Supplement of Ocean Sci., 20, 217–239, 2024
https://doi.org/10.5194/os-20-217-2024-supplement

Ocean Science Open Access EGU

Supplement of

Linking satellites to genes with machine learning to estimate phytoplankton community structure from space

Roy El Hourany et al.

Correspondence to: Roy El Hourany (roy.elhourany@univ-littoral.fr), Marina Levy (marina.levy@locean.ipsl.fr), and Chris Bowler (cbowler@biologie.ens.fr)

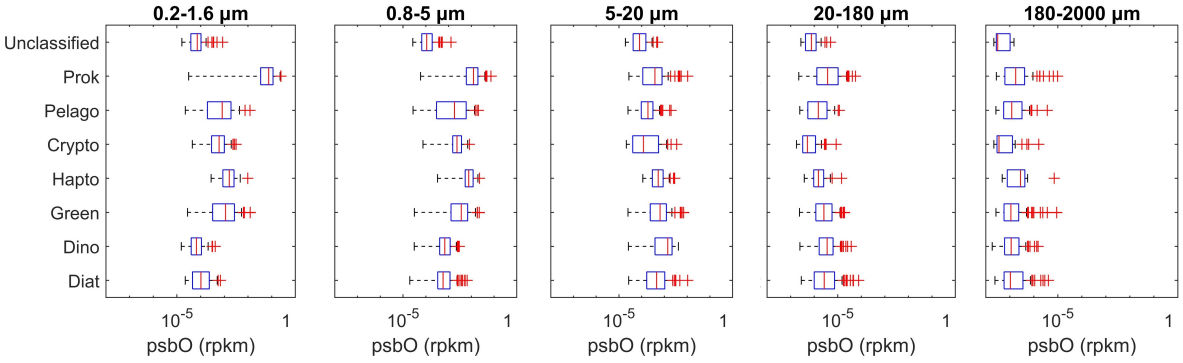

Figure S1. The *psbO* abundance expressed in rpkm (reads per kilobase covered per million of mapped reads). There is a decrease in rpkm values towards the larger size fractions, probably explained by the increase in genome size and complexity in larger size fractions. In addition, prokaryotes are dominant in the smaller size fractions while the larger fractions are characterized by the higher prevalence of eukaryotic phytoplankton

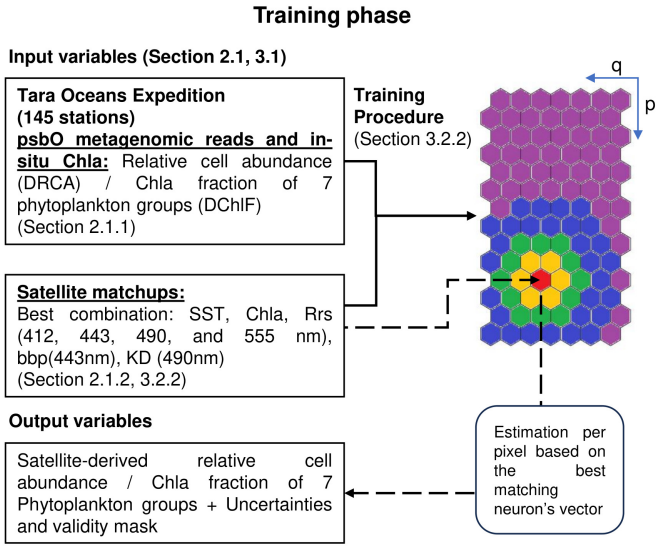

Figure S2. Flowchart describing the steps of the training process of SOMRCA and SOMChlF. Further information on the training procedure is elaborated in the flowchart Fig. S3

