Seeing the Unseen: Mining an ocean of data for retrieving the vertical distribution of Chlorophyll-a from Satellite Ocean Color Data

Jinghui Wu
Lamont Doherty Earth Observatory
at Columbia University
61 Route 9W, Palisades, New York
jw3794@coulmbia.edu

Joaquim Goes
Lamont Doherty Earth Observatory
at Columbia University
61 Route 9W, Palisades, New York
jig@ldeo.columbia.edu

Yingying Wu
Department of Mathematics at
The University of Texas at Austin
2515 Speedway, RLM 8.100, Austin
ywu@math.utexas.edu

ABSTRACT

Ocean color measurements from space have made it possible to obtain synoptic views of phytoplankton biomass (measured as chlorophyll-a, Chl-a) distribution in the surface ocean at unprecedented space and time scales. A long-standing challenge, however, has been to determine the vertical distribution of Chl-a (VDCHL) of the global oceans. Over the years, there have been several attempts to infer the VDCHL solely from sea-surface Chl-a concentration measurements. Parametric curve fitting routines of varied complexities, as well as neural network classification algorithms, have been used to derive the VDCHL with limited success. Here we present a novel application of the Finite Metric Space Characterization (FMSC) algorithm [1], called FMSC Ocean Color Algorithm (FOCAL) to first derive surface Chl-a. In this project, we apply the FMSC approach to a vast global oceanographic dataset being collected by autonomous floats (Bio-Argo) to accurately retrieve the VDCHL. FOCAL allows us to retrieve surface (2D) Chl-a from Level-2 remote sensing reflectance ($R_{rs}(\lambda)$, sr^{-1}) products generated by NOAA-Visible Infrared Imaging Radiometer Suite (VIIRS). Compared to in-situ Chl-a from NASA Bio-Optical Marine Algorithm Dataset (NOMAD), the adjusted R-squared of FOCAL retrieved Chl_{BL} (R^2=0.90) is much better than the NASA standard Chl-a products (R^2=0.82) that are based on band ratio algorithms. We believe that the FMSC would provide a more accurate way of determining the VDCHL from the three-dimensional (3D) oceanographic datasets.

1. INTRODUCTION

The biogeochemical activity of the oceans and its carbon cycle are two parts of a complex feedback system that are involved in regulating the build-up of CO_2 in our planet's atmosphere. The 'biological pump' in which oceanic phytoplankton play a key role, mediates the flux of carbon to the interior of the ocean by interactions between the components of the vertically-structured pelagic ecosystem of the photic zone. The VDCHL is a result of multiple forces and not only a simple indicator of autotrophic biomass or production, but it also represents ecologically significant features of the planktonic ecosystem. The shapes and concentration of VDCHL in lakes, coastal seas, and open oceans are highly variable, and to this day, it has been difficult to understand what regulates the VDCHL. A large part of the problem is the paucity of data traditionally collected on research ships, which has prevented in-depth studies on the VDCHL and its evolution in space and time.

Over the years, oceanographic satellites have made it possible to obtain quasi-continuous measurements of the global ocean, but only at the surface (Fig. 1). The high-resolution 2D images provided by this coverage contain information on physical or bio-geophysical variables at the surface but not in the deeper photic layer. There have been several attempts to develop methods for retrieving VDCHL from observed sea-surface satellite images [2, 3]. Initial studies consider the VDCHL to be uniform throughout the water column, but subsequent studies revealed that phytoplankton always migrates up and down in the water column, resulting in an inhomogeneous vertical distribution. The model by Antoine and Morel (1996) [2] goes further by assuming that the VDCHL changes its shape according to the concentration of the surface layer. On the basis of these findings, attempts have been made to infer the VDCHL, using parametric curve fitting routines [3]. All these methods usually require numerical modeling of the relationship between Chl-a with VDCHL. Such models are, however, often faced with problems of non-linearity, complexity and incomplete knowledge of the mechanisms that govern these profiles.

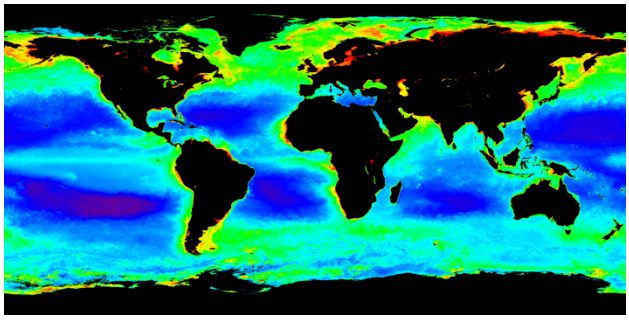

Fig. 1 Annual VIIRS-SNPP Chlorophyll Concentration in 2018, based on OCx Algorithm, download from NASA website.

More recently there have been attempts to infer the VDCHL using neural network classification algorithms. These include Artificial Neural Networks (ANNs), Hidden Markov Models and Self-Organizing Topological Maps (SOM) analysis [4]. The important inputs of the neural network are surface Chl-a from satellite imagery and the VDCHL from in-situ discrete depth

data. There also exist some databases of extrapolated VDCHL calculated by biogeochemical models such as the MERCATOR-VERT and NEMO-PISCES models. Despite the progress, most approaches fall short in their use over global scales, because a handful of parameters that are obtained at local scales are not universal in their applicability, and at times are inadequate descriptors of the shape of the profile especially if they are different.

