



Assessing the uncertainties of model estimates of primary productivity in the tropical Pacific Ocean

Marjorie A.M. Friedrichs^{a,*}, Mary-Elena Carr^{b,1}, Richard T. Barber^c, Michele Scardi^d, David Antoine^e, Robert A. Armstrong^f, Ichio Asanuma^g, Michael J. Behrenfeld^h, Erik T. Buitenhuisⁱ, Fei Chai^j, James R. Christian^k, Aurea M. Ciotti^l, Scott C. Doney^m, Mark Dowellⁿ, John Dunne^o, Bernard Gentili^e, Watson Gregg^p, Nicolas Hoepffnerⁿ, Joji Ishizaka^q, Takahiko Kameda^r, Ivan Lima^m, John Marra^s, Frédéric Mélinⁿ, J. Keith Moore^t, André Morel^e, Robert T. O'Malley^h, Jay O'Reilly^u, Vincent S. Saba^a, Marjorie Schmeltz^b, Tim J. Smyth^v, Jerry Tjiputra^w, Kirk Waters^x, Toby K. Westberry^h, Arne Winguth^y

^a Virginia Institute of Marine Science, College of William and Mary, Gloucester Point VA, 23062-1346, USA

^b Jet Propulsion Laboratory, California Institute of Technology, 4800 Oak Grove Drive, Pasadena, CA 91101-8099, USA

^c Duke University Marine Lab, 135 Duke Marine Lab Road, Beaufort, NC 28516, USA

^d Department of Biology, University of Rome 'Tor Vergata', Via della Ricerca Scientifica, 00133 Roma, Italy

^e Laboratoire d'Océanographie de Villefranche, CNRS et Université Pierre et Marie Curie, Paris 6, France

^f School of Marine and Atmospheric Sciences, State University of New York at Stony Brook, Stony Brook, NY 11794, USA

^g Tokyo University of Information Sciences, 4-1-1, Onaridai, Wakaba, Chiba, 265-8501, Japan

^h Department of Botany and Plant Pathology, Cordley Hall 2082, Oregon State University, Corvallis, OR 97331-2902, USA

ⁱ Laboratory for Global Marine and Atmospheric Chemistry, School of Environmental Sciences, University of East Anglia, Norwich NR4 7TJ, United Kingdom

^j School of Marine Sciences, University of Maine, Orono, ME 04469, USA

^k Fisheries and Oceans Canada, Victoria, BC, Canada V8L 4B2

^l UNESP-Campus do Litoral Paulista, Praça Infante Dom Henrique S/N, São Vicente, São Paulo CEP 11330-900, Brazil

^m Department of Marine Chemistry and Geochemistry, Woods Hole Oceanographic Institution, 266 Woods Hole Road, Woods Hole, MA 02543, USA

ⁿ European Commission-Joint Research Centre, 21020 Ispra, Italy

^o Geophysical Fluid Dynamics Laboratory, Princeton, NJ 08540, USA

^p NASA Global Modeling and Assimilation Office, Goddard Space Flight Center, Greenbelt, MD 20771, USA

^q Faculty of Fisheries, Nagasaki University, 1-14 Bunkyo, Nagasaki, 852-8521, Japan

^r Group of Oceanography, National Research Institute of Far Seas Fisheries, 5-7-1 Shimizu-Orido, Shizuoka 424-8633, Japan

^s Geology Department, Brooklyn College of the City University of New York, 2900 Bedford Ave., Brooklyn, NY 11210, USA

^t Department of Earth System Science, 3214 Croul Hall, University of California Irvine, Irvine, CA 92697-3100, USA

^u NOAA/NMFS Narragansett Laboratory, 28 Tarzwell Drive, Narragansett, RI, 02887 USA

^v Plymouth Marine Laboratory, Prospect Place, Plymouth, Devon PL1 3DH, United Kingdom

^w Bjerknes Center for Climate Research, Allegaten 70, 5007 Bergen, Norway

^x NOAA Coastal Services Center, 2234 South Hobson Ave., Charleston, SC 29405-2413, USA

^y Department of Earth & Environmental Sciences, University of Texas at Arlington, Box 19049, Arlington, TX 76019-0049, USA

* Corresponding author.