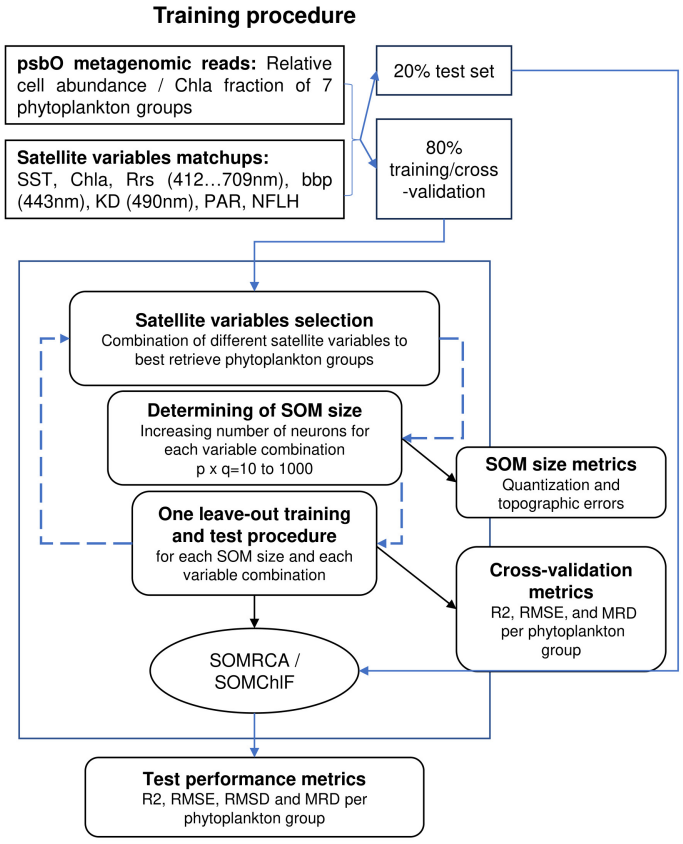

Figure S3. Flowchart describing the steps of the training procedure while testing combinations of satellite predictors and SOM map size (number of neurons) for both SOMRCA and SOMChlF. For each combination of satellite predictors and map size, a set of performance metrics are generated to allow an objective choice of satellite parameters and map size to best estimate the phytoplankton community structure.

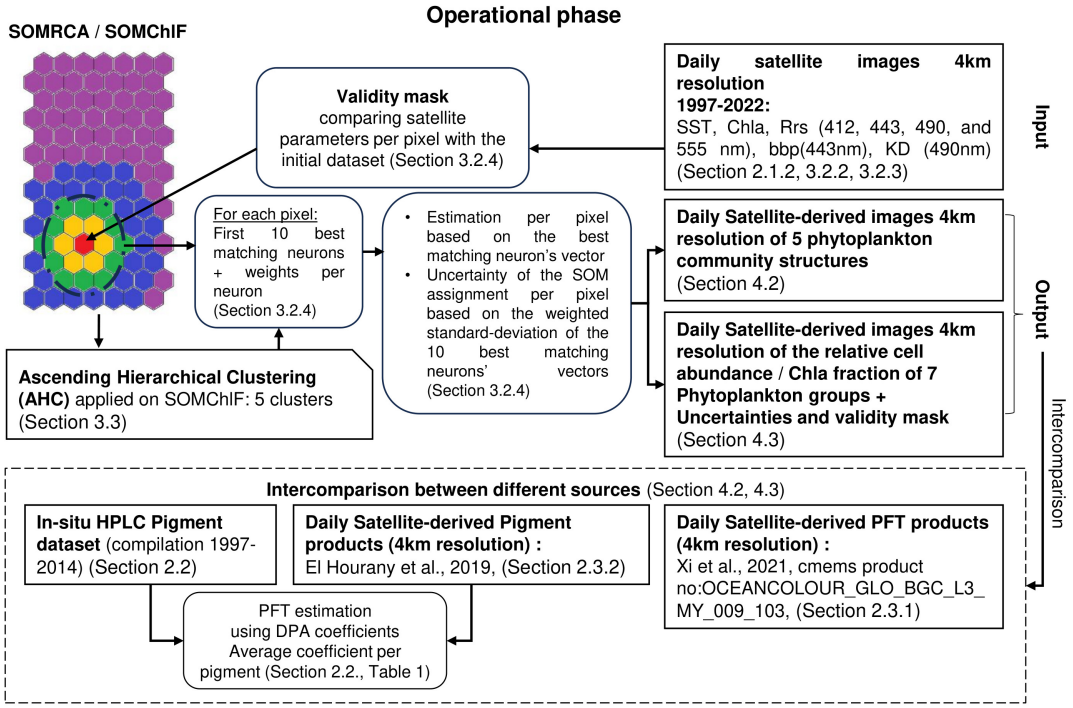

Figure S4. Flowchart describing the operational phase of the SOMRCA and SOMChlF, alongside the calculation of quality control (reliability index) and uncertainty metrics.

