MACHINE LEARNING MODELLING OF MARINE PHYTOPLANKTON PRIMARY PRODUCTION

Ph.D. Student: Francesco Mattei/ *Supervisor*: Prof. Michele Scardi

34th Cycle - A.Y. 2018/2019

Oceans cover roughly 70% of the Earth's surface and therefore enhancing the understanding of their physical and biological dynamics is paramount. In this context, marine phytoplankton plays a crucial role since it accounts for the bulk of the whole oceanic primary production. Accordingly, it is intimately linked to the carbon cycle and it influences the ocean-atmosphere interactions through a net carbon dioxide uptake by the water column. Moreover, a share of this fixed carbon is removed from the fast carbon cycle through sinking of organic matter in a process known as biological pump. Furthermore, since the phytoplankton constitutes the first trophic level in the marine environment, the production expressed by these photosynthetic organisms represents the main source of energy supporting the marine food webs. Therefore, phytoplankton primary production deeply influences not only the characteristics of the marine ecosystems, and in turn the respective ecosystem services, but also Earth's climate. Indirect approaches for estimating the phytoplankton production have raised in popularity as field measurements are expensive and time-consuming especially on a large spatial scale. Modelling attempts range from empirical to mechanistic estimators. In this context, we developed a depth-resolved Artificial Neural Network (ANN) model for the estimation of phytoplankton primary production. The rationale behind the use of Machine Learning (ML) techniques, such as ANNs, lies in the advantages they provide in handling complex non-linear relations, which dominate natural systems. Moreover, ML approaches enhance the exploitation of the available information which can be retrieved from several heterogeneous sources, e.g. the remote sensing. Indeed, no *a priori* explicit mathematical formulation of the link between predictive variables and the output is needed as the relationships between them are derived from the data during the model development (training, in ML jargon). The goal of my Ph.D. project is threefold: firstly, we aim at extend the area of applicability of our model by retrieving more data for training better ANNs; secondly, we want to enhance the quality of our estimates by both retrieving new predictive variables and embedding ecological knowledge into the training procedure; finally, we plan to exploit our production estimates in an ecosystem services context e.g. an emergetic assessment of the marine phytoplankton production.

Mattei, F., Franceschini, S., Scardi, M., 2018. A depth-resolved artificial neural network model of marine phytoplankton primary production. Ecol. Model., 382: 51-62.

Mattei, F., Scardi, M., 2018. A constrained depth-resolved artificial neural network model of marine phytoplankton primary production. 10th International Conference on Ecological Informatics, 24-28 September 2018, Jena, Germany.

Franceschini S, Tancioni L, Lorenzoni M, Mattei F, Scardi M, 2019 An ecologically constrained procedure for sensitivity analysis of Artificial Neural Networks and other empirical models. PLoS ONE 14(1): e0211445.