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

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

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

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

[3]Oroboros Instruments, Schöpfstraße 8, 6020, Innsbruck, Austria

[4]Departamento de Ecología y Biología Animal, Universidade de Vigo, 36310 Vigo, Spain

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

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

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

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

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

*corresponding author: aregaudi@ifremer.

Abstract

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

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

1. Introduction

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

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

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

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

2. Methods

2.1. Study area

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

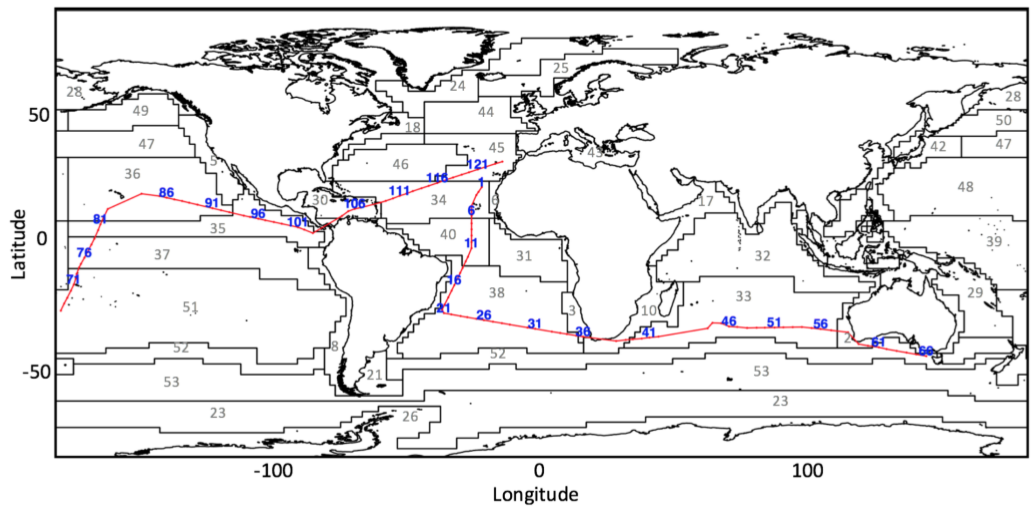


Figure 1. Location of the sampled station during the MCE and Longhurst biogeochemical ocean provinces (Lonhgurst, 1998; 2006). The provinces where IPP^{is} were sampled were: (2) Australia-Indonesia Coastal Province, AUSW; (3) Benguela Current Coastal Province, BENG; (30) Caribbean Province, CARB; (10)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

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

2.2. In situ measurement of chlorophyll-a

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

2.3. In situ mixed layer, nitracline and euphotic depths

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

The nitracline was determined from nitrate plus nitrite concentration data, measured on a segmented flow Skalar auto-analyser by standard methods (Grasshoff et al., 1999, Moreno-Ostos, 2012), as the depth, from the surface, where the first sustained increase of the concentration gradient is observed. It is a region of approximately maximum and steady concentration gradient in the first 200 m of the vertical profile, including 4 – 6 nitrate data points.

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

2.5. In situ measurement of primary production

Phytoplankton primary production was measured at 124 stations with the ^{14}C-uptake technique, following the procedures detailed in Marañón et al. (2000). Seawater was sampled from five depths in the euphotic zone corresponding to 100 % (ca. 3 m depth), 50 %, 20 %, 7 % and 1 % of incident Photosynthetically Active Radiation (PAR). For each depth, four 72 mL polystyrene bottles (three clear bottles and one dark bottle) were filled with unfiltered seawater, inoculated with 10 – 20 µCi $NaH^{14}CO_3$ and incubated on-deck from dawn to dusk. Temperature and irradiance in the incubators simulated the water temperature and the incident irradiance at the corresponding depth of each sample by using a combination of neutral density and blue filters (Mist Blue, ref. 061, Lee Filters ®). After incubation, samples from three of the five depths (100 %, 20 % and 1 % PAR) were sequentially filtered through 20, 2 and 0.2 µm polycarbonate filters while the other depths (50 % and 7 % PAR) were directly filtered by 0.2 µm. Immediately after filtering, filters were then exposed to concentrated HCl fumes at least 12 h to remove the non-fixed inorganic ^{14}C. Filters were placed in scintillation vials to which 5 mL of liquid scintillation cocktail was added. The radioactivity on each filter (disintegrations per minute, DPM) was determined using a Wallac scintillation counter. To compute the rate of photosynthetic carbon fixation, the dark-bottle DPM was subtracted from the light-bottle DPM values. A constant value of 24,720 µg L^{-1} (or 2,060 µmol L^{-1}) was assumed for the concentration of dissolved inorganic carbon for surface waters in tropical ocean (Key et al., 2004). A correction factor of 1.05 was applied to this constant value for discrimination isotopic. Total primary production was calculated as the sum of the primary production on each size class. For all triplicate measurements of total primary production conducted during MCE (n = 522), the mean coefficient of variation was 23 %. *In situ* depth-integrated primary production (IPP^{is}, mg C m^{-2} d^{-1}) was calculated by the trapezoidal integration of measurements from the surface to 1 % PAR depth. The IPP^{is} data set is available from Regaudie-de-Gioux et al. 2019. The original IPP^{is} measurements were reported as hourly rates and then were converted to daily rates multiplying the hourly rates by the corresponding day length at each sampled station. The highest hourly chl-*a*-specific primary production (P^B, mg C chl-a^{-1} h^{-1}) in the water column was defined as the observed *in situ* P^B_{opt} (Behrenfeld & Falkowski 1997b) for each station in this study. The variability in IPP^{is} along MCE transects is described in Pinedo-González et al. (2015).

