

1 Title: Multi-model remote sensing assessment of primary production in the subtropical
2 gyres

3

4 Authors: Regaudie-de-Gioux A.^{1,2*}, Huete-Ortega M.^{3,4}, Sobrino C.⁴, López-Sandoval D.
5 C.⁵, González N.⁶, Fernández-Carrera A.⁴, Vidal M.⁷, Marañón E.⁴, Cermeño P.⁸, Latasa,
6 M.⁹, Agustí S.^{5,1}, Duarte C. M.^{5,1}

7

8 ¹ Mediterranean Institute for Advanced Studies (IMEDEA), Calle Miquel Marques 21,
9 07190 Esporles, Spain

10 ² Present address: Ifremer, DYNECO PELAGOS, F-29280 Plouzané, France

11 ³Oroboros Instruments, Schöpfstraße 8, 6020, Innsbruck, Austria

12 ⁴Departamento de Ecología y Biología Animal, Universidade de Vigo, 36310 Vigo,
13 Spain

14 ⁵ King Abdullah University of Science and Technology (KAUST), Red Sea Research
15 Center (RSRC), Thuwal, 23955-6900, Saudi Arabia

16 ⁶ Área de Biodiversidad y Conservación, ESCET, Universidad Rey Juan Carlos,
17 Tulipán s/n., Móstoles 28933 Madrid, Spain

18 ⁷ Departament de Biologia Evolutiva, Ecologia i Ciències Ambientals. Universitat de
19 Barcelona. Diagonal, 643. 08028 Barcelona, Spain

20 ⁸ Institute of Marine Sciences (ICM-CSIC), Passeig Marítim de la Barceloneta, 37–49,
21 08003 Barcelona, Spain

22 ⁹ Spanish Institute of Oceanography, Avda. Principe de Asturias 70bis, 33212 Gijón,
23 Spain

24

25 *corresponding author: aregaudi@ifremer.

26 Abstract

27 The subtropical gyres occupy about 70 % of the ocean surface. While primary production
28 (PP) within these oligotrophic regions is relatively low, their extension makes their total
29 contribution to ocean productivity significant. Monitoring marine pelagic primary
30 production across broad spatial scales, particularly across the subtropical gyre regions, is
31 challenging but essential to evaluate the oceanic carbon budget. PP in the ocean can be
32 derived from remote sensing however *in situ* depth-integrated PP (IPP^{is}) measurements
33 required for validation are scarce from the subtropical gyres. In this study, we collected
34 more than 120 IPP^{is} measurements from both northern and southern subtropical gyres that
35 we compared to commonly used primary productivity models (the Vertically Generalized
36 Production Model, VGPM and six variants; the Eppley-Square-Root model, ESQRT; the
37 Howard–Yoder–Ryan model, HYR; the model of MARRA, MARRA; and the Carbon-
38 based Production Model, CbPM) to predict remote PP (PP_r) in the subtropical regions and
39 explored possibilities for improving PP prediction. Our results showed that satellite-
40 derived PP (IPP^{sat}) estimates obtained from the VGPM1, MARRA and ESQRT provided
41 closer values to the IPP^{is} (i.e., the difference between the mean of the IPP^{sat} and IPP^{is} was
42 closer to 0; |Bias| ~ 0.09). Model performance varied due to differences in satellite
43 predictions of *in situ* parameters such as chlorophyll a (chl-*a*) concentration or the optimal
44 assimilation efficiency of the productivity profile (P^{B_{opt}}) in the subtropical region. In
45 general, model performance was better for areas showing higher IPP^{is}, highlighting the
46 challenge of PP prediction in the most oligotrophic areas (i.e. PP < 300 mg C m⁻² d⁻¹).
47 The use of *in situ* chl-*a* data, and P^{B_{opt}} as a function of sea surface temperature (SST) and
48 the mixed layer depth (MLD) from gliders and floats in PP_r models would improve their
49 IPP predictions considerably in oligotrophic oceanic regions such as the subtropical gyres
50 where MLD is relatively low (< 60 m) and cloudiness may bias satellite input data.

51 Keywords: Primary production, remote PP model, skills, subtropical gyre

52

53 **1. Introduction**

54 Subtropical gyres are extensive regions that occupy about 70 % of the ocean
55 surface. While primary production per unit of surface within these regions is relatively
56 low (e.g. Jones 1996, Karl et al. 1996, Karl et al. 2001, Teira et al. 2002), their immense
57 size makes their total contribution to ocean productivity significant. In these regions,
58 phytoplankton growth rates and productivity show large variability with minimal net
59 changes in biomass (Laws et al. 1987, Marañón et al. 2000, Marañón et al. 2003).
60 Monitoring marine pelagic primary production across broad spatial scales, particularly
61 across the subtropical gyre regions, is indeed essential to evaluate its role in the oceanic
62 carbon budget and food webs (Volk & Hoffert 1985; Platt & Sathyendranath, 1988;
63 Longhurst et al. 1995; Field 1998; Duarte et al. 1999).

64 Over the last two decades, significant efforts have been made to derive models
65 that estimate marine primary production from remote sensing products (PP_r, e.g. Platt &
66 Sathyendranath, 1988; Lee et al., 1996; Behrenfeld & Falkowski, 1997; Behrenfeld et al.
67 2005; Westberry et al., 2008; Silsbe et al. 2016). PP_r models are able to estimate the
68 evolution of PP over different time scales (daily, monthly and annually) covering almost
69 all parts of the globe. PP_r models are mainly parameterized using remote sensing as input
70 data and may differ in their complexities when dealing with depth and irradiance
71 wavelength-dependent variability. However, the performance in reproducing *in situ*
72 depth-integrated PP (IPP^{is}) vary across regions, so evaluation of multiple models by
73 comparing satellite depth-integrated PP (IPP^{sat}) derived from PP_r models and IPP^{is} across
74 different regions is important to guide model selection (e.g. Behrenfeld & Falkowski
75 1997b; Campbell et al., 2002; Westberry et al., 2008; Friedrichs et al., 2009).

76 Performance assessment of PP_r models in the five subtropical gyre regions of the
77 ocean has been uneven (e.g. Campbell et al., 2002; Westberry et al., 2008; Friedrichs et
78 al., 2009), so that the evaluation of their performance for these areas is still insufficient.
79 Indeed, some studies covered a broad but still limited spatial area (e.g. ~ 50 stations
80 between North Pacific and South Pacific subtropical gyres in Friedrichs et al. 2009; ~ 30
81 stations between North Atlantic and South Atlantic subtropical gyres in Tilstone et al.
82 2009). Other studies have analyzed PP_r model skills using long-term time-series data,
83 however they only included data from two stations located in subtropical gyre regions,
84 specifically in the North Pacific subtropical gyre (ALOHA station of Hawaii Ocean Time
85 series, HOT) and the North Atlantic subtropical gyre (Bermuda Atlantic Time-series
86 Study, BATS) (Westberry et al. 2008, Saba et al. 2011; Ma et al. 2014). Therefore, the
87 performance of PP_r models in the North and South Pacific subtropical gyre regions remain
88 insufficiently evaluated, whereas no study has been conducted yet on the performance of
89 PP_r models in the Indian subtropical gyre region.

90 Due to the significant contribution of subtropical gyres to total oceanic primary
91 production, it is essential to improve our knowledge on the performance of PP_r models in
92 predicting PP in these regions. This requires the comparison between IPP^{is} and IPP^{sat} data
93 covering a broader spatial scale across the subtropical gyres so far reported. Here, we
94 provide more than 120 IPP^{is} measurements derived from the standard ^{14}C method along
95 the Malaspina Circumnavigation Expedition (MCE), which circumnavigated the
96 subtropical and tropical ocean between 2010 and 2011 (Duarte, 2015). It encompassed
97 fourteen Longhurst biogeochemical provinces (Longhurst, 1995), including four
98 subtropical gyre regions, and the poorly-sampled Indian subtropical gyre region. The
99 MCE lasted 7 months and was divided by 6 transects of which each could be considered
100 as an oceanic cruise on its own. These IPP^{is} data allowed to compare the performance of

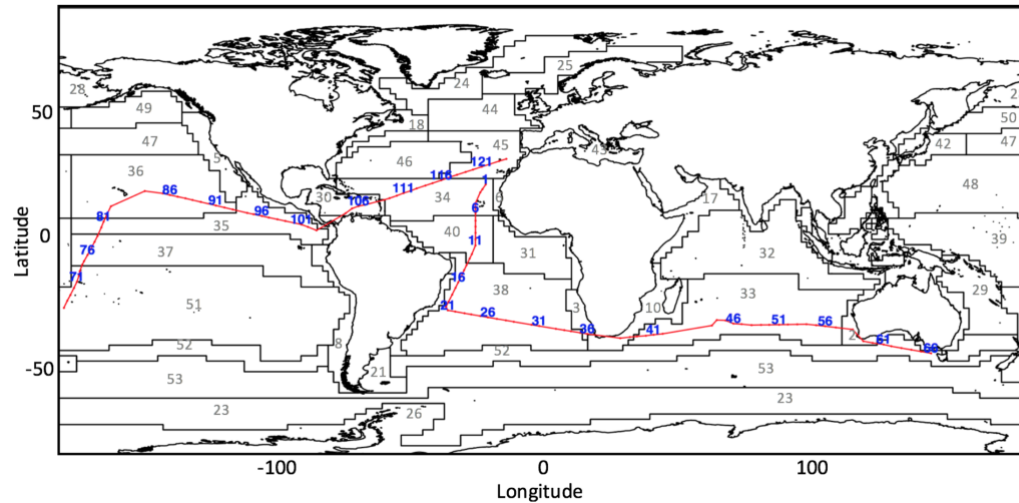
101 five commonly used remote PP models and explored afterward possibilities for improving
102 their performances to support improved remote sensing assessment of PP in subtropical
103 gyres.

