Supplementary Material

Reconstructing global Chlorophyll-a variations using a non-linear statistical approach

E. Martinez1†,2*, T. Gorgues1, M. Lengaigne3, C. Fontana2, R. Sauzède4†,2, C. Menkes5, J. Uitz6, E. Di Lorenzo7, R. Fablet8

1) Support Vector Regression

To assess the sensitivity of the SVR to those parameters, 0.2 % of the 1998-2010 complete dataset (a low value justified by computational limitations) is randomly selected, then randomly divided into a training set (80%) and a cross-validation set (20%). This random data selection is then repeated over a large number of C and gamma combinations. The resulting Mean Absolute Error (MAE) between the Chl product used to train the SVR and the reconstructed Chl remains low across the large range of C and gamma combinations (see for instance for ChlPISCES vs. ChlSVR-PISCES in Figure Supp1A), highlighting the SVR stability and low sensitivity to the choice of C and gamma. The best C and gamma values in terms of local MAE minimum were selected (1 and 0.1, respectively when the SVR is trained on ChlPISCES, and 2 and 0.3 when trained on ChlOC-CCI).

With these C and gamma settings, the sensitivity of the SVR to an increasing portion of the total number of samples (from 0.2 % to 9 % randomly selected) is tested. The SVR is trained each time over 80% of the sub-dataset, while the remaining 20% is kept for validation. For instance, the training is performed over 7.2% of the total samples (i.e., 475 000 samples) in the 9% experiment. The MAE steeply decreases when increasing the size of the data sample from 0.2 to 4 %, then
stabilizes for values higher than 6% (Figure Supp1B). It suggests that the SVR skills don’t improve much when further increasing the sample size in the training process. This observation combined with computational limitations lead us to use and present here the 9% experiment.

**Figure Supp1 |** **A.** Mean absolute error (MAE, in mg.m⁻³) of the Chl_PISCES vs. Chl_Svr-PISCES validation dataset of the 0.2% experiment over a large range of C and gamma parameters. **B.** MAE according to the portion of samples used to train the SVR. **C.** Scatter plot of Chl_PISCES vs. Chl_Svr-PISCES normalized monthly anomalies for the 9% experiment. This validation is performed from 20% of the 9% total samples. The Chl_PISCES vs. Chl_Svr-PISCES and the 1:1 regression lines are plotted in red and black, respectively. The figure is color-coded according to the density of observations.
2) Empirical Orthogonal Function analysis on SST$_{\text{NEMO}}$

**Figure Supp2** | A) 1st EOF derived from interannual SST$_{\text{NEMO}}$ over 1998-2010 and their corresponding PCs in the B) Pacific, C) Indian and D) Atlantic Oceans (black lines). The MEI and AMO index are reported in red (right y-axis) on B-D. The MEI is correlated with the PCs in the Pacific and Indian Oceans ($r=0.9$ and $0.67$, respectively), and the AMO in the Atlantic ($r=0.75$).
Figure Supp3 | A) 1st EOF derived from interannual SST_{NEMO} over 1979-2010, and their corresponding PCs in the B) Pacific, C) Indian, D) Atlantic and E) Austral Oceans (black lines). The IPO and AMO index are reported in red (right y-axis) on B-E. The IPO is correlated with the PCs in the Pacific, Indian and Austral Oceans (r=0.95, 0.52 and 0.65, respectively), and the AMO in the Atlantic (r=0.92).
3) Empirical Orthogonal Function analysis on Chl$_{PISCES}$ and Chl$_{SVR-PISCES}$

**Figure Supp4** | 2nd EOFs of interannual A) Chl$_{PISCES}$ and B) Chl$_{SVR-PISCES}$, and (lower rows) their associated PCs. Chl$_{PISCES}$ and Chl$_{SVR-PISCES}$ PCs are indicated as the black and blue lines, respectively.
4) SVR procedure on the radiometric experiment

In section 4, the SVR is trained on radiometric observations over 1998-2010, still at global scale. The same procedure as in the synthetic case is followed for the C and gamma parameter’ selection (Figure Supp 5A). Likewise, 80% of 0.2 % to 9 % of the total number of samples is randomly selected to train this SVR. Doing so, the MAE steeply decreases along with the increasing number of data used to train the SVR (Figure Supp5B). A first validation is performed for the 9% experiment over the 1998-2010 training period on the dedicated subset (i.e., 20% of 9% of the total data set) showing a high determination coefficient of 0.87 and RMSE of 0.37 between ChlOC-CCI and ChlSVR-CCI (Figure Supp 5C).

![Figure Supp5](image.png)

Figure Supp5 | A. Mean absolute error (MAE, in mg.m$^{-3}$) of the ChlOC-CCI vs. ChlSVR-CCI validation dataset of the 0.2% experiment over a large range of C and gamma parameters. B. MAE according to the portion of samples used to train the SVR. C. Scatter plot of ChlOC-CCI vs. ChlSVR-CCI normalized monthly anomalies for the 9% experiment. This validation is performed from 20% of the 9% total samples. The ChlOC-CCI vs. ChlSVR-CCI and the 1:1 regression lines are plotted in red and black, respectively. The figure is color-coded according to the density of observations.
Figure Supp6 | A) 2nd EOF of interannual ChlSVR­CCI and their associated PCs for the B) Pacific (7.72% of the total variance), C) Indian (8.4% of the total variance), D) Atlantic (8% of the total variance) and E) Austral Oceans (8.3% of the total variance).