data. Note the relative "flattening" of the tail at the origin of the Gaussian curve. This feature is exploited to gain the maximal rate of information gain, as seen in (B). This distance, or spectral resolution, at which $d\gamma(h)/dh$ is maximized is interpreted as the optimal spectral sampling frequency.

Note that the variance decreases significantly as h approaches zero [Fig. 2(a)], i.e. the curve flattens out and little information is gained with every nanometer increase in spectral resolution. This "flattening" suggests that the minimum resolution of the spectral data is sufficient to resolve the variability of the underlying patterns. The degree to which this variance changes with every nm increase in resolution will change as a function of the kurtosis of the Gaussian curve. To quantify this, the rate of change in variance, $d\gamma(h)/dh$, can be plotted as a function of spectral resolution, emphasizing the sampling frequency at which more/less information is gained [Fig. 2(b)]. The maximum value of this plot shows the location of the maximum rate of information gain. Below this point (higher spectral resolution), there is less relative information gained per nanometer of spectral resolution increase, which indicates that increasing the spectral resolution (at the potential cost of reduced SNR) beyond $d\gamma(h)/dh$ MAX may not be an optimal sampling design. This maximal value was used to determine the optimal spectral sampling frequency across discrete segments of the UV/visible/NIR spectra.

For each individual spectrum measured, a series of empirical variograms (and subsequent analyses, as described above) were constructed from at least 50 nm discrete segments of data across a spectral range of 375–725 nm. For the phytoplankton absorbance spectra derived from laboratory cultures, these 50 nm windows used to run the statistical analysis were centered and sampled in 25 nm increments (e.g. 375–425 nm, 400–450 nm, 425–475 nm … 675–725 nm), ensuring the entire spectral range of data were covered and slightly oversampled. The R_{RS} variograms were also constructed from 50 nm windows of data, but were sampled at 15 nm increments (e.g. 375–425 nm, 390–440 nm, 405–455 nm … 675–725 nm). The R_{RS} data were sampled at higher incremental frequency due to increased likelihood of overlapping peaks in a mixed natural assemblage. Results are reported in this manuscript as the center wavelength of each 50 nm window. Relatively large intervals, $N(h)$, are used for this analysis to ensure that sufficient data points exist to yield robust statistical results. As a general rule of thumb, a minimum of 30 pairs should be examined when calculating empirical variograms, as an increase in data points enhances the accuracy of variogram estimates. The minimum number of pairs that can be examined to yield a reliable variogram estimate is equivalent to half of the maximum distance of the field [26]. In this case, 50 nm is the maximum distance of the field, therefore the information retrieved from variograms is less reliable at resolutions of 25 - 50 nm, and more reliable at 1 – 24 nm.

It should be noted that not all variance is resolved at the maximal point of $d\gamma(h)/dh$ that is calculated from each spectrum. Since the value of $\gamma(h)$ at any given point represents the variance at a corresponding spectral resolution, a ratio of this value against the sill (total

variance of the random field) represents the relative percentage of total resolved variance at that resolution. This was computed at $\gamma(h)$ points corresponding to the maximal point of $d\gamma(h)/dh$ at each sampling increment in this analysis to quantify how much total variance is resolved at the optimal sampling frequency.

In addition, a sensitivity analysis was performed on the full analyzed spectrum (375–725 nm) of the 21 R_{RS} spectra (converted to normalized water leaving reflectance), in order to assess how much noise can be added to the spectra before the utility of the data is compromised, and thus examine the viability of using derivative-based algorithms with potentially noisy satellite data. The sensitivity analysis is performed by adding 0.25% increments of Gaussian white noise and identifying the noise level at which $\gamma(5nm) = \gamma(10$ nm) over the whole spectrum. This is essentially a measure of when 5 nm data is not distinguishable from 10 nm data, therefore a 10 nm continuous spectrum has as much utility as a 5 nm continuous spectrum. This analysis is performed again after performing a Savitsky-Golay smoothing function with a polynomial order of 4 and frame size of 13 (chosen as the closest proxy between 10 and 15 nm smoothing). The Savitsky-Golay filter is frequently used for derivative analysis and was chosen for its capacity to preserve relative minima and maxima inflections in the spectrum [27].