104

105 **2. Methods**

106 *2.1. Study area*

107 Seawater samples were collected during the Malaspina Circumnavigation
108 Expedition on board the R/V *Hespérides* from December 2010 to July 2011 (Fig. 1). The
109 MCE was divided in seven transects during which IPP^{is} was measured: 1) from Cádiz,
110 Spain to Rio de Janeiro, Brazil (Station 6 – 26) from December 2010 to January 2011; 2)
111 from Rio de Janeiro to Cape Town, South Africa (Station 27 – 44) from January 2011 to
112 February 2011; 3) from Cape Town to Perth Australia (Station 46 – 68) from February
113 2011 to March 2011; 4) from Perth to Sydney, Australia (Station 69 – 76) in March 2011;
114 5) from Auckland, New Zealand to Honolulu, Hawaii (Station 83 –99) from April 2011
115 to May 2011; 6) from Honolulu to Panama, Panama (Station 104 – 126) from May 2011
116 to June 2011; and 7) from Cartagena de Indias, Colombia to Cartagena, Spain (Station
117 127 – 147) from June 2011 to July 2011. Sampled stations were grouped into different
118 provinces following Longhurst classification (Longhurst, 1995): the North Atlantic gyre
119 region (NAGR) comprises all sampling sites located between the North Atlantic
120 Subtropical and Tropical Gyral Provinces (NATR and NASE); the South African Coastal
121 region (SACR) comprises all the stations found in the Benguela current (BENG) and the
122 East African coastal current (EAFC); the West Australian Current region (WACR)
123 comprises all stations located in the Western Australian and Indonesian coasts (AUSW
124 and SSTC, respectively).



125

126 **Figure 1.** Location of the sampled station during the MCE and Lohngurst biogeochemical ocean provinces (Lohngurst, 1998; 2006). The
 127 provinces where IPP^{is} were sampled were: (2) Australia-Indonesia Coastal Province, AUSW; (3) Benguela Current Coastal Province, BENG;
 128 (30) Caribbean Province, CARB; (10) E. Africa Coastal Province, EAFR; (33) Indian S. Subtropical Gyre Province, ISSG; (45) N. Atlantic
 129 Subtropical Gyral Province (East), NASE; (34) N. Atlantic Tropical Gyral Province, NATR; (36) N. Pacific Tropical Gyre Province, NPTG;
 130 (37) Pacific Equatorial Divergence Province, PEQD; (35) N. Pacific Equatorial Countercurrent Province, PNEC; (38) South Atlantic Gyral
 131 Province, SATL; (51) S. Pacific Subtropical Gyre Province, SPSG; (52) S. Subtropical Convergence Province, SSTC; (40) Western Tropical
 132 Atlantic Province, WTRA. Grey numbers represent the provinces referenced by Lohngurst, 1998 and 2006. For a complete list of provinces,
 133 please to the Table S1 in Pinedo-González et al. (2015). Blue numbers represent the station number during MCE and red solid lines represent
 134 the MCE transects.

135 We maintained the same provinces codes as Longhurst (Longhurst, 1995) for the stations
136 located in the South Atlantic Gyral Province (SATL), in the Indian South Subtropical
137 Gyre Province (ISSG), in the South Pacific Subtropical Gyre Province (SPSG), in the
138 North Pacific Tropical Gyre Province (NPTG) and in the North Pacific Equatorial
139 Countercurrent Province (PNEC).

140

141 *2.2. In situ measurement of chlorophyll-a*

142 Chlorophyll-*a* concentration was measured by High-Performance Liquid
143 Chromatography (HPLC) as described in Zapata et al. (2000) with minor modifications
144 (Latasa, unpublished). 2 L of seawater were filtered onto 25 mm glass fiber filters
145 (Whatman GF/F) and frozen at - 80 °C until analysis by HPLC. Pigments were extracted
146 with 2.5 mL acetone 90 % containing trans- β -apo-8'-carotenal as internal standard,
147 sonicated and stored at - 20 °C for 24 h. A large volume (720 – 1400 μ L) of extract was
148 injected onto an Agilent 1200 HPLC system and analyzed following the procedure
149 described by Latasa (2014). The analytical precision of the method is better than 1 %
150 (Latasa 2014).

151

152 *2.3. In situ mixed layer, nitracline and euphotic depths*

153 The mixed layer depth (MLD) was estimated from CTD data (SBE911plus, Sea-
154 Bird Electronics) using the threshold method with a finite difference criterion, as the
155 depth at which the potential density changed by 0.125 kg m⁻³ relative to the one at a near-
156 surface reference level (usually 6 m), according to Monterey & Levitus (1997).

157 The nitracline was determined from nitrate plus nitrite concentration data,
158 measured on a segmented flow Skalar auto-analyser by standard methods (Grasshoff et
159 al., 1999, Moreno-Ostos, 2012), as the depth, from the surface, where the first sustained

160 increase of the concentration gradient is observed. It is a region of approximately
161 maximum and steady concentration gradient in the first 200 m of the vertical profile,
162 including 4 – 6 nitrate data points.

163 At each station, vertical profiles of underwater solar radiation were measured at
164 noon (between 11 am to 1 pm local time) using a PRR-800 Underwater Profiling
165 Radiometer (Biospherical Instruments). The profiling submarine radiometer measured
166 underwater radiation in the ultraviolet and -visible bands. The euphotic zone depth (Z_{eu})
167 was determined as the depth at which the light intensity reached the 1 % of its intensity
168 at the surface.

169

170 *2.5. In situ measurement of primary production*

171 Phytoplankton primary production was measured at 124 stations with the ^{14}C -
172 uptake technique, following the procedures detailed in Marañón et al. (2000). Seawater
173 was sampled from five depths in the euphotic zone corresponding to 100 % (ca. 3 m
174 depth), 50 %, 20 %, 7 % and 1 % of incident Photosynthetically Active Radiation (PAR).
175 For each depth, four 72 mL polystyrene bottles (three clear bottles and one dark bottle)
176 were filled with unfiltered seawater, inoculated with 10 – 20 $\mu\text{Ci NaH}^{14}\text{CO}_3$ and
177 incubated on-deck from dawn to dusk. Temperature and irradiance in the incubators
178 simulated the water temperature and the incident irradiance at the corresponding depth of
179 each sample by using a combination of neutral density and blue filters (Mist Blue, ref.
180 061, Lee Filters ®). After incubation, samples from three of the five depths (100 %, 20
181 % and 1 % PAR) were sequentially filtered through 20, 2 and 0.2 μm polycarbonate filters
182 while the other depths (50 % and 7 % PAR) were directly filtered by 0.2 μm . Immediately
183 after filtering, filters were then exposed to concentrated HCl fumes at least 12 h to remove
184 the non-fixed inorganic ^{14}C . Filters were placed in scintillation vials to which 5 mL of

185 liquid scintillation cocktail was added. The radioactivity on each filter (disintegrations
186 per minute, DPM) was determined using a Wallac scintillation counter. To compute the
187 rate of photosynthetic carbon fixation, the dark-bottle DPM was subtracted from the light-
188 bottle DPM values. A constant value of $24,720 \mu\text{g L}^{-1}$ (or $2,060 \mu\text{mol L}^{-1}$) was assumed
189 for the concentration of dissolved inorganic carbon for surface waters in tropical ocean
190 (Key et al., 2004). A correction factor of 1.05 was applied to this constant value for
191 discrimination isotopic. Total primary production was calculated as the sum of the
192 primary production on each size class. For all triplicate measurements of total primary
193 production conducted during MCE ($n = 522$), the mean coefficient of variation was 23
194 %. *In situ* depth-integrated primary production (IPP^{is} , $\text{mg C m}^{-2} \text{d}^{-1}$) was calculated by the
195 trapezoidal integration of measurements from the surface to 1 % PAR depth. The IPP^{is}
196 data set is available from Regaudie-de-Gioux et al. 2019. The original IPP^{is} measurements
197 were reported as hourly rates and then were converted to daily rates multiplying the
198 hourly rates by the corresponding day length at each sampled station. The highest hourly
199 chl-*a*-specific primary production (P^{B} , $\text{mg C chl-}a^{-1} \text{h}^{-1}$) in the water column was defined
200 as the observed *in situ* $\text{P}^{\text{B}}_{\text{opt}}$ (Behrenfeld & Falkowski 1997b) for each station in this
201 study. The variability in IPP^{is} along MCE transects is described in Pinedo-González et al.
202 (2015).

203

204 2.6. Input data variables for IPP^{sat} : Satellite-derived and modelled variables

205 Ocean color models typically use Level-3, monthly or 8-day, satellite-derived
206 input data. In this study, daily standard level 3 (i.e. mapped processed to surface
207 quantities) products of PAR, ocean color index (OCI)-based chl-*a*, diffuse attenuation at
208 490 nm ($K_d(490)$), sea surface temperature (SST) and particulate backscatter coefficient
209 at 443 nm ($b_{\text{bp}}(443)$ from GSM model) were provided by the OceanColor Web

210 (<https://oceancolor.gsfc.nasa.gov>) and were calculated from the Moderate Resolution
211 Imaging Spectroradiometer aboard the Aqua NASA spacecraft (MODISA). Spatial
212 resolution of all products was ~ 9 km at the Equator. Additionally, the mixed layer depth
213 (MLD) and the nitracline depths (Z_{NO_3}) were modelled variables for IPP^{sat}. The daily
214 product of MLD was provided by the Ocean Productivity Online Data
215 (<https://orca.science.oregonstate.edu>). Z_{NO_3} were calculated from monthly climatological
216 nutrient fields reported in the World Ocean Atlas 2013 (Garcia et al., 2014) at 1-degree
217 resolution, and defined as the first depth at which nitrate + nitrite exceeded 0.5 μ M. All
218 variables were extracted from 1 pixel radius windows (i.e. 3 x 3-pixel box) centered at
219 the pixel nearest to the *in situ* sample and we calculated the average of each window
220 (Bailey & Werdell 2006). Satellite variables were excluded when more than 70 % were
221 masked. We used this matchup criteria to increase the number of matchups, particularly
222 in subtropical areas where cloudiness is important increasing satellite masked data.