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

Ocean color models typically use Level-3, monthly or 8-day, satellite-derived input data. In this study, daily standard level 3 (i.e. mapped processed to surface quantities) products of PAR, ocean color index (OCI)-based chl-*a*, diffuse attenuation at 490 nm ($K_d(490)$), sea surface temperature (SST) and particulate backscatter coefficient at 443 nm ($b_{bp}(443)$ from GSM model) were provided by the OceanColor Web (https://oceancolor.gsfc.nasa.gov) and were calculated from the Moderate Resolution Imaging Spectroradiometer aboard the Aqua NASA spacecraft (MODISA). Spatial resolution of all products was ~ 9 km at the Equator. Additionally, the mixed layer depth (MLD) and the nitracline depths (Z_{NO3}) were modelled variables for IPP^{sat}. The daily product of MLD was provided by the Ocean Productivity Online Data (https://orca.science.oregonstate.edu). Z_{NO3} were calculated from monthly climatological nutrient fields reported in the World Ocean Atlas 2013 (Garcia et al., 2014) at 1-degree resolution, and defined as the first depth at which nitrate + nitrite exceeded 0.5 µM. All variables were extracted from 1 pixel radius windows (i.e. 3 x 3-pixel box) centered at the pixel nearest to the *in situ* sample and we calculated the average of each window (Bailey & Werdell 2006). Satellite variables were excluded when more than 70 % were masked. We used this matchup criteria to increase the number of matchups, particularly in subtropical areas where cloudiness is important increasing satellite masked data.

2.7. Satellite algorithms

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

Model	Remote variable input							P^B_{opt} *	Z_{eu} **	Reference
	chl-a	SST	PAR	MLD	K_d (490)	Z_{NO3}	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) Berhenfeld and Falkowski (1997); 2) Antoine and Morel (1996); 3) Kameda and Ishizaka (2005)

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

Table 1 – Model descriptions

First, we used the most widely utilized PP_r model, the Vertically Generalized Production Model (VGPM) based on chl-*a* (Behrenfeld & Falkowski 1997). The VGPM uses remote inputs of chl-*a*, SST and PAR. Here we proposed several VGPM variants and alternative methods to estimate P^B_{opt}. The first VGPM variant (here called VGPM1) is the original VGPM as in Behrenfeld & Falkowski 1997 where the P^B_{opt} was obtained from a 7th-order polynomial SST regression (here called $P^B_{opt}1$). For the VGPM2 variant, P^B_{opt} was estimated after Eppley (1972) as implemented by Antoine and Morel (1996) as an exponential function of temperature (here called $P^B_{opt}2$). In the VGPM3 variant, P^B_{opt} was estimated after Kameda and Ishizaka (2005) as inversely proportional to phytoplankton size (here called $P^B_{opt}3$). Additionally, we modified these three VGPM variants with an alternative method to estimate Z_{eu}, originally estimated from chl-*a* concentration (Morel & Berthon, 1989). The three extra variants of the VGPM models cited above (called hereafter, VGPM11, VGPM22 and VGPM33) included a modified Z_{eu} estimated from the diffuse attenuation coefficient of PAR (m^{-1}) (following Mobley, 2004).

Additionally, we used the simplest PP_r model, the Eppley-Square-Root model (ESQRT; Eppley et al., 1985). The ESQRT model uses only chl-*a* as remote inputs assuming that the standing stock is the sole determinant of photosynthetic rate. It ignores any external forcing or changes in physiological state. We also used the original Howard–Yoder–Ryan model (HYR; Howard & Yoder 1997) which for many years was used as a standard MODIS algorithm. The maximum growth rate is parameterized here as a function of SST according to Eppley (1972). PP_r is integrated here to the MLD rather than the euphotic depth. The HYR model uses remote inputs as chl-*a*, SST, PAR and MLD. Furthermore, we used the PPr model described by Marra et al. (2003) (MARRA) that is based on chlorophyll-specific absorption parameterized by SST and maximum quantum yield. The MARRA model uses chl-*a*, SST and PAR as remote inputs. Finally, we used the Carbon-based Production Model (CbPM; Westberry et al. 2008). The CbPM uses remote inputs as chl-*a*, $b_{bp}(443)$, PAR, $K_d(490)$, MLD and Z_{NO3}. The CbPM utilizes $b_{bp}(443)$ to derive phytoplankton carbon biomass.

2.8. Model validation

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

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

where $\Delta(i) = \log_{10}[IPP^{sat}(i)]-\log_{10}[IPP^{is}(i)]$ and N is the total number of paired data. The model performance and predictive skills increase as RMSD values become closer to 0. RMSD captures a model's ability to represent both the mean and the variability of *in situ* data and thus, is composed by the bias (i.e. the difference between the means, B) and the unbiased RMSD (i.e. representing the difference of variability, uRMSD):

$$\mathrm{RMSD}^2 = B^2 + \mathrm{uRMSD}^2 \quad (2)$$

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

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

Target diagram (Jolliff et al. 2009) will be used to illustrate model performances. This diagram allows visualizing bias, uRMSD, and total RMSD for all models on a single plot. For that, these quantities are normalized by the standard deviation of $\log_{10}$ IPP^{is} (σ_d = 0.26):

$$B^* = {}^{B}/_{\sigma_d} \tag{4}$$

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

where σ_m is the standard deviation of $\log_{10}$ IPP^{sat}.