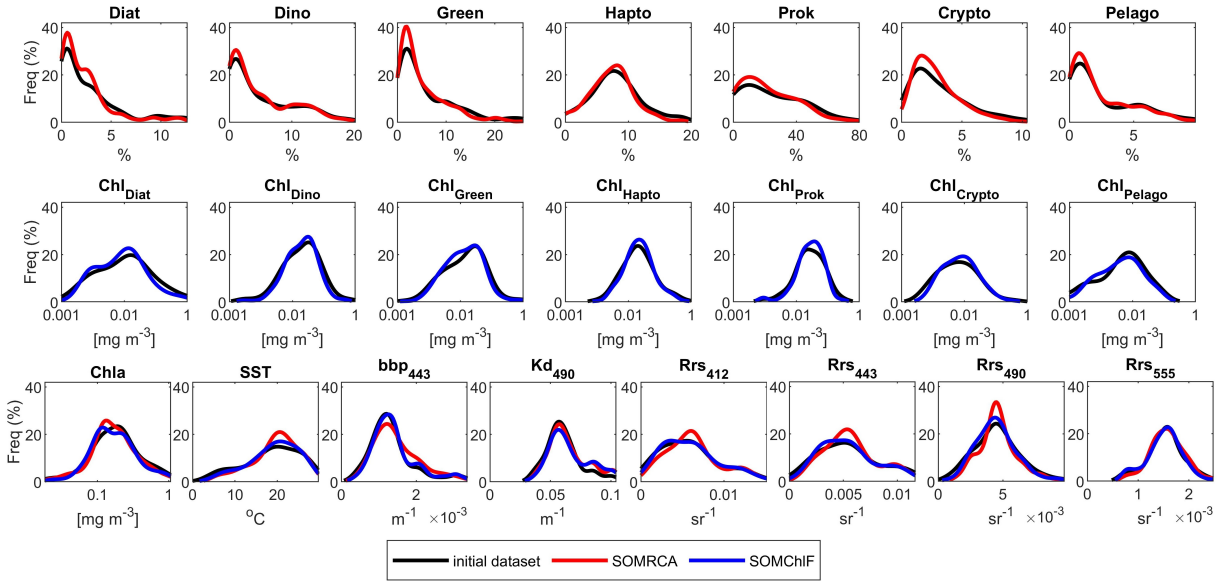

Figure S5. Evaluation of the preservation of the initial dataset's characteristics; distribution of the values within the initial dataset (D_{RCA} and D_{ChlF}) and the SOMRCA and SOMChlF neurons.

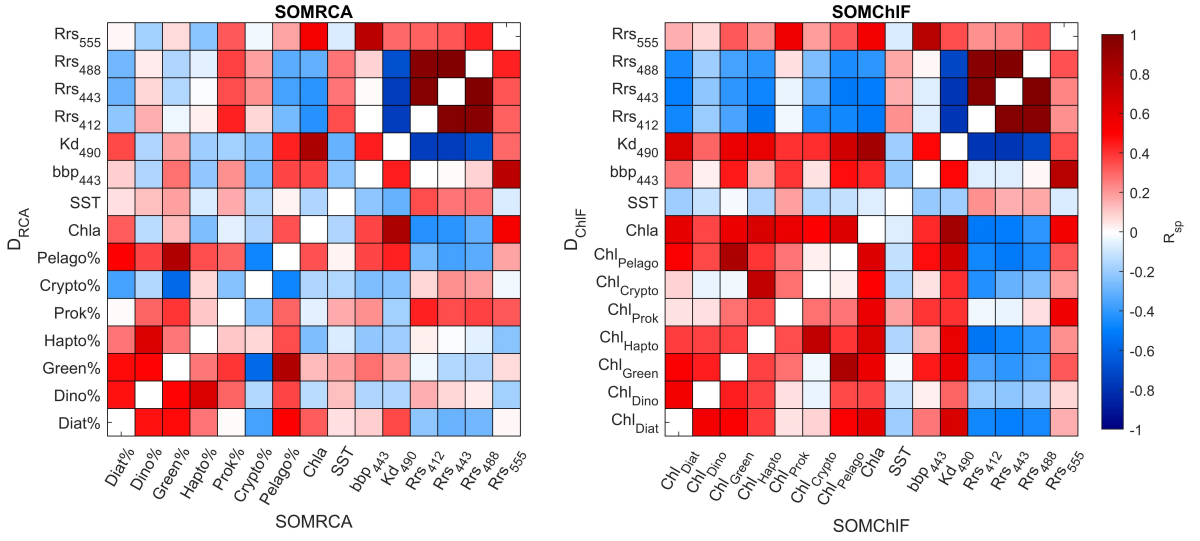

Figure S6. Evaluation of the preservation of the initial dataset's characteristics; Spearman correlation coefficient matrix comparing intra-correlations in the initial datasets DRCA and DChlF and the SOMRCA and SOMChlF neurons respectively.

Table S1. List of HPLC datasets and campaigns compiled in this study

Dataset/cruise	Source/Reference
Marine Ecosystem Data (MAREDAT)	Peloquin et al. (2013)
NASA bio-Optical Marine Algorithm Dataset (NOMAD)	Werdell and Bailey (2005)
SeaWiFS Bio-optical Archive and Storage System (SeaBASS)	seabass.gsfc.nasa.gov, Werdell et al. (2003)
Géochimie, Phytoplanton, et Couleur de l'Océan (GeP&CO)	https://www.obs-vlfr.fr/proof/php/y_graph_stat_op_0.php?v1=gepco , Dandonneau et al. (2004)
Palmer LTer long term research station	https://portal.lternet.edu/nis/home.jsp, Kozlowski et al. (2011)
RV Polarstern campaigns	Bracher (2015a) Bracher (2015b) Bracher (2015c)
AESOP-CSIRO database	http://aesop.csiro.au/, https://portal.aodn.org.au/, de Salas et al. (2011); Wright et al. (2010)

References

Bracher, A.: Phytoplankton pigment concentrations during POLARSTERN cruise ANT-XXIV/1, PANGAEA, https://doi.org/10.1594/PANGAEA.848583, in supplement to: Bracher, Astrid; Taylor, Marc H; Taylor, Bettina B; Dinter, Tilman; Röttgers, Rüdiger; Steinmetz, Francois (2015): Using empirical orthogonal functions derived from remote-sensing reflectance for the prediction of phytoplankton pigment concentrations. Ocean Science, 11(1), 139-158, https://doi.org/10.5194/os-11-139-2015, 2015a.