E-mail addresses: marjy@vims.edu (M.A.M. Friedrichs), mcarr@ei.columbia.edu (M.-E. Carr), rbarber@duke.edu (R.T. Barber), mcardi@mclink.it (M. Scardi), antoine@obs-vlfr.fr (D. Antoine), rarmstrong@notes.cc.sunysb.edu (R.A. Armstrong), asanuma@rsch.tuis.ac.jp (I. Asanuma), mjb@science.oregonstate.edu (M.J. Behrenfeld), martinburo@email.com (E.T. Buitenhuis), fchai@umaine.edu (F. Chai), Jim.Christian@ec.gc.ca (J.R. Christian), ciotti@csv.unesp.br (A.M. Ciotti), sdoney@whoi.edu (S.C. Doney), mark.dowell@jrc.it (M. Dowell), jpd@gfdl.noaa.gov (J. Dunne), gentili@obs-vlfr.fr (B. Gentili), watson.gregg@nasa.gov (W. Gregg), nicolas.hoepffner@jrc.it (N. Hoepffner), ishizaka@nagasaki-u.ac.jp (J. Ishizaka), takukame@affrc.go.jp (T. Kameda), ilima@whoi.edu (I. Lima), jfm7780@brooklyn.cuny.edu (J. Marra), frederic.melin@jrc.it (F. Mélin), jkmoore@uci.edu (J.K. Moore), morel@obs-vlfr.fr (A. Morel), omalleyr@science.oregonstate.edu (R.T. O'Malley), jay.oreilly@noaa.gov (J. O'Reilly), vssaba@vims.edu (V.S. Saba), tjism@pml.ac.uk (T.J. Smyth), jtj061@bjerknes.uib.no (J. Tjiputra), kirk.waters@noaa.gov (K. Waters), toby.westberry@science.oregonstate.edu (T.K. Westberry), awinguth@uta.edu (A. Winguth).

¹ Now at Columbia Climate Center, The Earth Institute at Columbia University, 2910 Broadway, New York, NY 10025, USA.

ARTICLE INFO

Article history:

Received 2 November 2007

Received in revised form 12 February 2008

Accepted 20 May 2008

Available online 29 May 2008

Keywords:

Primary production

Modeling

Remote sensing

Satellite ocean color

Statistical analysis

Tropical Pacific Ocean (15°N to 15°S and 125°E to 95°W)

ABSTRACT

Depth-integrated primary productivity (PP) estimates obtained from satellite ocean color-based models (SatPPMs) and those generated from biogeochemical ocean general circulation models (BOGCMs) represent a key resource for biogeochemical and ecological studies at global as well as regional scales. Calibration and validation of these PP models are not straightforward, however, and comparative studies show large differences between model estimates. The goal of this paper is to compare PP estimates obtained from 30 different models (21 SatPPMs and 9 BOGCMs) to a tropical Pacific PP database consisting of ~1000 ^{14}C measurements spanning more than a decade (1983–1996). Primary findings include: skill varied significantly between models, but performance was not a function of model complexity or type (i.e. SatPPM vs. BOGCM); nearly all models underestimated the observed variance of PP, specifically yielding too few low PP ($<0.2 \text{ g C m}^{-2} \text{ d}^{-1}$) values; more than half of the total root-mean-squared model–data differences associated with the satellite-based PP models might be accounted for by uncertainties in the input variables and/or the PP data; and the tropical Pacific database captures a broad scale shift from low biomass-normalized productivity in the 1980s to higher biomass-normalized productivity in the 1990s, which was not successfully captured by any of the models. This latter result suggests that interdecadal and global changes will be a significant challenge for both SatPPMs and BOGCMs. Finally, average root-mean-squared differences between in situ PP data on the equator at 140°W and PP estimates from the satellite-based productivity models were 58% lower than analogous values computed in a previous PP model comparison 6 years ago. The success of these types of comparison exercises is illustrated by the continual modification and improvement of the participating models and the resulting increase in model skill.

© 2008 Elsevier B.V. All rights reserved.

1. Introduction

Marine primary productivity is a large and highly variable component of the global carbon cycle and drives ocean biogeochemical cycles of major chemical elements such as silicon, nitrogen and phosphorus. From an ecological standpoint, primary productivity provides the upper bound for production at higher trophic levels and defines ecosystem carrying capacity, a key factor for the design of marine protected areas. Awareness of bottom-up forcing even to understand the dynamics of the upper trophic levels that form the most important commercial fisheries has increased (Pauly and Christensen, 1995; Ware and Thomson, 2005). Calculating accurate primary productivity (PP) estimates over large areas is thus a primary step for ecosystem models charged with the task of assessing trophic dynamics. Reliable estimates of PP are also necessary for multiple other applications, including quantifying the flux of carbon dioxide (e.g. Bianchi et al., 2005), assessing export production (e.g. Boyd and Trull, 2007) and estimating production of climate-active gases such as dimethyl sulfide (e.g. Larsen, 2005). Estimating accurate PP on global scales is also essential to understanding the consequences of climate change on phytoplankton growth (Behrenfeld et al., 2006).

Although the broad spatial and temporal patterns of productivity were elucidated on the basis of compilations of in situ measurements (Koblentz-Mishke et al., 1970; Berger, 1989), global dynamics cannot be quantified without the quasi-synoptic view afforded by satellites. Consequently, there have been many efforts to develop models that use satellite-derived information (e.g. surface chlorophyll, chl_0 ; sea surface temperature, SST; and photosynthetically available radiation, PAR) to estimate PP (SatPPMs hereafter; these and other acronyms are defined in Table 1) (Platt and Sathyendranath, 1993; Behrenfeld and Falkowski, 1997a).

Although there is considerable understanding of the photosynthetic process and knowledge of the ocean optics that determine ocean color signals, SatPPMs often have limited success reproducing the observed variability of PP data (Siegel et al., 2001; McClain et al., 2002).

