EGU21-11974, updated on 24 Oct 2021
https://doi.org/10.5194/egusphere-egu21-11974
EGU General Assembly 2021

Deep learning approach to reconstruct satellite ocean color time series in the global ocean

Joana Roussillon[1], Ronan Fablet[2], Lucas Drumetz[2], Thomas Gorgues[1], and Elodie Martinez[1]
[1]Laboratoire d'Océanographie Physique et Spatiale, CNRS/IFREMER/IRD/UBO, Institut Universitaire Européen de la Mer, Plouzané, France (joana.roussillon@etudiant.univ-brest.fr)
[2]IMT Atlantique, UMR CNRS LabSTICC, Technopole Brest Iroise, 29239 Brest, France

Phytoplankton plays a key role in the carbon cycle and constitutes the basis of the marine food web. Its seasonal and interannual cycles are relatively well-known on a global scale thanks to continuous ocean color satellite observations acquired since 1997. The satellite-derived chlorophyll-a concentrations (Chl-a, a proxy of phytoplankton biomass) time series are still too short to investigate phytoplankton biomass low-frequency variability. However, it is a vital prerequisite before being able to confidently detect anthropogenic signals, as natural decadal variability can accentuate, weaken or even mask out any anthropogenic trends. Machine learning appears as a promising tool to reconstruct Chl-a past signals (including periods before satellite Chl-a era), and deep learning models seem particularly relevant to explore the spatial and/or temporal structure of the data.

Here, different neural network architectures have been tested on a 18-year satellite and re-analysis dataset to infer Chl-a from physical predictors. Their ability to reconstruct spatial and temporal (seasonal and interannual) variations on a global scale will be presented. Convolutional neural networks (CNN) better capture Chl-a spatial fields than models that do not account for the structure of the data, such as multi-layer perceptrons (MLPs). We also assess how the selection of training period may affect the reconstruction performance. This is a necessary step before being able to reconstruct any past Chl-a multi-decadal time series with confidence, which is the ultimate goal of this work.

Our study also addresses the carbon footprint associated with the use of GPU resources when training the CNN. GPUs are energy intensive, and their use in geosciences is expected to grow fast. Systematically reporting the computational energy costs in the geoscience community studies would provide an overview of models energy-efficiency on different kinds of datasets and may encourage actions to reduce consumption when possible.