Inversion of vertical profiles of phytoplankton from satellite images

Supervised by Dominique Béréziat* and Anastase Charantonis†

Feb 2023

Summary

Retrieving the vertical profiles of bio-geophysical parameters is crucial in understanding the impacts of the changing climate on the ocean's biodiversity. Such retrieval is an arduous task given that in-situ observations of the vertical distributions are very sparse, and satellite imaging contains proxy information integrated over the top layer of the oceans. There exist numerical models that are plugged into global circulation models that simulate this evolution. These models are parametrized to fit specific time series of such vertical profiles, but they are highly uncertain. As such, the development of new methods that use deep learning advancements in order to combine all sources of information: in-situ, satellite, and numerically modeled, in order to reconstruct the vertical distributions are of big importance to both the bio-geophysics and data science community.

Scientific Context

The current density of satellite observations has allowed a quasi-continuous observation of the global ocean surface. The two-dimensional images provided by this coverage contain information on physical or bio-geophysical parameters but not on their vertical profiles [3]. Inverting the sea-surface data remotely sensed by satellite to obtain the vertical profiles of bio-geophysical parameters, however, requires numerical modeling of their relations. Such models are, however, often faced with problems of nonlinearity, complexity, and incomplete knowledge of the mechanisms that govern these profiles.

The biogeochemical activity of the oceans and the carbon cycle are two parts of a complex feedback system. A change in climate and an increase in the amount of available carbon can affect oceanic primary production. In return, a change in the biogeochemical activity affects the climate, by modifying the albedo and carbon fixation rates, as well as the atmospheric and oceanic carbon concentrations. It is therefore essential to be able to determine the oceanic primary production, of which chlorophyll-a is a proxy.

In recent years, many algorithms have been developed to infer the chlorophyll-a concentration in ocean surface layers through satellite imaging [1]. It has also been proved that the vertical chlorophyll-a distribution is related to various types of sea-surface data [5] and can be retrieved locally through machine learning approaches [2, 4].

Since the cost of determining the vertical distribution of chlorophyll-a by in situ measurements is prohibitively high, due to the lack of complete databases of such measurements such methods are very useful. The previous methodology however bases itself on large databases of extrapolated vertical profiles of chlorophyll-a calculated by biogeochemical models such as the MERCATOR-VERT and NEMO-PISCES models that tack a life-cycle module onto global circulation models. While it is generally accepted that these models are able to reproduce the dynamic processes that govern the evolution of the vertical profiles of chlorophyll-a, these models are tuned to specific locations on the globe and are optimized to best perform in these locations.

*HDR, associate professor at LIP6, Sorbonne Université, dominique.bereziat@lip6.fr
†Associate professor at LaMME, ENSIEE,

Since the PROFHMM method was developed a slew of new deep learning techniques have emerged, while the size of the global dataset of in-situ measurements of vertical profiles of biogeochemical parameters has increased by a significant amount (ARGO floats, Gliders, and oceanographic campaigns such as the TARA oceans mission).

Objectives of the thesis

The main objective of the thesis is the reconstruction of the vertical profiles of bio-geophysical parameters (such as Chlorophyll-a, or phytoplankton concentration) from satellite imaging. This is a highly interdisciplinary project, bringing members of the bio-geophysics community (such as Roy El Hourani and Sakina Ayata who will be involved in the project) in contact with deep-learning and data assimilation specialists. The problems imposed by the nature of the data are of interest to the later community (represented by the two main advisors of this project) and require the conception of new and dedicated algorithms that can then be generalized to other fields (for example inversion of the atmospheric column).

We will consider many methodologies applicable to the stated problem. For instance, attention networks and transformers have the biggest potential in modeling the temporal dynamics that can help constrain the multidimensional reconstruction of vertical fields of biogeochemical parameters from satellite observations.

One of the main obstacles to using Deep Learning techniques with in-situ data is the quality of these datasets. Even in their increased state, they lack the fullness required to fully train a deep architecture. As such during the thesis, there will be a need to remediate this problem.

One research avenue is would be to first learn the dynamics through a twin experiment using the outputs of the numerical model to approximate its function, then transfer the learned mechanics onto the existing in-situ dataset and adjust the model to better represent the observed behaviors.

The candidate will at the outcome of the thesis have a solid understanding of the research process in data science and bio-geophysical models and could reorient themselves in an academic or industry position in either field of study.

References

[1] J. Brajard, J. Cédric, M. Cyril, and S. Thiria. se of a neuro-variational inversion for retrieving oceanic and atmospheric constituents from satellite ocean color sensor: Application to absorbing aerosols neural networks. *Earth Sciences and Environmental Applications of Computational Intelligence*, 19(2), 2006.

[2] A. A. Charantonis, F. Badran, and 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.

[3] E. Dinnat, J. Boutin, G. Caudal, J. Etcheto, and P. Waldteufel. Influence of sea surface emissivity model parameters in l-band for the estimation of salinity. *Int. J. Rem. Sensing*, 23:5117–5122, 2002.

[4] H. Su, X. Lu, Z. Chen, H. Zhang, W. Lu, and W. Wu. Estimating coastal chlorophyll-a concentration from time-series olci data based on machine learning. *Remote Sensing*, 13(4):576, 2021.

[5] J. Uitz, H. Claustre, A. Morel, and S.B. Hooker. ertical distribution of phytoplankton communities in open ocean: An assessment based on surface chlorophyll. *J. Geophys. Res*, 111(C08005), 2006.