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

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

2.9. Uncertainty analysis

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

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

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

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

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

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

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

3. Results

3.1. In situ data

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

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

	IPP^{is}	**chl-*a***	**Zeu**	**P^{B}_{opt}**	**SST**	**MLD**	**Nitracline**
	mg C m^{-2} d^{-1}	mg m^{-3}	m	mg C mg chl-*a*$^{-1}$ h^{-1}	ºC	m	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)
total	244.8 (± 146.7)	0.11 (± 0.08)	113 (± 28)	4.20 (± 2.75)	25.4 (± 3.2)	54 (± 22)	114 (± 48)

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

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

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

	R^2	Slope	r	N	Median	SIQR	MPD
Chl-*a*	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
MLD	0.56	0.48 ± 0.04	0.75	107	0.65	0.12	35.03

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

3.3. Model phytoplankton physiology variable

The three P^B_{opt} variables modelled from satellite-derived data ($P^B_{opt}1$, $P^B_{opt}2$ and $P^B_{opt}3$) presented weak agreements with observed P^B_{opt} data ($R^2 < 0.12$) and showed the highest spreading values (SIQR = 0.57, 0.65 and 1.13, respectively). These P^B_{opt} data showed the highest overall bias in comparison with *in situ* inputs (median $P^{Bopt}1 = 1.49$, median $P^{Bopt}2 = 1.80$ and median $P^{Bopt}3 = 2.69$).

While $P^B_{opt}1$ and $P^B_{opt}2$ were described as function of SST, $P^B_{opt}3$ and *in situ* P^B_{opt} did not follow any correlation with SST (Fig. 2). $P^B_{opt}3$ presented a wider value range (from 5 to > 30 mg C mg Chl-a^{-1} d^{-1}) than *in situ* P^B_{opt} (from 0.14 to < 15 mg C mg Chl-a^{-1} d^{-1}) between 21 and 28 ºC.

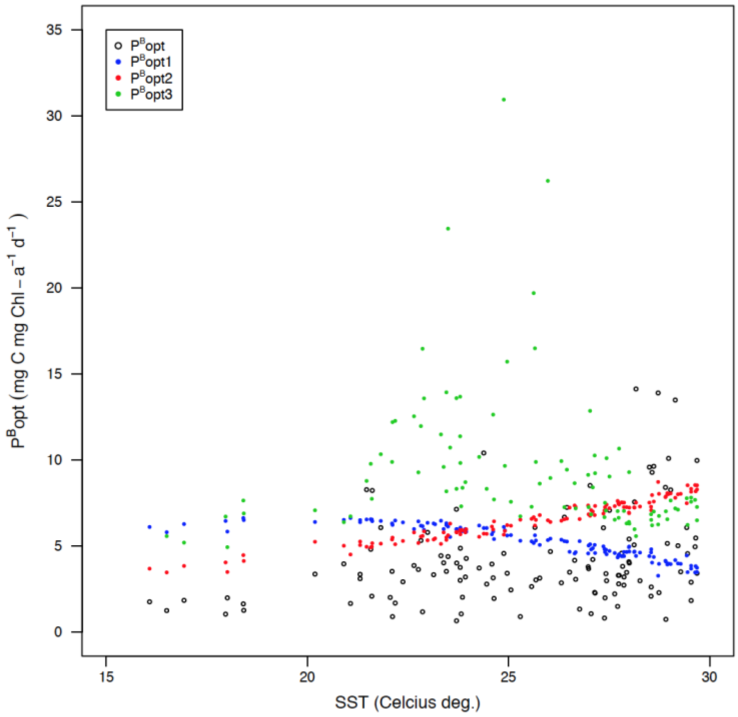


Figure 2. Representation of the three $P^B{}_{opt}$ algorithms used in this study and *in situ* $P^B{}_{opt}$ in function of *in situ* SST.

3.4. Model performance across all regions

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

Table 4 – Number of match-ups, linear regression parameters (intercept, slope and R^2; Fig. S2), RMSD, B and uRMSD for each participating model relative to IPP^{is}.

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

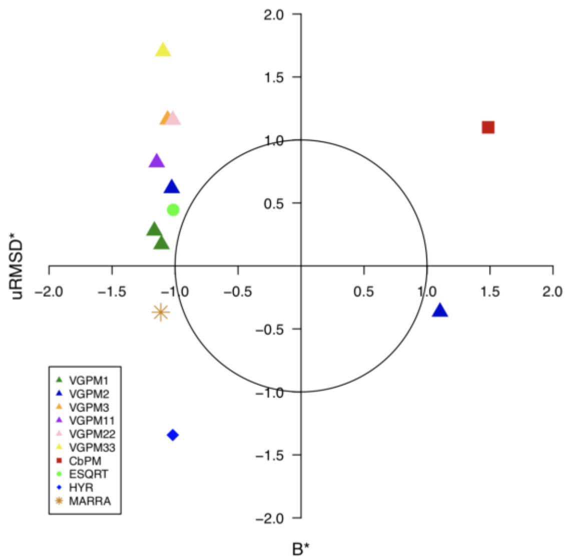


Figure 3. Target diagram displaying B* (Eq. 4) and uRMSD* (Eq. 5) for the 5 models and VGPM variants relative to the IPP^{is}. The solid circle is the normalized standard deviation of the IPP^{is}.