Another approach to quantify global patterns of photosynthesis is to use coupled biogeochemical ocean general circulation models (BOGCMs), which, as a result of increased computational resources, can now be run globally at adequate horizontal and vertical resolution. While these models parameterize photosynthesis similarly to SatPPMs, BOGCMs additionally have explicit compartments for different nutrients, detritus, and one or more functional or size groups of phytoplankton and zooplankton and incorporate mechanistic knowledge of nutrient uptake and physical transport of biomass. Whereas SatPPMs require surface chlorophyll and

Table 1

Acronym	Definition
PP	Primary productivity
SatPPM	Satellite-based productivity model
BOGCM	Biogeochemical ocean general circulation model
chl_0	Surface chlorophyll
SST	Sea-surface temperature
PAR	Photosynthetically available radiation
MLD	Mixed layer depth
PPARR	Primary Productivity Algorithm Round Robin
HNLC	High-nutrient low chlorophyll
DI (DR)	Depth-integrated (depth-resolved)
WI (WR)	Wavelength-integrated (wavelength-resolved)
RMSD	Root mean square difference
RMSD_{CP}	Centered-pattern root mean square difference
B	Bias
PCA	Principal component analysis
VGPM	Vertically Generalized Productivity Model

temperature as input variables, BOGCMs explicitly compute these fields. Although surface chlorophyll fields can be assimilated into BOGCMs (Friedrichs, 2002; Tjiputra et al., 2007; Gregg et al., 2009–this volume), the computational cost of doing so is high, and can entail reductions in horizontal and vertical grid resolution.

While diverse approaches for estimating productivity from ocean color and with coupled ecosystem general circulation models are desirable as this field grows and develops, it is extremely important to quantify the performance of these various methods relative to observations and to elucidate the reasons underlying the similarities/differences in model output. To this end, the Primary Productivity Algorithm Round Robin (PPARR) series has provided a context in which the performance of primary productivity models can be quantified. In addition, the PPARR exercises continue to help model developers and users understand the conditions under which each productivity model is most applicable (a summary of past PPARRs is given in Section 2). For any model, a vital element of model skill is the ability to reproduce in situ observations; in the case of PP models, measurements of PP. If observations are representative and the data have undergone careful quality control, firm conclusions can be reached regarding the environmental conditions that challenge model skill. These challenging conditions, in turn, can be taken into account by model developers and end-users to improve model formulation and/or application.

This paper compares output of models of productivity with a large quality-controlled database (ClimPP) spanning more than a decade in the tropical Pacific Ocean. Since tropical and oligotrophic regions cover a large percentage of ocean surface area, it is critical that our ability to model them is improved. The size, temporal and spatial range, and consistent quality of this database help us meet the data requirements mentioned above for data–model intercomparisons. Although it is not possible to make conclusive statements concerning global model performance based on a single regional comparison, improving PP estimates in the tropical Pacific will increase the skill of global models because (1) this region represents a large fraction of the global ocean, and (2) this region represents one of the greatest current challenges to PP modelers (Campbell et al., 2002; McClain et al., 2002).

The goal of this effort is thus to compare the skill of various models (including both SatPPMs and BOGCMs) in estimating PP within the tropical Pacific Ocean. Assessing the overall skill of the BOGCMs is far beyond the scope of this paper; here, only the skill of the BOGCMs in estimating PP is assessed. Additionally, we note up front that the identification of a single best satellite-based PP model, or consensus model, is not the goal of the PPARR exercises. Here we implement a variety of different skill assessment methodologies, and identify which models perform well according to each criterion. These comparative results are relevant for those who wish to choose a single PP model to implement for a given study. In addition, we highlight specific problems that tend to characterize the current generation of PP models. This information is of use to PP model developers, as they continue to adjust and improve their model formulations.

The background of the PPARR exercises is described in the next section (Section 2), and is followed by an introduction to the observational dataset and the methodologies employed to assess model performance (Section 3). Skill assessment results

are presented using Taylor and target diagrams as well as cumulative distribution functions. The impact of uncertainties in the input variables and in situ PP measurements are also quantified with error perturbation and principal component analyses (Section 4). Correlations between model errors and environmental variables are then discussed (Section 5), and we conclude with a summary in Section 6.

2. PPARR background

2.1. PPARR/PPARR2

For over a decade, NASA has supported research aiming to improve our ability to quantify marine photosynthesis from satellites in the form of a series of round robin experiments for evaluation and comparison. The first two Primary Productivity Algorithm Round Robin (PPARR) exercises used in situ measurements of PP to quantify the ability of participating models to predict PP based on information accessible via remote sensing. While PPARR1 tested the approach with data from 25 stations, PPARR2 compared the output of twelve models using data from 89 stations with wide geographic coverage (Campbell et al., 2002). The models that performed best were within a factor of 2.4 (based on one standard deviation in log-difference errors) of the in situ ^{14}C measurements. Of the eight regions represented in the comparison, the most serious biases were found in the equatorial Pacific, where all algorithms underestimated in situ measurements by more than a factor of 2. If biases, which in all cases contributed significantly to absolute model–data misfit, could be corrected, then ten of the twelve models were within a factor of two of the in situ data. Model–data misfit was lowest in regions that have historically contributed the most data for parameterization, i.e. the Atlantic, whereas misfit was high in both the equatorial Pacific and the Southern Ocean.