3. Results

3.1 Phytoplankton absorbance/absorption

The absorbance of one phytoplankton species (*Thalassiosira weissflogii*) at different spectral resolutions (0.3125, 0.625, and 1.00 nm) yielded near-identical variogram results, at the cost of increased noise sensitivity (higher nugget effect and data scatter) with increased spectral resolution [Fig. 3]. The absorbance from four separate phytoplankton species, including one diatom (*Thalassiosira weissflogii*), one eustigmatophyte (*Nannochloropsis sp.*), one cyanobacteria (*Synechococcus sp.*), one coccolithophore (*Emiliana huxleyi*), and one mixed culture with all four species, encompasses spectral variability scales which include a variety of overlapping pigment peaks, including phycocyanin, phycoerythrobilin, phycoerythrin, chlorophylls, carotenoids, etc. [Fig. 4(a)]. The optimal spectral sampling frequency was rarely greater than 7 nm spectral resolution for individual or mixed cultures, and was as low as 15 nm spectral resolution [Fig. 4(b)]. On average, the percent resolved variance at the optimal spectral sampling frequency was 74 ± 3% for *Thalassiosira weissflogii*, 78 ± 6% for *Nannochloropsis sp.,* 72 ± 3% for *Synechococcus sp.*, 72 ± 5% for *Emiliana huxleyi*, and 74 ± 4% for the mixed assemblage. As a point of reference for comparison, an analysis run on nine natural seawater samples [Fig. 4(c)] with higher spectral sampling (0.2 nm) spanning a spectral range of 400–480 nm and 600–680 nm (areas with highest suggested spectral sampling frequency from Fig. 4(b)), yielded an average optimal resolution of 7.0 ± 0.4 nm (90 ± 6% resolved variance) and 6.5 ± 0.3 nm (84 ± 6% resolved variance), respectively.
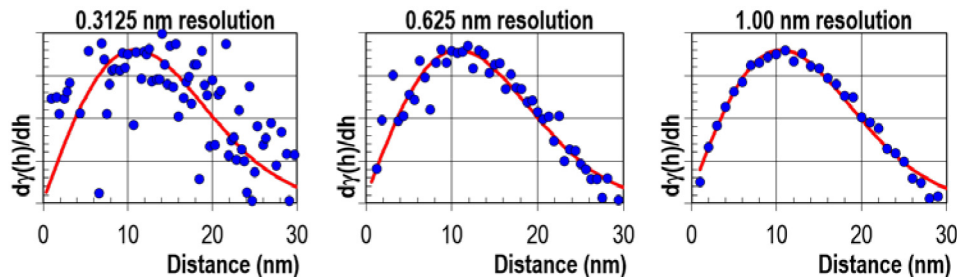

Fig. 3. The first derivative of empirical variograms run on different resolutions of data from one phytoplankton absorption sample across the entire integrated spectrum (342–750 nm). The theoretical variogram (red line) helps resolve variability in presence of noise, yielding near-identical results between resolutions.
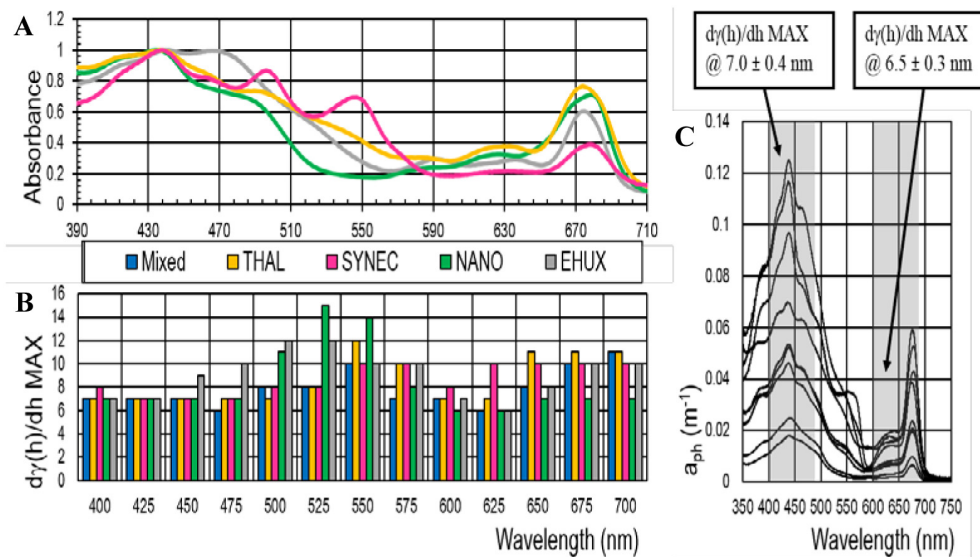
Fig. 4. (A) Continuous 1 nm resolution spectra of four species of phytoplankton used in this analysis, *Thalassiosira weissflogii* (THAL), *Synechococcus sp.* (SYNEC), *Nannochloropsis sp.* (NANO), *Emiliana huxleyi* (EHUX), representing a diverse collection of pigment expressions. (B) The corresponding dγ(h)/dx max in 25 nm increments, indicating the optimal spectral resolution at which the most information is obtained within each increment. Results show most features are resolved between 6 and 15 nm spectral resolution. (C) Phytoplankton absorption from natural assemblages collected from the Korean Strait, highlighting the areas with the highest frequency of absorptions peaks/inflections. Within these highlighted ranges, the optimal spectral frequency is around 6.5-7.0 nm.