On account of the limited ability of satellites to directly measure any subsurface properties, as well as the inadequacy of research cruise data collections, to provide data at the spatial and temporal scales necessary for climate research, unmanned technologies have been developed for collecting oceanographic data autonomously with great success. One such technology is the Argo profiling floats, which are part of a broad-scale array of temperature and salinity profiling floats. More recently, biogeochemical Argo floats have been developed that are capable of capturing the VDCHL (Figs. 2a-b).

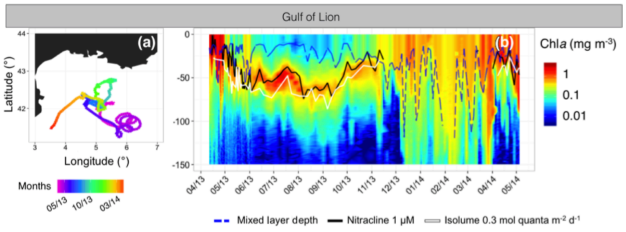

Fig.2 (a) Trajectory and (b) Profiles of chlorophyll concentration versus time for the float deployed in the Gulf of Lion [5].

In the past, we have used the non-parametric statistical methods to infer in-situ depth profiles of salinity and temperature in the oceans using data from shipboard and ARGO floats. The results show that profiles of both temperature and salinity exhibit unique features that can be easily captured with non-parametric statistical methods. As a first order approximation, we plan to use non-parametric splines, local linear regression type fitting procedures for discerning the VDCHL, because it is generally accepted that this method can reproduce the dynamic processes that govern the evolution of the vertical profiles of Chl-a. We are not sure about the sensitivity of the non-parametric method to the data quality. As part of this study, we will attempt to compare the performance of non-parametric statistical measurements against the FMSC method to infer the VDCHL.

Research Project.

Fit with the KDD Ecosystem. As the basis of this study, we developed a FOCAL algorithm based on the FMSC algorithm, which characterizes features of the profile shape by encoding the shape of the vertical data. FOCAL is an extension of FMSC algorithm, which is applicable to data mining, knowledge discovery and machine learning in research fields that produce massive nonlinear and high-dimensional data like the VDCHL.

2. MAIN INNOVATIONS

Retrieval the evolution of 3D Chl-a. We will demonstrate the efficacy of the FOCAL method with Bio-Argo profile floating Chl-a data not only for surface Chl-a but for inferring the VDCHL. Further, by using the FOCAL retrieved VDCHL in conjunction with VIIRS sea-surface data, we will demonstrate the evolution of 3D distribution of Chl-a concentration in the global ocean. The R-squared results show that the surface Chl-a retrieval achieved by FOCAL is more accurate than contemporary algorithms. On the other hand, the relationship between the FOCAL retrieved VDCHL with VIIRS Chl-a obtained from this FOCAL model will then be utilized to infer the profiles of Chl-a at different locations in the world's oceans and at different times of the year with small root mean squared error (RMSE).

3. DEMONSTRATION

The VIIRS sensor on the Suomi National Polar-orbiting Partnership (SNPP) has 22 spectral bands (from visible to infrared), which enables a new generation of operational moderate resolution-imaging capabilities following the legacy of the AVHRR on NOAA and MODIS on Terra and Aqua satellites. The VIIRS datasets can be found at https://www.star.nesdis.noaa.gov/sod/mecb/color/. The concept of global robotic biogeochemical measurements was articulated in a Community White Paper (Gruber et al., 2007) that was supported by the International Ocean Carbon Coordinating Project (IOCCP) and the US Ocean Carbon and Biogeochemistry Program (US-OCB). More details in http://biogeochemical-argo.org/#.

In our demonstration, we will exhibit further visualization of the features extracted from FOCAL, the 3D distribution of Chl-a and the evolution analysis. Based on these 3D oceanographic databases, we will demonstrate that the fast computing FMSC method can be extended to provide near real-time, global, high spatial and temporal resolution 3D data of key variables such as sea surface temperature (SST), sea surface salinity (SSS), sea surface height (SSH), winds and ocean circulation, etc.

4. ACKNOWLEDGEMENTS

The authors thank Dr. Kui Ren, Dr. Andrew J. Blumberg, Dr. Mu-Tao Wang, 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] Y. Wu, Y. Liu, R. Y.-H. Tsai, and S.-T. Yau. Method for finite metric space characterization, Apr. 18 2019. U.S. Provisional Patent.

[2] D. Antoine, A. Morel. Oceanic primary production: 1. Adaptation of a spectral light-photosynthesis model in view of application to satellite chlorophyll observations. *Global biogeochemical cycles*, 10(1): 43-55, 1996.

[3] J. Uitz, H. Claustre, A. Morel, S. Hooker. Vertical distribution of phytoplankton communities in open ocean: An assessment based on surface chlorophyll. *Journal of Geophysical Research: Oceans*, 111(C8), 2006.

[4] A. A. Charantonis, F. Badran, S. Thiria. Retrieving the evolution of vertical profiles of Chlorophyll-a from satellite observations using Hidden Markov Models and Self-Organizing Topological Maps. *Remote Sensing of Environment*, 163: 229-239, 2015.

[5] M. Barbieux, J. Uitz, B. Gentili, et al. Bio-optical characterization of subsurface chlorophyll maxima in the Mediterranean Sea from a Biogeochemical-Argo float database. *Biogeosciences*, 16(6): 1321-1342, 2019.