Bracher, A.: Phytoplankton pigment concentrations during POLARSTERN cruise ANT-XXIV/4, PANGAEA, https://doi.org/10.1594/PANGAEA.848584, in supplement to: Bracher, Astrid; Taylor, Marc H; Taylor, Bettina B; Dinter, Tilman; Röttgers, Rüdiger; Steinmetz, Francois (2015): Using empirical orthogonal functions derived from remote-sensing reflectance for the prediction of phytoplankton pigment concentrations. Ocean Science, 11(1), 139-158, https://doi.org/10.5194/os-11-139-2015, 2015b.

Bracher, A.: Phytoplankton pigment concentrations during POLARSTERN cruise ANT-XXVI/4, PANGAEA, https://doi.org/10.1594/PANGAEA.848585, in supplement to: Bracher, Astrid; Taylor, Marc H; Taylor, Bettina B; Dinter, Tilman; Röttgers, Rüdiger; Steinmetz, Francois (2015): Using empirical orthogonal functions derived from remote-sensing reflectance for the prediction of phytoplankton pigment concentrations. Ocean Science, 11(1), 139-158, https://doi.org/10.5194/os-11-139-2015, 2015c.

Dandonneau, Y., Deschamps, P.-Y., Nicolas, J.-M., Loisel, H., Blanchot, J., Montel, Y., Thieuleux, F., and Bécu, G.: Seasonal and interannual variability of ocean color and composition of phytoplankton communities in the North Atlantic, equatorial Pacific and South Pacific, Deep Sea Research Part II: Topical Studies in Oceanography, 51, 303–318, https://doi.org/10.1016/j.dsr2.2003.07.018, 2004.

de Salas, M. F., Eriksen, R., Davidson, A. T., and Wright, S. W.: Protistan communities in the Australian sector of the Sub-Antarctic Zone during SAZ-Sense, Deep Sea Research Part II: Topical Studies in Oceanography, 58, 2135–2149, 2011.

Kozlowski, W. A., Deutschman, D., Garibotti, I., Trees, C., and Vernet, M.: An evaluation of the application of CHEMTAX to Antarctic coastal pigment data, Deep Sea Research Part I: Oceanographic Research Papers, 58, 350–364, 2011.

Peloquin, J., Swan, C., Gruber, N., Vogt, M., Claustre, H., Ras, J., Uitz, J., Barlow, R., Behrenfeld, M., Bidigare, R., Dierssen, H., Ditullio, G., Fernandez, E., Gallienne, C., Gibb, S., Goericke, R., Harding, L., Head, E., Holligan, P., Hooker, S., Karl, D., Landry, M., Letelier, R., Llewellyn, C. A., Lomas, M., Lucas, M., Mannino, A., Marty, J.-c., Mitchell, B. G., Muller-Karger, F., Nelson, N., Prezelin, B., Repeta, D., Smith Jr, W. O., Smythe-Wright, D., Stumpf, R., Subramaniam, A., Suzuki, K., Trees, C., Vernet, M., Wasmund, N., and Wright, S.: The MAREDAT global database of high performance liquid chromatography marine pigment measurements, Earth Syst. Sci. Data, 5, 109–123, https://doi.org/10.5194/essd-5-109-2013, 2013.

Werdell, P. J. and Bailey, S. W.: An improved in-situ bio-optical data set for ocean color algorithm development and satellite data product validation, Remote Sensing of Environment, 98, 122–140, https://doi.org/10.1016/j.rse.2005.07.001, 2005.

Werdell, P. J., Bailey, S., Fargion, G., Pietras, C., Knobelspiesse, K., Feldman, G., and McClain, C.: Unique data repository facilitates ocean color satellite validation, Eos, Transactions American Geophysical Union, 84, 377–387, 2003.

Wright, S. W., van den Enden, R. L., Pearce, I., Davidson, A. T., Scott, F. J., and Westwood, K. J.: Phytoplankton community structure and stocks in the Southern Ocean (30–80 E) determined by CHEMTAX analysis of HPLC pigment signatures, Deep Sea Research Part II: Topical Studies in Oceanography, 57, 758–778, 2010.