

# Developing Zoo & Phytoplankton EOVS Products in Blue-Cloud

## Authors:

Patricia Cabrera, Lennert Schepers, Steven Pint, Gert Everaert (VLIZ), Renosh Pannimpullath Remanan (CNRS-LOV), Alexander Barth (GHER-Ullege) and Federico Drago (Trust-IT Services)



As highlighted by Principle 5 of [Ocean Literacy](#), the ocean supports a great diversity of life and ecosystems, which make up the vast majority of all life on our planet and are crucial for biological cycles as well as for economies all around the world.

Comprehending this wealth of biodiversity, collecting reliable marine biodiversity data, being able to perform high-quality research across disciplines and countries, are all critical steps towards a sustainable management of the precious resources that the ocean provides.

Marine plankton are at the base of the marine food web and play an important role in the functioning of coastal and open ocean ecosystems. Understanding how plankton changes through time and space is of key importance to assess the state of the marine plankton ecosystems and their response to climate change. The Blue-Cloud demonstrator [Zoo- and Phytoplankton EOV Products](#), is creating phyto- and zooplankton biomass and diversity products, working on data from leading European marine data infrastructures such as EMODnet, SeaDataNet, and Copernicus. In this article, we are going to explore the main results of the demonstrator, and its intended evolution.

The Zoo and Phytoplankton EOV Products demonstrator is developed by the [Flanders Marine Institute \(VLIZ\)](#), in collaboration with the [Faculty of Science and Engineering at Sorbonne University](#) and [GeoHydrodynamics and Environment Research \(GHER\)](#) at the University of Liège. A dedicated Virtual Lab was developed in the Blue-Cloud Virtual Research Environment powered by D4Science, and introduced through a public webinar in [February 2021](#) describing its scope, key features and the potential benefits for the ocean science community.

## What can researchers do with it?

The Zoo and Phytoplankton EOV demonstrator consists of 3 different products:

### ■ *Global ocean 3D Chlorophyll a product*

Chlorophyll a (Chla) is the key pigment associated with phytoplankton photosynthesis and it is widely used as a proxy for the phytoplankton biomass in the ocean. This product generates (3D) vertical distribution of Chla concentrations using machine learning methods. The workflow can be executed on a Jupyter notebook, which includes a script to combine satellite-based inputs from [CMEMS](#) and [GlobColour](#) with outputs from the BioGeoChemical ([BGC-Argo database](#)), allowing its reproduction by users with different data.

### ■ *Methodology*

The method developed follows the approach of Sauzède et al. (2016), which uses a neural network-based method that merges ocean color satellite imagery (measuring Chla) and in situ Biogeochemical (BGC)-Argo floats data to retrieve the vertical distribution of Chla at the global scale. The neural network used is the Multi Layer Perceptron (MLP). The training of the MLP model consists of 1 input layer, 2 hidden layers and 1 output layer.

The input layer includes:

- › Surface component composed of satellite data:
  - › Remote sensing reflectances (412, 443, 490, 555, and 670 nm)
  - › Sea level anomaly
  - › Photosynthetically available radiation (PAR)
- › Vertically resolved physical parameters derived from BGC-Argo floats data:
  - › Temperature profiles, Principal Component Analysis (PCA) transformed
  - › Salinity. PCA transformed
  - › Spiciness PCA transformed
  - › Mixed Layer Depth

> Space-time components:

- > Day of the year (in cycle transformation) and longitude and latitude (in cartesian transformation) of the considered satellite-to-Argo matchup

The output layer is profiles of Chla concentrations.