2.2. PPARR3, phases 1 and 2

Phases 1 and 2 of the third primary production algorithm round robin consisted of an intercomparison and sensitivity study of global primary productivity fields computed from 24 SatPPMs and 7 BOGCMs, but included no comparison with in situ PP measurements (Carr et al., 2006). The participating models, which are not the same suite as those participating in the current comparison, were provided with common input fields of PAR, SST, chl_a, and mixed layer depth (MLD), and produced global average PP fields for 8 months of 1998 and 1999. Maximum PP within the Atlantic and Pacific Oceans occurred in the equatorial band (10°S – 10°N ; $>0.5\text{ g C m}^{-2}\text{ d}^{-1}$). The simulated global average PP varied by a factor of two between models, with model results diverging most at low surface temperatures ($<10^{\circ}\text{C}$), at high chlorophyll concentrations ($>1\text{ mg chl m}^{-3}$), and within HNLC regions. The models were grouped based on pair-wise correlations and the model clustering was independent of model complexity and primarily depended on the form in which temperature was parameterized within the model.

3. Current PPARR methodology

In the earlier phases of PPARR3, global productivity fields computed from participating PP models were compared among

themselves and also with a 'mean-model' productivity field (Carr et al., 2006). This third and final phase of PPARR3 differs significantly from the Carr et al. (2006) study in that here simulated productivity fields were compared with shipboard measurements of ^{14}C uptake, as in PPARR1 and PPARR2. This effort expands significantly on PPARR2, in which only a small number of models participated (ten versus thirty in this study), only a small dataset was used (<90 stations versus >900 stations in this study), participation was almost exclusively from the USA, BOGCMs were not invited to participate, and only root mean square error was presented. We note that the present comparison exercise does not attempt to assess the overall skill of the participating biogeochemical models (Matsumoto et al., 2004; Doney et al., 2004; Najjar et al., 2007; Friedrichs et al., 2006, 2007). Instead, our goal is to compare the skill of the BOGCMs in estimating PP, allowing relative comparison with that of the SatPPMs. To our knowledge, a PP comparison on the scale of this current exercise (21 SatPPMs and 9 BOGCMs) has not heretofore been conducted.

3.1. ClimPP dataset

The PP dataset used here (ClimPP; available as an electronic supplement to this paper) was prepared by Barber and collaborators. The dataset was not made available to the model participants, and only a small subset of the ClimPP data were publicly available, e.g. the data from the Joint Global Ocean Flux Study equatorial Pacific Process study (Barber et al., 1996) which makes up only a small subset of the ClimPP database.

The ClimPP database comprises ~1000 stations (Fig. 1) in the tropical Pacific between 15°S and 15°N (Chavez and Barber, 1991; Barber and Chavez, 1991; Le Borgne et al., 2002). Collected between 1983 and 1996, these data consist of chlorophyll and productivity profiles acquired from 31 U.S. and international research cruises. Productivity values were integrated to the 1% light depth. Integrations to the 0.1% light depth were on average only ~3.5% higher than those integrated to the 1% light depth, and the portion of the photosynthetic profile from the 1% light depth to the 0.1% light depth showed very little spatial or temporal variation (Barber et al., 2001).

On-deck PP incubations were carried out in seawater-cooled Plexiglas incubator boxes. Light in the incubators was attenu-

ated to varying percentages of the surface light field, using neutral density screening and blue Plexiglas. Trace-metal clean procedures (Sanderson et al., 1995) were used in all data collected. Specifically, the black rubber closing springs of the Niskin or Go-Flo bottles that were found to be a major source of trace metal inhibition were replaced with Teflon coated metal springs. These retrofitted bottles were mounted on a Teflon coated rosette (Chavez and Barber, 1991; Sanderson et al., 1995; Barber et al., 1996).

The largest source of variation in estimating PP is the process of determining the attenuation of light and of assigning a depth to each percent light level (Barber and Chavez, 1991). Barber et al. (1997) have outlined a procedure that uses the Morel (1988) model to estimate the PAR attenuation coefficient and assigns light depths using observed chlorophyll profiles. This method removes important sources of variation due to project, ship or investigator, and has proven to be an efficient and accurate way to reprocess historical on deck incubation data to make basinwide and/or interannual and decadal comparisons (Le Borgne et al., 2002).

Both oligotrophic and mesotrophic areas are represented within this dataset. Productivities integrated to the 1% light level range from ~0.05 $\text{g C m}^{-2} \text{d}^{-1}$ in the western Pacific warm pool, to nearly 1.80 $\text{g C m}^{-2} \text{d}^{-1}$ in the HNLC region (Fig. 2a). Thirty-six percent of the data are from the 1990s, with the remainder having been collected in the 1980s. During each of these decades, data are available from all portions of the tropical Pacific domain: 15°N to 15°S and 125°E to 95°W (Section 5). These data thus provide more than a decade of comparable observations for a region of the ocean where interannual and interdecadal variability is relatively well studied (McPhaden and Zhang, 1999).

