



**HAL**  
open science

## Reconstructing Global Chlorophyll-a Variations Using a Non-linear Statistical Approach

Elodie Martinez, Thomas Gorgues, Matthieu Lengaigne, Clément Fontana, Raphaëlle Sauzède, Christophe E. Menkès, Julia Uitz, Emanuele Di Lorenzo, Ronan Fablet

► **To cite this version:**

Elodie Martinez, Thomas Gorgues, Matthieu Lengaigne, Clément Fontana, Raphaëlle Sauzède, et al.. Reconstructing Global Chlorophyll-a Variations Using a Non-linear Statistical Approach. *Frontiers in Marine Science*, 2020, 7, 10.3389/fmars.2020.00464 . hal-02941302

**HAL Id: hal-02941302**

<https://imt-atlantique.hal.science/hal-02941302v1>

Submitted on 4 Jan 2021

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# 1           **Reconstructing global Chlorophyll-a variations using a non-linear statistical** 2   **approach**

3 **E. Martinez<sup>1†,2\*</sup>, T. Gorgues<sup>1</sup>, M. Lengaigne<sup>3</sup>, C. Fontana<sup>2</sup>, R. Sauzède<sup>4†,2</sup>, C. Menkes<sup>5</sup>, J.**  
4 **Uitz<sup>6</sup>, E. Di Lorenzo<sup>7</sup>, R. Fablet<sup>8</sup>**

5   <sup>1</sup> Laboratoire d'Océanographie Physique et Spatiale (LOPS), IUEM, IRD, Univ. Brest, Ifremer,  
6 CNRS, F-29280, Brest, France

7   <sup>2</sup> Ecosystèmes Insulaires Océaniens (EIO), IRD, Ifremer, UPF and ILM, Tahiti, French  
8 Polynesia

9   <sup>3</sup> Sorbonne Universités (UPMC, Univ Paris 06) -CNRS-IRD-MNHN, LOCEAN Laboratory,  
10 IPSL, Paris, France

11   <sup>4</sup> Sorbonne Universités, CNRS-INSU, Institut de la Mer de Villefranche,181 Chemin du Lazaret,  
12 06230 Villefranche-Sur-Mer, France

13   <sup>5</sup>UMR ENTROPIE (IRD—Université de La Réunion—CNRS), Laboratoire d'Excellence Labex-  
14 CORAIL, Institut de 7 Recherche pour le Développement, BP A5, 98848 Nouméa Cedex, New  
15 Caledonia

16   <sup>6</sup> CNRS and Sorbonne Université, Laboratoire d'Océanographie de Villefranche, LOV, 06230  
17 Villefranche-sur-mer, France

18   <sup>7</sup> Program in Ocean Science & Engineering, Georgia Institute of Technology, Atlanta, USA

19   <sup>8</sup> IMT-Atlantic, Lab-STICC, Brest, France

20

21   \* Corresponding author: elodie martinez ([elodie.martinez@ird.fr](mailto:elodie.martinez@ird.fr))

22   † current addresses

23

## 24           **Key Points:**

25       • A machine learning approach is applied to reconstruct the surface phytoplankton biomass at  
26       global scale over three decades.

27       • Chlorophyll variability derived from this statistical approach accurately reproduces satellite  
28       observations (possibly better than biogeochemical models).

- 29 • The sole use of surface predictors allows to accurately reproduce chlorophyll variability, in  
30 spite of its known sensitivity to three-dimensional processes.

31

32           **Abstract**

33           Monitoring the spatio-temporal variations of surface chlorophyll-a concentration (Chl, a  
34 proxy of phytoplankton biomass) greatly benefited from the availability of continuous and global  
35 ocean color satellite measurements from 1997 onwards. These two decades of satellite  
36 observations are however still too short to provide a comprehensive description of Chl variations  
37 at decadal to multi-decadal timescales. This paper investigates the ability of a machine learning  
38 approach (a non-linear statistical approach based on Support Vector Regression, hereafter SVR)  
39 to reconstruct global spatio-temporal Chl variations from selected surface oceanic and atmospheric  
40 physical parameters. With a limited training period (13 years), we first demonstrate that Chl  
41 variability issued from a 32-years global physical-biogeochemical simulation can generally be  
42 skillfully reproduced with an SVR using the model surface variables as input parameters. We then  
43 apply the SVR to reconstruct satellite Chl observations using the physical predictors from the  
44 above numerical model and show that the Chl reconstructed by this SVR more accurately  
45 reproduces some aspects of observed Chl variability and trends compared to the model simulation.  
46 The SVR is able to reproduce the main modes of interannual Chl variations depicted by satellite  
47 observations in most regions, including El Niño signature in the tropical Pacific and Indian Ocean.  
48 In stark contrast with the trends simulated by the biogeochemical model, it also accurately captures  
49 spatial patterns of Chl trends estimated by satellites data, with a Chl increase in most extratropical  
50 regions and a Chl decrease in the center of the subtropical gyres, although the amplitude of these  
51 trends are underestimated by half. Results from our SVR reconstruction over the entire period  
52 (1979-2010) also suggest that the Interdecadal Pacific Oscillation drives a significant part of  
53 decadal Chl variations in both the tropical Pacific and Indian Ocean. Overall, this study  
54 demonstrates that non-linear statistical reconstructions can be complementary tools to *in situ* and

55 satellite observations as well as conventional physical-biogeochemical numerical simulations to  
56 reconstruct and investigate Chl decadal variability.

## 57 1 INTRODUCTION

58           Phytoplankton—the microalgae that populate the upper lit layers of the ocean—fuels the  
59 oceanic food web and regulates oceanic and atmospheric carbon dioxide levels through  
60 photosynthetic carbon fixation. The launch of the "Coastal Zone Color Scanner" (CZCS) onboard  
61 the Nimbus-7 spacecraft in October 1978 (Hovis et al., 1980) provided the first synoptic view of  
62 near-surface chlorophyll-a concentration (Chl, a proxy of phytoplankton biomass). Although  
63 primarily focusing on coastal regions, CZCS also provided global pictures of Chl distribution and  
64 a new perspective on phytoplankton biomass seasonal variability (Campbell and Aarup, 1992;  
65 Longhurst et al., 1995; Yoo and Son, 1998; Banse and English, 2000).

66           After the failure of CZCS in 1986, ocean color observations were not available for more  
67 than a decade. The launch of the modern radiometric Sea-viewing Wide Field-of-View Sensor  
68 (SeaWiFS; McClain et al., 2004) in late 1997 followed later by other satellites allowed monitoring  
69 and understanding the spatio-temporal Chl variations at global scale over the past two decades.  
70 For instance, it revealed that El Niño events induce a Chl decrease in the central and eastern  
71 equatorial Pacific in response to reduced upwelled nutrients to the surface layers (e.g., Chavez et  
72 al., 1999; Wilson and Adamec, 2001; McClain et al., 2002; Radenac et al., 2012) but also a Chl  
73 signature outside the tropical Pacific through atmospheric teleconnections (Behrenfeld et al., 2001;  
74 Yoder and Kennelly, 2003; Dandonneau et al., 2004; Messie and Chavez, 2012). It also allowed  
75 identifying the Indian Ocean Dipole (IOD; Saji et al., 1999) as the main climate mode driving Chl  
76 interannual variations in the Indian Ocean (e.g. Murtugudde et al., 1999; Wiggert et al., 2009;  
77 Currie et al., 2013) and monitoring a Chl increase in the subpolar North Atlantic related to the  
78 positive phase of the North Atlantic Oscillation (NAO) (Martinez et al., 2016). Aside from the  
79 robust Chl decrease monitored in the mid-ocean gyres over the first decade of the XXI<sup>st</sup> century

80 (Polovina et al., 2008; Irwin and Oliver, 2009; Vantrepotte and Melin, 2009; Signorini and  
81 McClain, 2012), the reliability of the long-term trends derived from these satellite data are more  
82 questionable and led to conflicting results in the past literature (Behrenfeld et al., 2006;  
83 Vantrepotte and Mélin, 2011; Siegel et al., 2013; Gregg and Rousseaux, 2014). These  
84 discrepancies suggest that detection of robust global trend may require several decades of  
85 continuous observations (Beaulieu et al. 2013).

86         The production of longer, consistent ocean color time series can partly alleviate this issue.  
87 The combination of the global CZCS and SeaWiFS datasets provided an insight on the Chl  
88 response to natural decadal climate variations (Martinez et al., 2009; D’Ortenzio et al., 2012), such  
89 as the Pacific Decadal Oscillation (PDO; Mantua et al., 1997) and the Atlantic Multidecadal  
90 Oscillation (AMO; Enfield et al., 2001). However, blending these two archives or reconstructing  
91 them using compatible algorithms also led to contrasting results (Gregg and Conkright, 2002;  
92 Antoine et al., 2005).

93         The time span of the modern radiometric observations (~20 years), as well as the CZCS-  
94 SeaWiFS reprocessed time series, are still too short to investigate Chl decadal variations and  
95 longer-term trends. Longer, continuous and consistent records are required. *In situ* biogeochemical  
96 observatories can provide such long and continuous records, but their inhomogeneous spatial  
97 distribution and varying record length prevent a confident assessment of Chl long-term changes at  
98 the scale of a basin (Henson et al., 2016).

99         Coupled physical-biogeochemical ocean model simulations can provide additional,  
100 valuable information’s in areas with limited observational coverage. These models resolve  
101 reasonably well the seasonal to interannual biogeochemical variability (Dutkiewicz et al., 2001;  
102 Wiggert et al., 2006; Aumont et al., 2015). They can however diverge in capturing Chl variations

103 at a timescale of a decade (Henson et al., 2009a,b; Patara et al., 2011), in particular phytoplankton  
104 regime shifts (Henson et al., 2009b). Different biological models are often coupled to different  
105 physical models, which renders the attribution of the different modelled responses to their physical  
106 or biological components difficult. The decadal or longer variability of the simulated primary  
107 producers should then be interpreted cautiously.

108 In this context, statistical methods reconstructing past Chl variations may be useful  
109 alternatives to overcome limitations associated with both observations and numerical models.  
110 While statistical reconstructions are now commonly used to extend physical variables back in time  
111 (e.g. Smith et al., 2012; Huang et al., 2017; Nidheesh et al., 2017), reconstructions of surface Chl  
112 are still in their infancy. Phytoplankton distribution is strongly controlled by physical processes,  
113 such as mixing and uplifting, fueling nutrients in the upper-lit layer (i.e., bottom up processes).  
114 Thus, relevant physical variables may allow to reconstruct Chl past variations. To our knowledge,  
115 a single study allowed the derivation of spatio-temporal surface Chl variations over the 1958-2008  
116 period in the tropical Pacific (Schollaert et al., 2017). This reconstruction used a linear canonical  
117 correlation analysis on Sea Surface Temperature (SST) and Sea Surface Height (SSH) to improve  
118 the description of the Chl response to the diversity of observed El Niño events and decadal climate  
119 variations in the tropical Pacific.

120 The objective of the present study is to explore the potential of an alternative statistical  
121 technique to reconstruct Chl at global scale over a 32-year time-series (i.e., 1979-2010). The  
122 considered machine learning technique is based on a Support Vector Regression (SVR) which  
123 accounts for non-linearities between predictors and Chl. First, the SVR is trained over 1998-2010  
124 on a self-consistent dataset of physical and Chl variables, all extracted from a forced ocean model  
125 simulation that includes a biogeochemical component (i.e., the NEMO-PISCES model). Then,



126 modelled physical variables are used to reconstruct Chl over 1979-2010. The feasibility and  
127 robustness of the proposed reconstruction process is assessed through the comparison of modelled  
128 *vs.* reconstructed Chl. In a second step, this framework is applied to satellite ocean color  
129 observations.

## 130 **2 DATA AND METHODS**

### 131 **2.1 The NEMO-PISCES simulation**

132 In this study, we used the “Nucleus for European Modelling of the Ocean” (NEMO)  
133 modelling framework (Madec, 2008). The NEMO configuration used displays a coarse resolution  
134 with 31 vertical levels and a 2° horizontal grid with a refined 0.5° resolution in the equatorial band.  
135 The model includes a biogeochemical component, the Pelagic Interaction Scheme for Carbon and  
136 Ecosystem Studies (PISCES; Aumont et al., 2015). PISCES is a model of intermediate complexity  
137 designed for global ocean applications (Aumont and Bopp, 2006), which uses 24 prognostic  
138 variables and simulates biogeochemical cycles of oxygen, carbon and the main nutrients  
139 controlling phytoplankton growth (nitrate, ammonium, phosphate, silicic acid and iron). It  
140 simulates the lower trophic levels of marine ecosystems distinguishing four plankton functional  
141 types based on size: two phytoplankton groups (small=nanophytoplankton and large=diatoms) and  
142 two zooplankton groups (small=microzooplankton and large=mesozooplankton). Chl from  
143 PISCES (hereafter referred to as Chl<sub>PISCES</sub>) is defined as the sum of the simulated diatoms and  
144 nanophytoplankton Chl content.

145 The NEMO-PISCES simulation is forced with atmospheric fields from the interannual  
146 Drakkar Forcing Set 5 (DFS5.2, Dussin et al., 2014) for wind, air temperature and humidity,  
147 precipitation, shortwave and longwave radiations. It is initialized with the World Ocean Atlas 2005

148 (WOA05) climatology for temperature, salinity, phosphate, nitrate and silicate (Garcia et al.,  
149 2006), while iron initial state is similar to the model climatology employed by Aumont and Bopp  
150 (2006). The model simulation was spun up using 3 repetitions of the 30 years' DFS5.2 forcing set,  
151 and finally ran over 1979-2010.

152 Although successfully used in a variety of biogeochemical studies (e.g. Bopp et al., 2005;  
153 Gehlen et al., 2006; Lengaigne et al., 2007; Schneider et al., 2008; Steinacher et al., 2010;  
154 Tagliabue et al., 2010; Séférian et al., 2013; Aumont et al., 2015; Keerthi et al., 2017; Parvathi et  
155 al., 2017 and references therein), the ability of the PISCES model to reproduce satellite surface  
156 Chl is briefly illustrated in section 3.1.

## 157 **2.2 Chl derived from satellite radiometric observations**

158 Satellite surface Chl for Case I waters is provided by the Ocean Colour – Climate Change  
159 Initiative (OC-CCI, hereafter referred to as Ch<sub>OC-CCI</sub>) from the European Space Agency  
160 (<http://www.esaoceancolour-cci.org/>). This product combines multi-sensor, global, ocean-color  
161 products while attempting to reduce inter-sensor biases for climate research (Storm et al., 2013).  
162 OC-CCI extends the time series beyond that provided by single satellite sensors and perform better  
163 in terms of long-term consistency than other products from multi-mission initiatives (Belo Couto  
164 et al., 2016).

165 Only deep oceanic areas (depth>200 m) are considered to avoid coastal waters where  
166 specific non-case-1 waters products are required. The Chl Level-3 product is binned on a regular  
167 1° grid with a monthly resolution over January 1998-December 2010. This time period does not  
168 extend beyond 2010 to be consistent with the NEMO-PISCES simulation. Ch<sub>OC-CCI</sub> is used to

169 evaluate the PISCES model performances in Section 3.1, and to train the statistical method in  
 170 Section 4.

### 171 **2.3 Predictors and Chl variables**

172 The variability of phytoplankton biomass is driven in many regions of the world ocean and  
 173 at many timescales by physical processes (e.g., Wilson and Adamec, 2002; Wilson and Coles,  
 174 2005; Kahru et al., 2010; Feng et al., 2015; Messie and Chavez, 2015). Our statistical architecture  
 175 relates to 12 predictors and one biological variable (Chl). A sample thus refers to 13 variables. The  
 176 12 predictors (7 physical variables from NEMO-DFS5.2, 2 temporal and 3 spatial parameters) are  
 177 detailed in **Table 1**, including their influence on Chl variations and the references supporting this  
 178 influence.

