



## **A global neural network-based parameterization of biogeochemical water mass properties and processes based on GLODAP data**

Henry C. Bittig (1), Raphaëlle Sauzède (1,2), Hervé Claustre (1), Orens Pasqueron de Fommervault (1,3), Jean-Pierre Gattuso (1,4), Louis Legendre (1), and Ken Johnson (5)

(1) Sorbonne Universités, UPMC Univ Paris 06, CNRS-INSU, Observatoire Océanologique de Villefranche, Laboratoire d'Océanographie de Villefranche, Villefranche-Sur-Mer, France, (2) Ecosystemes Insulaires Océaniques (EIO, UMR-241), IRD, Ifremer, UPF and ILM, Tahiti, French Polynesia, (3) Departamento de Oceanografía Física, Centro de Investigación Científica y de Educación Superior de Ensenada, Ensenada, Mexico, (4) Institute for Sustainable Development and International Relations, Sciences Po, Paris, France, (5) Monterey Bay Aquarium Research Institute, Moss Landing, CA, USA

Global data collections like GLODAP are an extensive source of biogeochemical and hydrological data. However, data are irregularly distributed in space and time with varying parameter-coverage. This poses a challenge to data analysis of, e.g., the global distribution of stoichiometric ratios or temporal trends. Here we utilize a neural network-based approach called CANYON to estimate carbonate system parameters ( $C_T$ ,  $A_T$ , pH, and  $pCO_2$ ) and nitrate, phosphate, and silicate concentrations from commonly measured quantities (P, T, S,  $O_2$ , location, and date). CANYON was derived using GLODAPv2 data but can be applied to any set of input quantities (e.g., observations from autonomous platforms like Biogeochemical-Argo floats with accurate  $O_2$  measurements). In essence, CANYON provides a mapping of water mass properties and biogeochemical relations for those parameters based on the multidecadal, global observations collected in GLODAPv2. It can thus provide biogeochemical context and fill observational gaps, e.g., where nutrient or carbonate system measurements are unavailable. As an example, float-based surface CTD- $O_2$  observations together with the CANYON parameterization are used to obtain surface  $pCO_2$  estimates in the Southern Ocean, complementing sparse surface underway  $pCO_2$  data collected in SOCAT. Moreover, it can shed light on global variations of, e.g., Redfield ratios of nitrate, phosphate, oxygen, and carbon. We believe that this parametrization provides a useful alternative to scattered data points or a mapped climatology to facilitate utilization and exploitation of the unique GLODAP data collection.