223

224 2.7. Satellite algorithms

225 In the present study, we did not focus on the comparison of the Primary
226 Production Algorithm Round Robin (PPARR) as it has already been thoroughly assessed
227 (e.g. Campbell et al., 2002; Westberry et al., 2008; Friedrichs et al., 2009). Instead, we
228 used several well-known PP_r models (Table 1) that are commonly used to estimate
229 satellite IPP (IPP^{sat}) and assessed their performance in the subtropical gyres.

230

Model	Remote variable input							P_{opt}^B *	Z_{eu} **	Reference
	chl-a	SST	PAR	MLD	K_I (490)	Z_{NO_3}	b_{bp} (443)			
VGPM1	X	X	X					1	1	Behrenfeld and Falkowski (1997)
VGPM11	X	X	X					1	2	Behrenfeld and Falkowski (1997)
VGPM2	X	X	X					2	1	Behrenfeld and Falkowski (1997)
VGPM22	X	X	X					2	2	Behrenfeld and Falkowski (1997)
VGPM3	X	X	X					3	1	Behrenfeld and Falkowski (1997)
VGPM33	X	X	X					3	2	Behrenfeld and Falkowski (1997)
ESQRT	X									Eppley et al. (1985)
HYR	X	X	X	X						Howard and Yoder (1997)
MARRA	X	X	X							Marra et al. (2003)
CbPM	X	X	X	X	X	X	X		1	Westberry et al. (2008)

* 1) Behrenfeld and Falkowski (1997); 2) Antoine and Morel (1996); 3) Kameda and Ishizaka (2005)

** 1) Morel and Berthon (1989); 2) Mobley (2004)

231

232 **Table 1** – Model descriptions

233 First, we used the most widely utilized PP_r model, the Vertically Generalized
234 Production Model (VGPM) based on chl-*a* (Behrenfeld & Falkowski 1997). The VGPM
235 uses remote inputs of chl-*a*, SST and PAR. Here we proposed several VGPM variants
236 and alternative methods to estimate P_{opt}^B . The first VGPM variant (here called VGPM1)
237 is the original VGPM as in Behrenfeld & Falkowski 1997 where the P_{opt}^B was obtained
238 from a 7th-order polynomial SST regression (here called P_{opt}^B1). For the VGPM2 variant,
239 P_{opt}^B was estimated after Eppley (1972) as implemented by Antoine and Morel (1996) as
240 an exponential function of temperature (here called P_{opt}^B2). In the VGPM3 variant, P_{opt}^B
241 was estimated after Kameda and Ishizaka (2005) as inversely proportional to
242 phytoplankton size (here called P_{opt}^B3). Additionally, we modified these three VGPM
243 variants with an alternative method to estimate Z_{eu} , originally estimated from chl-*a*
244 concentration (Morel & Berthon, 1989). The three extra variants of the VGPM models
245 cited above (called hereafter, VGPM11, VGPM22 and VGPM33) included a modified
246 Z_{eu} estimated from the diffuse attenuation coefficient of PAR (m^{-1}) (following Mobley,
247 2004).

248 Additionally, we used the simplest PP_r model, the Eppley-Square-Root model
249 (ESQRT; Eppley et al., 1985). The ESQRT model uses only chl-*a* as remote inputs
250 assuming that the standing stock is the sole determinant of photosynthetic rate. It ignores

251 any external forcing or changes in physiological state. We also used the original Howard–
 252 Yoder–Ryan model (HYR; Howard & Yoder 1997) which for many years was used as a
 253 standard MODIS algorithm. The maximum growth rate is parameterized here as a
 254 function of SST according to Eppley (1972). PP_r is integrated here to the MLD rather
 255 than the euphotic depth. The HYR model uses remote inputs as chl-*a*, SST, PAR and
 256 MLD. Furthermore, we used the PP_r model described by Marra et al. (2003) (MARRA)
 257 that is based on chlorophyll-specific absorption parameterized by SST and maximum
 258 quantum yield. The MARRA model uses chl-*a*, SST and PAR as remote inputs. Finally,
 259 we used the Carbon-based Production Model (CbPM; Westberry et al. 2008). The CbPM
 260 uses remote inputs as chl-*a*, $b_{bp}(443)$, PAR, $K_d(490)$, MLD and Z_{NO_3} . The CbPM utilizes
 261 $b_{bp}(443)$ to derive phytoplankton carbon biomass.

262

263 2.8. Model validation

264 Performance of each PP_r model was analyzed using the total root mean square
 265 difference (RMSD; Campbell et al., 2002):

$$266 \text{ RMSD} = \left(\frac{1}{N} \sum_{i=1}^N \Delta(i)^2 \right)^{1/2} \quad (1)$$

267 where $\Delta(i) = \log_{10}[\text{IPP}^{\text{sat}}(i)] - \log_{10}[\text{IPP}^{\text{is}}(i)]$ and N is the total number of paired data. The
 268 model performance and predictive skills increase as RMSD values become closer to 0.

269 RMSD captures a model’s ability to represent both the mean and the variability of *in situ*
 270 data and thus, is composed by the bias (i.e. the difference between the means, B) and the
 271 unbiased RMSD (i.e. representing the difference of variability, $u\text{RMSD}$):

$$272 \text{ RMSD}^2 = B^2 + u\text{RMSD}^2 \quad (2)$$

$$273 B = \overline{\log_{10}(\text{IPP}^{\text{sat}})} - \overline{\log_{10}(\text{IPP}^{\text{is}})} \quad (3)$$

274 When B is negative or positive, model underestimates or overestimates IPP^{is},
275 respectively. Model estimation is closer to IPP^{is} when B is closer to 0. The differences in
276 the variability of IPP^{is} and IPP^{sat} are smaller when uRMSD is closer to 0.

277 Target diagram (Jolliff et al. 2009) will be used to illustrate model performances. This
278 diagram allows visualizing bias, uRMSD, and total RMSD for all models on a single plot.

279 For that, these quantities are normalized by the standard deviation of log₁₀ IPP^{is} ($\sigma_d =$
280 0.26):

$$281 \quad B^* = B/\sigma_d \quad (4)$$

$$282 \quad \text{uRMSD}^* = \text{sign}(\sigma_m - \sigma_d) \text{uRMSD}/\sigma_d \quad (5)$$

283 where σ_m is the standard deviation of log₁₀ IPP^{sat}.

284 The target diagram provides information about whether the model standard deviation is
285 larger (uRMSD* > 0) or smaller (uRMSD* < 0) than the *in situ* standard deviation and
286 on the presence of positive (B* > 0) or a negative (B* < 0) bias. The distance of each
287 point from the origin is the standard deviation normalized total RMSD, RMSD*. Any
288 points greater than RMSD* = 1 may be considered poor performers.

289 The empirical cumulative distribution function (ECDF) illustrates the distribution of the
290 data among values and orders the data from the smallest to the largest data.

291

292 2.9. Uncertainty analysis

293 Considering that ocean color models use satellite-derived input variables, it is
294 important to estimate how these data can affect their derived IPP^{sat}. For that, we compared
295 each station *in situ* variables with its respective satellite-derived variable when available:
296 *in situ* chl-*a*, SST, P^B_{opt}, Z_{eu}, MLD and nitracline depth with daily satellite chl-*a*, SST,
297 P^B_{opt}, Z_{eu}, MLD and nitracline depth data (for details see 2.6), respectively.

298 The median value of the ratio satellite to *in situ* inputs points to the overall bias. The semi-
299 interquartile range (SIQR) provides insight on the spreading data and is defined as
300 followed:

$$301 \quad SIQR = \frac{Q3-Q1}{2} \quad (6)$$

302 Where $Q1$ is the 25th percentile and $Q3$ is the 75th percentile of each series of satellite to
303 *in situ* inputs ratio.

304 The median percent difference (MPD) was calculated to measure how accurately the
305 satellite inputs agree with *in situ* inputs. It is defined as the median of the individual
306 absolute percent differences (PD) as followed:

$$307 \quad PD_i = 100 \frac{|X_i - Y_i|}{Y_i} \quad (7)$$

308 where Y_i is the *in situ* inputs and X_i is the corresponding satellite-derived inputs.
309 Parameters of linear regression between *in situ* and satellite-derived inputs were also
310 evaluated.

311

312 **3. Results**

313 *3.1. In situ data*

314 During the MCE, IPP^{is} ranged from 42.4 mg C m⁻² d⁻¹ in the Indian Ocean (station
315 54) to 877.6 mg C m⁻² d⁻¹ in the Pacific Ocean (station 102). The region with the highest
316 mean IPP^{is} was the PNEC with 434.6 mg C m⁻² d⁻¹ while the region with the lowest mean
317 IPP^{is} was the ISSG with 125 mg C m⁻² d⁻¹ (Table 2). The region with the highest variability
318 of IPP^{is} was SACR, while the region with the lowest variability of IPP^{is} was NPTG (Table
319 2). The mean input variables ranged as followed (Table 2): chl-*a* from 0.06 mg m⁻³
320 (SATL) to 0.17 mg m⁻³ (SACR and WACR, respectively); Z_{eu} from 73 m (PNEC) to 139
321 m (SATL); P_{opt}^B from 2.01 mg C mg chl-*a*⁻¹ d⁻¹ (WACR) to 9.66 mg C mg chl-*a*⁻¹ d⁻¹

322 (PNEC); SST from 19 °C (WACR) to 29 °C (both SPSG and PNEC); MLD from 27 m
 323 (PNEC) to 75 m (NPTG); and nitracline depth from 30 m (PNEC) to 147 m (SATL).