179 **Table 1:** physical predictors, the drivers of which they are proxy and associated references.

| Proxy used as predictors                                | Relevance to Chl variations   | References  |
|---|---|---|
| <b>SST</b>  | -Vertical mixing and upwelling<br><br>-Impacts on phytoplankton metabolic rates                   | -Behrenfeld et al. (2006); Polovina et al. (2008); Martinez et al. (2009); Thomas et al. (2012)<br><br>-Lewandowska et al. (2014) |
| <b>Sea Level Anomaly</b>                                | Thermocline/pycnocline depths   | Wilson and Adamec (2001; 2002); Radenac et al. (2012)   |
| <b>Zonal and meridional wind components</b>             | Surface momentum flux forcing and vertical motions driven by Ekman pumping                        | Martinez et al. (2011); Thomas et al. (2012)  |
| <b>Zonal and meridional surface current components</b>  | Horizontal advective processes  | Messie and Chavez (2012); Radenac et al. (2013)   |
| <b>Short-Wave radiations</b>                            | Photosynthetically Active Radiation   | Sakamoto et al. (2011)  |
| <b>Month</b> (cos and sin)                              | Periodicity of the day of the year (day 1 is very similar to day 365 from a seasonal perspective) | Sauzède et al. (2015)   |
| <b>Longitude</b> (cos and sin)<br><b>Latitude</b> (sin) | Periodicity (longitude 0°= longitude 360°)  | Sauzède et al. (2015)   |

180

181 We purposely limited the predictors to surface variables because our objectives are 1) to  
182 reconstruct Chl from physical observations, which are mainly available through remotely sensed  
183 surface data (oceanic observations below the surface are indeed usually not accessible at large  
184 spatial-scales or interannual time-scales); 2) to build a statistical scheme that can complement more  
185 complex numerical models (here, NEMO-PISCES) which simulate complex three-dimensional  
186 processes and are costly to run.

187 First, the SVR is trained on physical predictors from NEMO and DFS5.2 vs. Chl<sub>PISCES</sub>. The  
188 reconstructed Chl time-series is referred to as Chl<sub>SVR-PISCES</sub>. Second, the SVR is trained using the  
189 same physical predictors but vs. satellite Chl observations (Chl<sub>OC-CCI</sub>). The reconstructed Chl time-  
190 series is referred to as Chl<sub>SVR-CCI</sub>.

## 191 **2.4 Climate indices**

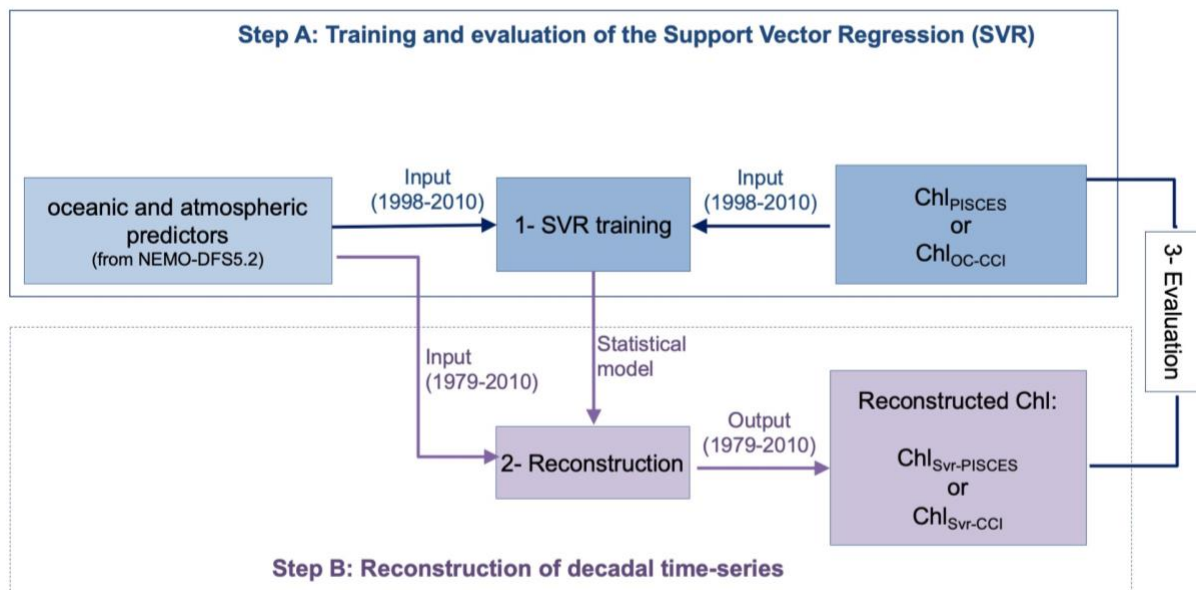
192 Climate indices are provided by the National Oceanic and Atmospheric Administration  
193 website ([www.esrl.noaa.gov/psd](http://www.esrl.noaa.gov/psd)): the AMO, the Multivariate El Niño Southern Oscillation  
194 (ENSO) Index (MEI) and the Interdecadal Pacific Oscillation (IPO).

## 195 **2.5 Support Vector Regression**

196 The statistical reconstruction technique is based on a SVR. This method belongs to kernel  
197 methods in Statistical Learning Theory and relates to the Support Vector Machine (SVM, Vapnik,  
198 1998). SVM is a kernel-based supervised learning method (Vapnik, 2000) developed for  
199 classification purpose in the early 1990s and then extended for regression by Vapnik (1995). The  
200 basic idea behind SVR is to map the variables into a new non-linear space using the kernel  
201 function, so that the regression task becomes linear in this space. The learning step estimates the  
202 parameters of the regression model according to a linear quadratic optimization problem, which

203 can be solved efficiently. SVR also uses a robust error norm based on the principle of structural  
 204 risk minimization, where both the error rates and the model complexity should be minimized  
 205 simultaneously. Because SVR can efficiently capture complex non-linear relationships, it has been  
 206 used in a variety of fields, and more specifically for [SEP] oceanographic, meteorological and climate  
 207 impact studies (Aguilar-Martinez and Hsieh, [SEP]2009; Descloux et al., 2012; Elbisy, 2015; Neetu  
 208 et al., 2020), as well as in marine bio-optics (Kim et al., 2014; Hu et al., 2018; Tang et al., 2019).

209 Predictors and Chl are normalized by removing their respective average and dividing them  
 210 by their standard deviations. Two SVR are trained over 1998-2010: one on Chl<sub>PISCES</sub> and one on  
 211 Chl<sub>OC-CCI</sub> (Step A in **Figure 1**). This time period has been chosen as 1998 is the first complete year  
 212 of the satellite Chl<sub>OC-CCI</sub> time-series, and 2010 is the last year available of the modelled Chl<sub>PISCES</sub>.  
 213 The two resulting SVR schemes are applied on the NEMO-DFS5.2 physical predictors over 1979-  
 214 2010. Finally, the annual means and standard deviations initially removed are applied to perform  
 215 the back transformation and reconstruct either Chl<sub>SVR-PISCES</sub> or Chl<sub>SVR-CCI</sub> (Step B in **Figure 1**).



216

217 **Figure 1** | Steps performed to train the SVR and reconstruct Chl time-series

218

219 Considering a Gaussian kernel, SVR only involves the selection of two hyperparameters: the  
220 penalty parameter  $C$  of the error term and the kernel coefficient  $\gamma$ , driving the reduction of  
221 the cost function.  $C$  and  $\gamma$  values are 1 and 0.1, respectively when the SVR is trained on  
222  $\text{Chl}_{\text{PISCES}}$ , and 2 and 0.3 when trained on  $\text{Chl}_{\text{OC-CCI}}$  (See details in the Supplementary material, and  
223 **Figure Supp 1A**). Sensitivity tests to an increasing portion of the sample total number (from 0.2  
224 % to 9 % of the full dataset) used in the training process are performed (See Supplementary  
225 materials and **Figure Supp 1B**). The mean absolute error stabilizes for a sample number higher  
226 than 6 %, suggesting that the SVR skills don't improve much afterward. This observation  
227 combined with computational limitations lead us to present the 9% experiment hereafter.

## 228 **2.6 Empirical Orthogonal Function analysis**

229 The SVR skills to reconstruct Chl interannual to decadal variations are investigated  
230 performing Empirical Orthogonal Function analysis on  $\text{Chl}_{\text{PISCES}}$ ,  $\text{Chl}_{\text{OC-CCI}}$ ,  $\text{Chl}_{\text{SVR-PISCES}}$  and  $\text{Chl}_{\text{SVR-CCI}}$ .  
231 First, Chl data are centered and reduced (i.e., the monthly climatology is removed and the  
232 induced anomalies are divided by their standard deviations) to avoid an overly dominant  
233 contribution of high values on the analysis (Emery and Thompson, 1997) over the periods of  
234 interest (i.e., 1998-2010 or 1979-2010). A 5-month running mean is applied to focus on the  
235 interannual/decadal signal. The analysis is separately performed for the Atlantic, Pacific and  
236 Indian Oceans north of 40°S until 60°N, and for the 40°S-60°S region hereafter referred to as the  
237 Austral Ocean. Indeed, the large area covered by the Pacific Ocean and its dominant modes in  
238 climate variability (i.e., ENSO/IPO), could regionally dampen other modes of variability. Basin-  
239 scale spatial maps are then gathered to a global one, refer to as EOF. The associated time-series  
240 refer to as the Principal Components (PCs).

### 241 **3 Synthetic reconstruction from a physical-biogeochemical ocean** 242 **model**

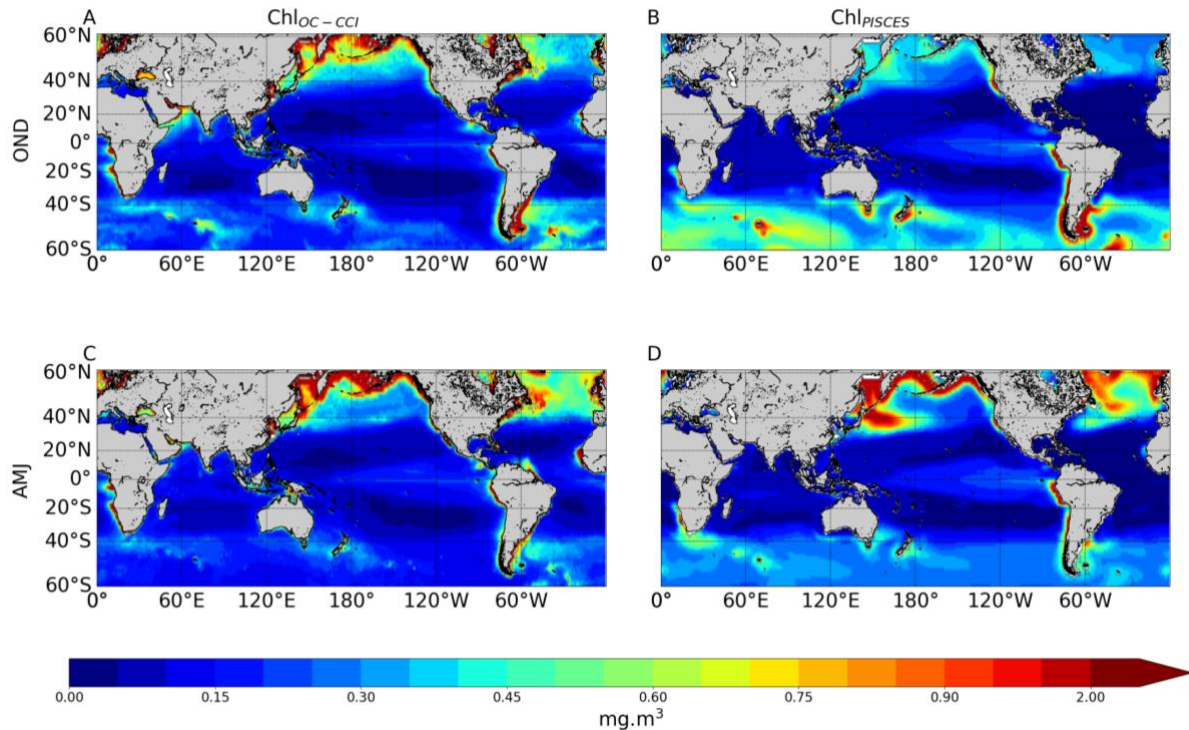
243 This section assesses the reliability and robustness of the SVR approach using a complete  
244 and coherent dataset extracted from a global simulation performed with a coupled physical-  
245 biogeochemical ocean model. The SVR is first trained over 1998-2010 on  $\text{Chl}_{\text{PISCES}}$ , and  $\text{Chl}_{\text{SVR-}}$   
246  $\text{PISCES}$  is reconstructed over 1979-2010.  $\text{Chl}_{\text{PISCES}}$  and  $\text{Chl}_{\text{SVR-PISCES}}$  are then compared over 32 years  
247 to evaluate the consistency of the proposed data-driven reconstruction scheme.

#### 248 **3.1 Evaluation of $\text{Chl}_{\text{PISCES}}$ at global scale**

249 The ability of the NEMO-PISCES model to reproduce the satellite Chl over 1998-2010 is  
250 briefly presented here. Boreal winter and summer climatology from  $\text{Chl}_{\text{PISCES}}$  compare reasonably  
251 well with those of  $\text{Chl}_{\text{OC-CCI}}$  (**Figures 2A vs. B, and C vs. D**). The model correctly represents the  
252 main spatial patterns with, for instance, higher Chl and a stronger seasonal cycle at high latitudes,  
253 despite an overestimated biomass in the Southern Ocean (Launois et al., 2015). The model also  
254 captures low Chl in the subtropical gyres, with some underestimation. This discrepancy may be  
255 explained by the lack of acclimation dynamics to oligotrophic conditions or by the assumption of  
256 constant stoichiometry either in phytoplankton or in organic matter in the model (Aumont et al.,  
257 2015; Ayata et al., 2013). The model underestimates Chl values in the equatorial Atlantic and  
258 Arabian Sea. In this latter region, mesoscale and submesoscale processes unresolved by the model  
259 have been shown to be of critical importance (Hood et al., 2003; Resplandy et al., 2011). Finally,  
260 the parameterization of nitrogen-fixing organisms not explicitly modelled in that PISCES version  
261 could explain the  $\text{Chl}_{\text{PISCES}}$  underestimation in the western Pacific in austral summer (Dutheil et  
262 al., 2018).

263 High Chl are accurately simulated in the eastern boundary upwelling systems. In two of  
264 the three main High Nutrient Low Chlorophyll (HNLC) regions, i.e. the equatorial Pacific and the  
265 eastern subarctic Pacific, the model successfully reproduces the moderate  $Ch_{loc-CCI}$ . However, the  
266 model overestimates  $Ch_{loc-CCI}$  east of Japan because of an incorrect representation of the Kuroshio  
267 current trajectory. This common bias in coarse resolution models (i.e., Gnanadesikan et al., 2002;  
268 Dutkiewicz et al., 2005; Aumont and Bopp, 2006) is potentially related to too deep mixed layer  
269 simulated in winter inducing very strong spring blooms (Aumont et al., 2015). In the Southern  
270 Ocean, the third and largest main HNLC region, the model overestimates  $Ch_{loc-CCI}$  values,  
271 especially during summer. However, the standard satellite algorithms that deduce Chl from  
272 reflectance tend to underestimate *in situ* observations by a factor of about 2 to 2.5, especially for  
273 intermediate concentrations (e.g., Dierssen and Smith, 2000; Kahru and Mitchell, 2010). It is to  
274 note that Chl in physical-biogeochemical coupled models is commonly overestimated in the  
275 Southern Ocean, and systematically underestimated in the oligotrophic gyres (Séférian et al.,  
276 2013).





