Blue-Cloud

Applying machine learning methods to ocean patterns and ocean regimes indicators

Authors:
Andrea Garcia Juan, Kevin Balem,
Loic Bachelot (LOPS, IFREMER),
Massimiliano Drudi (CMCC Foundation),
Federico Drago (Trust-IT Services)

Blue-Cloud has received funding from the European Union's Horizon programme call BG-07-2019-2020, topic: [A] 2019 - Blue Cloud services, Grant Agreement No.862409.

The global challenges that humankind is called to face highlight the need for establishing innovative algorithms and technologies to enable the transition from data to knowledge, and foster the consolidation of a science-informed decision-making process.

For a successful implementation of this value chain, the development of science-based algorithms clearly represents a crucial phase. We will analyse the latest updates on the application of machine learning methods to ocean patterns and the ocean regimes indicators in the context of Blue-Cloud.

The Blue-Cloud demonstrator "Marine Environmental Indicators" has a specific focus on data related to the marine environment. Its development is led by the CMCC Foundation, in collaboration with IFREMER, Mercator Ocean International, the Royal Netherlands Meteorological Institute (KNMI), and the University of Bergen.

Its dedicated Virtual Lab was created in the Blue-Cloud Virtual Research Environment powered by D4Science, and introduced in a public webinar in December 2020 outlining its scope, key features and the potential benefits for the ocean science community.

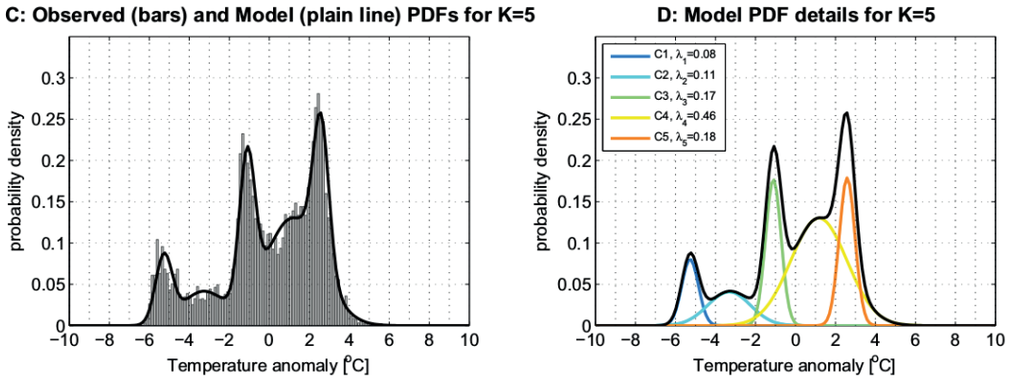
Developing marine environmental indicators

As part of their work on this demonstrator, the team has recently developed the Ocean Patterns and the Ocean Regimes Indicators, which constitute an easy way of applying machine learning methods to ocean profiles and ocean time series, respectively.

- The **Ocean Patterns** Indicator consists in applying a clustering method called GMM (Gaussian Mixture Model) to a dataset of ocean profiles. Any type of variable can be used: temperature profiles, chlorophyll profiles… The ocean profiles are assembled into several clusters depending on their vertical structure. When plotting the different classes, spatial and temporal coherences can be revealed, that is what we define as the Ocean Patterns Indicator.
- The **Ocean Regimes** Indicator involves applying the same clustering method to a dataset of ocean time series (Chlorophyll-a, SST…). The ocean time series are gathered into clusters depending on their seasonal variability. For this indicator, spatial coherences can be revealed when plotting the different classes in a map.