Although the target diagram gave information about uRMSD for each model, it does not allow assessing whether a given uRMSD results from getting the correlation or the variability wrong. The Taylor diagram gives additional information about the variance of IPP^{sat} (the distance from the origin is the standard deviation) and the correlation between IPP^{sat} and IPP^{is} (the azimuth angle) (Fig. 4). Correlation coefficients between modelled and observation estimates ranged between 0.42 and 0.67. Model standard deviations ranged from < 100 mg C m^{-2} d^{-1} (HYR model) to > 300 mg C m^{-2} d^{-1} (CbPM) (Fig. 4). As the target diagram showed that none of the present models estimated IPP more accurately than using the mean of the observed data (Fig. 3), the Taylor diagram showed that VGPM1, VGPM3, VGPM11, VGPM33 and ESQRT were better at reproducing the magnitude of IPP^{is} variance (i.e. closer to the standard deviation of IPP^{is} data) than the other models (Fig. 4). Furthermore, the Taylor diagram showed that the models with the highest correlations did not reproduce well the variability in IPP^{is} (VGPM2, VGPM22, MARRA; Fig. 4).

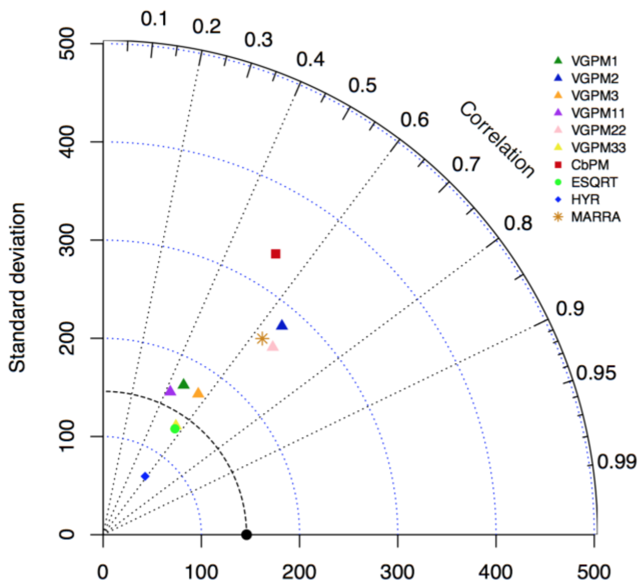


Figure 4. Taylor diagram of IPP. The black dot represents IPP^{is} data. Blue dashed lines represent arcs along the standard deviation axes and the black dashed line represents the standard deviation of IPP^{is}. Symbols falling close to the black dashed line indicate the best models at reproducing the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

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

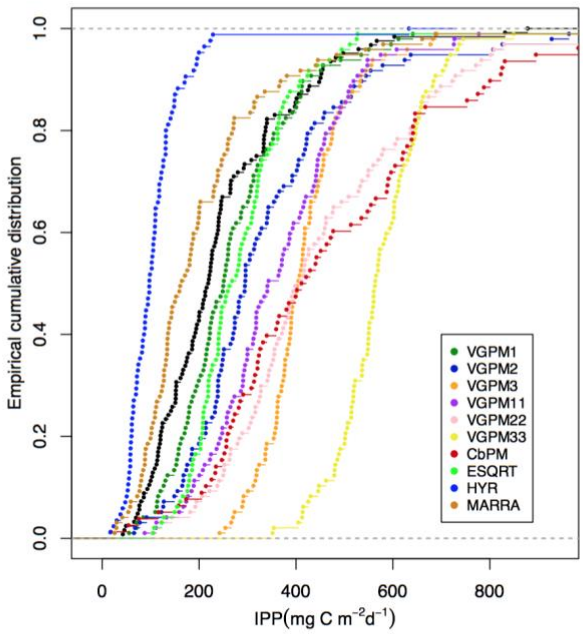


Figure 5. Empirical cumulative distribution function of IPP for the seven models and the observed data (black symbols).

3.5. Region-specific model performance

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

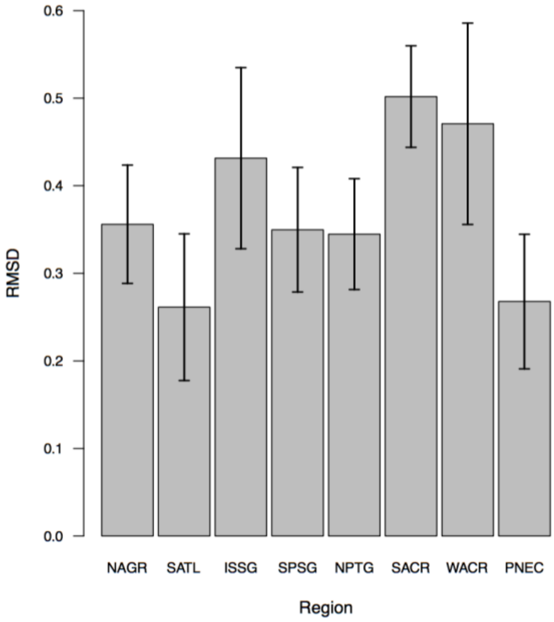


Figure 6. Average RMSD for all 5 models and VGPM variants at each region. The error bars are 2x standard error.