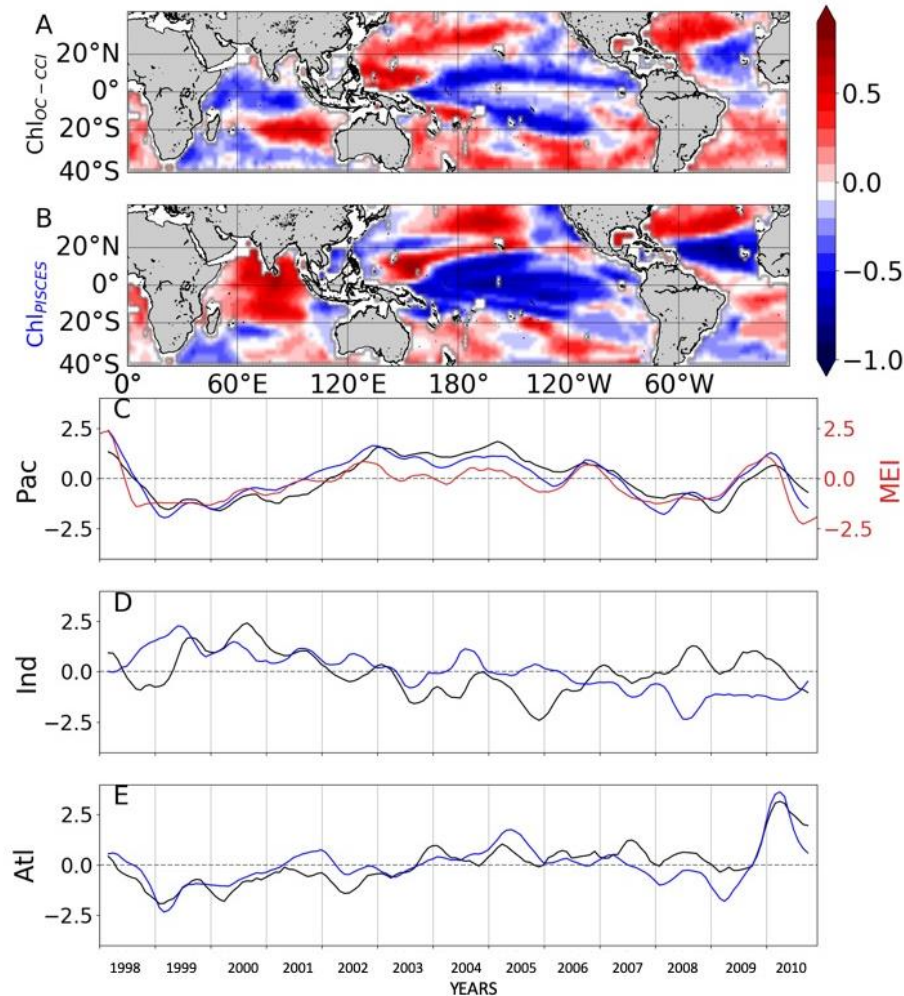
277

278 **Figure 2** | Surface seasonal mean of Chl ( $\text{mg}\cdot\text{m}^{-3}$ ) over 1998-2010 derived from satellite (left panels) and the PISCES  
 279 model (right panels), in October-November-December (A-B) and April-May-June (C-D).

280

281 The 1<sup>st</sup> mode of the EOF analysis performed on interannual Chl displays close percent of  
 282 total variance for Chl<sub>OC-CCI</sub> and Chl<sub>PISCES</sub> (16.6% vs. 21.1%, respectively). Their PCs in the Pacific  
 283 Ocean are well correlated with the MEI ( $r = 0.71$  and  $0.89$  with  $p=0.0015$  and  $p<0.001$ ,  
 284 respectively; **Figure 3C**). PCs show the greatest positive values in January 1998 during the peak  
 285 of the strong 1997/1998 El Niño event and the greatest negative values during the following La  
 286 Niña beginning of 1999. The associated EOFs display a Chl horseshoe pattern (**Figure 3A and**  
 287 **3B**), reminiscent of the ENSO pattern on SST from NEMO ( $\text{SST}_{\text{NEMO}}$ ; **Figure Supp 2**; Messie  
 288 and Chavez, 2012). While the tropical Pacific experiences a Chl decrease during El Niño events,  
 289 the North and South Pacific display a Chl increase, and inversely during La Niña. This typical  
 290 ENSO pattern is also related to remote Chl anomalies outside the Pacific induced by atmospheric  
 291 teleconnections, such as a Chl decrease in the tropical North Atlantic and in the South Indian Ocean

292 during El Niño. Although the Atlantic and Indian Ocean's PCs are not correlated with the MEI  
 293 (0.14 and 0.05, respectively), their EOFs are similar to those obtained from analysis performed at  
 294 global scale (*vs.* basin scale here) and which have been largely discussed in the past (e.g.,  
 295 Behrenfeld et al., 2001, 2006; Chavez et al., 2011; Yoder and Kennelly, 2003). Chl<sub>PISCES</sub>  
 296 reasonably well captures the first mode of Chl<sub>oc-cci</sub> interannual variability over 1998-2010 in the  
 297 Pacific and Atlantic Oceans, with 0.89 and 0.77 ( $p < 0.001$ ) correlations between their PCs,  
 298 respectively, but not in the Indian Ocean, where the PCs correlation is far weaker (0.13) and  
 299 insignificant (**Figure 3C-E**).



300

301 **Figure 3** | 1<sup>st</sup> mode of basin-scale EOFs of interannual **A**) Chl<sub>OC-CCI</sub> and **B**) Chl<sub>PISCES</sub>, and their corresponding PCs over  
302 1998-2010 in the **C**) Pacific, **D**) Indian and **E**) Atlantic Oceans. Chl<sub>OC-CCI</sub> and Chl<sub>PISCES</sub> PCs are represented by the black and blue  
303 lines, respectively. The MEI index is reported in red (right y-axis) on C.  
304

## 305 **3.2 Evaluation of the SVR method trained on synthetic data only**

### 306 3.2.1 Statistical performances

307 A first evaluation of the SVR applied on the synthetic dataset (i.e., both physical and  
308 biogeochemical model outputs) is provided for the dedicated subset (i.e., 20% of 9% of the total  
309 data set) over the 1998-2010 training time period. Chl<sub>PISCES</sub> and Chl<sub>SVR-PISCES</sub> datasets display a  
310 determination coefficient of 0.95 and a root mean square error (RMSE) of 0.22 (see **Figure Supp**  
311 **1C**), indicating at first glance a very good ability of the SVR to reconstruct Chl<sub>PISCES</sub>. The SVR  
312 reconstruction is very accurate when comparing the full modelled and reconstructed Chl for (i) the  
313 1998-2010 training time period, (ii) the 1979-1997 fully independent dataset and (iii) the 1979-  
314 2010 whole dataset, both at global and basin scales (**Table 2** and **Figure 4**). For each oceanic  
315 basin, determination coefficients between both datasets exceed 0.85, except in the Austral Ocean  
316 where they get down to 0.71. RMSE are lower than 0.14 and associated with a slope ranging from  
317 0.84 in the Austral to 0.97 in the Atlantic (**Figure 4**). In addition, the quality of the reconstructed  
318 Chl<sub>SVR-PISCES</sub> over the 1979-1997 independent time period is only marginally degraded compared  
319 to the 1998-2010 training period or the 1979-2010 full period.

320

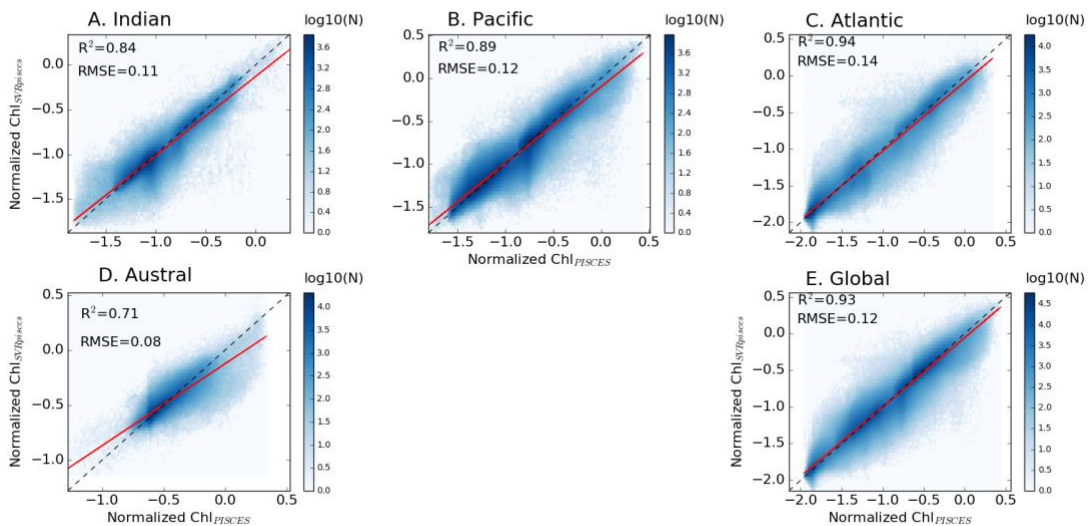
321

322

323 **Table 2** | Statistical performances between  $Chl_{PISCES}$  vs.  $Chl_{SVR-PISCES}$  normalized monthly anomalies for the global ocean  
 324 and the 4 oceanic basins over the 1998-2010, 1979-1997 and the whole 1979-2010 time period. The determination coefficient,  
 325 NRMSE and the slope of the  $Chl_{PISCES}$  vs.  $Chl_{SVR-PISCES}$  regression line are indicated.

|                         | 1998-2010 |      |       |                | 1979-1997 |      |       |                | 1979-2010 |      |       |                |
|-------------------------|-----------|------|-------|----------------|-----------|------|-------|----------------|-----------|------|-------|----------------|
|                         | $R^2$     | RMSE | slope | number of bins | $R^2$     | RMSE | slope | number of bins | $R^2$     | RMSE | slope | number of bins |
| <i>Global 60°S-60°N</i> | 0.96      | 0.1  | 0.95  | 4 487 457      | 0.94      | 0.12 | 0.95  | 6 562 272      | 0.94      | 0.11 | 0.95  | 11 049 729     |
| <i>Pacific</i>          | 0.92      | 0.1  | 0.93  | 1 945 173      | 0.88      | 0.12 | 0.9   | 2 843 501      | 0.9       | 0.11 | 0.91  | 4 788 674      |
| <i>Indian</i>           | 0.86      | 0.11 | 0.85  | 612 612        | 0.85      | 0.11 | 0.89  | 895 356        | 0.85      | 0.11 | 0.87  | 1 507 968      |
| <i>Austral</i>          | 0.81      | 0.06 | 0.8   | 1 026 876      | 0.71      | 0.08 | 0.75  | 1 506 260      | 0.76      | 0.07 | 0.77  | 2 533 136      |
| <i>Atlantic</i>         | 0.96      | 0.12 | 0.96  | 902 796        | 0.94      | 0.14 | 0.94  | 1 317 155      | 0.94      | 0.13 | 0.95  | 2 219 951      |

326



327

328 **Figure 4** | Scatter plots of  $Chl_{PISCES}$  vs.  $Chl_{SVR-PISCES}$  normalized monthly anomalies over 1979-1997, **A-D**) for each basin  
 329 and **E**) at global scale between 60°S and 60°N. The  $Chl_{PISCES}$  vs.  $Chl_{SVR-PISCES}$  and the 1:1 regression lines are plotted as the red and  
 330 black lines, respectively. The figure is color-coded according to the density of observations.

331

332 3.2.2 Evaluation of the reconstructed Chl spatio-temporal variability

333 The Normalized Root-Mean-Square-Error (NRMSE, i.e. RMSE normalized by the average  
 334 of Chl used to train the SVR) between  $Chl_{PISCES}$  and  $Chl_{SVR-PISCES}$  filtered with a 5-month running

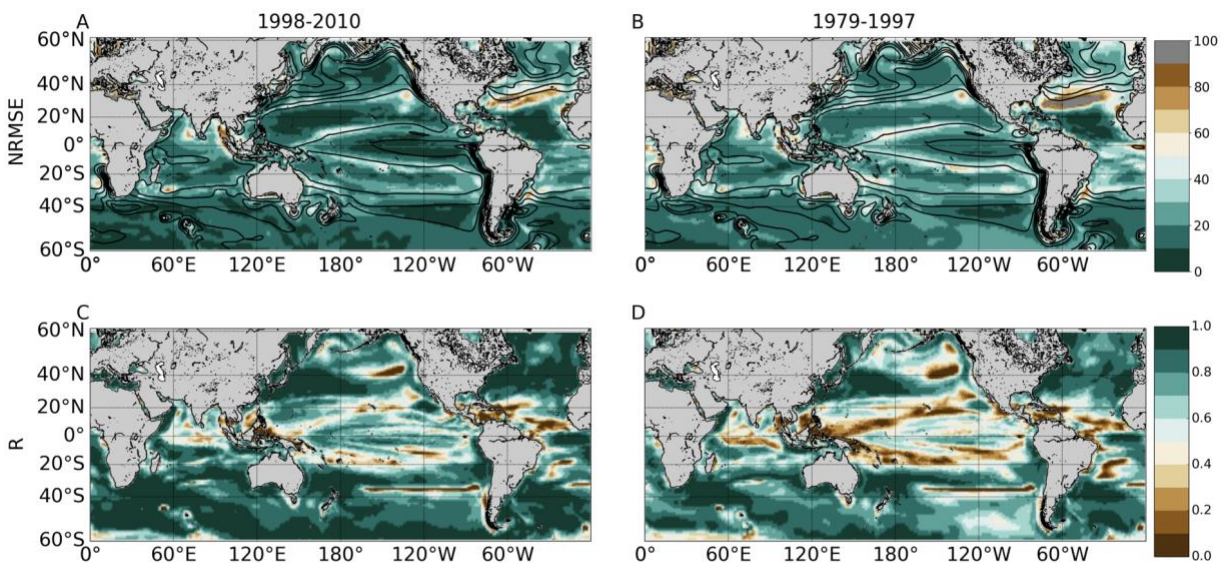
335 mean (to discard the high frequency signal) shows an error ranging between 10 and 20% over  
336 1998-2010 (**Figure 5A**). Their correlation exceeds 0.7 ( $p < 0.001$ ) over most of the global ocean  
337 (**Figure 5B**). At mid-latitudes they are generally larger than 0.8, and they range between 0.6 and  
338 0.9 in the equatorial Pacific. This accurate reconstruction demonstrates that a strong relationship  
339 exists between physical processes and Chl at global scale. However, the reconstructed Chl field  
340 can be regionally less accurate. For instance, the edges of the oligotrophic gyres (delimited by the  
341  $0.1 \text{ mg.m}^{-3}$  contour in **Figure 5A**) exhibit the highest NRMSE and lowest correlations. Large  
342 NRMSE are also evident in the Gulf Stream region while the western tropical Atlantic exhibits  
343 lower correlations than 0.5.

344 Those discrepancies could be due first to the zooplankton grazing pressure (top-down  
345 control) which is often overestimated in PISCES simulations. It results in an underestimated  
346 nanophytoplankton biomass in the oligotrophic gyres, emphasized along their edges (Laufkötter  
347 et al., 2015). Because the top-down control is not accounted for by the SVR, Chl variability  
348 induced by the overgrazing in these areas might not be captured. Second, in the equatorial Pacific  
349 Ocean, a minimum iron threshold value has been imposed ( $0.01 \text{ nmol.L}^{-1}$ ) in the biogeochemical  
350 model. Without that threshold Chl is too low on both sides of the equator, resulting in a strong  
351 accumulation of macronutrients and a spurious poleward migration of the subtropical gyre  
352 boundaries (Aumont et al., 2015). While the existence of such a threshold suggests that a minor  
353 but regionally important source of iron is missing in PISCES, it also suggests the inability of the  
354 SVR in reproducing ecosystem dynamics related to such artificial input of micro-nutrient. Finally,  
355 atmospheric input of iron through desert dust deposition is known to be stronger in the Atlantic  
356 than in the Pacific Ocean (Jickells et al., 2005). Such signal cannot be accounted for by the SVR



357 with the given predictors, which might (with meso – and sub-mesoscale activities) explain the  
 358 higher NRMSE in the north western Atlantic than in the north-western Pacific.

359 As expected, areas of high NRMSE and low correlations between  $Chl_{PISCES}$  and  $Chl_{SVR-PISCES}$   
 360  $PISCES$  identified over 1998-2010 (**Figure 5, left column**) extend and strengthen over 1979-1997  
 361 (**Figure 5, right column**). Indeed, the correlations significantly decrease in the tropical Pacific  
 362 while they slightly decrease in mid-latitudes between the two periods. Correlations remain high  
 363 and NRMSE low in the North-West Pacific, North and South-West Atlantic, and South Indian  
 364 Oceans as well as over a large part of the Southern Ocean providing confidence for analyses  
 365 extended beyond the training period of the SVR.



366

367 **Figure 5 | A-B)** NRMSE (in %) and **C-D)** correlation between  $Chl_{PISCES}$  vs.  $Chl_{SVR-PISCES}$  after applying a 5 month-running  
 368 mean on both time-series. These 2 diagnostics are calculated over 1998-2010 (left column) and 1979-1997 (right column). Contours  
 369 on the upper panels show their respective 1998-2010 Chl time average (every 0.1  $mg.m^{-3}$ ).

370

371 The analysis is now extended to the 1979-2010 time-period to investigate the skills of the  
 372 SVR in reproducing phytoplankton interannual/decadal cycles. The 1<sup>st</sup> EOFs of  $Chl_{PISCES}$  vs.