3.2. Participating models

The 30 models used to generate PP estimates for comparison (Table 2; Appendix A) include 21 SatPPMs and 9 BOGCMs. Of the SatPPMs, the first twelve (Models 1–12) represent examples of depth-integrated, wavelength-integrated (DI/WI) models. Five of these models (#8–12) are variants of the Vertically Generalized Productivity Algorithm (VGPM; Behrenfeld and Falkowski, 1997b). The next three models (#13–15) are depth-resolved but wavelength-integrated (DR/WI), and the remaining six (Models

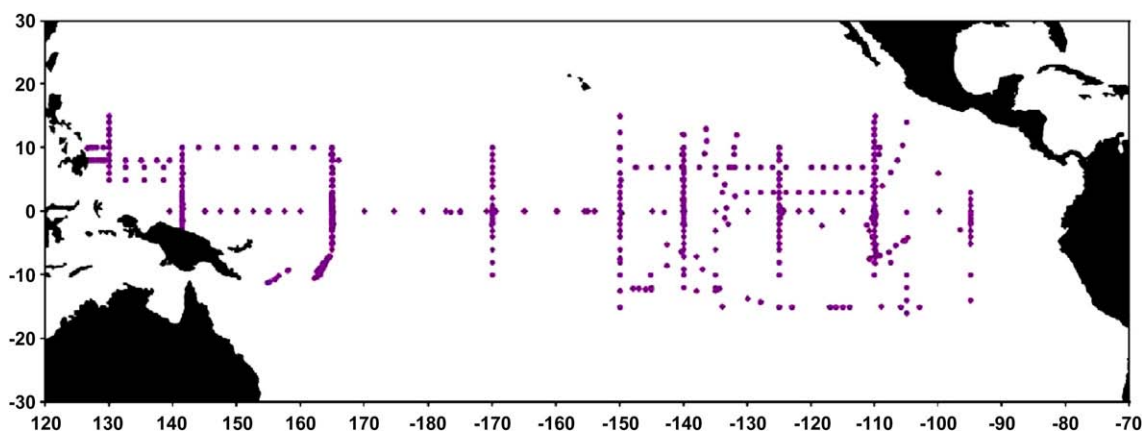


Fig. 1. Map of stations in the ClimPP database.

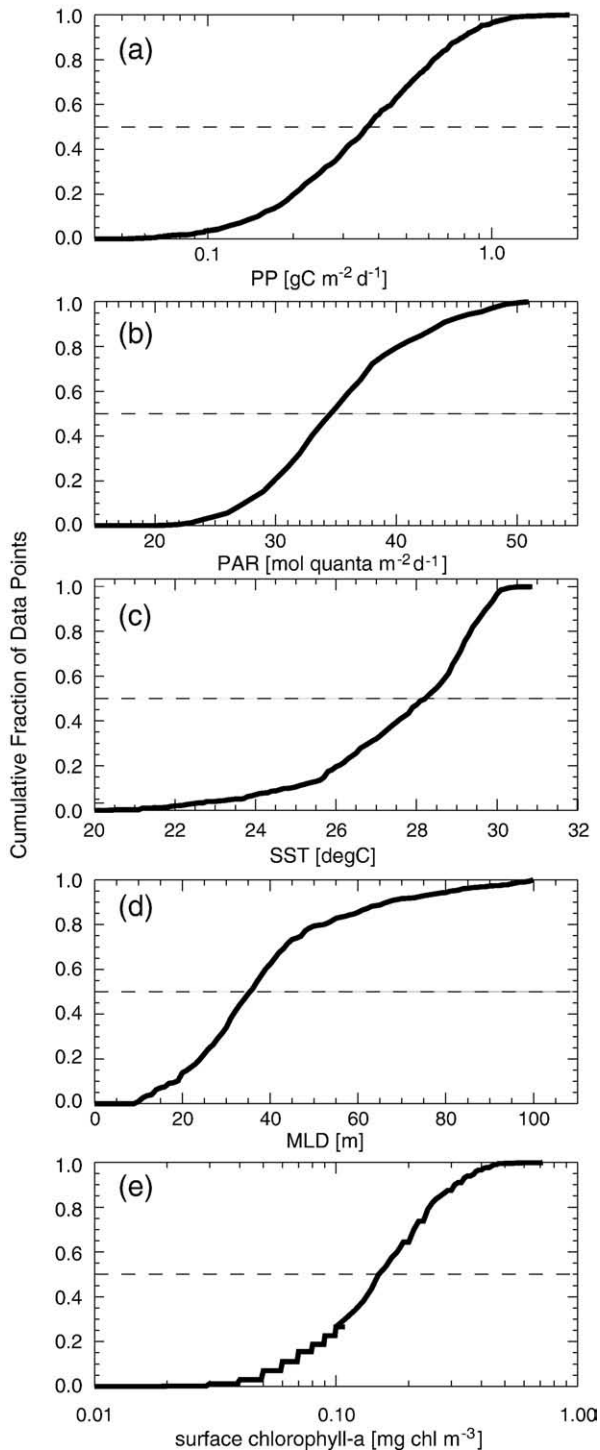


Fig. 2. Cumulative distribution function of the ClimPP data and the input variables provided to the SatPPMs: (a) PP integrated to the 1% light level, (b) PAR, (c) SST, (d) MLD and (e) surface chlorophyll. Median value is illustrated by the intersection of the solid lines with the dashed 0.5 line.