3.2 Remote sensing reflectance

The analysis of remote sensing reflectance derived from six distinct water types, ranging from turbid freshwater/coastal waters to blue, oligotrophic waters showed that most spectral features were optimally resolved at a spectral resolution of 4–15 nm [Figs. 5(a)-5(f)], depending on the location within the spectrum and water type. At the defined spectral resolutions within the figures, this accounts for 60–87% of the total resolved variance.

In the case of the turbid monoculture *Microcystis* bloom, most features were resolved with a 6 nm continuous spectrum, with some sloping portions of the spectrum tolerating >12 nm spectral resolution [Fig. 5(a)]. The coastal and shelf waters along the Gulf of Mexico/U.S. East Coast show a spectral sampling frequency of 5 nm may be required to optimally resolve the spectral features centered around 685 nm [Figs. 5(c) and 5(d)]. The analysis shows that the same regions in the low-signal oligotrophic Gulf Stream and Bahamas water [Figs. 5(e) and 5(f)], as well as the high-signal *Microcystis* and turbid river plume waters [Figs. 5(a) and 5(b)] are optimally resolved with a 6 nm semi-continuous spectrum.

Other persistent features included a region of spectral sensitivity (4 – 5 nm spectral sampling frequency) centered around 535 nm, which was present in all the spectra, besides the *Microcystis* bloom [Figs. 