373 Chl<sub>SVR-PISCES</sub> have the same sign of variability over 72% of the global ocean (**Figures 6A-B**). Both  
374 EOFs are similar in the Pacific and Atlantic Oceans and their PCs are highly correlated over 1979-  
375 2010 (**Table 3; Figures 6C,6E**). In the Pacific, these EOFs strongly resemble the typical horseshoe  
376 pattern of IPO with SST anomalies of opposite polarities in the tropical and extra-tropical Pacific  
377 regions (**Figure Supp 3**). Correlations between Chl<sub>PISCES</sub> and Chl<sub>SVR-PISCES</sub> 1<sup>st</sup> PCs and the IPO  
378 index are high (0.94 and 0.95 with  $p < 0.001$ , respectively; blue and black vs. red lines in **Figure**  
379 **6C**). It highlights that the 1<sup>st</sup> mode of Chl variability in the Pacific is strongly driven by the IPO.  
380 In the Atlantic, both PCs are strongly correlated with the AMO (-0.8 for Chl<sub>SVR-PISCES</sub> and -0.85 for  
381 Chl<sub>PISCES</sub> with  $p < 0.001$ ; **Figure 6E**). The AMO shifts from a cold to a warm phase in the mid-  
382 1990's (**Figure Supp3**), and is associated with a decrease in Chl (**Figures 6A-B**).

383 The 1<sup>st</sup> two modes explain a similar percent variance for Chl<sub>PISCES</sub> and Chl<sub>SVR-PISCES</sub> in the  
384 four oceanic basins, with the exception of the 1<sup>st</sup> mode in the Atlantic Ocean (see **Table 3**). In this  
385 basin Chl<sub>SVR-PISCES</sub> percent variance is underestimated by a factor 2 compared to Chl<sub>PISCES</sub>, while  
386 their 1<sup>st</sup> EOFs and PCs are well correlated. One explanation might be that the AMO is the climate  
387 cycle with the longest period (80 years) when compared to the IPO. Thus, it might be the most  
388 difficult signal to reproduce as the SVR is trained over a relatively “short” 12 years’ time-period.

389 The agreement between Chl<sub>PISCES</sub> and Chl<sub>SVR-PISCES</sub> 1<sup>st</sup> EOFs is not as good in the Austral  
390 and Indian Oceans when compared to the Atlantic and Pacific Oceans (**Table 3** and **Figures 6A-**  
391 **B**). In the Indian Ocean, the Chl<sub>PISCES</sub> EOF exhibits a maximum positive variability along the  
392 western Arabian Sea, while it is located north-east of Madagascar for Chl<sub>SVR-PISCES</sub>. In the Austral  
393 Ocean, Chl<sub>PISCES</sub> and Chl<sub>SVR-PISCES</sub> EOFs roughly follow a zonal distribution.

394 A strong correspondence between SST and Chl has been previously reported over a large  
395 part of the global ocean (Behrenfeld et al., 2006; Martinez et al., 2009; Siegel et al., 2013),

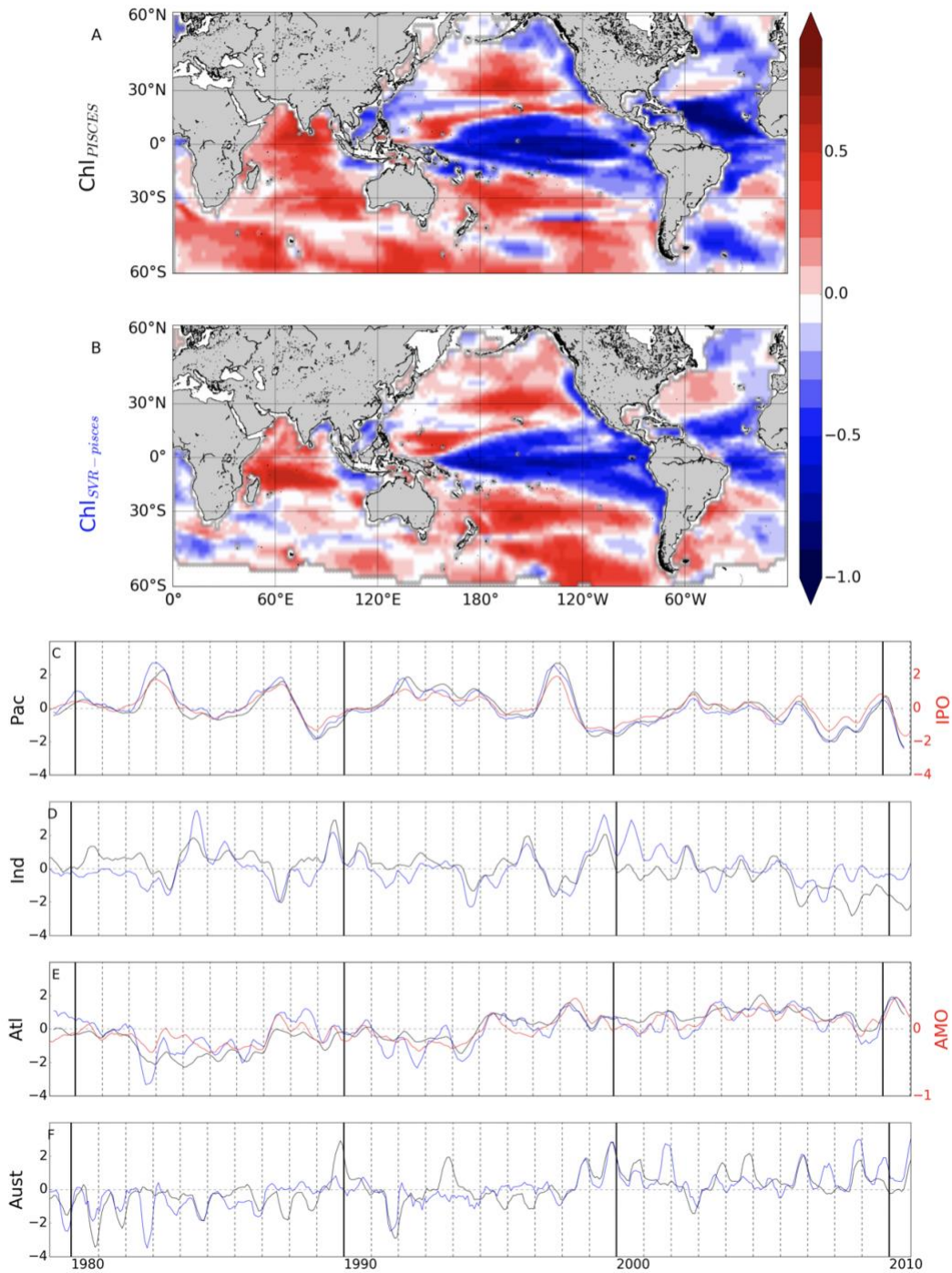
396 demonstrating the close interrelationship between ocean biology and climate variations.  
 397 Consequently, it is not surprising to observe strong correlations between  $\text{Chl}_{\text{PISCES}}$  or  $\text{Chl}_{\text{SVT-PISCES}}$   
 398 and climatic indexes mostly built on SST anomalies (**Figure Supp 3**).

399 **Table 3** | Percent variance explained by the first two modes of the Empirical Orthogonal Function analysis performed on  
 400  $\text{Chl}_{\text{PISCES}}$  and  $\text{Chl}_{\text{SVT-PISCES}}$  for each oceanic basin over 1979-2010. The correlation ( $r$ ) between the  $\text{Chl}_{\text{SVT-PISCES}}$  and  $\text{Chl}_{\text{PISCES}}$  PCs  
 401 is also reported with a significant level of  $p < 0.001$  (\*) and  $p < 0.002$  (\*\*).

|                 | <i>1<sup>st</sup> mode</i>   |                                  |       | <i>2<sup>nd</sup> mode</i>   |                                  |        |
|-----------------|------------------------------|----------------------------------|-------|------------------------------|----------------------------------|--------|
|                 | $\text{Chl}_{\text{PISCES}}$ | $\text{Chl}_{\text{SVT-PISCES}}$ | $r$   | $\text{Chl}_{\text{PISCES}}$ | $\text{Chl}_{\text{SVT-PISCES}}$ | $r$    |
| <i>Pacific</i>  | 19.7                         | 23.5                             | 0.95* | 7.8                          | 6.7                              | 0.6*   |
| <i>Indian</i>   | 13.1                         | 14.1                             | 0.58* | 10.5                         | 9.4                              | 0.78*  |
| <i>Austral</i>  | 13.8                         | 12.1                             | 0.62* | 11                           | 8.4                              | 0.47** |
| <i>Atlantic</i> | 23.2                         | 13.9                             | 0.81* | 9.6                          | 10.9                             | 0.73*  |

402





403

404

405

406

407

**Figure 6** | 1<sup>st</sup> mode of basin-scale EOFs of interannual **A)** Chl<sub>PISCES</sub> and **B)** Chl<sub>SVR-PISCES</sub>, and their corresponding PCs over 1979-2010 in the **C)** Pacific, **D)** Indian **E)** Atlantic and **F)** Austral Oceans (black and blue lines, respectively). Climate indices are reported in red (right y-axis).

408 The 2<sup>nd</sup> mode of variability of Chl<sub>PISCES</sub> is also well reproduced by the SVR. The percent  
409 variances are close (**Table 3**) as well as their spatio-temporal variability in the four oceanic basins  
410 (**Figure Supp 4**). The high correlations between the first two modes of Chl<sub>PISCES</sub> vs. Chl<sub>SVR-PISCES</sub>  
411 highlight the SVR ability to relatively well reproduce the Chl<sub>PISCES</sub> low-frequency variability.

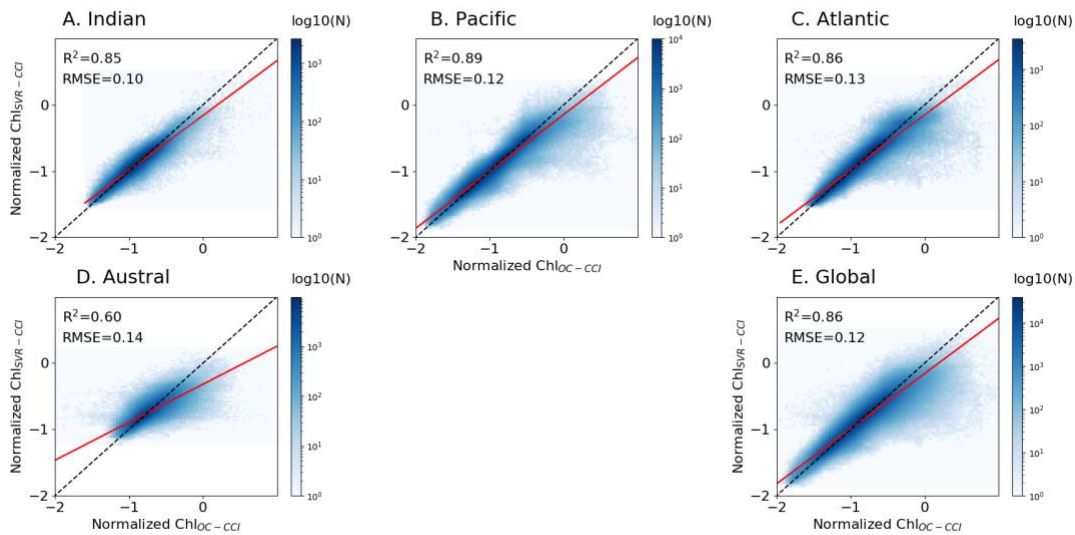
## 412 **4 Application to satellite radiometric observations**

### 413 **4.1 SVR statistical performances and sensitivity tests**

414 In this section, the SVR uses the same physical predictors from NEMO-DFS5.2 as in  
415 Section 3, but it is trained on satellite radiometric observations (e.g., Chl<sub>OC-CCI</sub>). The same  
416 procedure is followed (See **Figure Supp 5A-B**). A first validation is performed for 20% of 9% of  
417 the full data set and over the 1998-2010 training period showing a high determination coefficient  
418 of 0.87 and RMSE of 0.37 between Chl<sub>OC-CCI</sub> and Chl<sub>SVR-CCI</sub> (**Figure Supp 5C**).

419 As expected, the regression lines of Chl<sub>OC-CCI</sub> vs. Chl<sub>SVR-CCI</sub> for each oceanic basin and at  
420 global scale are farther away from the 1:1 line than for the synthetic study over the training period,  
421 but still remain close (higher slope than 0.8, except in the Austral Ocean; **Figure 7**). The SVR  
422 trained on NEMO-DFS5.2 predictors vs. satellite Chl is expected to be less efficient than the SVR  
423 trained on the coherent NEMO-DFS5.2-PISCES physical-biogeochemical dataset. Some of the  
424 biological interactions/processes (such as the diversity of the prey-predator relationships, the  
425 complexity of photoacclimation phenomena) are not yet optimally formulated by model equations  
426 inducing that Chl derived from numerical modelling is oversimplified compared to the complexity  
427 of the real ocean. Not to mention that satellite Chl may itself be partially affected by other  
428 components that are not Chl, such as colored dissolved organic matter (CDOM; Morel and Gentili,  
429 2009) and suspended particulate matter (SPM). Phytoplankton can also adjust their intracellular

430 Chl according to light and nutrient availability (e.g., Laws and Bannister, 1980; Behrenfeld et al.,  
 431 2015). The induced Chl changes are no longer ascribed to changes in biomass. All these signatures  
 432 on satellite Chl could explain  $\text{Chl}_{\text{SVR-CCI}}$  underestimation. Nevertheless, correlations between  
 433  $\text{Chl}_{\text{SVR-CCI}}$  and  $\text{Chl}_{\text{OC-CCI}}$  remain high over the training time period (0.92, 0.94 and 0.93 for the  
 434 Indian, Pacific and Atlantic Oceans, respectively, **Figure 7**).



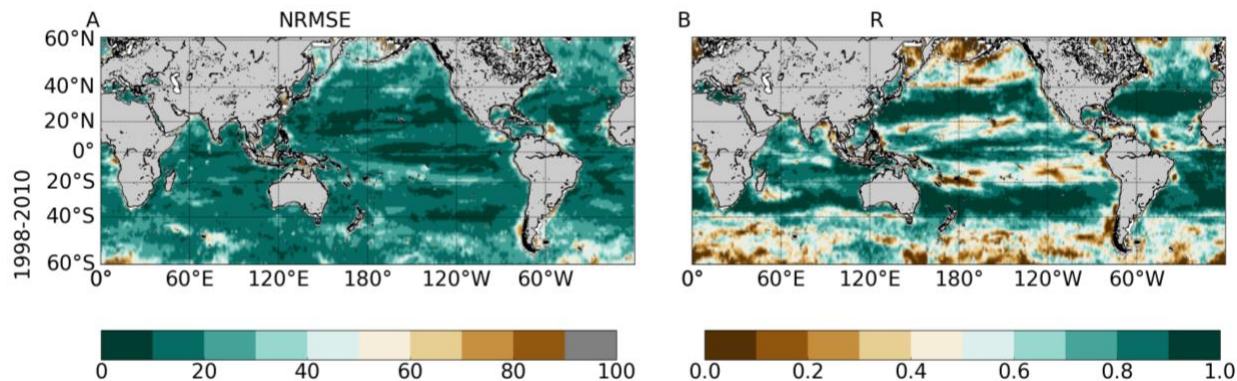
435

436 **Figure 7** | Same as **Figure 4** for  $\text{Chl}_{\text{OC-CCI}}$  vs.  $\text{Chl}_{\text{SVR-CCI}}$  normalized monthly anomalies over 1998-2010.

437

438 The NRMSE between  $\text{Chl}_{\text{OC-CCI}}$  vs.  $\text{Chl}_{\text{SVR-CCI}}$  is lower than 20% over most of the global  
 439 ocean (**Figure 8A**). Correlations higher than 0.9 ( $p < 0.001$ ) are evident over large subtropical areas  
 440 in the Atlantic, Indian and Pacific Oceans as well as in the Equatorial Pacific (**Figure 8B**).  
 441 Interestingly, the SVR generally does a better job at reconstructing the satellite Chl than the  
 442 modeled one (**Figure 5A-B** vs. **Figure 8**). NRMSE are higher at high latitudes and along the  
 443 oligotrophic area boundaries, although to a less extent than for  $\text{Chl}_{\text{PISCES}}$ . Because  $\text{Chl}_{\text{OC-CCI}}$  can  
 444 only be retrieved under clear sky conditions, gaps in satellite observations (especially during

445 wintertime) likely alters the SVR learning and could explain such a degradation of  $\text{Chl}_{\text{SVR-CCI}}$  as  
 446 moving towards high latitudes.