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

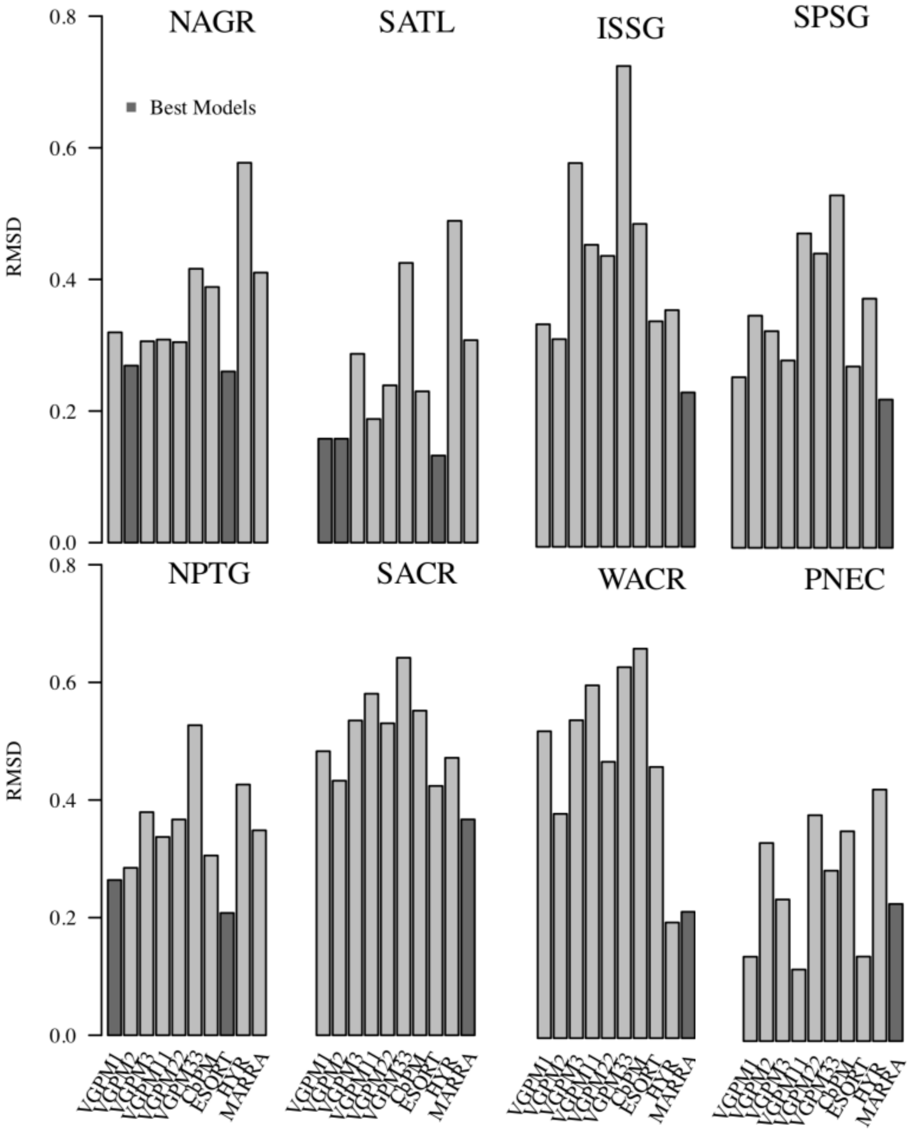


Figure 7. Model RMSD for each model at each region. Dark grey bars indicate models with better performance.

3.6. PP model's adjustments

We further explored each model performance by replacing the variables derived by remotely sensing (i.e. chl-a, SST, Z_{eu}, P^{B}_{opt}, MLD, Z_{NO3}) for our *in situ* data. From the five models tested here, we observed that when run with *in situ* values, model performance was improved only for two models: VGPM (RMSD = 0.18 and B = 0.06) and MARRA (B = -0.01) (Table 5). For the rest of PP_r models, the average performance did not show significant improvement (Table 5). Regionally, the improvement was not significant for all models (data not shown here). As explained previously, satellite-derived P^B_{opt} presented the weakest agreement with observed P^B_{opt} data and might partly cause the poor agreement between IPP^{sat} from VGPM and IPP^{is}. Hence, adjusting parameterized P^B_{opt} to match *in situ* P^B_{opt} data improved VGPM performance.

