A Novel Approach for Retrieving Chlorophyll-a Concentrations from Satellite Ocean Color Data

Yingying Wu
Department of Mathematics
The University of Texas at Austin, Austin, TX
ywu@math.utexas.edu

Joaquim Goes
Lamont Doherty Earth Observatory
Columbia University, New York, NY
jig@ldeo.columbia.edu

ABSTRACT

Remote sensing reflectance (Rrs(λ), sr^{-1}), measurements by sensors on board ocean color satellites are an important means of studying the oceans from space. One of the most important products of ocean color remote sensing is Chlorophyll-a (Chl-a), which is a widely accepted proxy for phytoplankton biomass and vital input for measuring primary productivity of ocean ecosystems from space. Most optical approaches for measuring Chl-a from space rely on the variance of spectral shapes of Rrs measured either in-situ or from space. Spectral shape features are then characterized based on either the absolute values, differences, ratios or slopes of Rrs(λ) at certain Chl-a sensitive wavebands, which are then used in algorithms for retrieving Chl-a. Despite much progress, most algorithms provide only local features of spectral shapes and do not consider global properties. Here we present a novel application of Finite Metric Space Characterization (FMSC) algorithm [2] that encodes the spectral shape of Rrs to retrieve Chl-a based on geometrical and topological invariants. The FMSC Ocean Color Algorithm (FOCAL) was developed and validated using Moderate Resolution Imaging Spectroradiometer-Aqua (MODIS-Aqua) Rrs and Chl-a data from NASA Bio-Optical Marine Algorithm Dataset (NOMAD). The adjusted R-squared of FOCAL Chl-a retrievals was 0.9023, as compared to the adjusted R-squared value of 0.8204 achieved using the standard OC3/OC4 (OCx) band ratio algorithm merged with the color index algorithm. Therefore, FOCAL outperforms other widely used algorithms for estimating Chl-a using remotely sensed multispectral and hyperspectral Rrs data.

1. INTRODUCTION

Since the launch of the Coastal Zone Color Scanner (CZCS), the first ocean color satellite in the 1970s, optical oceanographers have relied on quantitative relationships between the apparent optical properties (AOPs) and the inherent optical properties (IOPs) to derive synoptic fields of ocean biogeochemical properties that have helped revolutionize how biological oceanographers, ocean biogeochemists and climate scientists study the oceans. One of the most important satellite ocean color products is Chl-a, which is a widely used proxy for studying phytoplankton biomass distribution, ocean primary productivity, and for assessments of the capacity of the oceans to sequester atmospheric CO_2. One of NASA's currently orbiting, flagship ocean color satellite

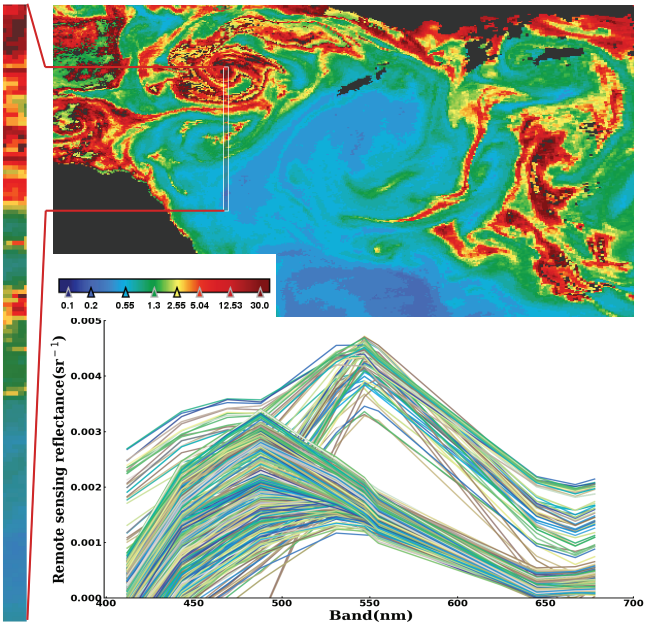

Figure 1: The Chl-a and Rrs product of MODIS Aqua from NASA (Left: the white rectangle area. Top: remote sensing image from Aqua Satellite. Bottom: Rrs extracted from the Rrs product of the white rectangle area).

MODIS Aqua launched in 2002 has provided vital daily data on Chl-a distribution in the oceans and has been used extensively for studying phytoplankton dynamics and phytoplankton productivities on local, regional and global scales, exhibited as Fig. 1. To support the development of algorithms and satellite data product validation activities, NASA Ocean Biology Processing Group at Goddard Space Flight Center, Maryland, USA has provided a bio-optical database called NOMAD [1].

On the basis of the specific spectral features of optically active constituents (OACs) in waters, a series of algorithms were developed to retrieve Chl-a and other products from Rrs. In a spectral line made of Rrs of multiple bands, the reflectance magnitudes that increase rapidly are called peaks; rapid reflectance decreases are called valleys or dips. These are generally referred to as features of spectral shapes. The behavior of the spectral shapes results from the combination of OACs and has a quantitative relationship with the concentrations of individual OACs, including Chl-a and other phytoplankton pigments.