324

	IPP ^{is} mg C m ⁻² d ⁻¹	chl- <i>a</i> mg m ⁻³	Z _{eu} m	P ^B _{opt} mg C mg chl- <i>a</i> ⁻¹ h ⁻¹	SST °C	MLD m	Nitracline m
NAGR	272.2 (± 102.9)	0.07 (± 0.06)	110 (± 20)	5.70 (± 2.52)	25.5 (± 2.5)	49 (± 16)	132 (± 43)
SATL	224.8 (± 62)	0.06 (± 0.02)	139 (± 15)	3.77 (± 1.20)	25.3 (± 2.2)	51 (± 14)	147 (± 37)
SACR	288.2 (± 232.9)	0.17 (± 0.18)	79 (± 20)	3.26 (± 2.32)	22.7 (± 2.4)	54 (± 11)	69 (± 26)
ISSG	125 (± 62)	0.06 (± 0.02)	124 (± 27)	2.82 (± 1.07)	23.9 (± 1.3)	50 (± 16)	130 (± 40)
WACR	156.5 (± 63.9)	0.17 (± 0.10)	119 (± 31)	2.01 (± 0.96)	19.1 (± 2.9)	58 (± 17)	91 (± 34)
SPSG	210.6 (± 120.8)	0.10 (± 0.04)	134 (± 27)	2.97 (± 1.63)	28.9 (± 1.1)	59 (± 8)	137 (± 23)
NPTG	183.2 (± 58.2)	0.08 (± 0.02)	115 (± 22)	2.97 (± 1.23)	24.2 (± 1.6)	75 (± 28)	128 (± 39)
PNEC	434.6 (± 193.5)	0.15 (± 0.05)	73 (± 21)	9.66 (± 3.58)	28.8 (± 0.6)	27 (± 12)	30 (± 17)
325 <i>total</i>	244.8 (± 146.7)	0.11 (± 0.08)	113 (± 28)	4.20 (± 2.75)	25.4 (± 3.2)	54 (± 22)	114 (± 48)

326 **Table 2** – Means and standard deviations of *in situ* IPP, chl-*a*, Z_{eu}, P^B_{opt}, SST, MLD and
 327 Nitracline depth for each regional group and for the whole MCE.

328

329 3.2. Comparison between *in situ* data and input data variables for IPP^{sat}

330 Input data variables for IPP^{sat} showed variable agreements with their
 331 corresponding *in situ* observed variables. Satellite SST showed very good agreement with
 332 *in situ* SST ($R^2 = 0.95$, $r = 0.98$, MPD = 2 %) and presented the lowest spreading values
 333 (SIQR = 0.02) (Table 3). The rest of the input data variables for IPP^{sat} showed reasonable
 334 agreement with their corresponding *in situ* variables with R^2 ranging from 0.16 to 0.56,
 335 SIQR from 0.09 to 0.23 and MPD from 19 % to 35 % (Table 3). The lowest overall bias
 336 of input data variables for IPP^{sat} in comparison with *in situ* inputs was observed for MLD
 337 (median MLD = 0.65) (Table 3).

338

	R ²	Slope	r	N	Median	SIQR	MPD
Chl- <i>a</i>	0.16	0.72 ± 0.18	0.39	93	0.89	0.23	29.87
SST	0.95	0.96 ± 0.02	0.98	121	0.98	0.02	2.25
Nitracline	0.53	0.76 ± 0.07	0.73	113	0.83	0.21	23.94
Z _{eu} 1	0.32	0.34 ± 0.05	0.56	81	0.81	0.09	19.49
Z _{eu} 2	0.23	0.56 ± 0.11	0.48	81	1.08	0.16	20.16
338 MLD	0.56	0.48 ± 0.04	0.75	107	0.65	0.12	35.03

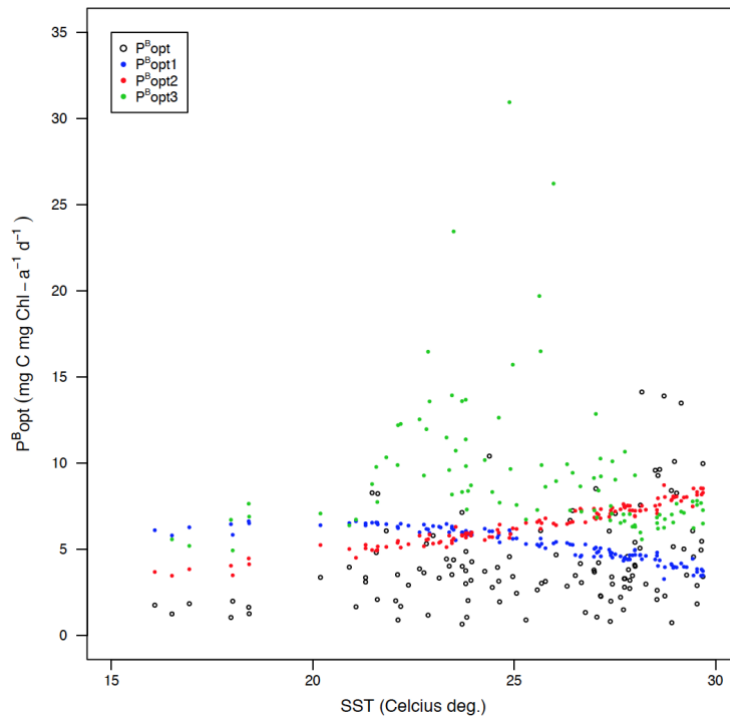
339 **Table 3** – Uncertainty analysis on differences between *in situ* data and input data
340 variables for IPP^{sat} with the statistics of linear regression (R^2 and slope \pm 95 % CI; Fig.
341 S1), the Pearson correlation coefficient (r), the number of match-ups, the median value
342 of the ratio satellite to *in situ* data (Median), the semi-interquartile range of satellite to *in*
343 *situ* inputs ratio (SIQR) and the median percent difference between satellite and *in situ*
344 inputs data (MPD).

345

346 3.3. Model phytoplankton physiology variable

347 The three $P^{B_{opt}}$ variables modelled from satellite-derived data ($P^{B_{opt1}}$, $P^{B_{opt2}}$ and
348 $P^{B_{opt3}}$) presented weak agreements with observed $P^{B_{opt}}$ data ($R^2 < 0.12$) and showed the
349 highest spreading values (SIQR = 0.57, 0.65 and 1.13, respectively). These $P^{B_{opt}}$ data
350 showed the highest overall bias in comparison with *in situ* inputs (median $P^{B_{opt1}} = 1.49$,
351 median $P^{B_{opt2}} = 1.80$ and median $P^{B_{opt3}} = 2.69$).

352 While $P^{B_{opt1}}$ and $P^{B_{opt2}}$ were described as function of SST, $P^{B_{opt3}}$ and *in situ* $P^{B_{opt}}$ did not
353 follow any correlation with SST (Fig. 2). $P^{B_{opt3}}$ presented a wider value range (from 5 to
354 $> 30 \text{ mg C mg Chl-}a^{-1} \text{ d}^{-1}$) than *in situ* $P^{B_{opt}}$ (from 0.14 to $< 15 \text{ mg C mg Chl-}a^{-1} \text{ d}^{-1}$)
355 between 21 and 28 °C.



356

357 **Figure 2.** Representation of the three P^B_{opt} algorithms used in this study and *in situ* P^B_{opt}
 358 in function of *in situ* SST.

359

360 *3.4. Model performance across all regions*

361 The estimation of model biases allowed observing that two models
 362 underestimated IPP^{is} (i.e. $B < 0$; HYR and MARRA models) while the rest of the models
 363 overestimated IPP^{is} (i.e. $B > 0$). Furthermore, IPP^{sat} estimated from VGPM1 and MARRA
 364 models provided the closest to IPP^{is} ($B = 0.07$ and $B = -0.09$, respectively; Table 4).
 365 RMSD showed significant variability among the different models ranging from 0.28
 366 (ESQRT) to 0.52 (VGPM33) (Table 4). On contrary, uRMSD did not show significant
 367 variability among the different models ranging from 0.26 to 0.30 (Table 4). All models
 368 showed relatively poor agreement with IPP^{is} with R^2 ranging from 0.18 to 0.45 (Table 4).

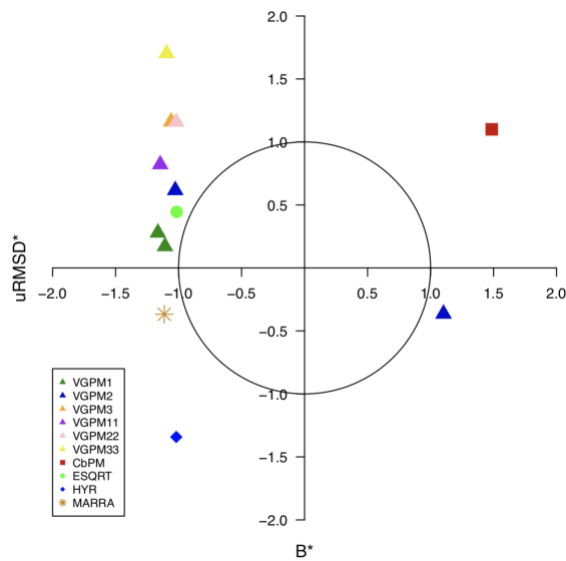
Model	N	intercept	slope	R ²	RMSD	B	uRMSD
VGPM1	97	143.79 ± 30.26	0.58 ± 0.11	0.22	0.31	0.07	0.30
VGPM2	97	51.70 ± 42.40	1.28 ± 0.15	0.42	0.31	0.16	0.26
VGPM3	90	269.61 ± 28.47	0.68 ± 0.10	0.31	0.40	0.30	0.27
VGPM11	97	255.35 ± 28.89	0.48 ± 0.10	0.18	0.36	0.21	0.29
VGPM22	97	178.59 ± 37.92	1.21 ± 0.14	0.45	0.40	0.30	0.26
VGPM33	90	457.10 ± 22.01	0.52 ± 0.08	0.31	0.52	0.44	0.28
ESQRT	97	169.36 ± 21.40	0.51 ± 0.08	0.32	0.28	0.11	0.26
HYR	85	38.38 ± 12.51	0.29 ± 0.04	0.34	0.43	-0.34	0.26
MARRA	97	-47.72 ± 39.67	1.14 ± 0.14	0.40	0.30	-0.09	0.28
CbPM	78	204.22 ± 62.03	1.20 ± 0.22	0.27	0.48	0.28	0.38

369

370 **Table 4** – Number of match-ups, linear regression parameters (intercept, slope and R²;
371 Fig. S2), RMSD, B and uRMSD for each participating model relative to IPP^{is}.