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

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

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

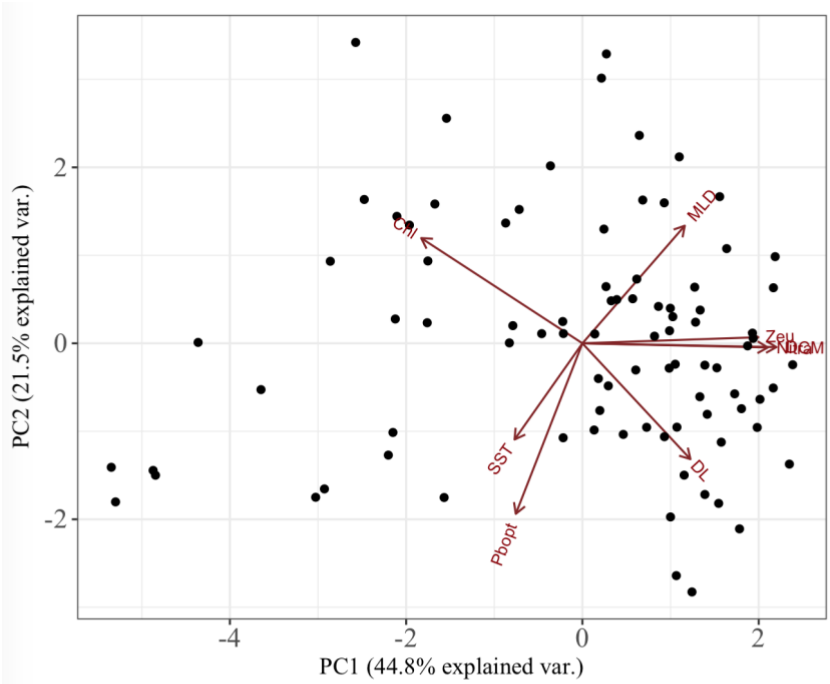


Figure 8. Biplot of *in situ* parameters (SST, sea surface temperature; Chl, chl-*a*; MLD, mixed layer depth; Zeu, Z_{eu}; DCM, deep chlorophyll maximum; DL, daylength; Nitra, Z_{NO3}; Pbopt, $P^B{}_{opt}$).

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

4. Discussion

Here, we compared here about 100 IPP^{is} with IPP^{sat} values derived from five of the most commonly used PP_r models for the four subtropical gyre regions sampled, including the Indian subtropical gyre region. IPP^{is} results presented here were consistent as the methodology was coherent and consistent from the first to the last transect allowing to dissipate any uncertainties about model prediction variations resulting from the methodology. This comparison allowed us to estimate model performances and explore pathways to improve them. From the five models and variants tested here, we observed that most of them did not derive a good representation of the IPP^{is} variability. Only VGPM1, MARRA and ESQRT models were better at estimating IPP^{is} with IPP^{sat} closer to IPP^{is} (|B| ~ 0.09) than for the other models (|B| ~ 0.29).

Although HYR model had been used for many years as a standard MODIS algorithm, we observed that it showed low performance to predict IPP^{is}. The original HYR model is extended to MLD and in general, MLD was less than Z_{eu} in the studied regions (Table 2). Thus, we estimated the performance of HYR model extended to Z_{eu} (data not shown here) and we observed that B and RMSD were lower for HYR model extended to Z_{eu} (using $Z_{eu}1$, B = -0.09 and RMSD = 0.30; using $Z_{eu}2$, B = 0.04 and RMSD = 0.29) than for HYR model extended for MLD. Hence, HYR models extended to Z_{eu} derived IPP^{sat} values closer to IPP^{is} than the original in subtropical gyre regions where MLD is shallower. However, a Taylor diagram revealed that they were not better at reproducing the magnitude of IPP^{is} variance than the original HYR model that (data not shown here), and that they had low correlation with IPP^{is}.

To understand the limitations of the models used here to estimate IPP^{is} accurately, we examined whether these limitations were caused by the input data variables for IPP^{sat} or by the model itself. Indeed, PP_r models strongly rely on chl-*a* and P^B_{opt} and a weak agreement between satellite-derived chl-*a* and P^B_{opt} with *in situ* data may explain the poor model performance. Although satellite-derived and modelled data inputs such as SST, MLD and Z_{NO3} had a relatively good agreement with *in situ* data, chl-*a* and, especially P^B_{opt}, had poor agreements with *in situ* data (R^2 = 0.16 and $0.01<R^2<0.11$, SIQR = 0.23 and 0.57<SIQR<1.13, respectively; Table 3). When satellite-derived and modelled data were substituted by *in situ*-derived data (when available), we found that, over the best performing models, the VGPM and MARRA models improved model-data linear regression statistics (R^2) by 68 % and 10 %. Total RMSD declined about the half for VGPM and a large decline was observed in VGPM and MARRA biases (76 % and 86 % respectively; Table 5).

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

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

The majority of MCE stations where IPP^{is} was estimated encompassed subtropical gyre regions with low IPP^{is} (i.e. more than 70 % of IPP^{is} < 300 mg C m^{-2} d^{-1}). Model performances were generally better for high values of IPP (above 500 mg C m^{-2} d^{-1}); when model-data misfit was in general lower (Fig. 4). This observation confirms the challenge to predict IPP^{is} in the ultra-oligotrophic regions encompassed by the oligotrophic gyres. We believe that further efforts are required to improve the performance of ocean color models such as VGPM to be applied to highly oligotrophic regions such as subtropical gyres, where IPP is relatively low (< 300 mg C m^{-2} d^{-1}), the MLD is shallower and cloudiness may bias satellite input data. Hence, efforts to improve the algorithms and parameters used in PP_r models (such as those provided in this work) specific for the oligotrophic subtropical gyre regions are essential to further understand the reasons for the poor predictions made by the existing models. The development of improved and robust satellite-based algorithms to predict oceanic primary production in subtropical gyres requires additional efforts to obtain *in situ* estimates of net primary production. This sampling effort is particularly necessary for some of the subtropical gyre regions, like in the three gyres located in the Southern Hemisphere (e.g. Marañon et al. 2000; Poultron et al. 2006; Regaudie-de-Gioux et al. 2012; this study). Because this ocean bioma comprises 70 % of the ocean, improving PP_r estimates therein is an imperative to progress toward a global ocean observing system.