5(b)-5(f)]. Two other notable features include a region in the coastal green water [Fig. 5(c)] centered around 625 nm that suggests an optimal spectral sampling frequency of 4 nm, in addition to a spectral feature centered at 475 nm appearing in the oligotrophic Gulf Stream and Bahamas water [Figs. 5(e) and 5(f)], which shows an optimal spectral sampling frequency of 5 nm. With a few exceptions, the remainder of the spectral features in all natural waters typically require less than 10 nm spectral resolution to optimally resolve the underlying spectral features.
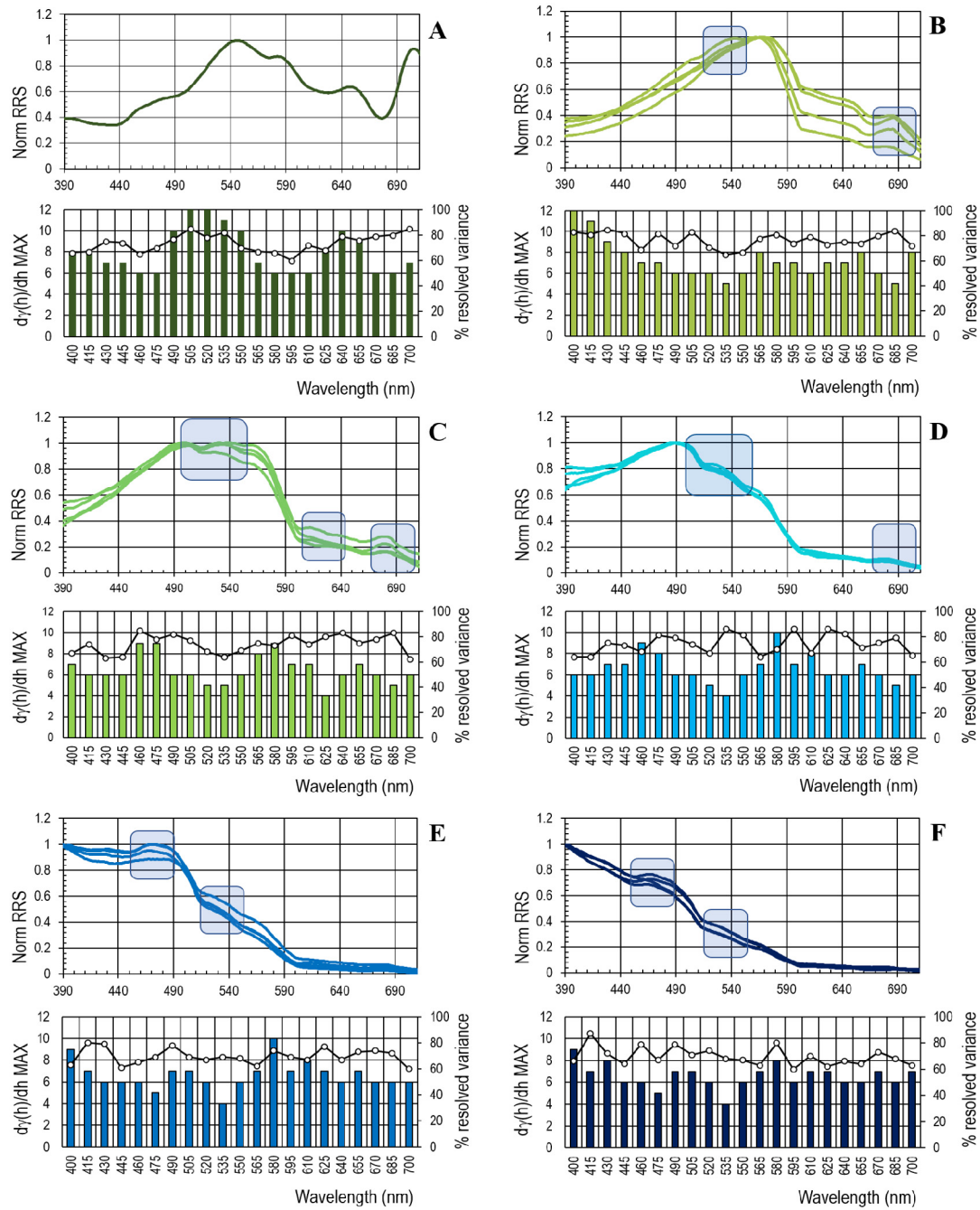
Fig. 5. Continuous 1 nm spectra and corresponding bar graphs (below) indicating the optimal spectral sampling frequency (dγ(h)/dh maximum; bars) as well as the percent resolved variance (dots + line) at given band centers for varying water types. The blue boxes highlight regions that are optimally resolved at 5 nm or less. Water types: (A) One Microcystis bloom spectrum collected from Lake Erie, MI., (B) Four spectra from the turbid river plume waters of the Gulf of Mexico, (C) Four spectra from green coastal waters in the Gulf of Mexico and U.S. East Coast, (D) Four spectra from the shelf waters in the Gulf of Mexico and U.S. East Coast, (E) Four spectra from the open blue waters of the Gulf Stream along the U.S. East Coast, and (F) Four spectra from clear oligotrophic waters in the Bahamas.