16–21) are the most complex, resolving both depth and wavelength (DR/WR).

The simplest model (#1) estimates integrated primary productivity (in units of $g\ C\ m^{-2}\ d^{-1}$) as the square root of

surface chlorophyll (in units of $mg\ chl\ m^{-3}$). This purely empirical relationship was suggested by Eppley et al. (1985) and serves as a reference model to illustrate the role of chlorophyll in PP estimation. In PPARR3 Phase 1, PP estimated by this model had far lower temporal variability than the other models, but time averages were not that different from those of other models, except in areas of extreme light or temperature conditions (Carr et al., 2006). All 21 SatPPMs require chl_0 fields, Models 2–21 use PAR in their calculation, 15 of the 21 models use SST, and only seven models use MLD (Table 2).

The nine BOGCMs participating in this intercomparison vary widely, though each has between two and four

Table 2
Model descriptions

Model #	Contributor	Type	Input variables used				Reference
			chl_0	SST	PAR	MLD	
1	Friedrichs	DI,WI ^a	x				Eppley et al. (1985)
2	Carr	DI,WI	x	x	x		Carr (2002)
3	Carr	DI,WI	x	x	x	x	Howard and Yoder (1997)
4	Carr	DI,WI	x		x	x	
5	Scardi	DI,WI	x	x	x	x	Scardi (2001)
6	Dowell	DI,WI	x	x	x		
7	Westberry/Behrenfeld	DI,WI	x	x		x	Behrenfeld et al. (2005)
8	Behrenfeld/Westberry	DI,WI	x	x	x		Behrenfeld and Falkowski (1997b)
9	Behrenfeld/Westberry	DI,WI	x	x	x		Behrenfeld and Falkowski (1997b)
10	Kameda/Ishizaka	DI,WI	x	x	x		Kameda and Ishizaka (2005)
11	Ciotti	DI,WI	x	x	x		
12	Behrenfeld/Westberry	DI,WI	x			x	
13	Armstrong	DR,WI	x	x	x		Armstrong (2006)
14	Asanuma	DR,WI	x	x	x		Asanuma (2006)
15	Marra/O'Reilly	DR,WI	x	x	x		Marra et al. (2003)
16	Smyth	DR,WR	x	x	x		Smyth et al. (2005)
17	Smyth	DR,WR	x	x	x		Smyth et al. (2005)
18	Mélin/Hoepffner	DR,WR	x		x		Mélin (2003)
19	Waters	DR,WR	x	x	x	x	Ondrusek et al. (2001)
20	Antoine/Morel/Gentili	DR,WR	x	x	x	x	Antoine and Morel (1996)
21	Westberry	DR,WR	x		x	x	Westberry et al. (2008)
22	Christian	BOGCM					Christian et al. (2002)
23	Christian	BOGCM					Christian et al. (in press)
24	Lima/Moore/Doney	BOGCM					Moore et al. (2004)
25	Dunne	BOGCM					Dunne et al. (2005)
26	Gregg	BOGCM					Gregg and Casey (2007)
27	Gregg	BOGCM					Gregg (2008)
28	Buitenhuis	BOGCM					Le Quéré et al. (2007)
29	Tjiputra/Winguth	BOGCM					Tjiputra et al. (2007)
30	Chai	BOGCM					Chai et al. (2003)

^a DI = depth-integrated, DR = depth-resolved, WI = wavelength-integrated, WR = wavelength-resolved.

phytoplankton functional groups, and each model contains either three or four nutrients. All BOGCMs except Model 29 provide interannual, rather than climatological, PP distributions. Although one BOGCM (#27) assimilates in situ National Ocean Data Center chlorophyll data, none of the BOGCMs utilize any of the provided input fields (Table 2).

All 30 models provided estimates of productivity integrated to the 1% light level, for each of the ~1000 ClimPP data points. Details regarding the structure of all 30 participating models are beyond the scope of this paper, but key features of these models are provided in Table 2 and Appendix A. In addition, a forthcoming paper concentrates on identifying specific similarities and differences between model parameterizations, which may be affecting model performance (Saba et al., in prep.).