## ■ Model validation

From around 90,000 profiles of the Global BGC database, we selected the Chla profiles that had matchups with Remote sensing reflectance ( $R_{rs}$ ) and Sea Level Anomaly (SLA) for the present study. The total numbers of Chla profiles considered were 21737, from which 80% of profiles (17390 profiles) were selected for the training and 20% profiles (4347 profiles) were selected for the validation. The red points shown in the map are the locations of the Chla profiles used for the training, and the blue points are the locations of the profiles used for the validation (left panel in Figure 1). The validation of the model was performed with Chla from MLP versus Chla from floats for 4347 profiles using density scatter plot (right panel in Figure 1). The linear regression comparison between the Chla derived from the model versus Chla from floats shows satisfactory results with a coefficient of determination ( $R^2$ ) value of 0.76, slope value of 0.93, and Mean Absolute Percentage Difference (MAPD) of 21.28%.

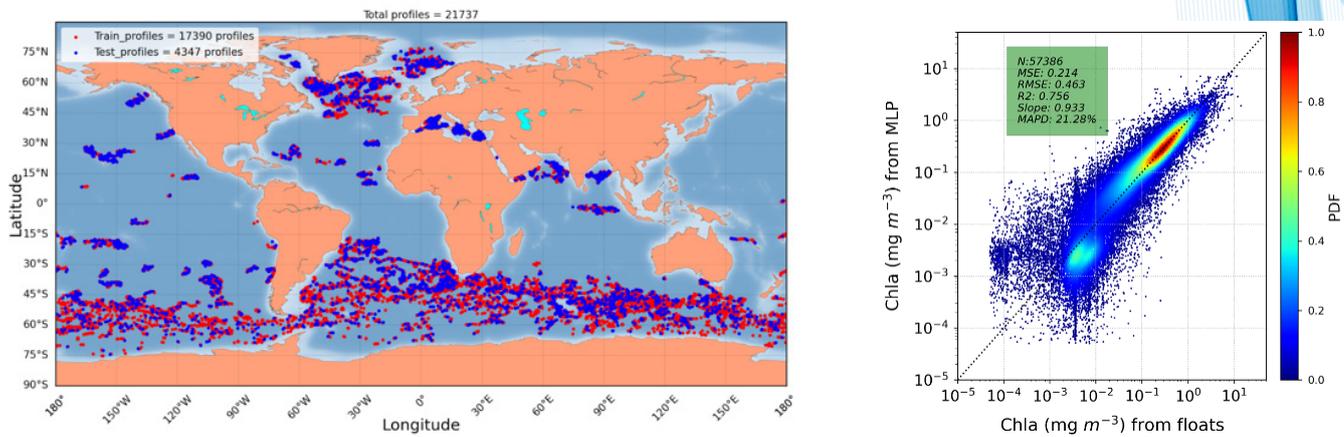


Figure 1: Left panel represents the geographical locations of Chla (Argo) profiles used for the present stud. Right panel represents a density scatter plot for the validation of the model.

## ■ Global 3D Chla product

Using the trained MLP model, we can generate Chla concentrations for 36 depths from surface to 1000 metres [0, 5, 10, 15, 20, 25, 30, 35, 40, 45, 50, 55, 60, 65, 70, 80, 90, 100, 125, 150, 175, 200, 225, 250, 275, 300, 350, 400, 450, 500, 550, 600, 700, 800, 900, 1000]. The jupyter notebooks provided in the Blue-Cloud demonstrator generate global 3D monthly averages of Chla at 36 depths for the year 2018. The global distributions of Chla for two specific seasons (typical winter and summer months) are shown in Figure 2.