372

373 The target diagram (Fig. 3) illustrates overestimation of observed productivity (B*
374 > 0) for all models except for MARRA and HYR. All models, except CbPM and MARRA
375 models, underestimated the variance of observed productivity (uRMSD* < 0) (Fig. 3).



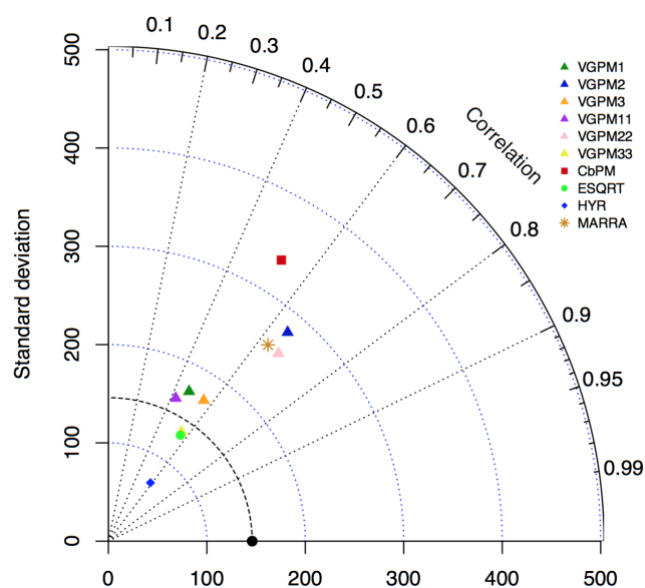
376

377 **Figure 3.** Target diagram displaying B* (Eq. 4) and uRMSD* (Eq. 5) for the 5 models
378 and VGPM variants relative to the IPP^{is}. The solid circle is the normalized standard
379 deviation of the IPP^{is}.

380

381 Although the target diagram gave information about uRMSD for each model, it does not
382 allow assessing whether a given uRMSD results from getting the correlation or the

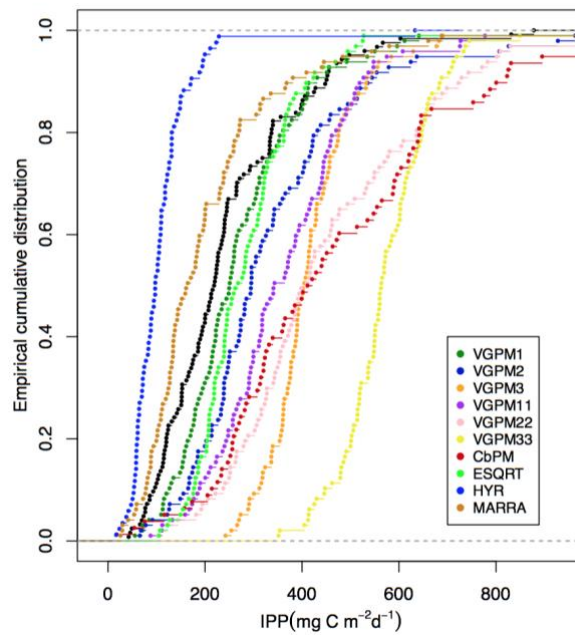
383 variability wrong. The Taylor diagram gives additional information about the variance of
 384 IPP^{sat} (the distance from the origin is the standard deviation) and the correlation between
 385 IPP^{sat} and IPP^{is} (the azimuth angle) (Fig. 4). Correlation coefficients between modelled
 386 and observation estimates ranged between 0.42 and 0.67. Model standard deviations
 387 ranged from $< 100 \text{ mg C m}^{-2} \text{ d}^{-1}$ (HYR model) to $> 300 \text{ mg C m}^{-2} \text{ d}^{-1}$ (CbPM) (Fig. 4).
 388 As the target diagram showed that none of the present models estimated IPP more
 389 accurately than using the mean of the observed data (Fig. 3), the Taylor diagram showed
 390 that VGPM1, VGPM3, VGPM11, VGPM33 and ESQRT were better at reproducing the
 391 magnitude of IPP^{is} variance (i.e. closer to the standard deviation of IPP^{is} data) than the
 392 other models (Fig. 4). Furthermore, the Taylor diagram showed that the models with the
 393 highest correlations did not reproduce well the variability in IPP^{is} (VGPM2, VGPM22,
 394 MARRA; Fig. 4).



395
 396 **Figure 4.** Taylor diagram of IPP. The black dot represents IPP^{is} data. Blue dashed lines
 397 represent arcs along the standard deviation axes and the black dashed line represents the
 398 standard deviation of IPP^{is} . Symbols falling close to the black dashed line indicate the
 399 best models at reproducing the magnitude of IPP^{is} variance.

400

401 The empirical cumulative distribution function (ECDF) illustrates the range of PP
402 from observed data and from the different models (Fig. 5). Here, we observed that
403 MARRA, VGPM1 and ESQRT models reproduced accurately the range of IPP from 300
404 – 400 mg C m⁻² d⁻¹. Below 300 – 400 mg C m⁻² d⁻¹, MARRA model tended to
405 underestimate the range of IPP and VGPM1 and ESQRT models tended to overestimate
406 the range of IPP (Fig. 5).

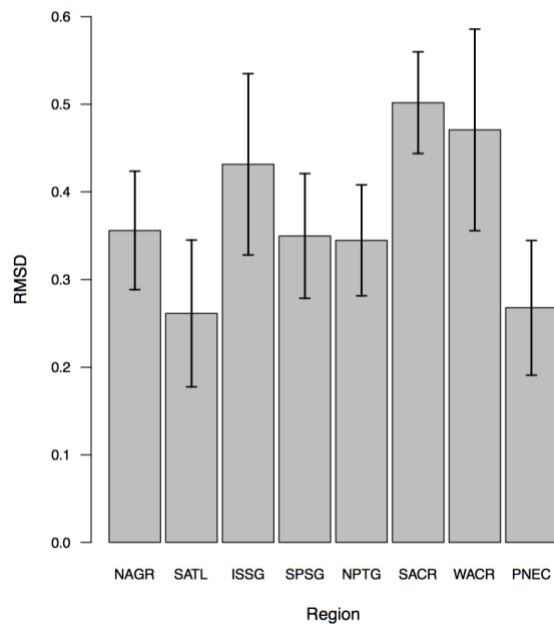


407
408 **Figure 5.** Empirical cumulative distribution function of IPP for the seven models and the
409 observed data (black symbols).

410
411 *3.5. Region-specific model performance*

412 The average performance of the five models and VGPM variants tested here
413 varied across regions, with RMSD varying from 0.26 at SATL to 0.50 at SACR (Fig. 6).
414 The average model performance was significantly lower at SATL (RMSD = 0.26) and
415 PNEC (RMSD = 0.27) than at ISSG (RMSD = 0.43), SACR (RMSD = 0.50) and WACR
416 (RMSD = 0.47) (t-test, P < 0.05). At SACR, the average model performance was

417 significantly higher than for the rest of the regions (t-test, $P < 0.05$), except WACR (t-
418 test, $P > 0.05$).