447  
 448 **Figure 8** | **A**) NRMSE (in %) and **B**) correlation between  $\text{Chl}_{\text{oc-CCI}}$  vs.  $\text{Chl}_{\text{svr-CCI}}$  over 1998-2010 after applying a 5  
 449 month-running mean on both time-series. Contours on the NRMSE show the 1998-2010  $\text{Chl}_{\text{oc-CCI}}$  time average (every 0.1  $\text{mg}\cdot\text{m}^{-3}$ ).  
 450 Correlations  $<0.73$  and  $0.6$  are significant with a  $p$ -value  $<0.001$  and  $0.01$ , respectively.

451

## 452 4.2 Reconstruction of satellite Chl interannual to decadal variability and trends

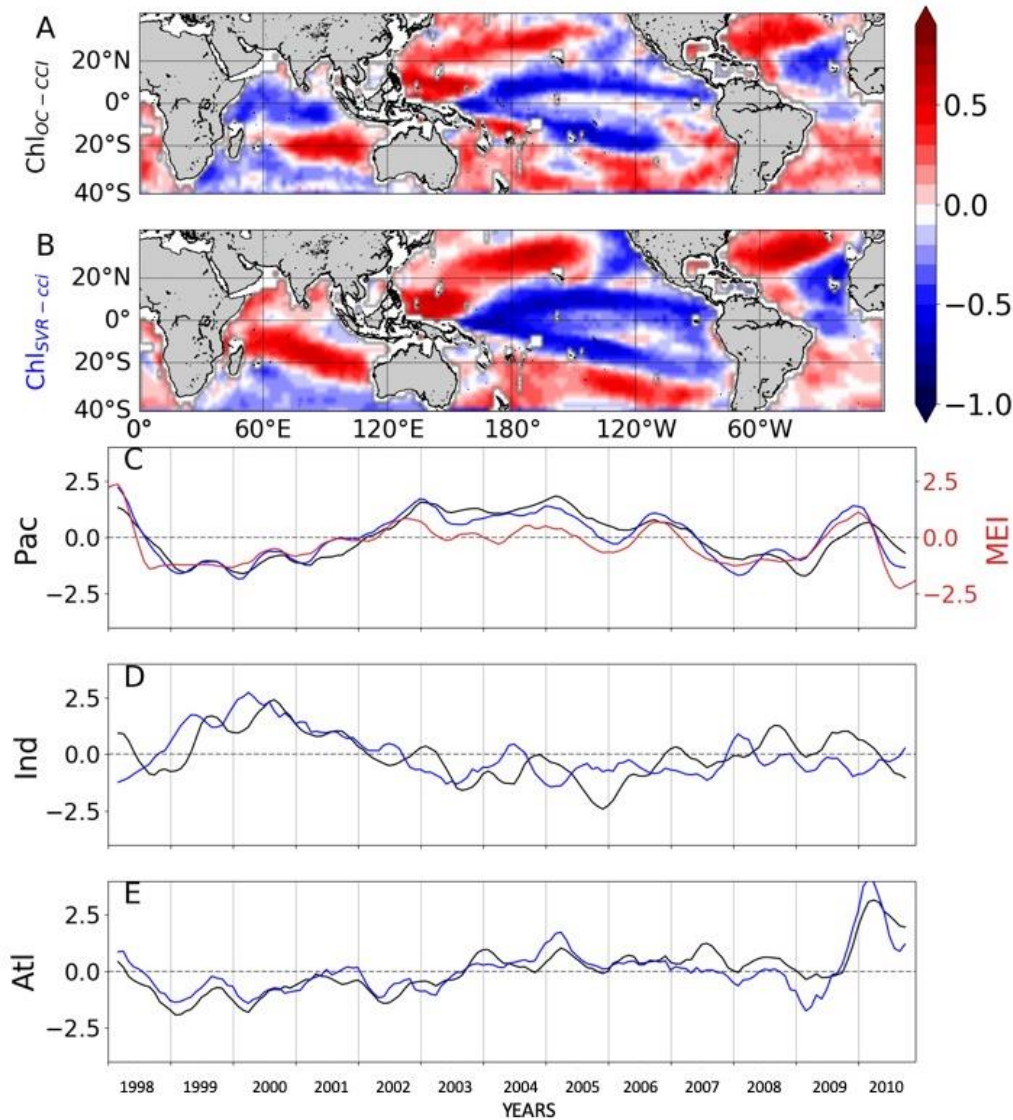
453 The SVR ability to replicate  $\text{Chl}_{\text{oc-CCI}}$  interannual variability is now investigated over  
 454 1998-2010 (**Figure 9**). In the Pacific Ocean,  $\text{Chl}_{\text{oc-CCI}}$  and  $\text{Chl}_{\text{svr-CCI}}$  1<sup>st</sup> EOFs are close (**Figure**  
 455 **9A** vs. **9B**), their PCs are highly correlated ( $r=0.89$ ,  $p<0.001$ ; **Figure 9C**), and their percent  
 456 variance are similar (**Table 4**). As presented in Section 3.1, this mode of Chl variability can be  
 457 attributed to ENSO, given their EOFs pattern as well as their PCs highly correlated with the MEI  
 458 ( $r_{\text{oc-CCI}/\text{MEI}}=0.71$  and  $r_{\text{svr-CCI}/\text{MEI}}=0.91$ , with  $p=0.0015$  and  $p<0.001$ , respectively). Interestingly,  
 459  $\text{Chl}_{\text{svr-CCI}}$  EOFs are closer to  $\text{Chl}_{\text{oc-CCI}}$  than  $\text{Chl}_{\text{PISCES}}$  in several areas such as in the north-western  
 460 Pacific, the south-western Atlantic and the Indian Ocean from Madagascar to the western coast of  
 461 Australia (**Figures 9A-B** vs. **Figure 3B**). Consistently, correlations between  $\text{Chl}_{\text{oc-CCI}}$  and  $\text{Chl}_{\text{svr-CCI}}$   
 462 PCs in the three basins and for the 1<sup>st</sup> two modes are higher than between  $\text{Chl}_{\text{oc-CCI}}$  and  
 463  $\text{Chl}_{\text{PISCES}}$  (**Table 4**).



464 **Table 4** | Percent variance explained by the first two modes of the Empirical Orthogonal Function analysis performed on  
 465  $Ch_{LOC-CCI}$ ,  $Ch_{SVR-CCI}$  and  $Ch_{PISCES}$  for each oceanic basin over 1998-2010. The correlation ( $r$ ) between  $Ch_{LOC-CCI}$  vs.  $Ch_{SVR-CCI}$  and  
 466 between  $Ch_{LOC-CCI}$  vs.  $Ch_{PISCES}$  is also reported with a significant level of  $P < 0.001$  (\*) and  $P < 0.02$  (\*\*).

|          | 1 <sup>st</sup> mode |                |               |                        |               | 2 <sup>nd</sup> mode |                |               |                        |               |
|----------|----------------------|----------------|---------------|------------------------|---------------|----------------------|----------------|---------------|------------------------|---------------|
|          | % of variance        |                |               | $r$ $Ch_{LOC-CCI}$ vs. |               | % of variance        |                |               | $r$ $Ch_{LOC-CCI}$ vs. |               |
|          | $Ch_{LOC-CCI}$       | $Ch_{SVR-CCI}$ | $Ch_{PISCES}$ | $Ch_{SVR-CCI}$         | $Ch_{PISCES}$ | $Ch_{LOC-CCI}$       | $Ch_{SVR-CCI}$ | $Ch_{PISCES}$ | $Ch_{SVR-CCI}$         | $Ch_{PISCES}$ |
| Pacific  | 16.6                 | 23.7           | 21.1          | 0.89*                  | 0.89*         | 10.7                 | 12.5           | 13.6          | 0.81*                  | 0.52**        |
| Indian   | 16.9                 | 16.6           | 17.3          | 0.57**                 | 0.13          | 11.8                 | 12.2           | 15.1          | 0.48                   | 0.36          |
| Atlantic | 14                   | 17.9           | 19.4          | 0.85*                  | 0.77*         | 10.7                 | 9.1            | 12.5          | 0.82*                  | 0.59**        |

467



468

469 **Figure 9:** 1<sup>st</sup> mode of basin-scale EOFs of interannual **A)**  $Ch_{loc-CCI}$  and **B)**  $Ch_{svr-CCI}$  and their associated PCs over 1998-  
470 2010 in the **C)** Pacific, **D)** Indian and **E)** Atlantic Oceans as the black and blue lines, respectively (left y-axis). The climate indices  
471 are reported in red on the right y-axis.

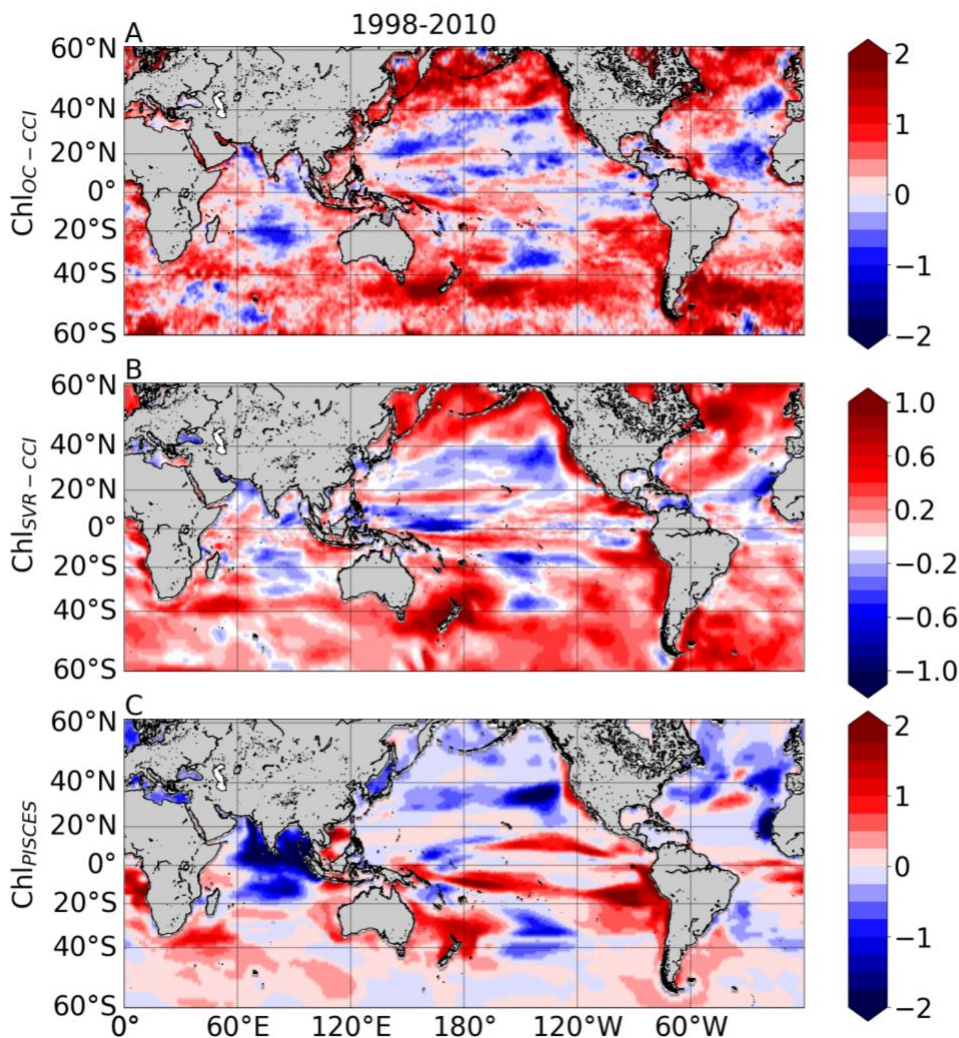
472

473  $Ch_{loc-CCI}$  linear trends over 1998-2010 exhibit large areas of increase or decrease (red and  
474 blue areas in **Figure 10A**, respectively). Productive regions at high latitudes and along the  
475 equatorial and upwelling areas generally exhibit positive  $Ch_{loc-CCI}$  trends, albeit many underlying  
476 regional nuances. Contrastingly, trends are generally negative in the center of the gyres. These  
477 regional trends are consistent with those extracted from the first 13 years of the SeaWiFS record  
478 and discussed by Siegel et al. (2013) (see their Figures 5b and 8b). The negative trends in the  
479 oligotrophic gyres were also reported by Signorini et al. (2015) who attributed this behavior to  
480 MLD shallowing trends. Surface water density variability induced by changes in temperature and  
481 salinity, combined with wind stirring, are effective drivers of vertical mixing, which in turn control  
482 the renewal of nutrients from the rich-deep layers toward the euphotic zone. Thus, shallower MLD  
483 would decrease nutrient uplift and phytoplankton growth in the oligotrophic areas.

484  $Ch_{svr-CCI}$  trends agree qualitatively well with those of  $Ch_{loc-CCI}$  at global scale (**Figures**  
485 **10B vs. 10A**, respectively). Indeed, decline of  $Ch_{svr-CCI}$  can be observed in the center of the gyres,  
486 while outside  $Ch_{svr-CCI}$  generally increases in a similar way to  $Ch_{loc-CCI}$ .  $Ch_{svr-CCI}$  accurately  
487 captures the largest  $Ch_{loc-CCI}$  increase observed in the Southern Ocean along the Antarctic  
488 Circumpolar Current. While Gregg and Casey (2004) reported a substantial negative bias in the  
489 SeaWiFS data for this region when compared to *in situ* observations, which could hamper the  
490 reliability of satellite trends discussed in this area (e.g. Siegel et al., 2013), the SVR remains able  
491 to reproduce the positive observed trend. Despite qualitative spatial agreements, it is noteworthy

492 that the SVR underestimates by half the magnitude of the satellite trend (see scales in **Figure 10A**  
 493 vs. **10B**).

494 Interestingly, trends in  $\text{Chl}_{\text{PISCES}}$  generally differs from  $\text{Chl}_{\text{OC-CCI}}$  (**Figure 10C**). This is  
 495 striking for the North Pacific and Atlantic high latitudes, but also in the equatorial Atlantic and  
 496 Arabian Sea with opposite trends when compared with  $\text{Chl}_{\text{OC-CCI}}$  and  $\text{Chl}_{\text{SVR-CCI}}$ , and in a more  
 497 mitigated manner in the Austral Ocean.



498  
 499 **Figure 10** | Linear trends (in  $\% \text{ year}^{-1}$ ) calculated over 1998-2010 from the monthly **A)**  $\ln(\text{Chl}_{\text{OC-CCI}})$ , **B)**  $\ln(\text{Chl}_{\text{SVR-CCI}})$ ,  
 500 **C)**  $\ln(\text{Chl}_{\text{PISCES}})$ . Note that the scale is divided by 2 for  $\ln(\text{Chl}_{\text{SVR-CCI}})$ .