Acknowledgements

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

References

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

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

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

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

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

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

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

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

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

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

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

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

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

Grasshoff, K., Kremling, K., Ehardt, M., 1999. Methods of seawater analysis. Weinheim: Wiley-VCH.

Karl, D.M., 1999. A sea of change: biogeochemical variability in the North Pacific subtropical gyre. Ecosystems 2, 181–194.

Key, R.M., Kozyr, A., Sabine, C.L., Lee, K., Wanninkhof, R., Bullister, J., Feely, R.A., Millero, F., Mordy, C., Peng, T.-H., 2004. A global ocean carbon climatology: Results from GLODAP. Global Biogeochemical Cycles 18, GB4031

Jolliff, J.K., et al., 2009. Summary diagrams for coupled hydrodynamic-ecosystem model skill assessment, Journal of Marine Systems 76, 64–82.

Latasa, M., 2014. A simple method to increase sensitivity for RP-HPLC phytoplankton pigment analysis Limnology and Oceanography: Methods, https://doi.org/10.4319/lom.2014.12.46.

Lee, D.-K., Niiler, P., Warnas-Varnas, A., 1994. Wind-driven secondary circulation in ocean mesoscale. Journal of Marine Research 5, 1–25.

Lee, Z.P., et al., 1996. Estimating primary production at depth from remote sensing. Applied Optics, https://doi.org/10.1364/AO.35.000463.

Longhurst, A., et al., 1995. An estimate of global primary production in the ocean from satellite radiometer data. Journal of Plankton Research 17, 1245–1271.

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

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

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

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

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

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

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

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

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

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

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

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

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

Regaudie-de-Gioux, A., Huete-Ortega, M., Sobrino, C., López-Sandoval, D.C., González, N., Fernández-Carrera, A., Vidal, M., Marañón, E., Cermeño, P., Latasa, M., Agustí, S., Duarte, C.M. 2019. Data set on xxxx [titulo]. Pange.de, doi:xxx [the data set will be made openly available in this repository upon acceptance of the paper]

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

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

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

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

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

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

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

Volk, T., Broecker, W.S., 1985. Ocean carbon pumps: analysis of relative strengths and efficiencies in ocean-driven atmospheric CO_2 changes. In: Walker, J.C.G, The carbon cycle and atmospheric CO_2: natural variations Archean to present. American Geophysical Union, pp. 99-110.

Westberry, T., et al., 2008. Carbon-based primary productivity modeling with vertically resolved photophysiology, Global Biogeochemical Cycles, https://doi.org/10.1029/2007GB003078.