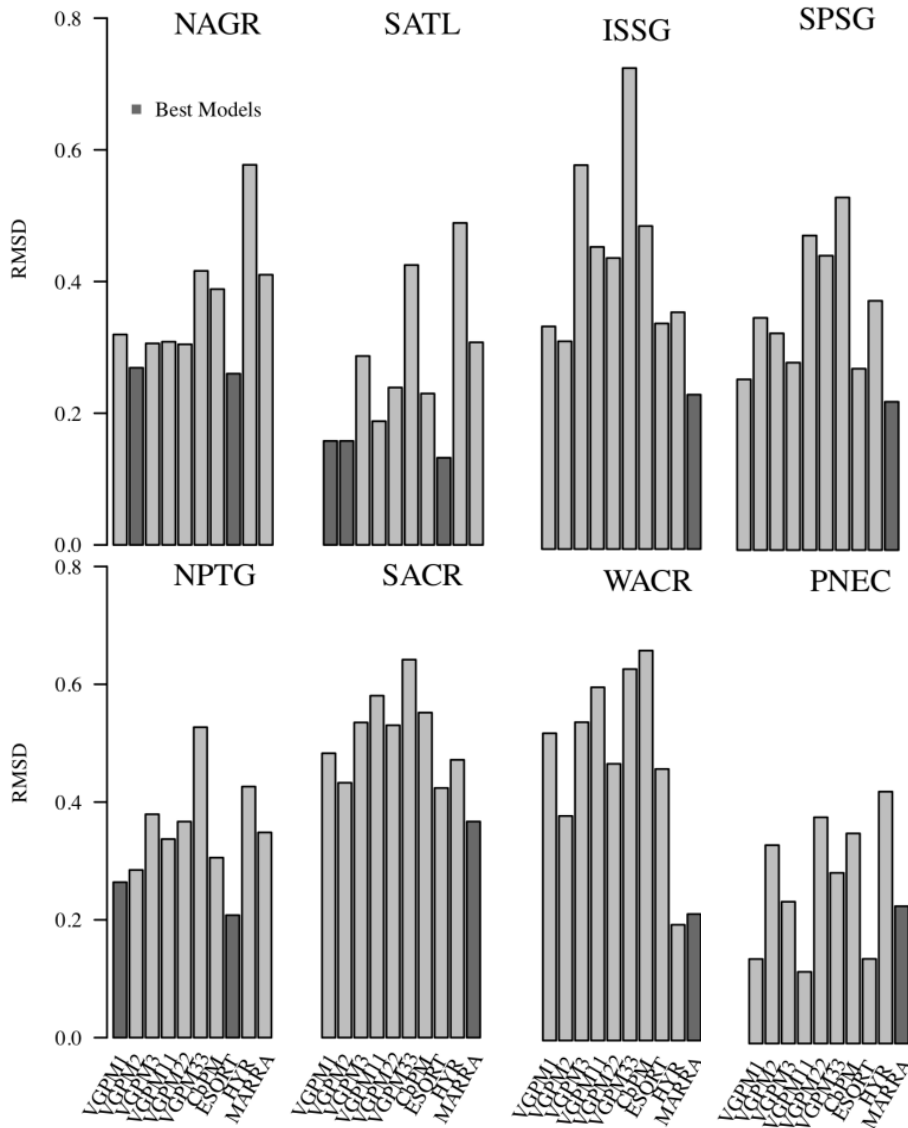


419

420 **Figure 6.** Average RMSD for all 5 models and VGPM variants at each region. The error
421 bars are 2x standard error.

422

423 Considering individual model skill, we observed that some models performed better than
424 others in specific regions (Fig. 7). In four regions (NAGR, SATL, NPTG, PNEC), the
425 ESQRT model showed the lowest RMSD and in the other four regions (ISSG, SPSG,
426 SACR, WACR), the model that mainly showed the lowest RMSD was the MARRA
427 model (Fig. 7).



428

429 **Figure 7.** Model RMSD for each model at each region. Dark grey bars indicate models
 430 with better performance.

431

432 3.6. PP model's adjustments

433 We further explored each model performance by replacing the variables derived
 434 by remotely sensing (i.e. chl-a, SST, Z_{eu} , P^B_{opt} , MLD, Z_{NO3}) for our *in situ* data. From the
 435 five models tested here, we observed that when run with *in situ* values, model
 436 performance was improved only for two models: VGPM (RMSD = 0.18 and B = 0.06)
 437 and MARRA (B = -0.01) (Table 5). For the rest of PP_r models, the average performance
 438 did not show significant improvement (Table 5). Regionally, the improvement was not

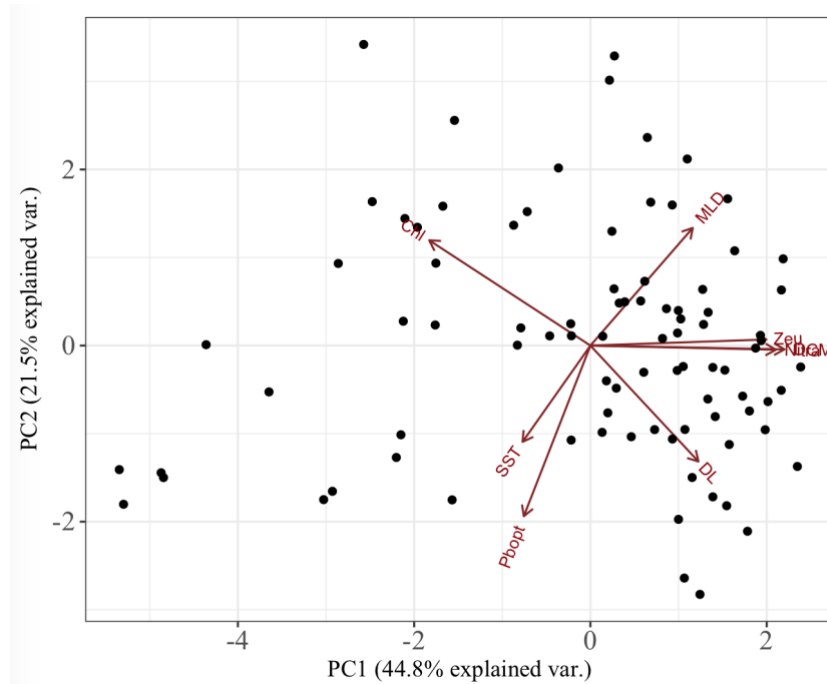
439 significant for all models (data not shown here). As explained previously, satellite-
 440 derived P_{opt}^B presented the weakest agreement with observed P_{opt}^B data and might partly
 441 cause the poor agreement between IPP^{sat} from VGPM and IPP^{is} . Hence, adjusting
 442 parameterized P_{opt}^B to match *in situ* P_{opt}^B data improved VGPM performance.

443 **Table 5** – Pearson correlation coefficient (r), statistics of linear regression (R^2), RMSD,
 444 B, uRMSD and N values for the five PP_r models tested here with *in situ* variables and for
 445 VGPM1 (here called VGPM1') and VGPM11 (here called VGPM11') using modelled
 446 P_{opt}^B (Eq. 8).

Model	r	R^2	RMSD	B	uRMSD	N
VGPMis	0.73	0.53	0.18	0.06	0.17	86
CbPMis	0.39	0.07	0.61	0.45	0.42	78
ESQRTis	0.51	0.26	0.28	0.15	0.24	120
HYRis	0.40	0.16	0.39	-0.28	0.27	110
MARRAis	0.60	0.36	0.28	-0.01	0.28	110
VGPM1'	0.63	0.39	0.25	-0.0008	0.25	83
VGPM11'	0.66	0.44	0.28	0.14	0.24	83

447
 448 Although previous studies (e.g. Friedrichs et al., 2009, Milutinovic et al. 2009, Jacox et
 449 al. 2013) tried to improve the P_{opt}^B estimate, our approach for improving the P_{opt}^B estimate
 450 involved testing first the possible correlations between *in situ* P_{opt}^B and the other *in situ*
 451 variables using a principal component analysis (PCA) and then, guided from the PCA
 452 results, formulating P_{opt}^B as a function of the variables with the highest correlation with
 453 P_{opt}^B . From PCA results (Fig. 8), we observed that P_{opt}^B had a strong positive correlation
 454 with SST and a strong negative correlation with MLD. Then, using multiple least-square
 455 regression, we estimated *in situ* P_{opt}^B as a function of *in situ* MLD and SST ($R^2 = 0.26$, P
 456 < 0.0001) through the fitted regression equation:

$$457 \quad P_{opt}^B = 10^{(1.2264 * \log_{10}(SST) - 0.5626 * \log_{10}(MLD) + 0.22812)} \quad (8)$$



458

459 **Figure 8.** Biplot of *in situ* parameters (SST, sea surface temperature; Chl, chl-*a*; MLD,
 460 mixed layer depth; Zeu, Z_{eu} ; DCM, deep chlorophyll maximum; DL, daylength; Nitra,
 461 Z_{NO3} ; P_{bopt} , P_{opt}^B).

462

463 Then we evaluated the VGPM performance (i.e. VGPM1) using this modelled P_{opt}^B (Eq.
 464 8). We observed that the VGPM had $RMSD = 0.25$ and $B = -0.0008$ (Table 5). We
 465 observed also that its normalized standard deviation was lower than the normalized
 466 standard deviation of the IPP^{is} , meaning that this model estimated IPP more accurately
 467 than using the mean of the observed data (data not shown here).

468

469 **4. Discussion**

470 Here, we compared here about 100 IPP^{is} with IPP^{sat} values derived from five of
 471 the most commonly used PP_r models for the four subtropical gyre regions sampled,
 472 including the Indian subtropical gyre region. IPP^{is} results presented here were consistent
 473 as the methodology was coherent and consistent from the first to the last transect allowing
 474 to dissipate any uncertainties about model prediction variations resulting from the

475 methodology. This comparison allowed us to estimate model performances and explore
476 pathways to improve them. From the five models and variants tested here, we observed
477 that most of them did not derive a good representation of the IPP^{is} variability. Only
478 VGPM1, MARRA and ESQRT models were better at estimating IPP^{is} with IPP^{sat} closer
479 to IPP^{is} ($|B| \sim 0.09$) than for the other models ($|B| \sim 0.29$).

480 Although HYR model had been used for many years as a standard MODIS
481 algorithm, we observed that it showed low performance to predict IPP^{is}. The original
482 HYR model is extended to MLD and in general, MLD was less than Z_{eu} in the studied
483 regions (Table 2). Thus, we estimated the performance of HYR model extended to Z_{eu}
484 (data not shown here) and we observed that B and RMSD were lower for HYR model
485 extended to Z_{eu} (using Z_{eu1} , $B = -0.09$ and $RMSD = 0.30$; using Z_{eu2} , $B = 0.04$ and $RMSD$
486 $= 0.29$) than for HYR model extended for MLD. Hence, HYR models extended to Z_{eu}
487 derived IPP^{sat} values closer to IPP^{is} than the original in subtropical gyre regions where
488 MLD is shallower. However, a Taylor diagram revealed that they were not better at
489 reproducing the magnitude of IPP^{is} variance than the original HYR model that (data not
490 shown here), and that they had low correlation with IPP^{is}.