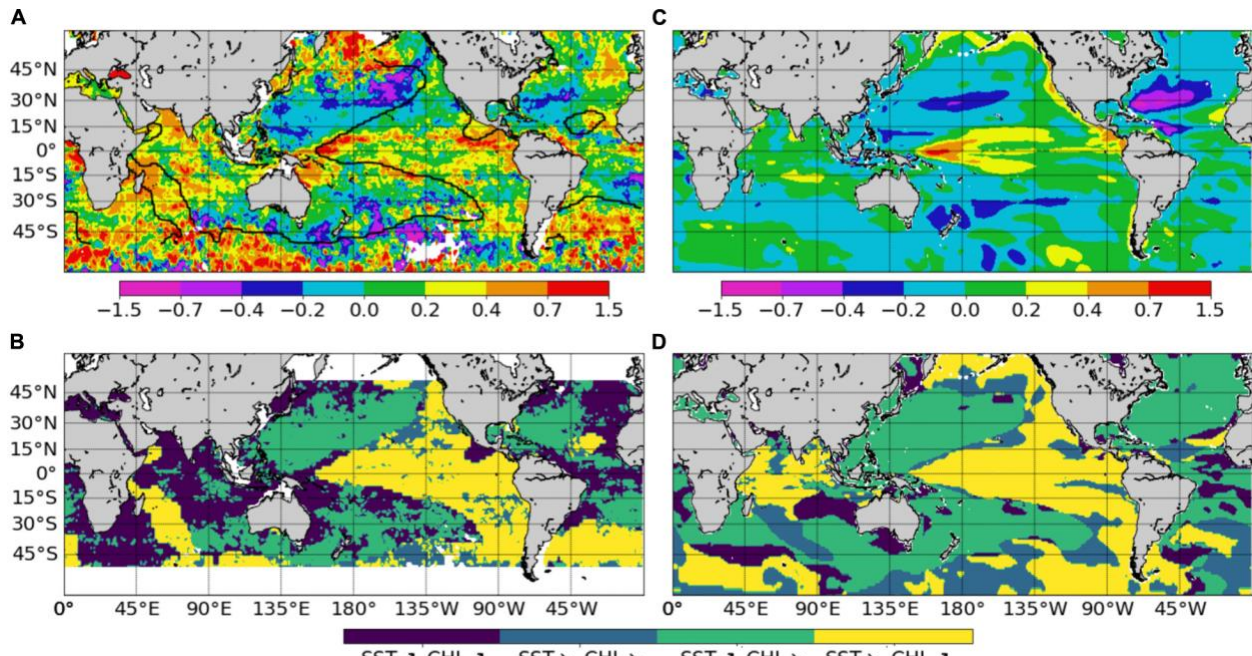
501

502  $\text{Chl}_{\text{SVR-CCI}}$  is also compared with the only historical consistent dataset built by Antoine et  
503 al. (2005) who reanalyzed ocean color time series from CZCS (1979–1983) and SeaWiFS (1998–  
504 2002). A 22% global mean increase of Chl between the two era was reported. It was mainly due  
505 to large increases in the intertropical areas and to a lesser extent in higher latitudes, while  
506 oligotrophic gyres displayed declining concentrations (**Figure 11A**). SST from the SODA  
507 reanalysis was used as a proxy of ocean stratification and opposite Chl and SST changes over 60%  
508 of the ocean between 50°S and 50°N was reported (light blue and yellow in **Figure 11B**, adapted  
509 from Martinez et al., 2009). This inverse relationship was used to hypothesized that multidecadal  
510 changes in global phytoplankton abundances were related to basin-scale oscillations of the ocean  
511 dynamics. Briefly, SST changes were related to a regime shift of the PDO (although the use of the  
512 basin-scale IPO would have been more appropriate) from a warm to a cold phase in the Pacific  
513 and Indian Oceans leading to an increase of Chl, and inversely in the Atlantic Ocean with a regime  
514 shift from a cold to a warm phase of the AMO leading to a Chl decrease.

515 Observed Chl changes over the last decades are accurately reproduced by  $\text{Chl}_{\text{SVR-CCI}}$ ,  
516 including a Chl increase in the equatorial Pacific and the southern tropical Indian Oceans, as well  
517 as a Chl decline in both the Atlantic and Pacific oligotrophic gyres (**Figure 11C**). However, the  
518 magnitude of the SVR reconstructed Chl is underestimated (note that the Chl ratio is multiplied by  
519 2 in **Figure 11C** to allow the comparison with **Figure 11A**). On average, the inverse relationship  
520 between  $\text{Chl}_{\text{SVR-CCI}}$  and  $\text{SST}_{\text{NEMO}}$  (**Figure Supp 4**) occurs over 69.4% of the global ocean between  
521 50°S and 50°N in a similar way to that reported by Martinez et al. (2009), especially in the Pacific  
522 Ocean (see **Figure 11D** vs. **Figure 11B**). In the Indian Ocean, although Chl mainly increases in  
523 both studies, it is here associated with a SST decrease. Interestingly, this inverse Chl-SST  
524 relationship in the Indian Ocean (yellow area in **Figure 11D**) was reported in Behrenfeld et al.



525 (2006) over the SeaWiFS era, suggesting that the SST dataset used in Martinez et al. (2009) may  
 526 have decadal discrepancies for this region.

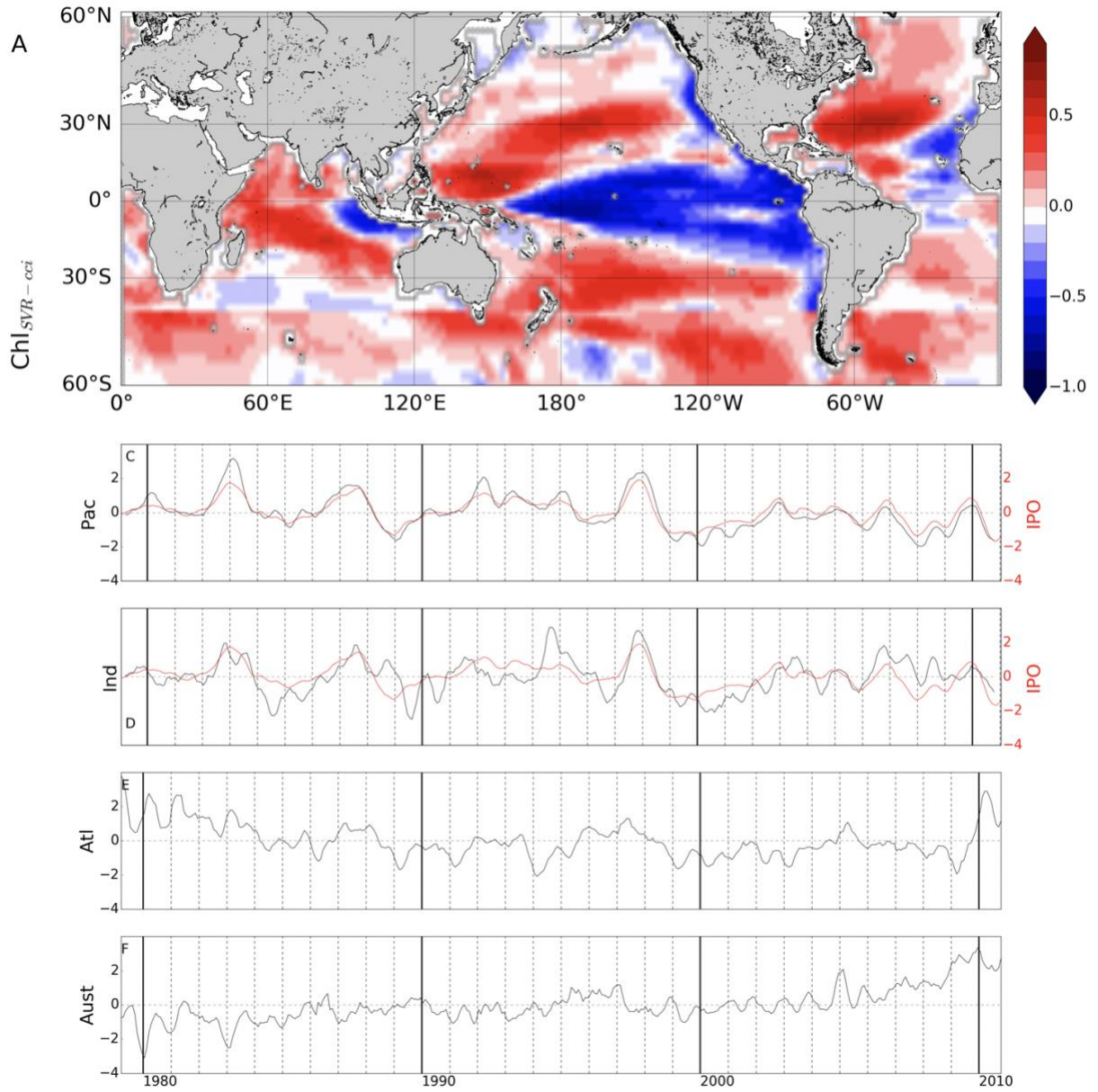


527  
 528 **Figure 11** | Chl change from the CZCS (1979–1983) to the SeaWiFS (1998–2002) era, expressed as the logarithm of the  
 529 ratio of the average values over the two time periods **A**) from satellite Chl adapted from Antoine et al. (2005), **C**) from  $Chl_{svr-CCI}$ .  
 530 Note that this ratio is multiplied by 2 to fit the same color bar as in **A**. Maps of areas with concomitant parallel or opposite changes  
 531 of Chl and SST **B**) from Chl satellite and SST from the SODA reanalysis adapted from Martinez et al. (2009) and **D**) from  $Chl_{svr-CCI}$   
 532 and  $SST_{NEMO}$ . The respective SST zero differences are shown on the maps as a thick black curve.  
 533

534 In their study, Martinez et al. (2009) analyzed two 5-year time periods apart each other by  
 535 15 years. They suggested that averaging observations separately over the two time-periods may  
 536 have dampen the effect of interannual variability and reveal the decadal one. Most of the changes  
 537 observed between the time periods covered by the two satellites are here confirmed based on the  
 538 reconstructed  $Chl_{svr-CCI}$ . However, the continuous 30-year time series of  $Chl_{svr-CCI}$  provides new  
 539 insights on the observed regime shifts (**Figure 12**). In the Pacific Ocean, the 1<sup>st</sup> EOF of  $Chl_{svr-CCI}$   
 540 (**Figure 12A**) is close to the Chl spatial patterns obtained from the CZCS to SeaWiFS era (**Figure**  
 541 **11C**) and the PC remains highly correlated with the IPO over 1979-2010 ( $r=0.94$  with  $p<0.001$ ,  
 542 **Figure 12B**). The Chl increase in the Indian Ocean, north-east of Madagascar toward the west

543 coast of Australia, between the 1980's and the 2000's also appears on the  $\text{Chl}_{\text{SVR-CCI}}$  EOF. These  
544 temporal changes might also be related to the IPO variability (correlation between the IPO index  
545 and  $\text{Chl}_{\text{SVR-CCI}}$  PC =0.6,  $p < 0.001$ ; **Figure 12C**).

546 In the Atlantic Ocean, CZCS-SeaWiFS Chl and  $\text{Chl}_{\text{SVR-CCI}}$  1<sup>st</sup> EOFs also share some  
547 similarities, including a decrease of Chl in the subtropical gyres and an increase in the  
548 equatorial/tropical regions. The associated PC (**Figure 12E**), exhibits a shift between 1979-1983  
549 and 1998-2002 consistently with Figure 2C of Martinez et al. (2009). In this latter study, this  
550 change is attributed to a regime shift of the AMO. However, the AMO index is not correlated with  
551 the 1<sup>st</sup>  $\text{Chl}_{\text{SVR-CCI}}$  PC ( $r=0.03$ ,  $p=0.43$ ) but rather with the 2<sup>nd</sup> mode ( $r = 0.43$  with  $p=0.003$ ), likely  
552 explaining the spatial discrepancies in **Figures 11A vs. 11C**. Although the detailed analysis of Chl  
553 decadal variability is beyond the scope of the present study, these initial findings underscore the  
554 importance of continuous time series at regional/global scales to combine spatial and temporal  
555 information's and properly investigate Chl long-term variability.



556

557

558

559

560

**Figure 12 | A)** 1<sup>st</sup> mode of basin-scale EOFs of interannual  $Chl_{SVR-CCI}$  over 1979-2010 and their corresponding PCs in the **B)** Pacific (23.2% of the total variance), **C)** Indian (15.2% of the total variance), **D)** Atlantic (13.5% of the total variance) and **E)** Austral Oceans (11.4% of the total variance). IPO is reported in red (right y-axis).

## 561 **5 Summary and conclusions**

562 In this paper, we assess the efficiency of a machine learning statistical approach based on  
563 support vector regression (SVR) to reconstruct surface Chl from oceanic and atmospheric  
564 variables. We first apply this strategy on self-consistent global dataset gathering physical  
565 predictors and Chl data simulated by a coupled physical-biogeochemical model simulation. Our  
566 results indicate that this nonlinear method accurately hindcasts interannual-to-decadal variations  
567 of the phytoplankton biomass simulated at global scale by the model, except at the boundaries of  
568 the subtropical gyres where the strong top-down control of zooplankton grazing in the numerical  
569 model is not accounted for by the SVR. Likewise, this statistical approach cannot yet reproduce  
570 Chl variability induced by nutrient inputs that are not directly related to our selected physical  
571 predictors, such as atmospheric iron deposit.

572 The SVR was then trained on satellite Chl observations. It accurately reproduces observed  
573 interannual Chl variations in most regions, including El Niño signature in the tropical Pacific and  
574 Indian Ocean as well as the main modes of Atlantic Chl variability. Despite an amplitude  
575 underestimation by half, it also accurately captures spatial patterns of Chl trends over the satellite  
576 period, with a Chl increase in most extratropical regions and a Chl decrease in the center of the  
577 subtropical gyres, as well as their changes between the CZCS and SeaWiFS era, while the model  
578 simulation generally fails to capture these long-term trends. Interestingly, while  $Chl_{PISCES}$   
579 magnitude is closer to  $Chl_{OC-CCI}$  than  $Chl_{SVR-CCI}$ , interannual variability and spatial trends of  
580  $Chl_{PISCES}$  are farther than  $Chl_{SVR-CCI}$  to  $Chl_{OC-CCI}$ . Equations representing the processes that govern  
581 the evolution of biogeochemical variables in a biogeochemical model are obviously less complex  
582 than the ones at play in the real ocean. We thus anticipated the modeled Chl to be easier to  
583 reconstruct than the satellite one. Additional complications were also expected through the

584 reconstruction of satellite Chl from the model oceanic and atmospheric predictors, which may be  
585 less realistic than physical parameters derived from satellite measurements. As a consequence, the  
586 SVR is indeed slightly less efficient at reproducing the major satellite Chl patterns compared to  
587 the model ones but is surprisingly more efficient at capturing observed Chl temporal variations.  
588 This results in a normalized root mean square error generally weaker when reconstructing satellite  
589 data compared to the model one, although the predictors used are identical.

590         Machine learning techniques are powerful tools to statistically model non-linear processes.  
591 They require a significant amount of data to be trained and are well-suited to analyze remote  
592 sensing data. While several attempts have been made over the last decade to retrieve oceanic Chl  
593 content (Kwiatkowska and Fargion, 2003; Zhan et al., 2003; Camps-Valls et al., 2009; Jouini et  
594 al., 2013; Blix and Eltoft, 2018), the present work is one of the first attempt to use such machine  
595 learning techniques to reconstruct past time series of phytoplankton biomass at global scale. To  
596 our knowledge only Schollaert et al. (2017) tried to reconstruct the Chl multi-decadal variability  
597 in the tropical Pacific using a canonical correlation analysis built only from SST and SSH. Our  
598 SVR approach leads to higher correlations between reconstructed and satellite Chl in the tropical  
599 Pacific, highlighting the strength of such non-linear machine-learning methods with multiple  
600 predictors. These results emphasize deep learning approaches as promising tools to reconstruct  
601 multidecadal Chl time series in the global ocean, based on the knowledge of physical conditions.  
602 The successful use of surface variables only in reproducing Chl variability which is influenced by  
603 3D-processes is here clearly noteworthy, and investigation of variable importance in the Chl  
604 reconstruction will deserve some future insights.

605         An obvious short-term perspective of the current study is to train a wider range of such  
606 statistical models with physical predictors derived from surface satellite observations but also

607 observations within the water column which could be derived from Argo data (i.e., mixed layer  
608 and thermocline depth). Including complementary variables such as satellite particulate  
609 backscattering coefficient (as a proxy of the Particulate Organic Carbon) in the  
610 training/reconstruction process should also be considered. It would allow to investigate the extent  
611 to which the Chl variability reflects changes in phytoplankton biomass *vs.* cellular changes in  
612 response to light (e.g., Siegel et al., 2005; Westberry et al., 2008; Behrenfeld et al., 2015). The use  
613 of longitude and latitude as predictors may limit the ability to capture long-term trends in the  
614 evolution of the biogeochemical province boundaries, such as the expansion of the oligotrophic  
615 areas (Irwin and Oliver, 2009; Polovina et al., 2008; Staten et al. 2018). Thus, exploring deep  
616 learning schemes which may not explicitly depend on longitude and latitude, especially  
617 convolutional representations (Lecun et al., 2015), are particularly appealing. Further efforts need  
618 also to be dedicated to alleviate the issue of the underestimation of the long-term Chl trends. For  
619 instance, it would be noteworthy to investigate secular trends such as the 30% Chl decrease  
620 reported at global scale over the last century by Boyce et al. (2010), which remains largely debated  
621 (Mackas, 2011; McQuatters-Gollop et al., 2011; Rykaczewski and Dunne, 2011).