3.3 Spectral noise thresholds

Assuming a 5 nm continuous spectra as a baseline for hyperspectral satellite missions, incremental Gaussian noise additions made on all 21 normalized water leaving reflectance spectra show that only ~3% noise could be tolerated before 5 nm (first derivative) data resolves the same amount of variability as 10 nm (first derivative) data [Fig. 6(b)] over the entire integrated visible spectrum. For these spectra, the 3% noise addition equated to an SNR of approximately 66. Since it is a common practice to utilize smoothing techniques prior to performing derivative analysis, a 13 nm Savitsky-Golay smoothing function was applied after each increment of noise, bringing the noise threshold up to approximately 13% [Fig. 6(c)]. Across the entire visible spectrum and across all water types, this equated to an average SNR of approximately 15. Using a 15 nm mean boxcar filter, the threshold was increased to nearly 20% (not shown), however, there were severe aberrations to the original spectrum, including a spectral shift in peaks.
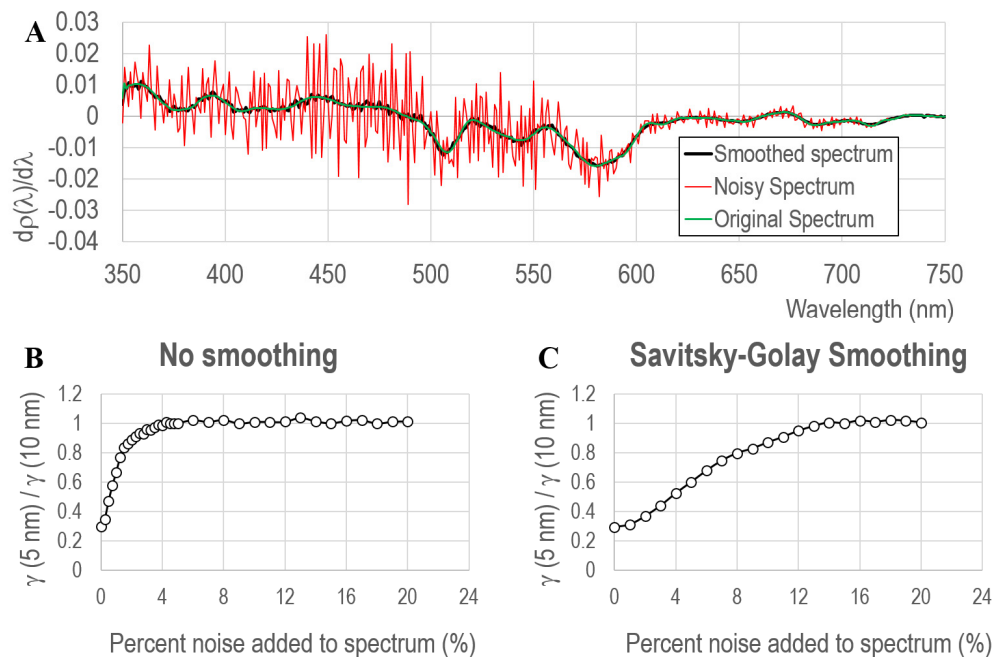

Fig. 6. (A) A single differential normalized reflectance spectrum (green line) taken from shelf waters of the Gulf of Mexico [Fig. 5(d)], shows aberrations after adding 3% Gaussian noise, with and without smoothing (black and red lines, respectively). Below, the ratio of γ(5 nm) to γ(10 nm) at incremental Gaussian noise additions is calculated for all spectra, indicating the threshold at which the variability of first derivative reflectance data at 5 nm spectral resolution meets or exceeds the variability of first derivative reflectance data at 10 nm spectral resolution, e.g. γ(5 nm)/γ(10 nm) = 1. (B) Without any spectral smoothing, the noise threshold is at ~3%. (C) With a Savitsky-Golay spectral smoothing function, this threshold is extended to 13%.

4. Discussion

The principle objective of this research is to evaluate the frequency of spectral sampling required to resolve subtle spectral features contained in hyperspectral data sets, in addition to nominally evaluating the efficacy of expected data product quality based on expected sensor performance. It is recognized that the data sets used for this analysis do not represent an inclusive measure of global ocean environments, however, this does not preclude the significance of results, as the data still represent diverse examples of bio-optical variability that will be detected by future ocean sensors. While further investigation is required to

expand the implications for deriving various ocean parameters in a wider range of natural environments, the methodology presented in this manuscript is adaptable in nature and can be applied to any high spectral resolution data set to determine optimal sampling frequency.

It should be understood in the interpretation of results that a first-order derivative transformation of the data increases the sensitivity of the analysis. The original, non-transformed data do not have enough of a signal over narrow spectral windows for the variogram technique to be useful, so the features were exaggerated with the use of a first derivative. This means, at 1 nm spectral sampling, a single peak is transformed into two peaks after the first-derivative, three peaks after the second derivative, etc. creating a higher spectral sampling frequency with every consecutive derivative. Therefore, the spectral sampling frequency obtained from the statistical analysis of the first derivative is relevant to accurately re-constructing the shape of spectral inflections, not just capturing the location of peak maxima/minima. This may not be a high priority for all applications, but is potentially important to resolving spectral features where absorption peaks may overlap closely, especially in waters with mixed phytoplankton community compositions, or, for resolving shifts in spectral peaks. The use of 2nd and 4th derivative analysis can be used to further accentuate these peaks, however, the application of the empirical variogram technique to higher order derivatives would yield frequency information required for the hyperbolic reconstruction of the absolute spectral shape of a high order differential spectra, an application which may be excessive when considering the optimization of sensor design.