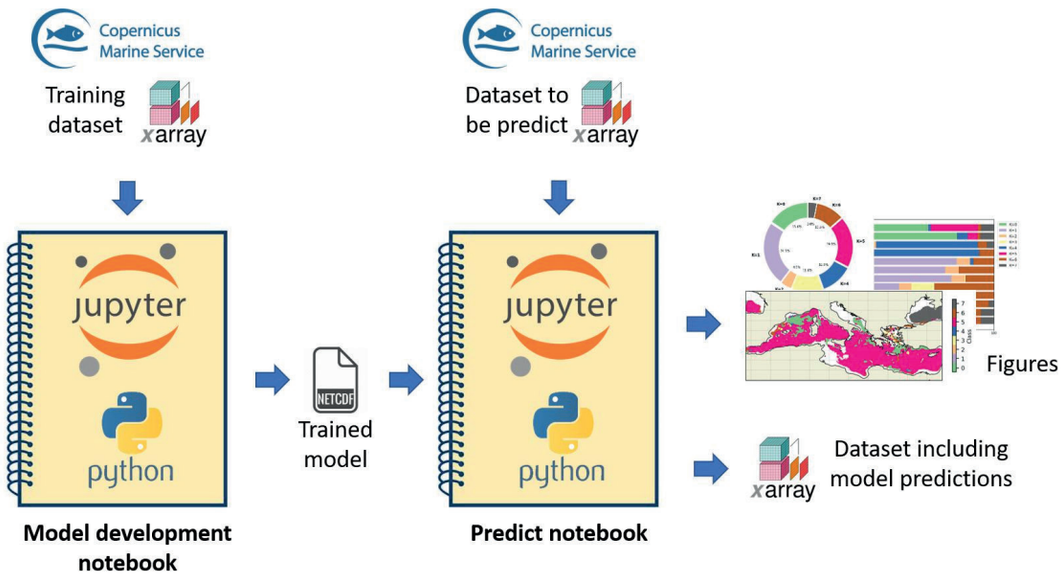
Method and workflow

The core method behind these two indicators is the Gaussian Mixture Model (GMM). This clustering method decomposes the probability density functions (PDF) of the dataset into a sum of gaussian PDF. Users should only choose the number of classes. In the input there is no spatial or temporal information so the classes depend only on the PDF of the dataset: the vertical structure similarities in the case of profiles (Ocean Patterns) and seasonal structure similarities in the case of time series (Ocean Regimes).



Maze et al, Prg.Oc, 2017

These indicators have been developed into Jupyter Notebooks available to users in the Marine Environmental Indicators Virtual Lab, included in the Blue-Cloud Virtual Research Environment. The workflow is structured into two notebooks for both indicators, a model development notebook, and a prediction notebook.

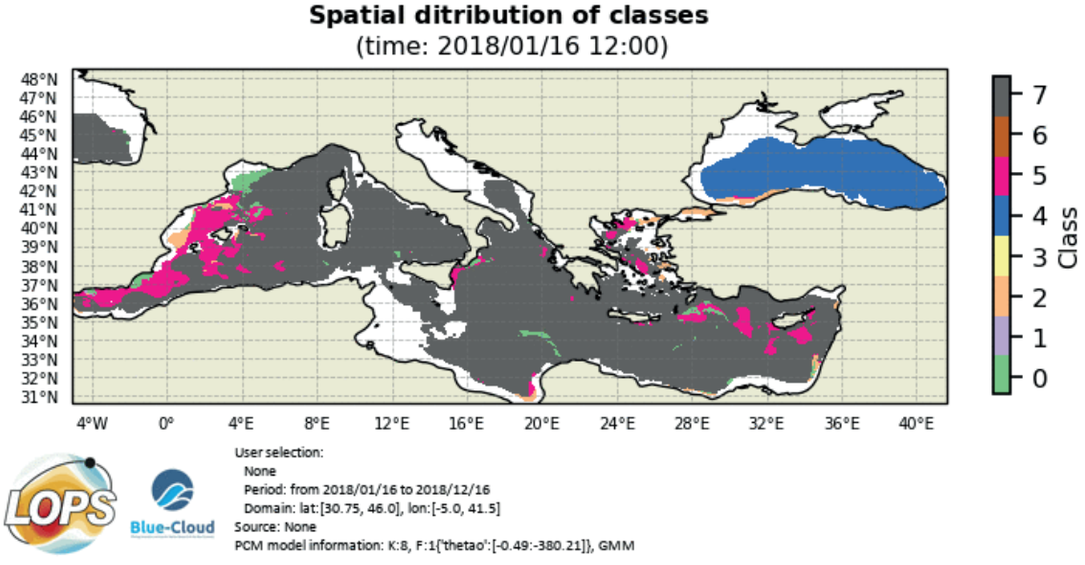


In the **model development** notebook, users can design a clustering model: they can choose the number of classes to apply, then the model is trained (fitted) with the training dataset. Some plots are available to adjust the optimal number of classes. Finally, the trained model is saved into a file, so that it can be used in the prediction method.