491 To understand the limitations of the models used here to estimate IPP^{is} accurately,
492 we examined whether these limitations were caused by the input data variables for IPP^{sat}
493 or by the model itself. Indeed, PP_r models strongly rely on chl-*a* and $P^{B_{opt}}$ and a weak
494 agreement between satellite-derived chl-*a* and $P^{B_{opt}}$ with *in situ* data may explain the poor
495 model performance. Although satellite-derived and modelled data inputs such as SST,
496 MLD and Z_{NO3} had a relatively good agreement with *in situ* data, chl-*a* and, especially
497 $P^{B_{opt}}$, had poor agreements with *in situ* data ($R^2 = 0.16$ and $0.01 < R^2 < 0.11$, SIQR = 0.23
498 and $0.57 < SIQR < 1.13$, respectively; Table 3). When satellite-derived and modelled data
499 were substituted by *in situ*-derived data (when available), we found that, over the best

500 performing models, the VGPM and MARRA models improved model-data linear
501 regression statistics (R^2) by 68 % and 10 %. Total RMSD declined about the half for
502 VGPM and a large decline was observed in VGPM and MARRA biases (76 % and 86 %
503 respectively; Table 5).

504 Several studies concluded P_{opt}^B to be the IPP model input parameter with the
505 weakest agreement with *in situ* data (Behrenfeld & Falkowski 1997, Behrenfeld et al.
506 2002; Siegel et al. 2001; Milutinović et al. 2011). These studies suggested that P_{opt}^B
507 cannot be derived adequately using only sea surface temperature (SST) as input,
508 considering that light and nutrient availability may have analogous physiological effects
509 on algal photosynthetic capacity and, thus, on P_{opt}^B . Milutinović et al. (2011) suggested
510 that the success of an SST-dependent P_{opt}^B will be variable over time and location. After
511 our parameterization of P_{opt}^B as a function of MLD and SST, we found that using satellite-
512 derived and modelled input data, VGPM had its model-data linear regression statistics
513 (R^2) improved by 32 %, its total RMSD reduced by 31 % and its bias reduced by 72 %
514 (Table 5). In this study, about 80 % of the dataset collected during the MCE was located
515 in subtropical regions where MLD is relatively shallow (< 60 m). We believe that the
516 parametrization of P_{opt}^B from SST and MLD in oceanic ecosystems where MLD is
517 shallower (< 60 m) should improve the estimation of P_{opt}^B for IPP^{sat} and thus, improve
518 PP_r models. Indeed, when modelled P_{opt}^B (Eq. 8) was substituted in the original VGPM1
519 model here, its performance was improved by 100 % ($B < 0.0008$, Table 5). Our results
520 are specific to a circumnavigation that lasted 7 months and cruised all the subtropical
521 oceans by 6 transects with consistent methodology. Although we believe that our results
522 present a good representation of subtropical gyres ecosystems, we are aware that further
523 sampling efforts are required to confirm the improvement of the parametrization of P_{opt}^B
524 using SST and MLD where $MLD < 60$ m.

525 The use of *in situ* variables, especially chl-*a*, in PP_r models improved remote PP
526 estimates and provided a pathway to improve their performance. Obtaining *in situ* chl-*a*
527 data across the oceans is now possible through use of autonomous technologies such as
528 gliders and profiling floats. Part of the Argo International Program (www.argo.ucsd.edu),
529 Bio-Argo aims to contribute to the development of profiling float equipped with bio-
530 optical sensors to measure chl-*a* and backscattering. Starting in 2011, Bio-Argo delivers
531 a series of 5-6 profiling floats on a yearly basis. Although the deployment of floats
532 equipped with bio-optical sensors did not achieve widespread coverage as yet, it is
533 expected that these will deliver a high quantity of long-range and months-long
534 deployments shortly.

535 The majority of MCE stations where IPP^{is} was estimated encompassed subtropical
536 gyre regions with low IPP^{is} (i.e. more than 70 % of $IPP^{is} < 300 \text{ mg C m}^{-2} \text{ d}^{-1}$). Model
537 performances were generally better for high values of IPP (above $500 \text{ mg C m}^{-2} \text{ d}^{-1}$); when
538 model-data misfit was in general lower (Fig. 4). This observation confirms the challenge
539 to predict IPP^{is} in the ultra-oligotrophic regions encompassed by the oligotrophic gyres.
540 We believe that further efforts are required to improve the performance of ocean color
541 models such as VGPM to be applied to highly oligotrophic regions such as subtropical
542 gyres, where IPP is relatively low ($< 300 \text{ mg C m}^{-2} \text{ d}^{-1}$), the MLD is shallower and
543 cloudiness may bias satellite input data. Hence, efforts to improve the algorithms and
544 parameters used in PP_r models (such as those provided in this work) specific for the
545 oligotrophic subtropical gyre regions are essential to further understand the reasons for
546 the poor predictions made by the existing models. The development of improved and
547 robust satellite-based algorithms to predict oceanic primary production in subtropical
548 gyres requires additional efforts to obtain *in situ* estimates of net primary production. This
549 sampling effort is particularly necessary for some of the subtropical gyre regions, like in

550 the three gyres located in the Southern Hemisphere (e.g. Marañón et al. 2000; Poultron et
551 al. 2006; Regaudie-de-Gioux et al. 2012; this study). Because this ocean bioma comprises
552 70 % of the ocean, improving PP_r estimates therein is an imperative to progress toward a
553 global ocean observing system.

554

555 **Acknowledgements**

556 This work is a contribution to the Malaspina Circumnavigation Expedition 2010, funded
557 by the INGENIO 2010 CONSOLIDER program (ref. CDS2008-00077) of the Spanish
558 Ministry of Economy and Competitiveness. We thank the crew of R/V Hespérides for
559 their invaluable support during the expedition, Dr A. López-Urrutia and Dr F. García-
560 García for their help with PP models and satellite data collection.

561 **References**

562 Antoine, D., André, J.M., Morel, A., 1996. Oceanic primary production: II. Estimation at
563 global scale from satellite (Coastal Zone Color Scanner) chlorophyll. *Global*
564 *Biogeochemistry Cycles* 10, 57– 69.

565

566 Bailey, S.W., Werdell, P.J., 2006. **A multi-sensor approach for the on-orbit validation**
567 **of ocean color satellite data products.** *Remote Sensing of Environment*, 102, 12–23

568

569 Barbini R., et al., 2003. Remotely sensed primary production in the western Ross Sea:
570 results of in situ tuned models. *Antarctic Science*,
571 <https://doi.org/10.1017/S095410200300107X>.

572

573 Behrenfeld, M.J., et al., 2005. Carbon-based ocean productivity and phytoplankton
574 physiology from space. *Global Biogeochemistry Cycles*,
575 <https://doi.org/10.1029/2004GB002299>.

576

577 Behrenfeld, M.J., et al., 2002. Photoacclimation and nutrient-based model of light-
578 saturated photosynthesis for quantifying oceanic primary production. *Marine Ecology*
579 *Progress Series* 228, 103–117.

580

581 Behrenfeld, M.J., Falkowski, P.G., 1997b. A consumer's guide to phytoplankton primary
582 productivity models. *Limnology and Oceanography* 42, 1479-1491.

583

584 Behrenfeld, M.J., Falkowski, P.G., 1997. Photosynthetic rates derived from satellite-
585 based chlorophyll concentration. *Limnology and Oceanography*,
586 <https://doi.org/10.4319/lo.1997.42.1.000>.

587

588 Campbell, J., et al., 2002. Comparison of algorithms for estimating ocean primary
589 production from surface chlorophyll, temperature, and irradiance. *Global*
590 *Biogeochemistry Cycles*, <https://doi.org/10.1029/2001GB001444>.

591

592 Duarte, C.M., 2015. Seafaring in the 21st Century: The Malaspina 2010 Circumnavigation
593 Expedition. *Limnology and Oceanography Bulletin*, <https://doi.org/10.1002/lob.10008>.

594

595 Eppley, R.W., 1972. Temperature and phytoplankton growth in the Sea. *Fishery Bulletin*
596 70, 1063–1085.

597

598 Field, C.B., et al., 1998. Primary production of the biosphere: Interacting terrestrial
599 and oceanic components. *Science*, <https://doi.org/10.1126/science.281.5374.237>.

600

601 Friedrichs, M.A.M., et al., 2009. Assessing the uncertainties of model estimates of
602 primary productivity in the tropical Pacific Ocean. *Journal of Marine Systems*,
603 <https://doi.org/10.1016/j.jmarsys.2008.05.010>.

604

605 Garcia, H.E., et al., 2014. Volume 4: Dissolved Inorganic Nutrients (phosphate, nitrate,
606 silicate). In: Levitus, S., Mishonov, A., *World Ocean Atlas 2013*. Technical Ed., NOAA
607 Atlas NESDIS, pp. 76.

608

609 Grasshoff, K., Kremling, K., Ehardt, M., 1999. Methods of seawater analysis. Weinheim:
610 Wiley-VCH.
611
612 Karl, D.M., 1999. A sea of change: biogeochemical variability in the North Pacific
613 subtropical gyre. *Ecosystems* 2, 181–194.
614
615 Key, R.M., Kozyr, A., Sabine, C.L., Lee, K., Wanninkhof, R., Bullister, J., Feely, R.A.,
616 Millero, F., Mordy, C., Peng, T.-H., 2004. A global ocean carbon climatology: Results
617 from GLODAP. *Global Biogeochemical Cycles* 18, GB4031
618
619
620 Jolliff, J.K., et al., 2009. Summary diagrams for coupled hydrodynamic-ecosystem model
621 skill assessment, *Journal of Marine Systems* 76, 64–82.
622
623 Latasa, M., 2014. A simple method to increase sensitivity for RP-HPLC phytoplankton
624 pigment analysis *Limnology and Oceanography: Methods*,
625 <https://doi.org/10.4319/lom.2014.12.46>.
626
627 Lee, D.-K., Niiler, P., Warnas-Varnas, A., 1994. Wind-driven secondary circulation in
628 ocean mesoscale. *Journal of Marine Research* 5, 1–25.
629
630 Lee, Z.P., et al., 1996. Estimating primary production at depth from remote sensing.
631 *Applied Optics*, <https://doi.org/10.1364/AO.35.000463>.
632
633 Longhurst, A., et al., 1995. An estimate of global primary production in the ocean from
634 satellite radiometer data. *Journal of Plankton Research* 17, 1245–1271.