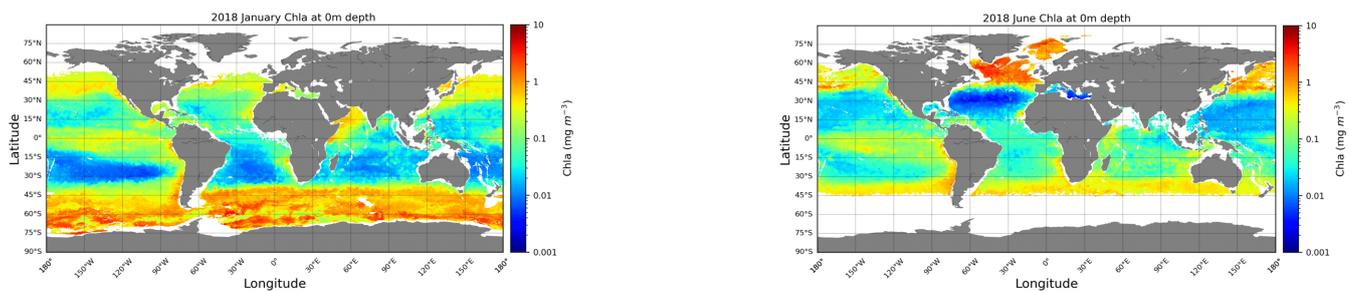


Figure 2: Global 3D monthly averages of Chla at 36 depths for January 2018 (a typical winter month) in the left panel and for June 2018 (a typical summer month) in the right panel.

## ■ Zooplankton distribution maps

In situ data in oceanography are often scarce and can also present inhomogeneous distributions. The spatial interpolation performed with the tool [DIVAnd](#) (Data Interpolating Variational Analysis in  $n$  dimensions; Barth et al. 2014, 2021) allows the use of irregular sample datasets to go from sparse in situ observations to gridded maps. The Continuous Plankton Recorder (CPR) dataset from the Marine Biological Association (MBA, UK), provides unique abundance data of plankton over a long-term period and a large spatial area with a heterogeneous distribution, therefore we used this dataset in this product.

## ■ Methodology

This product allows the development of interpolated abundance maps of the most abundant copepod species from the North East Atlantic: *Acartia* spp, *Calanus helgolandicus*, *Calanus finmarchicus*, *Metridia lucens*, *Oithona* spp and *Temora longicornis* (Figure 3). It is useful to detect spatial trends and long-term anomalies in zooplankton distributions. The method uses DIVAnd software tool and a neural network model in a combined loss function. The observed zooplankton abundances are complemented with environmental parameters as co-variables to improve the interpolation; temperature and salinity from SeaDataCloud, and the nutrients nitrate, silicate and phosphate from the World Ocean Atlas 2018.

The workflow can be executed in 2 Jupyter notebooks written in Julia. The first notebook allows us to do the analysis (covariables preparation, split of the data into a training and validation set and the DIVAnd interpolation and neural network analysis). The second notebook allows us to visualise the results (Figure 3).

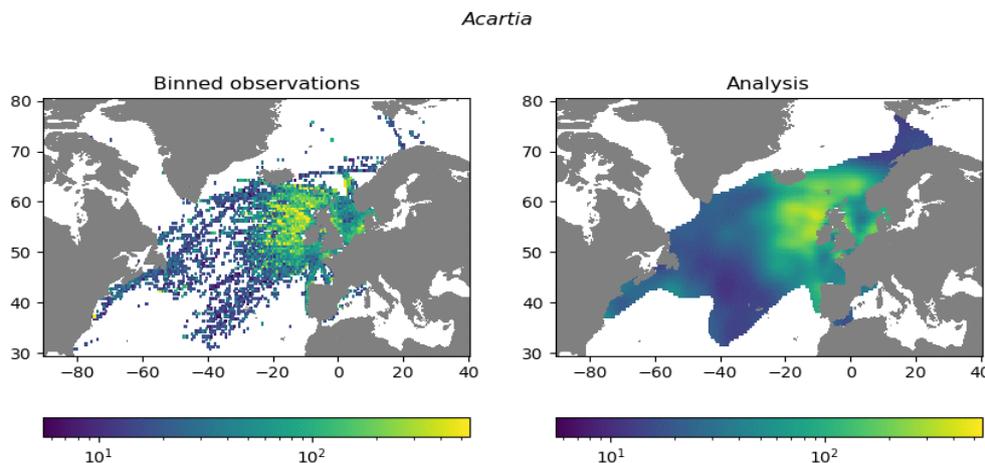


Figure 3: Interpolated abundance maps of the most abundant Copepod species from the North

East Atlantic. Areas with only very few data points are masked but the results over the full grid are computed.