622         Whatever the methodology used (i.e. numerical models, satellite or *in situ* observations),  
623 they all have both advantages and drawbacks. *In situ* observations are considered as ground truth  
624 (with some errors/uncertainties depending for instance on the field measurement protocols), but  
625 are heterogeneous in time and space. Satellite Chl data provide a spatio-temporal synoptic view  
626 but they have their own measurements issues and uncertainties (e.g., sensor radiometric and  
627 spectral properties, atmospheric corrections, water constituents and their optical properties) and  
628 are limited to 20 years in their record length. Biogeochemical models are useful tools to (i)  
629 interpolate or extrapolate in time and space biogeochemical tracers such as Chl and to (ii)

630 investigate complex three-dimensional processes responsible for their variations. However, those  
631 models suffer from biases and are farther from *in situ* data than satellite observations. They are  
632 also not straightforward to run and require large computing resources. Thus, machine learning  
633 statistical schemes could be seen as a complementary tool to the “interpolate/extrapolate” use of  
634 biogeochemical models in providing a long-term synoptic surface view built from observations  
635 (being aware of the uncertainties associated with the variables used in the training schemes). Such  
636 methods, applied on observations only, will then provide an independent tool that may either  
637 question or enforce conclusions drawn from model simulations. Comparison between both  
638 methods and observations will also help to improve biogeochemical models with acute  
639 quantification of model biases and identification of the most meaningful predictors that may point  
640 to missing processes in biogeochemical models. As a conclusion, machine learning is a versatile  
641 tool that, associated with biogeochemical models and observations, may greatly enhance our view  
642 of global biogeochemistry.

643

#### 644 **Acknowledgments, Samples, and Data**

645 This work was supported by CNES under contract n°160515/00 within the framework of  
646 the PhytoDev project. We thank the two reviewers who helped to improve this manuscript.

#### 647 **Authors contribution**

648 EM led the project, analyzed results and wrote the 1<sup>st</sup> draft of the manuscript. TG provided  
649 physical model outputs. TG, ML provided support in the analysis and the writing of the  
650 manuscript. CF processed the machine learning approach with support from R. Sauzede. R. Fablet  
651 provided feedbacks on the statistical approach. All authors contributed to the development of the  
652 manuscript and provided feedbacks throughout its many stages of preparation.

653

654

655           **References**

656   Aguilar-Martinez, S., and Hsieh, W. W. (2009). Forecasts of tropical pacific sea surface  
657   temperatures by neural networks and support vector regression. *Int. J. Oceanog.*, 13,  
658   doi:10.1155/2009/167239.

659   Antoine, D., Morel, A. , Gordon, H. R., Banzon, V. F., and Evans, R. H. (2005). Bridging ocean  
660   color observations of the 1980s and 2000s in search of long-term trends. *J. Geophys. Res.*  
661   *Oceans* 110, C06009.

662   Aumont, O., and Bopp, L. (2006). Globalizing results from ocean in situ iron fertilization  
663   studies. *Global Biogeochemical Cycles*, 20, GB2017, doi:10.1029/2005GB002591, 2006.

664   Aumont, O., Ethé, C., Tagliabue, A., Bopp, L., and Gehlen, M. (2015). PISCES-v2: an ocean  
665   biogeochemical model for carbon and ecosystem studies. *Geosci. Model Dev.*, 8, 2465–2513,  
666   2015, doi:10.5194/gmd-8-2465-2015.

667   Ayata, S. D., Lévy, M., Aumont, O., Sciandra, A., Sainte-Marie, J., Tagliabue, A., and Bernard,  
668   O. (2013). Phytoplankton growth formulation in marine ecosystem models: should we take  
669   into account photoacclimation and variable stoichiometry in oligotrophic areas? *J. Marine*  
670   *Syst.*, 125, 29–40, doi:10.1016/j.jmarsys.2012.12.010, 2013. 1433.

671   Banse, K., and English, D.C. (2000). Geographical differences in seasonality of CZCS-derived  
672   phytoplankton pigment in the Arabian Sea for 1978-1986. *Deep-Sea Res. II*, 47: 1623-1677.

673   Beaulieu, C., Henson, S. A., Sarmiento, J. L., Dunne, J. P., Doney, S. C., Rykaczewski, R. R., and  
674   Bopp, L. (2013). Factors challenging our ability to detect long-term trends in ocean  
675   chlorophyll. *Biogeosciences*, 10(4), 2711-2724.



- 676 Behrenfeld, M. J., Randerson, J. T., McClain, C. R., Feldman, G. C., Los, S. O., Tucker, C. J., et  
677 al. (2001). Biospheric primary production during an ENSO transition. *Science*, 291(5513),  
678 2594-2597.
- 679 Behrenfeld, M. J., O'Malley, R. T., Siegel, D. A., McClain, C. R., Sarmiento, J. L. et al. (2006).  
680 Climate driven trends in contemporary ocean productivity. *Nature*, 444(7120), 752–755.
- 681 Behrenfeld, M. J., O'Malley, R. T., Boss, E. S., Westberry, T. K., Graff, J. R., Halsey, K. H., et al.  
682 (2015). Revaluating ocean warming impacts on global phytoplankton. *Nat. Clim. Chang.* 6,  
683 323–330. doi:10.1038/nclimate2838.
- 684 Belo Couto, A., Brotas, V., Mélin, F., Groom, S., and Sathyendranath, S. (2016). Inter-comparison  
685 of OC-CCI chlorophyll-a estimates with precursor data sets. *International Journal of Remote*  
686 *Sensing*, 37(18), 4337-4355.
- 687 Blix, K., and Eltoft, T. (2018). Machine Learning Automatic Model Selection Algorithm for  
688 Oceanic Chlorophyll-a Content Retrieval. *Remote Sensing*, 10(5), 775.
- 689 Bopp, L., Aumont, O., Cadule, P., Alvain, S., and Gehlen, M. (2005). Response of diatoms  
690 distribution to global warming and potential implications: A global model study. *Geophys.*  
691 *Res. Lett.*, 32(19).
- 692 Boyce, D. G., Lewis, M. R., and Worm, B. (2010). Global phytoplankton decline over the past  
693 century. *Nature*, 466(7306), 591-596.
- 694 Campbell, J. W., and Aarup, T. (1992). New production in the North Atlantic derived from  
695 seasonal patterns of surface chlorophyll. *Deep-Sea Res. Part A.*, 39(10), 1669-1694.
- 696 Camps-Valls, G., Muñoz-Marí, J. L., Gómez-Chova, K. R., and Calpe-Maravilla, J. (2009).  
697 Biophysical Parameter Estimation With a Semisupervised Support Vector Machine. *IEEE*  
698 *Geoscience and Remote Sensing Letters*, 6 (2), 248 – 252.

- 699 Chavez, F. P., Strutton, P. G., Friederich, G. E., Feely, R. A., Feldman, G. C., Foley, D. G., and  
700 McPhaden, M. J. (1999). Biological and chemical response of the equatorial Pacific Ocean to  
701 the 1997-98 El Niño. *Science*, 286(5447), 2126-2131.
- 702 Chavez, F. P., Messié, M., and Pennington, J. T. (2011). Marine primary production in relation to  
703 climate variability and change. *Annu. Rev. Mar. Sci.*, 3, 227–260,  
704 doi:10.1146/annurev.marine.010908.163917.
- 705 Currie, J. C., Lengaigne, M., Vialard, J., Kaplan, D., Aumont, O., Naqvi, S. W. A., and Maury, O.  
706 (2013). Indian Ocean dipole and El Niño/southern oscillation impacts on regional chlorophyll  
707 anomalies in the Indian Ocean. *Biogeosciences*, 10(10), 6677-6698.
- 708 Dandonneau, Y., Deschamps, P. Y., Nicolas, J. M., Loisel, H., Blanchot, J., Montel, Y., et al.  
709 (2004). Seasonal and interannual variability of ocean color and composition of phytoplankton  
710 communities in the North Atlantic, equatorial Pacific and South Pacific. *Deep-Sea Res. Part II*,  
711 51(1-3), 303-318.
- 712 Descloux, E., Mangeas, M., Menkes, C. E., Lengaigne, M., Leroy, A., Tehei, T., et al. (2012).  
713 Climate-Based Models for Understanding and Forecasting Dengue Epidemics. *PLoS*  
714 *Neglected Tropical Diseases*, 6(2), e1470. doi:10.1371/journal.pntd.0001470.
- 715 Dierssen, H. M., and Smith, R. C. (2000). Bio-optical properties and remote sensing ocean color  
716 algorithms for Antarctic Peninsula waters. *J. Geophys. Res. Oceans*, 105, 26301–26312,  
717 doi:10.1029/1999JC000296, 2000. 1434, 1497.
- 718 D'Ortenzio, F., Antoine, D., Martinez, E., and Ribera d'Alcalà, M. (2012). Phenological changes  
719 of oceanic phytoplankton in the 1980s and 2000s as revealed by ocean-color remote-sensing  
720 observations. *Global Biogeochem. Cycles*, Vol. 26, GB4003.

- 721 Dussin, R., Barnier, B., Brodeau, L., and Molines, J. M. (2014). The making of Drakkar forcing  
722 set DFS5, DRAKKAR/MyOcean Rep.
- 723 Dutheil, C., Aumont, O., Gorguès, T., Lorrain, A., Bonnet, S., Rodier, M., et al. (2018). Modelling  
724 the processes driving *Trichodesmium* sp. spatial distribution and biogeochemical impact in  
725 the tropical Pacific Ocean. *Biogeosciences Discussions*, 1–34. [https://doi.org/10.5194/bg-](https://doi.org/10.5194/bg-2017-559)  
726 [2017-559](https://doi.org/10.5194/bg-2017-559).
- 727 Dutkiewicz, S., Follows, M., Marshall, J., and Gregg, W.W (2001). Interannual variability of  
728 phytoplankton abundances in the North Atlantic. *Deep Sea Res. II*, Volume 48, Pages 2323-  
729 2344.
- 730 Dutkiewicz, S., Follows, M. J., and Parekh, P. (2005). Interactions of the iron and phosphorus  
731 cycles: a three-dimensional model study. *Global Biogeochem. Cycles*, 19, GB1021,  
732 doi:10.1029/2004GB002342, 2005. 1412, 1430, 1433.
- 733 Elbisy, M. S. (2015). Sea Wave Parameters Prediction by Support Vector Machine Using a Genetic  
734 Algorithm. *J. Coastal Res.*, 31, 4, 892 –899.
- 735 Emery, W., and Thomson, R. (1997). *Data Analysis in Physical Oceanography* , 634 pp.,  
736 Pergamon, New York.
- 737 Enfield, D. B., Mestas Nunez, A. M., and Trimble, P. J. (2001). The Atlantic multidecadal  
738 oscillation and its relation to rainfall and river flows in the continental U.S. *Geophys. Res.*  
739 *Lett.*, 28, 2077–2080.
- 740 Feng, J., Durant, J. M., Stige, L. C., Hessen, D. O., Hjermann, D. Ø., Zhu, L., et al. (2015).  
741 Contrasting correlation patterns between environmental factors and chlorophyll levels in the  
742 global ocean. *Global Biogeochem. Cycles*, 29, doi:10.1002/2015GB005216.

- 743 Garcia, H. E., Locarnini, R. A., Boyer, T. P., and Antonov, J. I. (2006). World Ocean Atlas 2005,  
744 Volume 4: Nutrients (phosphate, nitrate, silicate). S. Levitus, Ed. NOAA Atlas NESDIS 64,  
745 U.S. Government Printing Office, Washington, D.C., 396 pp.
- 746 Gehlen, M., Bopp, L., Emprin, N., Aumont, O., Heinze, C., and Ragueneau, O. (2006).  
747 Reconciling surface ocean productivity, export fluxes and sediment composition in a global  
748 biogeochemical ocean model. *Biogeosciences, European Geosciences Union*, 2006, 3 (4),  
749 pp.521-537. hal-00330316
- 750 Gnanadesikan, A., Slater, R. J., Gruber, N., and Sarmiento, J. L. (2002). Oceanic vertical exchange  
751 and new production: a comparison between models and observations. *Deep-Sea Res. Pt. II*,  
752 49, 363–401, 2002. 1433.
- 753 Gregg, W. W., and Rousseaux, C. S. (2014). Decadal trends in global pelagic ocean chlorophyll:  
754 A new assessment integrating multiple satellites, in situ data, and models. *J. Geophys. Res.*  
755 *Oceans*, 119(9), 5921-5933.
- 756 Gregg, W. W., and Casey, N. W. (2004). Global and regional evaluation of the SeaWiFS  
757 chlorophyll data set. *Remote Sensing of Environment*, 93(4), 463-479.
- 758 Gregg, W. W., and Conkright, M. E. (2002). Decadal changes in global ocean chlorophyll.  
759 *Geophys. Res. Lett.*, 29(15), 20-1.
- 760 Henson, S. A., Dunne, J. P., and Sarmiento, J. L. (2009a). Decadal variability in North Atlantic  
761 phytoplankton blooms. *J. Geophys. Res. Oceans*, 114, C04013.
- 762 Henson, S. A., Raitso, D., Dunne, J. P., and McQuatters-Gollop, A. (2009b). Decadal variability  
763 in biogeochemical models: Comparison with a 50-year ocean colour dataset. *Geophys. Res.*  
764 *Lett.*, 36, L21061.

- 765 Henson, S. A., Beaulieu, C. and Lampitt, R. (2016). Observing climate change trends in ocean  
766 biogeochemistry: when and where. *Glob Change Biol*, 22: 1561-1571.  
767 doi:10.1111/gcb.13152.
- 768 Hood, R. R., Kohler, K. E., McCreary, J. P., and Smith, S. L. (2003). A four- dimensional  
769 validation of a coupled physical-biological model of the Arabian Sea. *Deep-Sea Res. Pt. II*,  
770 50, 2917–2945.
- 771 Hovis, W. A., Clark, D. K., Anderson, F., Austin, R. W., Wilson, W. H., Baker, E. T., et al. (1980).  
772 Nimbus-7 Coastal Zone Color Scanner: system description and initial  
773 imagery. *Science*, 210(4465), 60-63.
- 774 Hu, S., Liu, H., Zhao, W., Shi, T., Hu, Z., Li, Q., and Wu, G. (2018). Comparison of machine  
775 learning techniques in inferring phytoplankton size classes. *Remote Sensing*, 10(3), 191.
- 776 Huang, B., Thorne, P. W., Banzon, V. F., Boyer, T., Chepurin, G., Lawrimore, et al. (2017).  
777 Extended reconstructed sea surface temperature, version 5 (ERSSTv5): upgrades, validations,  
778 and intercomparisons. *Journal of Climate*, 30(20), 8179-8205.
- 779 Irwin, A. J., and Oliver, M. J. (2009). Are ocean deserts getting larger? *Geophys. Res. Lett.*, 36,  
780 L18609, doi:10.1029/2009GL039883.
- 781 Jickells, T. D., An, Z. S., Andersen, K. K., Baker, A. R., Berga-  
782 metti, G., Brooks, N., et al. (2005).  
783 Global iron connections between desert dust, ocean biogeochemistry, and climate. *Nature*,  
784 308, 67–71, doi:10.1126/science.1105959.
- 785 Jouini, M., Lévy, M., Crépon, M., and Thiria, S. (2013). Reconstruction of satellite chlorophyll  
786 images under heavy cloud coverage using a neural classification method. *Remote Sens.*  
*Environ.*, 131, 232-246.

- 787 Kahru, M. and Mitchell, B. G. (2010). Blending of ocean colour algorithms applied to the Southern  
788 Ocean. *Remote Sensing Letters*, 1, 119–124. 1434.
- 789 Kahru, M., Gille, S. T., Murtugudde, R., Strutton, P. G., Manzano-Sarabia, M., Wang, H., and  
790 Mitchell, B. G. (2010). Global correlations between winds and ocean chlorophyll. *J. Geophys.*  
791 *Res. Oceans* 115, C12040, doi:10.1029/2010JC006500.
- 792 Keerthi, M.G., Lengaigne, M., Levy, M., Vialard, J., and de Boyer Montegut, C. (2017). Physical  
793 control of interannual variations of the winter chlorophyll bloom in the northern Arabian Sea.  
794 *Biogeosciences*, 14(15), 3615, doi:10.5194/bg-14-3615-2017.
- 795 Kim, Y. H., Im, J., Ha, H. K., Choi, J. K., and Ha, S. (2014). Machine learning approaches to  
796 coastal water quality monitoring using GOCI satellite data. *GIScience & Remote*  
797 *Sensing*, 51(2), 158-174.
- 798 Kwiatkowska, E.J., and Fargion, G.S. (2003). Application of Machine-Learning Techniques  
799 Toward the Creation of a Consistent and Calibrated Global Chlorophyll Concentration  
800 Baseline Dataset Using Remotely Sensed Ocean Color Data. *IEEE Transactions on*  
801 *Geoscience and Remote Sensing*, 41 (12), 2844 – 2860.
- 802 Laufkötter, C., Vogt, M., Gruber, N., Aita-Noguchi, M., Aumont, O., Bopp, L., et al. (2015).  
803 Drivers and uncertainties of future global marine primary production in marine ecosystem  
804 models. *Biogeosciences*, 12, 6955-6984, <https://doi.org/10.5194/bg-12-6955-2015>.
- 805 Launois, T., Belviso, S., Bopp, L., Fichot, C. G., and Peylin, P. (2015). A new model for the global  
806 biogeochemical cycle of carbonyl sulfide–Part 1: Assessment of direct marine emissions with  
807 an oceanic general circulation and biogeochemistry model. *Atmospheric Chemistry and*  
808 *Physics*, 15(5), 2295-2312.