3.3. Input data variables for SatPPMs

Although none of the models were provided with the PP data (Fig. 2a), the SatPPMs were provided with four types of input variables: daily mean chl_a, PAR, SST, and MLD (Fig. 2b–e). These data, along with the integrated productivities from the ClimPP dataset are included in this paper as an electronic supplement. As in PPARR2, chl_a was obtained from in situ data, rather than remotely sensed data as in PPARR3 Phase 1. Although chlorophyll profiles were available for each station, only the surface values were provided to the modelers. The profiles were used to estimate the depth of maximum chlorophyll (Section 5), but not provided to the participants. NCEP/NCAR Reanalysis 1 data for daily mean downward solar radiation flux data were downloaded from the National Oceanic & Atmospheric Administration website for the Physical Sciences Division of their Earth System Research Laboratory. These NCEP/NCAR reanalysis data extend from 1948 until the present and are obtained from a state-of-the-art data assimilative analysis/forecast system. A factor of 0.43 was used to convert these daily shortwave fluxes into PAR and produced values that were in relatively good agreement (± 10 mol quanta m⁻² d⁻¹) with PAR measurements obtained on the Joint Global Ocean Flux Study equatorial Pacific Process Study cruises. SST was obtained from the Advanced Very High Resolution Radiometer. MLD was obtained from a reduced-gravity, primitive equation tropical Pacific Ocean model (Murtugudde et al., 1996; Murtugudde and Busalacchi, 1998) with a variable depth mixed layer overlying 19 sigma layers. In this model, mixed layer thickness is determined using a “hybrid” mixed layer model (Chen et al., 1994) that considers both wind stirring and shear instability, and produces interannual MLD estimates that are generally within ± 20 m of those derived from measurements collected on the Joint Global Ocean Flux Study equatorial Pacific Process Study cruises.

Two models (e.g. Models 7 and 21) require information on particulate backscatter at 443 nm (bbp), which is estimated from ocean color remote sensing data (Maritorena et al., 2002). Since these data were not available for this pre-SeaWiFS era ClimPP dataset, monthly 9 km climatological bbp values (Sep. 1997–Sep. 2007) were used for each given sample location.

3.4. Skill assessment strategies

The skill of the 30 participating models was assessed using several strategies outlined below (also see Stow et al., 2009–this volume). Campbell et al. (2002) concluded in PPARR2 that the

total root mean square difference (RMSD) summed over the N data points provides a valuable comparison of PP models

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

Here model–data misfit in log₁₀ space (Δ) is defined as:

$$\Delta(i) = \log(\text{PP}_m(i)) - \log(\text{PP}_d(i))$$

where $\text{PP}_m(i)$ is modeled PP, obtained either via SatPPMs or BOGCMs, and $\text{PP}_d(i)$ represents the ClimPP data.

RMSD is composed of two components, the bias (B) representing the difference between the means of the two fields, and the centered pattern RMSD (RMSD_{CP} ; sometimes referred to as the unbiased RMSD) representing the differences in the variability of the two fields:

$$\text{RMSD}^2 = B^2 + \text{RMSD}_{\text{CP}}^2 \quad (2)$$

The bias and RMSD_{CP} (Tables 3 and 4) provide measures of how well the mean and variability are modeled, respectively:

$$B = \overline{\log(\text{PP}_m)} - \overline{\log(\text{PP}_d)} \quad (3)$$

$$\text{RMSD}_{\text{CP}} = \left(\frac{1}{N} \sum_{i=1}^N \left(\left(\log \text{PP}_m(i) - \overline{\log \text{PP}_m} \right) - \left(\log \text{PP}_d(i) - \overline{\log \text{PP}_d} \right) \right)^2 \right)^{1/2} \quad (4)$$

Because the units of the above quantities are in decades of log₁₀ and not easily translated, a non-dimensional inverse transformed value for bias is presented:

$$F_{\text{med}} = 10^B \quad (5)$$

where F_{med} (Campbell et al., 2002) is the median value of the ratio $\frac{\text{PP}_m(i)}{\text{PP}_d(i)} = 10^{\Delta(i)}$ (Tables 3 and 4). Thus, if $F_{\text{med}} = 2.0$, the median value of $\text{PP}_m(i)$ is a factor of two larger than the median value of $\text{PP}_d(i)$; if $F_{\text{med}} = 0.5$, the median value of $\text{PP}_m(i)$ is a factor of two smaller than $\text{PP}_d(i)$.

The novel *target diagram* (Jolliff et al., 2009–this volume) is used to visualize bias, RMSD_{CP} , and total RMSD for the 30 models on a single plot. On the target diagram, these quantities are normalized by the standard deviation of $\log \text{PP}_d$ ($\sigma_d = 0.279$), i.e. normalized bias (B^*) is defined as:

$$B^* = B / \sigma_d, \quad (6)$$

Although RMSD_{CP} is inherently a positive quantity, normalized RMSD_{CP} (denoted as $\text{RMSD}_{\text{CP}}^*$) is defined as:

$$\text{RMSD}_{\text{CP}}^* = \text{sign}(\sigma_m - \sigma_d) \text{RMSD}_{\text{CP}} / \sigma_d, \quad (7)$$

where σ_m is the standard deviation of $\log \text{PP}_m$, and thus $\text{RMSD}_{\text{CP}}^*$ can be either positive (indicating that the model is overestimating the variance of the data) or negative (indicating that the model is underestimating the variance of the data). It is important to note here that when the total RMSD statistic is computed, models that underestimate the observed variance tend to result in lower total RMSD scores than those that overestimate the variance (Jolliff et al., 2009–this volume).

