Blue-Cloud Demonstrator: A machine learning approach to derive plankton biomass and diversity products from the Global Ocean.

Patricia Cabrera, Flanders Marine Institute (Belgium), patricia.cabrera@vliz.be
Viviana Otero, Flanders Marine Institute (Belgium) viviana.otero@vliz.be
Gert Everaert, Flanders Marine Institute (Belgium) gert.everaert@vliz.be
Raphaëlle Sauzède, Institut de la Mer de Villefranche, CNRS – Sorbonne Université (France), raphaelle.sauzede@imev-mer.fr
Renosh Pannimpullath Remanan, Laboratoire d'Océanographie de Villefranche – Institut de la Mer de Villefranche, CNRS – Sorbonne Université (France), renosh.pr@obs-vlfr.fr
Hervé Claustre, Laboratoire d'Océanographie de Villefranche – Institut de la Mer de Villefranche, CNRS – Sorbonne Université (France), claustre@obs-vlfr.fr
Julia Uitz, Laboratoire d'Océanographie de Villefranche – Institut de la Mer de Villefranche, CNRS – Sorbonne Université (France), julia.uitz@imev-mer.fr
Alexander Barth, GHER/ULiège (Belgium), a.barth@uliege.be
Charles Troupin, GHER/ULiège (Belgium), ctroupin@uliege.be
Lennert Schepers, Flanders Marine Institute (Belgium), lennert.schepers@vliz.be

The European marine data landscape and the Blue-Cloud project

Oceanographic and marine data management in the European landscape has progressed considerably during the last three decades, developing standards, services, and establishing dedicated infrastructures. As such, providing discovery and access to multidisciplinary data sets is becoming essential to provide analytical frameworks and ultimately acquire a better understanding of the ocean system. Large amounts of multidisciplinary marine data are available in an interoperable or harmonized manner in established leading European marine data infrastructures (e.g. SeaDataNet, EurOBIS, Euro-Argo, EMODnet, ELIXIR-ENA, EuroBioImaging, Copernicus Marine Environment Monitoring Service). As new technologies become available, there is a challenge to adopt new methodologies and optimise the technologies to expand the data management services. In this context, the operators of these European marine data infrastructures have joined forces to explore and demonstrate the power of a Blue Cloud virtual research environment.

Blue Cloud is a thematic European Open Science Cloud (EOSC) project that aims to implement a practical approach to explore and demonstrate the potential of cloud-based open science. The purpose is to exploit the already existing resources and to develop and deploy, through a pragmatic workplan, the Pilot Blue Cloud as a cyber platform bringing together and providing access to (1) multidisciplinary data from observations and models, (2) analytical tools, and (3) computing facilities essential for key blue science use cases. The Pilot Blue Cloud will support research to better understand and manage aspects of ocean sustainability, through a set of five pilot Blue-Cloud demonstrators. The demonstrators build upon a range of oceanographic data from multiple streams, made interoperable and integrated through Blue Cloud services. The Blue Cloud Services developed can contribute to unlocking the innovation potential of the Blue Cloud, showcasing its potential in promoting the blue economy shortening the time span between research and innovation in frontier fields, such as micro-organisms and genomics-enabled innovations.

Plankton EOV demonstrator

The demonstrator on plankton Essential Ocean Variables (EOVs) aims at rationalizing data collection, facilitate data dissemination and maximize data utilization. As biological EOVs on phytoplankton and zooplankton are still in conceptual or pilot phase, this demonstrator can contribute directly to their further development and application on large scales. Plankton plays an important role in the marine

ecosystem and comprises an integral and vital component at the base of the pelagic food web. Plankton communities provide an indication of the health of marine ecosystems and their response to anthropic stressors. Moreover, they are used within several descriptors of the Marine Strategy Framework Directive (MSFD); D1: Biodiversity; D4: Food webs; D6: Sea Floor Integrity; D5: Eutrophication. In particular, phytoplankton EOVs will contribute to the understanding of the environmental conditions and top-down factors that shape the global distribution of phytoplankton community biomass and diversity. Zooplankton EOVs showing spatial trends and long-term anomalies in abundances will provide the basis for studying the dynamics of food availability for commercially exploited fish species.

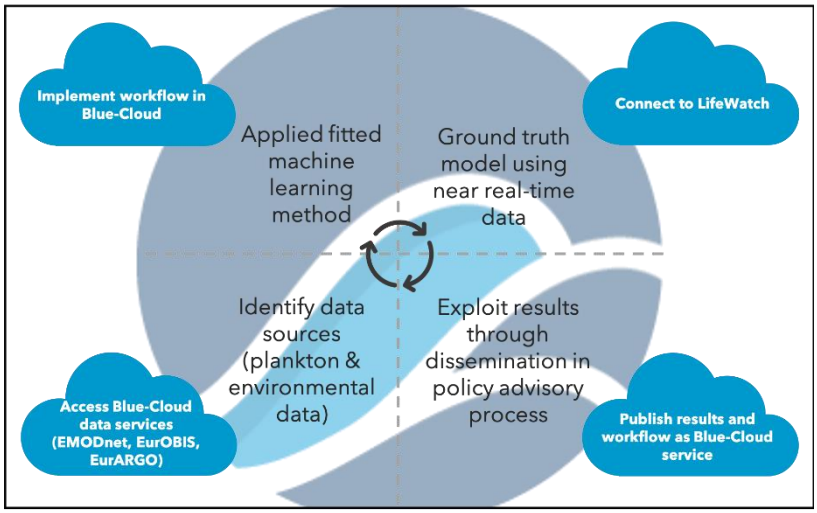

Figure 1: Plankton EOVs demonstrator workflow

The workflow of the plankton EOV demonstrator (Figure 1) consists of (1) compiling and processing several large-scale plankton and environmental data that are available from multiple European marine networks, (2) applying big data analysis and machine learning (e.g. neural networks) methods to create novel, synergistic ocean products, (3) including models using near real-time observations from LifeWatch data, (4) and exploiting the results that will be made available to a variety of users through a Blue Cloud Virtual Lab. The virtual EOVs produced will contribute to improve knowledge and quantitatively reduce uncertainty regarding the present state of the marine plankton ecosystems and their response to climate change. More specifically, the phytoplankton products will consist of global 4D fields of key phytoplankton products, namely EOVs of phytoplankton chlorophyll biomass and diversity, the latter being expressed as chlorophyll contribution of several phytoplankton groups (PFTs). These EOVs will be obtained from multidisciplinary (physics, biogeochemistry, biology) data acquired from oceanographic platforms with complementary space-time resolution and coverage (e.g. vessels, satellites, Argo/BGC-Argo floats). Whereas, the zooplankton products will show spatial trends and long-term anomalies in zooplankton distributions, focusing on the most abundant Copepod species from the North East Atlantic. A neural network model will be built with the observed zooplankton abundances complemented with environmental parameters (oxygen, temperature, salinity, chlorophyll, depth, nutrients) to create spatio-temporal products.

We verify our products using a nutrient–phytoplankton–zooplankton model and near-real time observations from LifeWatch data. This will enable us to detect and explain anomalies in long-term trends and quantify the relative contribution of the environmental drivers for these EOVs. The results from the demonstrator will be implemented in a Blue-Cloud Virtual Lab, providing its users access to blue multidisciplinary data for exploring the methodology and data used in the products. Ultimately, the products will help marine advisory process and serve the wider Blue community.