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

Supplement Material

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

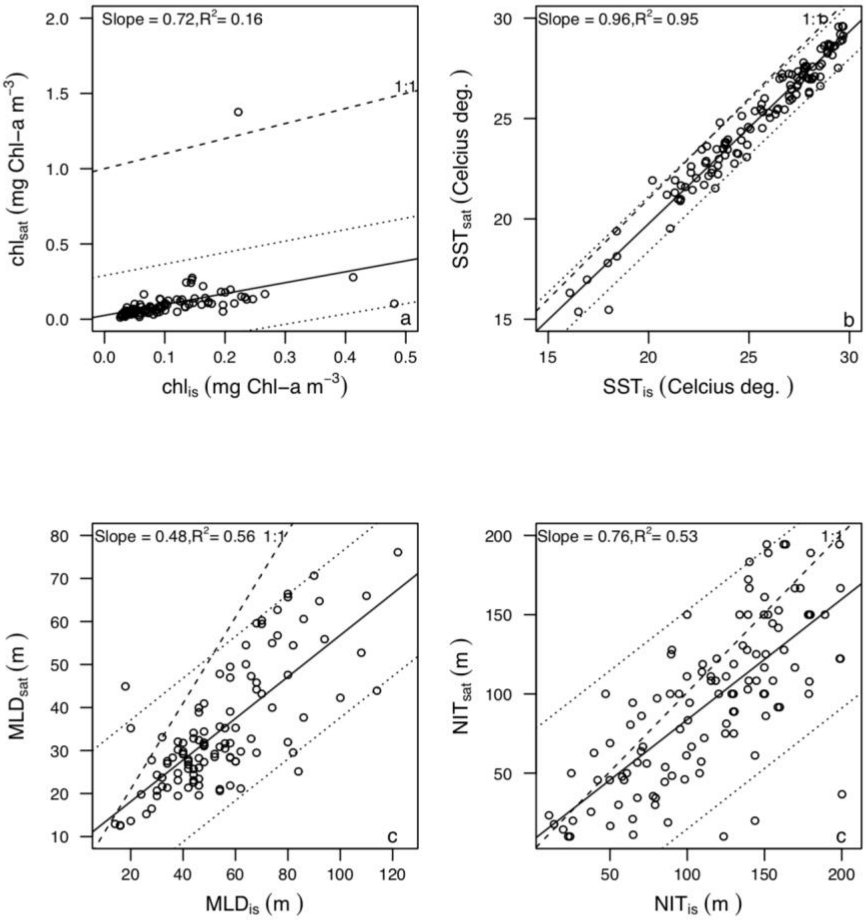
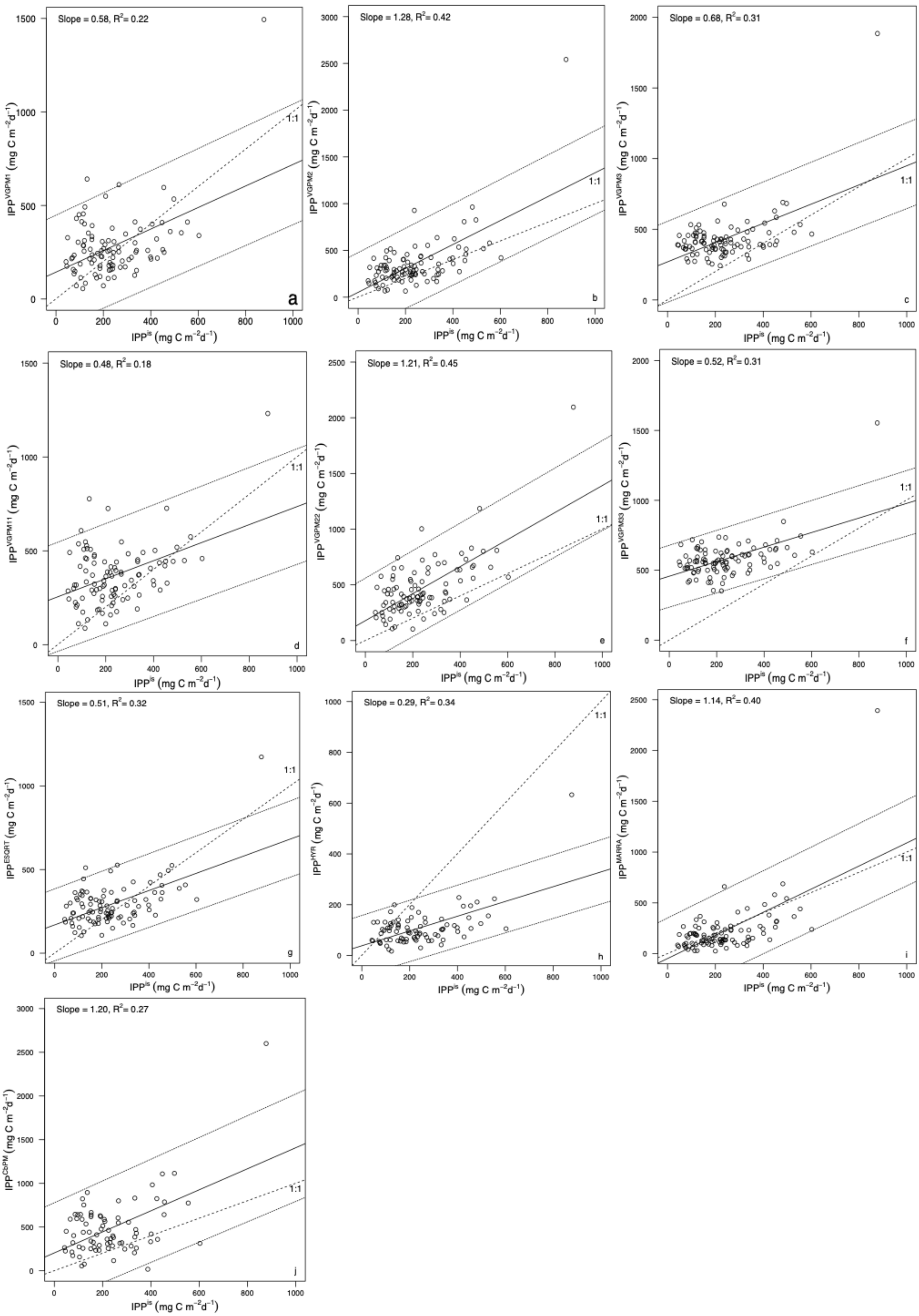


Figure S2. Multi-paneled regression plots of (a) VGPM1 *vs*. IPP^{is}, (b) VGPM2 *vs*. IPP^{is}, (c) VGPM3 *vs*. IPP^{is}, (d) VGPM11 *vs*. IPP^{is}, (e) VGPM22 *vs*. IPP^{is}, (f) VGPM33 *vs*. IPP^{is}, (g) ESQRT *vs*. IPP^{is}, (h) HYR *vs*. IPP^{is}, (i) MARRA *vs*. IPP^{is}, (j) CbPM *vs*. IPP^{is}.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


Slope = 0.58, R2= 0.22
IPPVGPM1 (mg C m-2d-1)
1:1
a
Slope = 1.28, R2= 0.42
IPPVGPM2 (mg C m-2d-1)
1:1
b
Slope = 0.68, R2= 0.31
IPPVGPM3 (mg C m-2d-1)
1:1
c
Slope = 0.48, R2= 0.18
IPPVGPM11 (mg C m-2d-1)
1:1
d
Slope = 1.21, R2= 0.45
IPPVGPM22 (mg C m-2d-1)
1:1
e
Slope = 0.52, R2= 0.31
IPPVGPM33 (mg C m-2d-1)
1:1
f
Slope = 0.51, R2= 0.32
IPPESQRT (mg C m-2d-1)
1:1
g
Slope = 0.29, R2= 0.34
IPPHYR (mg C m-2d-1)
1:1
h
Slope = 1.14, R2= 0.40
IPPMARRA (mg C m-2d-1)
1:1
i
Slope = 1.20, R2= 0.27
IPPCbPM (mg C m-2d-1)
1:1
j
IPPis (mg C m-2d-1)