Most existing algorithms do not fully exploit the valuable

information provided by the spectral line and the spectral shapes, such as the magnitudes, width, the center of the peaks, and the central band of the dips. Instead, the correlation between some Rrs wavebands, the values of Rrs, the differences between the values of two Rrs bands divided by the difference in wavelength (nm), Rrs ratios, and the transformed Rrs of many sensitive wavebands have been used in many empirical and semi-analytical algorithms to retrieve Chl-a and other biogeochemical products. The Quasi-Analytical Algorithm, a widely used semi-analytical algorithm, considers empirical relationships from the spectral shapes. To avoid the instability caused by the correlation of some bands Rrs, many algorithms use methods such as principal component analysis to reduce the number of input Rrs, which result in losing some spectral data information. The spectral shape algorithm has been applied for cyanobacterial bloom detection [3]; also, derivative analysis or continuum removal methods, which eliminate the tangent to each member of the spectral curves have been used to enhance the contrast between the bands of the spectral curve and highlight the absorption and reflection characteristics within the spectral curve.

In some methods for analyzing the spectral shape, such as the spectral angle mapper and different distances, including Euclidean or cosine distance have been used for quantifying the similarity of spectral pattern or shape, for classification and clustering spectra, or estimating Chl-a. Most Rrs spectra derived from ocean color satellites are based on only a few wavebands, and most current algorithms use only a fraction of the total wavebands for algorithm development, which can result in a derived biogeochemical product with significant uncertainties.

Research Project. The spectral lines convey geometric meaning because the spectral bands are ordinal. Albeit the frequent appearance in literature, few attempts were made to quantify the association between spectral shapes and Chl-a. In this project, we used the FMSC algorithm to encode the spectral shapes. The algorithm extracts interpretable features for statistical models and classifying algorithms. The algorithm computes the geometrical and topological invariants based on mathematical tools derived from finite element methods, the theory of curves from differential geometry, and topological data analysis. FOCAL characterizes the spectral shapes to understand how the Rrs curve varies with OACs according to dependent variables extracted by FMSC. FOCAL forecasts OAC concentrations based on ocean color, infers in-situ depth profiles from remote sensing images, and enables a geometric empowered transferred learning to train a deep learning model.

2. MAIN INNOVATIONS

Accurate quantitative retrieval. We designed FOCAL based on the FMSC algorithm to model Chl-a using the spectral data of Rrs curves. We evaluated the effectiveness of FOCAL by the NOMAD dataset. The R-squared of our models shows that the Chl-a retrieval achieved by FOCAL is more accurate than contemporary algorithms. A table of R-squared and adjusted R-squared between the standard OCx band ratio algorithm merged with the color index used by NASA, the Spectral Polygon (SP), NASA + SP blended algorithm, and FOCAL are presented in Table 1. The total number of data points used in this analysis is 958. As shown in Table 1, the adjusted R-squared of FOCAL Chl-a retrieval model is 0.9023, as compared to an adjusted R-squared of 0.8204 using NASA's algorithm. The adjusted R-squared of the SP algorithm and the model combining NASA's algorithm and the SP algorithm are 0.8200 and 0.8530, respectively. In comparison to the algorithm used by NASA, the SP-based algorithm is extendable to more bands Rrs.

Table 1: Comparison of R-squared between FOCAL and the blended algorithm from NASA.

	NASA	SP	NASA+SP	FOCAL
R-squared	0.8206	0.8202	0.8533	0.9052
Adjusted R-squared	0.8204	0.8200	0.8530	0.9023

3. DEMONSTRATION

The MODIS spectroradiometer is currently operating on both the Terra and Aqua spacecrafts. It has a viewing swath width of 2,330 km and views the entire surface of the Earth every one to two days. Its detectors measure 36 spectral bands between 0.405 μm and 14.385 μm, and it acquires data at three spatial resolutions: 250 m, 500 m, and 1,000 m. The MODIS dataset can be found at `https://modis.gsfc.nasa.gov/data/`. In our demonstration, we will exhibit visualization of the MODIS data, features extracted from FOCAL, feature significance, and estimating Chl-a concentrations. FOCAL can be extended to mine a wealth of information from both current multispectral sensors like the NOAA-Visible Infrared Imaging Radiometer Suite (VIIRS), European Space Agency (ESA) Sentinel-2 and Sentinel-3, Japan Space Agency (JAXA), Global Line Imager (GLI), and planned hyperspectral Plankton Aerosol Cloud ocean Ecosystem (PACE) mission.

4. ACKNOWLEDGEMENTS

The authors thank Dr. Kui Ren, Dr. Andrew J. Blumberg, Dr. Mu-Tao Wang, and Dr. Chiu-Chu Melissa Liu for discussions on geometric analysis. Wu was supported in part by NIH grant 5U54CA193313, NIH grant GG010211-R01-HIV, and AFOSR research grant FA9550-15-1-0302.

5. REFERENCES

[1] S. Bailey et al. An improved bio-optical data set for ocean color algorithm development and satellite data product variation. *Remote Sensing of Environment*, 98:122–140, 2005.

[2] Y. Wu, Y. Liu, R. Y.-H. Tsai, and S.-T. Yau. Investigating the role of eye movement and electrodermal activity signals in search satisfaction prediction using geometric analysis. *Journal of the Association for Information Science and Technology*, 2019.

[3] T. Wynne, R. Stumpf, M. Tomlinson, R. Warner, P. Tester, J. Dyble, and G. Fahnenstiel. Relating spectral shape to cyanobacterial blooms in the laurentian great lakes. *International Journal of Remote Sensing*, 29(12):3665–3672, 2008.