## ■ Validation of results

The uncertainty of the output is captured by the relative error (stored in the output NetCDF files) and the validation statistic (JSON files) that is the mean squared error between the analysis and the data set aside for validation. The validation statistics represents the accuracy of the results on average in absolute terms while the relative error indicates how much the analysis is influenced by observations close-by.

## ■ Modelling phyto-zooplankton interactions

Modelling phyto-zooplankton interactions helps us to understand how primary production changes through time and space in the ocean. The Nutrient, Phytoplankton, Zooplankton and Detritus (NPZD) ecosystem model used in this demonstrator is based on Soetaert and Herman (2009). The model in this demonstrator uses near real-time

data of phyto- and zooplankton abundances (accessible from EMODnet Biology) and environmental parameters from LifeWatch, except sea surface temperature for the offshore station from the Flemish Monitoring Network.

## Methodology

It is useful to know what factors drive phytoplankton abundance and how these factors change in space and time. The model simulates daily changes in phytoplankton abundances based on zooplankton grazing, abiotic parameters (temperature and Photosynthetically Active Radiation, PAR), and nutrients (Dissolved Inorganic Nitrogen, DIN), Phosphate ( $PO_4$ ) and silicate,  $SiO_4$ ), in three different locations in the Belgian part of the North Sea (BPNS).

The workflow can be executed in the Rstudio server of the Blue-Cloud VRE via a R Markdown file. The workflow includes the following steps:

- Applying Generalized Additive Models (GAM) on available (usually monthly and seasonally) observed time series to complete daily time series for the input data of the NPZD model;
- Running the NPZD Model to calculate and visualize phyto- and zooplankton abundances;
- Validating the NPZD model based on observed Chlorophyll-a and zooplankton abundances
- Calculating and visualizing the drivers that limit phytoplankton abundances (Figure 4)