- 809 Laws, E. A., and Bannister, T. T. (1980). Nutrient-and light-limited growth of *Thalassiosira*  
810 *fluviatilis* in continuous culture, with implications for phytoplankton growth in the ocean.  
811 *Limnology and Oceanography*, 25(3), 457-473.
- 812 LeCun, Y., Bengio, Y. and Hinton, G. (2015). Deep learning. *Nature* 521, 436–444,  
813 <https://doi.org/10.1038/nature14539>.
- 814 Lengaigne, M., Menkes, C., Aumont, O., Gorgues, T., Bopp, L., André, J.-M., and Madec, G.  
815 (2007). Influence of the oceanic biology on the tropical Pacific climate in a Coupled General  
816 Circulation Model. *Climate Dynamics*, 28: 503-516, doi:10.1007/s00382-006-0200-2.
- 817 Lewandowska, A. M., Hillebrand, H., Lengfellner, K., and Sommer, U. (2014). Temperature  
818 effects on phytoplankton diversity—The zooplankton link. *J. Sea Res.*, 85, 359-364.
- 819 Longhurst, A., Sathyendranath, S., Platt, T., and Caverhill, C. (1995). An estimate of global  
820 primary production in the ocean from satellite radiometer data. *J. Plankton Res.*, 17(6), 1245-  
821 1271.
- 822 Mackas, D. L. (2011). Does blending of chlorophyll data bias temporal trend? *Nature*, 472(7342),  
823 E4-E5.
- 824 Madec, G. (2008). NEMO reference manual, ocean dynamics component: NEMO-OPA.  
825 Preliminary version. Note du Pole de modélisation, Institut Pierre-Simon Laplace (IPSL),  
826 France, (27), 1288-161.
- 827 Mantua, N.J., Hare, S.R., Zhang, Y., Wallace, J.M., and Francis, R.C. (1997). A Pacific  
828 Interdecadal Climate Oscillation with Impacts on Salmon Production. *B. Amer. Meteo. Soc.*,  
829 78, 1069.
- 830 Martinez, E., Antoine, D., D'Ortenzio, F., and Gentili, B. (2009). Climate-driven basin-scale  
831 decadal oscillations of oceanic phytoplankton. *Science*, 36, 1253-1256.

- 832 Martinez, E., Antoine, D., D’Ortenzio, F., and de Boyer Montégut, C. (2011). Phytoplankton  
833 spring and fall blooms in the North Atlantic in the 1980s and 2000s. *J. Geophys. Res. Oceans*,  
834 116, C11029.
- 835 Martinez, E., Raitzos, D., and Antoine, D. (2016). Warmer, deeper and greener mixed layers in the  
836 north Atlantic subpolar gyre over the last 50 years. *Global Change Biology*, 22, 604–612, doi:  
837 10.1111/gcb.13100.
- 838 McClain, C. R., Christian, J. R., Signorini, S. R., Lewis, M. R., Asanuma, I., Turk, D., and Dupouy-  
839 Douchement, C. (2002). Satellite ocean-color observations of the tropical Pacific  
840 Ocean. *Deep-Sea Res. Pt. II: Topical Studies in Oceanography*, 49(13-14), 2533-2560.
- 841 McClain, C. R., Feldman, G., and Hooker, S. (2004). An overview of the SeaWiFS project and  
842 strategies for producing a climate research quality global ocean bio-optical time series, *Deep*  
843 *Sea Res. Part II*, 51, 5–42, doi:10.1016/j.dsr2.2003.11.001.
- 844 McQuatters-Gollop, A., Reid, P. C., Edwards, M., Burkill, P. H., Castellani, C., Batten, S., et al.  
845 (2011). Is there a decline in marine phytoplankton? *Nature*, 472(7342), E6-E7.
- 846 Messié, M., and Chavez, F. P. (2012). A global analysis of ENSO synchrony: The oceans’  
847 biological response to physical forcing. *J. Geophys. Res. Oceans*, 117, C09001,  
848 doi:10.1029/2012JC007938.
- 849 Messié, M., and Chavez, F. P. (2015). Seasonal regulation of primary production in eastern  
850 boundary upwelling systems. *Progress in Oceanography*, 134, 1-18.
- 851 Morel, A., and Gentili, B. (2009). The dissolved yellow substance and the shades of blue in the  
852 Mediterranean Sea. *Biogeosciences*, 6(11).



- 853 Murtugudde, R. G., Signorini, S. R., Christian, J. R., Busalacchi, A. J., McClain, C. R., and Picaut,  
854 J. (1999). Ocean color variability of the tropical Indo-Pacific basin observed by SeaWiFS  
855 during 1997–1998. *J. Geophys. Res. Oceans*, 104, 18351–18366.
- 856 Neetu S., Lengaigne, M., M. Mangeas, J. Vialard, J. Leloup, C. Menkes, et al. (2020). Quantifying  
857 the benefits of non-linear methods for global statistical hindcasts of tropical cyclones intensity,  
858 *Monthly Weather Review*, doi:10.1175/WAF-D-19-0163.1
- 859 Nidheesh, A. G., Lengaigne, M., Vialard, J., Izumo, T., Unnikrishnan, A. S., Meyssignac, B., et al.  
860 (2017). Robustness of observation-based decadal sea level variability in the Indo-Pacific  
861 Ocean. *Geophys. Res. Lett.*, 44(14), 7391-7400.
- 862 Parvathi, V., Suresh, I., Lengaigne, M., Ethé, C., Vialard, J., Levy, M., et al. (2017). Positive Indian  
863 Ocean Dipole events prevent anoxia along the west coast of India. *Biogeosciences*, 14: 1541-  
864 1559, doi:10.5194/bg-14-1541-2017.
- 865 Patara, L., Visbeck, M., Masina, S., Krahnemann, G., and Vichi, M. (2011). Marine biogeochemical  
866 responses to the North Atlantic Oscillation in a coupled climate model, *J. Geophys. Res.*  
867 *Oceans*, 116, C07023.
- 868 Polovina, J. J., Howell, E. A., and Abecassis, M. (2008). Ocean's least productive waters are  
869 expanding. *Geophys. Res. Lett.*, 35, L03618, doi:10.1029/2007GL031745.
- 870 Radenac, M. H., Léger, F., Singh, A., and Delcroix, T. (2012). Sea surface chlorophyll signature  
871 in the tropical Pacific during eastern and central Pacific ENSO events. *J. Geophys. Res.*  
872 *Oceans*, 117(C4).
- 873 Radenac, M. H., Messié, M., Léger, F., and Bosc, C. (2013). A very oligotrophic zone observed  
874 from space in the equatorial Pacific warm pool. *Remote sensing of environment*, 134, 224-  
875 233.

- 876 Resplandy, L., Lévy, M., Madec, G., Pous, S., Aumont, O., and Kumar, D. (2011). Contribution  
877 of mesoscale processes to nutrient budgets in the Arabian Sea. *J. Geophys. Res. Oceans*,  
878 116(C11).
- 879 Rykaczewski, R. R., and Dunne, J. P. (2011). A measured look at ocean chlorophyll trends. *Nature*,  
880 472(7342), E5-E6.
- 881 Saji, N. H., Goswami, B. N., Vinayachandran, P. N., and Yamagata, T. (1999). A dipole mode in  
882 the tropical Indian Ocean. *Nature*, 401(6751), 360.
- 883 Sakamoto, T., Gitelson, A. A., Wardlow, B. D., Verma, S. B., and Suyker, A. E. (2011). Estimating  
884 daily gross primary production of maize based only on MODIS WDRVI and shortwave  
885 radiation data. *Remote Sensing of Environment*, 115(12), 3091-3101.
- 886 Sauzède, R., Claustre, H., Jamet, C., Uitz, J., Ras, J., Mignot, A., and D'Ortenzio, F. (2015).  
887 Retrieving the vertical distribution of chlorophyll a concentration and phytoplankton  
888 community composition from in situ fluorescence profiles: A method based on a neural  
889 network with potential for global-scale applications. *J. Geophys. Res. Oceans*, 120, 451-470,  
890 doi:10.1002/2014JC010355.
- 891 Schneider, B., Bopp, L., Gehlen, M., Segschneider, J., Frölicher, T. L., Cadule, P., et al. (2008).  
892 Climate-induced interannual variability of marine primary and export production in three  
893 global coupled climate carbon cycle models. *Biogeosciences, European Geosciences Union*,  
894 2008, 5 (2), pp.597-614. hal-00330705
- 895 Schollaert Uz, S., Busalacchi, A. J., Smith, T. M., Evans, M. N., Brown, C. W., Hackert, E. C., et  
896 al. (2017). Interannual and Decadal Variability in Tropical Pacific Chlorophyll from a  
897 Statistical Reconstruction: 1958–2008. *J. Clim.* 30, 7293–7315. doi:10.1175/JCLI-D-16-  
898 0202.1.

- 899 S  f  rian, R., Bopp, L., Gehlen, M., Orr, J. C., Eth  , C., Cadule, P., et al. (2013). Skill assessment  
900 of three earth system models with common marine biogeochemistry. *Clim. Dynam.*, 40, 2549–  
901 2573, doi:10.1007/s00382-012-1362-8. 1379.
- 902 Siegel, D. A., Behrenfeld, M. J., Maritorena, S., McClain, C. R., Antoine, D., Bailey, S. W., et al.  
903 (2013). Regional to global assessments of phytoplankton dynamics from the SeaWiFS  
904 mission. *Remote Sensing of Environment*, 135, 77-91.
- 905 Siegel, D. A., Maritorena, S., Nelson, N. B., and Behrenfeld, M. J. (2005). Independence and  
906 interdependencies among global ocean color properties: Reassessing the bio-optical  
907 assumption. *J. Geophys. Res. Oceans*, 110(C7).
- 908 Signorini, S. R., Franz, B. A., and McClain, C. R. (2015). Chlorophyll variability in the  
909 oligotrophic gyres: mechanisms, seasonality and trends. *Frontiers in Marine Science*, 2, 1.
- 910 Signorini, S. R., and McClain, C. R. (2012). Subtropical gyre variability as seen from satellites.  
911 *Remote Sensing Letters*, 3(6), 471-479.
- 912 Smith, T. M., Arkin, P. A., Ren, L., and Shen, S. S. (2012). Improved reconstruction of global  
913 precipitation since 1900. *J. Atmosph. and Ocean. Tech.*, 29(10), 1505-1517.
- 914 Staten, P. W., Lu, J., Grise, K. M., Davis, S. M., and Birner, T. (2018). Re-examining tropical  
915 expansion. *Nature Climate Change*, 8(9), 768-775.
- 916 Steinacher, M., Joos, F., Fr  licher, T. L., Bopp, L., Cadule, P., Cocco, V., et al. (2010). Projected  
917 21st century decrease in marine productivity: a multi-model analysis. *Biogeosciences*, 7(3),  
918 979-1005.
- 919 Storm, T., Boettcher, M., Grant, M., Z  hlke, M., Fomferra, N., Jackson, T., and Sathyendranath,  
920 S. (2013). Product User Guide, Ocean Colour Climate Change Initiative. 51. [www.esa-](http://www.esa-oceancolour-cci.org/?q=webfm_send/317)  
921 [oceancolour-cci.org/?q=webfm\\_send/317](http://www.esa-oceancolour-cci.org/?q=webfm_send/317).

- 922 Tagliabue, A., Bopp, L., Dutay, J. C., Bowie, A. R., Chever, F., Jean-Baptiste, P., et al. (2010).  
923 Hydrothermal contribution to the oceanic dissolved iron inventory. *Nature Geoscience*, 3(4),  
924 252-256.
- 925 Tang, W., Li, Z., and Cassar, N. (2019). Machine Learning Estimates of Global Marine Nitrogen  
926 Fixation. *J. Geophys. Res.: Biogeosciences*, 124(3), 717-730.
- 927 Thompson, D., and Wallace, J. M. (2000). Annular modes in the extratropical circulation. Part I:  
928 Month-to-month variability. *J. Climate*, 13, 1000–1016.
- 929 Thomas, A. C., Strub, P. T., Weatherbee, R. A., and James, C. (2012). Satellite views of Pacific  
930 chlorophyll variability: Comparisons to physical variability, local versus nonlocal influences  
931 and links to climate indices. *Deep-Sea Res. Part II*, 77, 99-116.
- 932 Vantrepotte, V., and Mélin, F. (2009). Temporal variability of 10-year global SeaWiFS time-series  
933 of phytoplankton chlorophyll a concentration. *ICES J. Marine Science*, 66(7), 1547-1556.
- 934 Vapnik, V. (2000). Statistics for engineering and information science. The Nature of Statistical  
935 Learning Theory.
- 936 Vapnik, V. (1998). The support vector method of function estimation. *In Nonlinear Modeling* (pp.  
937 55-85). Springer, Boston, MA.
- 938 Vapnik, V. N. (1995). The Nature of Statistical Learning Theory. New York: Springer2 Verlag.
- 939 Westberry, T., Behrenfeld, M. J., Siegel, D. A., and Boss, E. (2008). Carbon-based primary  
940 productivity modeling with vertically resolved photoacclimation. *Global Biogeochem. Cycles*,  
941 22(2).
- 942 Wiggert, J. D., Murtugudde, R. G., and Christian, J. R. (2006). Annual ecosystem variability in  
943 the tropical Indian Ocean: Results of a coupled bio-physical ocean general circulation model.  
944 *Deep Sea Res. Part II*, Volume 53, Issues 5–7, Pages 644-676.

- 945 Wiggert, J. D., Vialard, J., and Behrenfeld, M. J. (2009). Basin wide modification of dynamical  
946 and biogeochemical processes by the positive phase of the Indian Ocean Dipole during the  
947 SeaWiFS era. In: Indian Ocean Biogeochemical Processes and Ecological Variability, vol.  
948 185, edited by: J. D. Wiggert, R. R. Hood, S. Wajih, A. Naqvi, K. H. Brink, and S. L. Smith,  
949 p. 350, 2009.
- 950 Wilson, C., and Adamec, D. (2001). Correlations between surface chlorophyll and sea surface  
951 height in the tropical Pacific during the 1997-1999 El Nino-Southern event. *J. Geophys. Res.*  
952 *Oceans*, 106(C12),31175-31188.
- 953 Wilson, C., and Adamec, D. (2002). A global view of bio-physical coupling from SeaWiFS and  
954 TOPEX satellite data, 1997 – 2001. *Geophys. Res. Lett.*, 29(8), 1257,  
955 doi:10.1029/2001GL014063.
- 956 Wilson, C., and Coles, V. J. (2005). Global climatological relationships between satellite  
957 biological and physical observations and upper ocean properties. *J. Geophys. Res. Oceans*,  
958 110, C10001, doi:10.1029/2004JC002724.
- 959 Yoder, J. A., and Kennelly, M. A. (2003). Seasonal and ENSO variability in global ocean  
960 phytoplankton chlorophyll derived from 4 years of SeaWiFS measurements. *Global*  
961 *Biogeochem. Cycles*, 17(4), 1112, doi:10.1029/2002GB001942.
- 962 Yoo, S. and Son, S.-H. (1998). Seasonal patterns of phytoplankton biomass change in the  
963 Northwestern Pacific Marginal Seas observed by CZCS. *Ocean Res.*, 20: 29-42.
- 964 Zhan, H., Shi, P., and Chen, C. (2003). Retrieval of Oceanic Chlorophyll Concentration Using  
965 Support Vector Machines. *IEEE Transactions on Geoscience and Remote Sensing*, 41 (12),  
966 2947–2951.