Additionally, in examining these results, it is noteworthy that there is some inherent limitation associated with the absolute resolution of the radiometric instrumentation to be able to detect narrow spectral features. While a 3 nm (or wider) bandwidth sampled at 1 nm could feasibly retrieve a signal from a 1 nm wide spectral feature (if the signal is strong enough), it is possible that such a feature would not be incorporated into the variogram analysis. Regardless, each variogram analysis consistently showed an increase in resolved variance as a function of increased the spectral resolution, down to the 1 nm sampling frequency [Fig. 2(a)]. In other words, while the ASD sensor may measure at a larger spectral bandwidth than the absolute width of a given feature, the variogram still shows a reduction in variance (gain in spectral information) at 1 nm relative to 3 nm sampling intervals. However, this analysis is not intended to examine the exact resolution required to detect all spectral features present, per say, it is otherwise designed as a method to determine the optimal spectral resolution. The variograms are cast over a wide spectral window (50 + nm), examining the rate at which the highest amount of information is gained, on average, over this spectral window. Even with multiple 1 nm features present, for instance, the variogram would not likely display 1 nm as the optimal spectral resolution. There is some corroboration that 1 nm spectral features are not ubiquitous, as the laboratory-based absorbance measurements (absolute resolution 1 nm or less) did not show features present at 1 nm among multiple PFTs [Fig. 3(a)], however, absorbance is only one dimension of a multi-faceted *in situ* or satellite-based reflectance measurement. Finally, we note that other derivative analysis in the literature specifically aimed at extracting information on PFTs and pigments [6,14] have utilized a 9 nm smoothing and band separation function to yield positive distinction of PFTs and/or pigments, therefore, the resolution and smoothing functions used in this study are likely relevant to these applications.

Based on this interpretation, several of the natural water spectra show that there are instances at which 5-6 nm spectral resolution may be required in order to fully resolve the various inflections within the spectra, such as around the chlorophyll fluorescence peak (~680 nm) in the reflectance data [Figs. 4(b)-4(d)]. While a relatively broad peak (~20 nm wide), a sampling frequency of 10 nm would only enable the capture of the relative magnitude of the peak, and makes a critical assumption that this peak location is static. However, fluorescence peaks have been known to shift by several nanometers, especially in waters with high concentrations of chlorophyll [28,29]. This suggests that in some instances, an even finer

spectral resolution than 5 nm could be useful to detect the true maxima of the fluorescence peak. In contrast, the absorbance data from the laboratory cultures do not include fluorescence and exhibit much broader absorption peaks in the red (650 – 700 nm) that are optimally resolved with 10 nm resolution, with the exception of the chlorophyll absorption peak seen in the *Nannochloropsis* culture, which may require a slightly higher sampling frequency (~7 nm) compared to other cultures in order to characterize the slightly irregular shaped peak [Fig. 3].

Elsewhere in the absorption and reflectance data, there are several instances of spectral variance on the order of 4-6 nm, such as around 625 nm, where multiple overlapping peaks and concentrations of chlorophylls-*a,b,c* are present over a broad spectral range, or around 425 – 500 nm, where multiple chlorophyll, photo-protective, and carotenoid pigment absorption peaks overlap [30]. There are also regions in the spectra which exhibit high spectral variance, but which are not traditionally recognized as significant to ocean color. For instance, a high amount of spectral variance centered at ~535 nm (510 – 560 nm) was common to most all reflectance spectra. While this is a region of strong phycoerythrobilin pigment absorption as well as cyanobacteria reflectance, it is unlikely ubiquitous in all the water types sampled. Since the nature of the variogram analysis requires large spectral windows (at least 50 nm window at 1 nm spectral resolution) to gain enough signal, there are likely a combination of spectral features spanning this window that are integrated into the results. This is supported by observations by Torrecilla et al. [6] showing that the 495 – 540 nm window has been recognized as a distinctly useful spectral region for discriminating phytoplankton pigment assemblages in open ocean waters. While the analysis is highly sensitive to spectral variability, the exact locations of these inflections within the 50 nm spectral window are more difficult to pinpoint with variograms, and are better resolved with studies examining zero-intercepts of the first and second derivative [13,14,16].