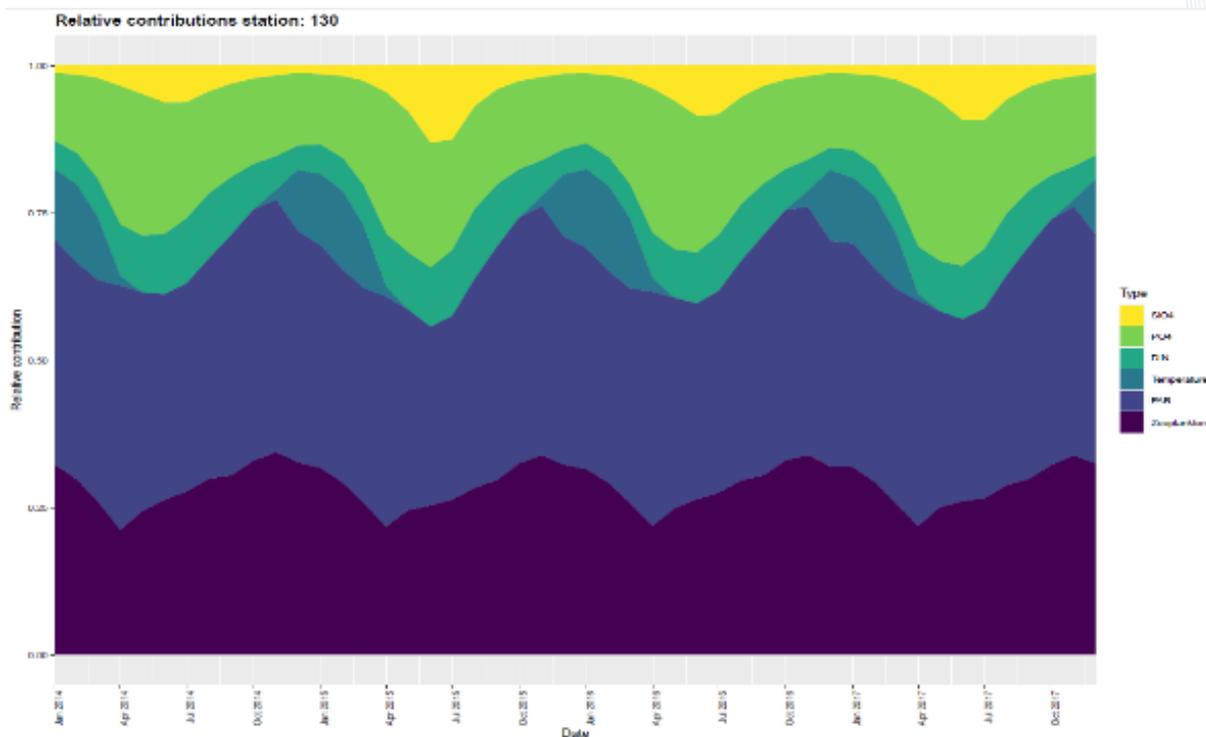


Figure 4: Average monthly relative contributions from 2014 to 2017 for each limitation factor (yellow= $SiO_4$ , light green= $PO_4$ , dark green=DIN, light blue=temperature, dark blue=PAR and purple=zooplankton grazing) in phytoplankton abundance in one of the stations at the BPNS.

The validation of the model is performed by comparing the model predictions with the actual field observations. By doing so, and by running the model for multiple parameterizations we are able to define confidence intervals around the model predictions.

## Multidisciplinary links and scaling up

New developments that are foreseen in the coming months include (1) the development of the phytoplankton community product as part of the Phytoplankton EOVs demonstrator, (2) to assess if a time component (e.g.,

monthly or seasonal) can be included in the Zooplankton EOVs demonstrator and (3) application of the NPZD model at other regional seas and with data used in the other parts of the demonstrator for validation.

In the coming months, we are also looking into how to link this demonstrator with the Blue-Cloud [plankton genomics](#) and [fisheries](#) demonstrators. In the future these products can be scaled up into indicators to inform policies regarding the number of loss species or ecosystem functionality. We could assess how our demonstrators can be scaled across space (from local/regional to global) and time (longer time series).

Finally we are also looking forward to the contribution of data experts in the upcoming Blue-Cloud Hackathon foreseen in 2022.

[🔗 Test the Zoo & Phytoplankton EOV Products Virtual Lab](#)

## References

1. Sauzède, R., Claustre, H., Uitz, J., Jamet, C., Dall'Olmo, G., d'Ortenzio, F., Gentili, B., Poteau C. & Schmechtig, C. (2016). A neural network-based method for merging ocean color and Argo data to extend surface bio-optical properties to depth: Retrieval of the particulate backscattering coefficient. *Journal of Geophysical Research: Oceans*, 121(4), 2552-2571, <https://doi.org/10.1002/2015JC011408>.
2. Soetaert, K., Herman, P.M.J. (2009). A practical guide to ecological modelling. *Using R as a Simulation Platform*. Springer-Verlag, New York, US, p. 54 - 58.
3. Barth, A., Beckers, J.-M., Troupin, C., Alvera-Azcárate, A., and Vandenbulcke, L. (2014): [DIVAnd-1.0: n-dimensional variational data analysis for ocean observations](#), *Geosci. Model Dev.*, 7, 225-241, doi:10.5194/gmd-7-225-2014.
4. Barth, A., Troupin, C., Reyes, E., Alvera-Azcárate, A., Beckers J.-M. and Tintoré J. (2021): [Variational interpolation of high-frequency radar surface currents using DIVAnd](#). *Ocean Dynamics*, 71, 293-308, doi: 10.1007/s10236-020-01432-x.