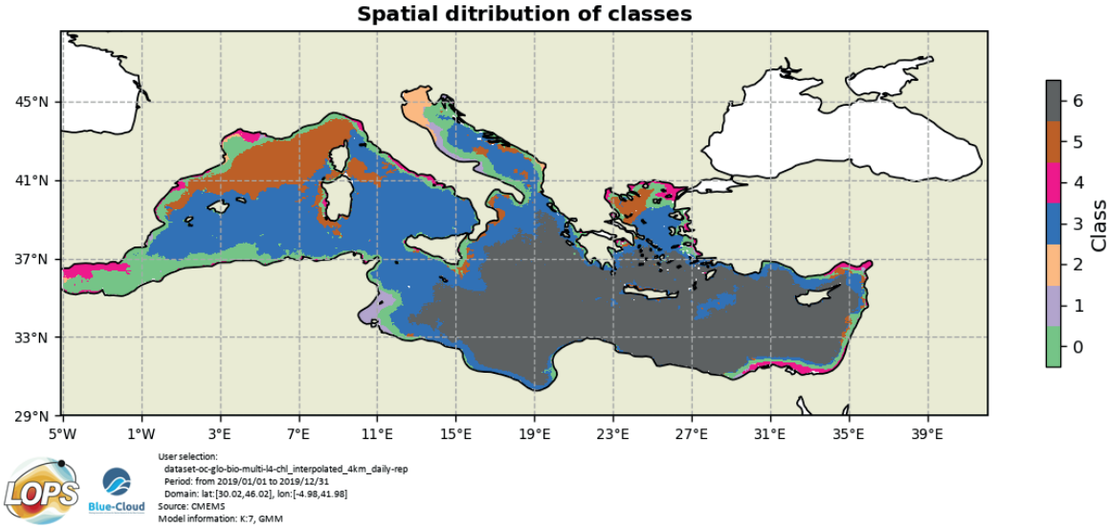
In the **prediction notebook**, a trained model is applied to some data, so that the profiles or time series from the input data selection will be sorted into clusters. Then different plots are proposed to analyse the results: spatial and temporal distributions, and median time series or profiles for each class.

Examples

For the Ocean Patterns Indicator, here is an example dataset of temperature profiles in the Mediterranean (GLOBAL_REANALYSIS_PHY_001_030 CMEMS product). Vertical profiles are classified into 8 classes. In general, one class is predominant in all Mediterranean Sea for each month: the classification shows the evolution of temperature profiles through one year: from mixture profiles in winter to more stratified ones in summer.



For the Ocean Regimes Indicator, the notebook shows an example dataset of chlorophyll-a time series in the Mediterranean Sea. The example below is based on the work of Fabrizio D'Ortenzio (D'Ortenzio and d'Alcalà, Biogeosciences, 2009) and Nicolas Mayot (Mayot et al, Biogeosciences, 2016). Spatial distribution of the classes highlights a "bloom" time series located in the gulf of Lyon and a structure in the Eastern basin corresponding to Rhodes Gyre.



Reaching out to the wider scientific community

The Blue-Cloud demonstrator "Marine Environmental Indicators" will continue to tackle the implementation of the value chain, ranging from Marine Data Infrastructures to the knowledge for enhancing decision-making processes. After the development of the algorithms for Ocean Patterns and Ocean Regimes Indicators, the following step is the integration into the production environment of this method, and therefore making this tool effectively available to a wider audience of end-users.

Furthermore, to foster the establishment of the FAIR principles, special attention will be dedicated to the assessment of the all the most relevant aspects to ensure not only the findability of the generated data and of the method itself, but also the interoperability with the many data infrastructures that could potentially provide valuable input data to this method.

Test the Marine Environmental Indicators Virtual Lab