Results derived from independent analysis specifically aimed to determine the spectral resolution required for PFT and/or phytoplankton pigment distinction are generally in line with the suggested optimal resolutions which are concluded from this study (5 – 7 nm), however, as previously mentioned, there is some discrepancy in the literature among specific recommendations for spectral resolutions. For instance, results from Roelke et al. [15] suggest a slightly higher spectral resolution than presented in this study, showing that the distinction of chlorophyll-*a* and chlorophyll-*b* peaks in green algae composite absorption data begin to deteriorate at 5 nm and 7 nm spectral resolutions, respectively. Wolanin et al. [16] focused on the distinction of three specific PFTs, and showed a small improvement of results when increasing from 10 nm to 5 nm continuous spectral resolution, and even less improvement when increasing from 5 nm to 1 nm spectral resolution. By contrast, Lee et al. [13] used 1st and 2nd derivative analysis to suggest 17 strategically placed bands based on the highest frequency of peak occurrences on over 400 spectra, however, this is optimized for the detection of the most frequently occurring peaks, and can change with variability of the data sets, potentially neglecting the detection of signal at the cost of reducing redundancy. The focus of this study is not to assign attribution of individual peak signals or detect specific PFTs, per say, but to present an unbiased methodology to detect the frequency of spectral inflections across the any region of the electromagnetic spectrum, regardless of the source of the signal. The adaptive nature of continuous hyperspectral data enables users to determine which bands are most important for specific applications, but it is imperative to first ensure that these spectral features are being detected with a reasonable SNR that will enable viable retrievals of downstream product applications. As suggested by Wolanin et al. [16], some redundancy in spectral information can increase the adaptably of data across multiple algorithms and water types.

The afore mentioned sensitivity of the empirical variogram analysis to variability makes it a useful tool to additionally quantify the impacts of noise on spectral signatures. This analysis showed that data viability for derivative analysis is compromised when only ~3% noise is

introduced (SNR ~66), however, this was significantly alleviated by smoothing the spectrum, which is a common treatment for derivative analysis. This finding is in line with a previous study of SNR in which an analysis of 2,500 absorption spectra containing mixed concentrations of *Prorocentrum minimum* and other algae showed that an SNR of ~34 was sufficient to properly distinguish and classify *P. minimum* as a dinoflagellate ($\geq$ 80% chlorophyll contribution) using 4th derivative analysis [15]. Various methods of smoothing can be implemented to reduce the possibility of leveling spectral features [31] beyond that used in this experiment, and may be useful for further reducing instrument noise effects.

Currently, the uncertainty thresholds for the NASA PACE mission are defined for normalized water leaving reflectance (mW cm^{-2} μm^{-1} sr^{-1}) as 5% or 0.001 from 400 – 600 nm, and 10% or 0.005 from 600 – 900 nm. The PACE Ocean Color Instrument (OCI) is projected to nominally sample at a spectral resolution of 5 nm, and with careful smoothing procedures, the current design for the PACE OCI should provide useful hyperspectral derivative data for use in PFT algorithm development and improved ocean color products. It should be noted, however, that this analysis is focused on the resolution required to detect ocean-specific features from *in situ* and laboratory data, but does not take into account very narrow spectral features introduced from vibrational Raman scattering and Fraunhofer Lines [32] or other spectral features that a satellite sensor will be subject to at the top of the atmosphere, such as absorbing gases [33], which can introduce uncertainty in retrievals and may require higher spectral resolution to correct for.

5. Conclusions

This study aimed to quantify the optimal spectral resolution and signal to noise thresholds required to discern subtle spectral features across various water types and phytoplankton functional groups. The results indicate that a continuous 5 nm spectrum with less than 13% uncertainty in the ocean signal is optimal in the resolution of ocean color variability from *in situ* derived remote sensing reflectance, with some regions of the spectrum potentially benefiting from enhanced spectral subsampling which would account for spectral shifts in peak maxima (e.g. chlorophyll fluorescence). At a 5 nm continuous spectral sampling frequency, the signal to noise ratio of the sensor will likely be the ultimate limiting factor for distinguishing fine scale peaks, but may be mitigated with various statistical smoothing techniques.