635

636 Longhurst, A., 1995. Seasonal cycles pelagic production and consumption. Progress in
637 Oceanography 36, 77–167.

638

639 Marañón, E., et al., 2003. High variability of primary production in oligotrophic waters
640 of the Atlantic Ocean: uncoupling from phytoplankton biomass and size structure. Marine
641 Ecology Progress Series 257, 1–11.

642

643 Marañón, E., et al., 2000. Basin-scale variability of phytoplankton biomass, production
644 and growth in the Atlantic Ocean. Deep Sea Research Part I,
645 [https://doi.org/10.1016/S0967-0637\(99\)00087-4](https://doi.org/10.1016/S0967-0637(99)00087-4).

646

647 Milutinović, S., Bertino, L., 2011. Assessment and propagation of uncertainties in input
648 terms through an ocean-color-based model of primary productivity. Remote Sensing
649 Environment 115, 1906–1917.

650

651 Mobley, C.D., et al., 2004. Optical modelling of ocean waters: Is the Case 1 - Case 2
652 classification still useful? Oceanography, <https://doi.org/10.5670/oceanog.2004.48>.

653

654 Monterey, G., Levitus, S., 1997. Seasonal variability of mixed layer depth for the world
655 ocean, NOAA Atlas NESDIS, Natl. Oceanic and Atmos. Admin., Silver Spring, Md.

656

657 Morel, A., et al., 2007. Optical properties of the « clearest » natural waters. Limnology
658 and Oceanography 52, 217–229.

659

660 Morel, A., Berthon, J.-F., 1989. Surface pigments, algal biomass profiles, and potential
661 production of the euphotic layer: relationships reinvestigated in view of remote sensing
662 applications. *Limnology and Oceanography* 34, 1545–1562.

663

664 Moreno-Ostos, E., 2012. Expedición de circunnavegación Malaspina 2010: Cambio
665 global y exploración de la biodiversidad del océano. Libro blanco de métodos y técnicas
666 de trabajo oceanográfico. Madrid, CSIC.

667

668 Palmer, S.C.J., et al., 2015. satellite remote sensing of phytoplankton phenology in Lake
669 Balaton using 10 years of MERIS observations. *Remote Sensing Environment* 158, 441–
670 452.

671

672 Pinedo-González, P., et al., 2015. Surface distribution of dissolved trace metals in the
673 oligotrophic ocean and their influence on phytoplankton biomass and productivity.
674 *Global Biogeochemical Cycles*, <https://doi.org/10.1002/2015GB005149>.

675

676 Platt, T., Sathyendranath, S., 1988. Oceanic primary production: Estimation by remote
677 sensing at local and regional scales. *Science* 241, 1613–1620.

678

679 Poultron, A.J., et al., 2006. Phytoplankton carbon fixation, chlorophyll-biomass and
680 diagnostic pigments in the Atlantic Ocean. *Deep-Sea Research Part II* 53, 1593–1610.

681

682 Regaudie-de-Gioux, A., Huete-Ortega, M., Sobrino, C., López-Sandoval, D.C.,
683 González, N., Fernández-Carrera, A., Vidal, M., Marañón, E., Cermeño, P., Latasa, M.,

684 Agustí, S., Duarte, C.M. 2019. Data set on xxxx [titulo]. [Pange.de](https://pange.de), doi:xxx [the data set
685 will be made openly available in this repository upon acceptance of the paper]
686

687 Regaudie-de-Gioux, A., Duarte, C.M., 2013. Global patterns in oceanic planktonic
688 metabolism. *Limnology and Oceanography* 58, 977–986.
689

690 Rousseaux, C.S., Gregg, W.W., 2015. Recent decadal trends in global phytoplankton
691 composition. *Global Biogeochemical Cycles* 29, 1674–1688.
692

693 Rousseaux, C.S., Hirata, T., Gregg, W.W., 2013. Satellite views of global phytoplankton
694 community distributions using empirical algorithm and a numerical model.
695 *Biogeosciences Discussion* 10.
696

697 Saba, V.S., et al., 2011. An evaluation of ocean color model estimates of marine primary
698 productivity in coastal and pelagic regions across the globe. *Biogeosciences*,
699 <https://doi.org/10.5194/bg-8-489-2011>.
700

701 Saulquin, B., et al., 2013. Estimation of the diffuse attenuation coefficient K_{dPAR} using
702 MERIS and application to seabed habitat mapping. *Remote Sensing Environment* 128,
703 224–233.
704

705 Siegel, D.A., et al., 2001. Bio-optical modeling of primary production on regional scale:
706 the Bermuda BioOptics project. *Deep-Sea Research Part II* 48, 1865–1896.
707

708 Tilstone, G., et al., 2009. Measured and remotely sensed estimates of primary production
709 in the Atlantic Ocean from 1998 to 2005. *Deep Sea Research Part II*,
710 <https://doi.org/10.1016/j.dsr2.2008.10.034>.

711

712 Volk, T., Broecker, W.S., 1985. Ocean carbon pumps: analysis of relative strengths and
713 efficiencies in ocean-driven atmospheric CO₂ changes. In: Walker, J.C.G, *The carbon*
714 *cycle and atmospheric CO₂: natural variations Archean to present*. American Geophysical
715 Union, pp. 99-110.

716

717 Westberry, T., et al., 2008. Carbon-based primary productivity modeling with vertically
718 resolved photophysiology, *Global Biogeochemical Cycles*, [https://doi.org/10.1029/](https://doi.org/10.1029/2007GB003078)
719 [2007GB003078](https://doi.org/10.1029/2007GB003078).

720

721 Zapata, M., Rodríguez, F., Garrido, J., 2000. Separation of chlorophylls and carotenoids
722 from marine phytoplankton: a new HPLC method using a reversed phase C8 column and
723 pyridine-containing mobile. *Marine Ecology Progress Series* 195, 29–45.

724

725

726

727

728

729

730

731

732

733

734

735

736

737

738

739

740

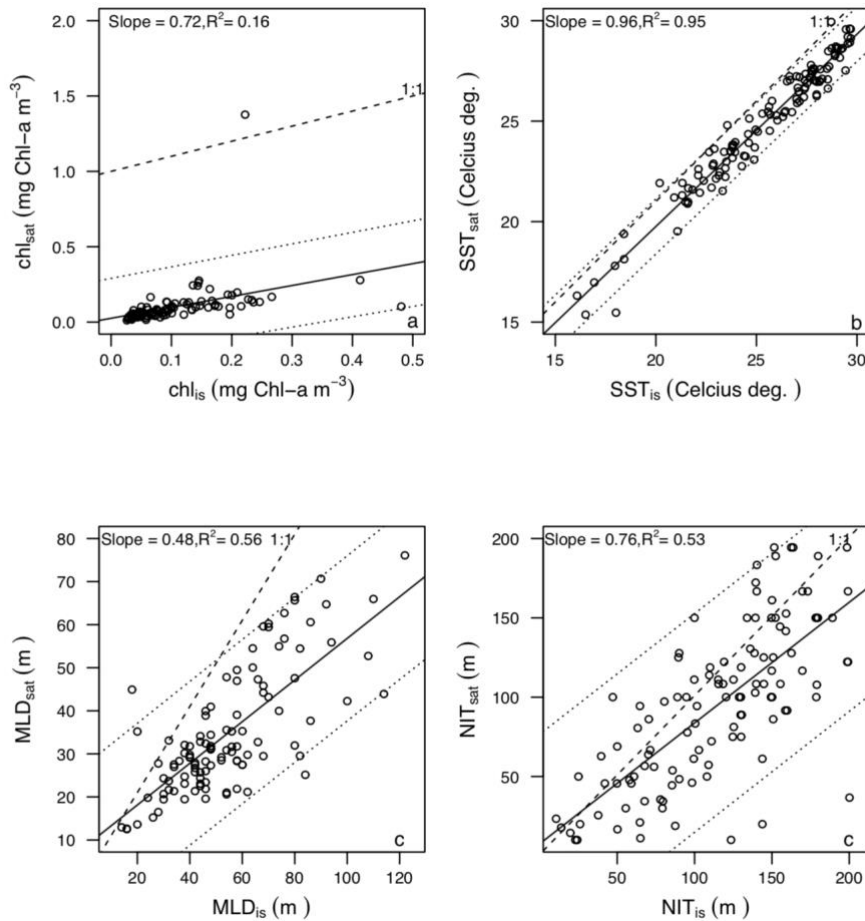
741

742

743 **Supplement Material**

744

745 Figure S1. Multi-paneled regression plots of satellite-derived vs. *in situ* data (a:
746 chlorophyll-a concentration; b: SST; c: MLD; d: nitracline depth). Solid lines represent
747 the linear regression. Dotted lines represent the 95 % confident intervals and the dashed
748 lines represent the 1:1 lines.
749



750

751

752

753

754

755

756

757

758

759

760

761

762

763

764

765

766

767

768 Figure S2. Multi-paneled regression plots of (a) VGPM1 *vs.* IPP^{is}, (b) VGPM2 *vs.* IPP^{is},
769 (c) VGPM3 *vs.* IPP^{is}, (d) VGPM11 *vs.* IPP^{is}, (e) VGPM22 *vs.* IPP^{is}, (f) VGPM33 *vs.*
770 IPP^{is}, (g) ESQRT *vs.* IPP^{is}, (h) HYR *vs.* IPP^{is}, (i) MARRA *vs.* IPP^{is}, (j) CbPM *vs.* IPP^{is}.
771 Solid lines represent the linear regressions. Dotted lines represent the 95 % confident
772 intervals and the dashed lines represent the 1:1 lines.
773