For a given value of RMSD_{CP} , a portion of the model–data misfit will result from phase differences between the simulated and observed fields, and a portion will result from differences

Table 3

Bias, F_{med} (see text), RMSD_{CP} , total RMSD, correlation and standard deviation computed for each participating model relative to the ClimPP dataset

Model # or type	Bias	F_{med}	RMSD_{CP}	Total RMSD	Correlation	Standard deviation
1	0.04	1.10	0.24	0.24	0.53	0.13
2	0.17	1.48	0.24	0.30	0.55	0.22
3	0.04	1.10	0.29	0.30	0.36	0.24
4	0.20	1.57	0.24	0.31	0.53	0.18
5	-0.09	0.80	0.25	0.27	0.51	0.21
6	0.16	1.43	0.25	0.29	0.57	0.26
7	0.05	1.13	0.27	0.27	0.49	0.25
8	-0.04	0.92	0.24	0.24	0.57	0.21
9	0.19	1.56	0.25	0.32	0.46	0.16
10	-0.21	0.62	0.24	0.31	0.54	0.18
11	0.02	1.06	0.24	0.24	0.56	0.23
12	-0.01	0.97	0.33	0.33	0.53	0.38
13	0.01	1.03	0.24	0.24	0.55	0.22
14	0.18	1.53	0.24	0.30	0.57	0.21
15	0.07	1.18	0.24	0.25	0.51	0.18
16	0.11	1.28	0.23	0.26	0.55	0.18
17	0.14	1.39	0.24	0.28	0.53	0.20
18	-0.04	0.91	0.23	0.23	0.60	0.20
19	0.30	2.01	0.24	0.39	0.55	0.10
20	0.17	1.48	0.24	0.29	0.54	0.15
21	0.25	1.79	0.30	0.39	0.36	0.26
22	-0.24	0.58	0.29	0.37	0.56	0.33
23	-0.05	0.89	0.29	0.29	0.50	0.30
24	0.20	1.58	0.24	0.31	0.54	0.19
25	0.35	2.22	0.24	0.42	0.55	0.22
26	0.12	1.33	0.24	0.27	0.59	0.26
27	0.12	1.31	0.24	0.27	0.60	0.26
28	0.32	2.08	0.27	0.42	0.23	0.08
29	0.00	1.01	0.32	0.32	0.37	0.29
30	0.27	1.87	0.24	0.36	0.50	0.12
DI/WI	0.10±0.02	1.26	0.26±0.01	0.28±0.01	0.52±0.02	0.22±0.02
DR/WI	0.09±0.05	1.23	0.24±0.002	0.26±0.02	0.54±0.02	0.20±0.01
DR/WR	0.17±0.04	1.48	0.25±0.01	0.31±0.03	0.52±0.03	0.18±0.02
SatPPM	0.12±0.02	1.32	0.25±0.01	0.29±0.01	0.52±0.01	0.21±0.01
BOGCM	0.19±0.04	1.53	0.26±0.01	0.34±0.02	0.49±0.04	0.23±0.03
All	0.14±0.02	1.38	0.25±0.01	0.30±0.01	0.51±0.01	0.21±0.01

F_{med} and correlation are non-dimensional; remaining statistics are in units of decades of \log_{10} . Standard deviation of ClimPP dataset is 0.28. In addition, the absolute means (\pm one standard error) of these statistics are reported for various model subsets.

between the amplitudes of the variations. Thus in addition to RMSD_{CP} values, correlation coefficients and variances are also computed (Tables 3 and 4) to understand the similarities and differences in model–data fit for each of the models participating in this comparison. In Section 4.2 *Taylor diagrams* (Taylor, 2001) are used to represent RMSD_{CP} , correlation, and standard deviation on a single plot.

The *cumulative distribution function* represents an alternative way to visualize model bias. Although B provides a succinct measure of the magnitude and sign of model bias, from this statistic alone it is not possible to determine whether positive biases result from overestimating high values, or low values, or both. A comparison of model and data cumulative distribution functions clearly reveals where in the spectrum of values the biases occur, and is an excellent method for visualizing median, maximum and minimum values.

In a comparison exercise such as this, it is critical to acknowledge that uncertainties exist in the input variables provided to the SatPPMs, which may affect the ability of the models to adequately reproduce the ^{14}C PP measurements. Furthermore, the ClimPP data themselves are associated with a level of uncertainty and need to be regarded as ranges rather than as exact values. To examine the effect of uncertainties in the input variables on modeled PP, an *error perturbation analysis* was conducted in which PP from each participating SatPPM was recalculated with individual and simultaneous perturbations of the four input variables. Each perturbation represented addition or subtraction of the expected level of observational uncertainty associated with that variable. Comparison between SST, PAR and MLD measured on the Joint Global Ocean Flux Study equatorial Pacific Process Study cruises (a subset of the ClimPP database) and the input fields described above (Section 3.3) revealed that the following uncertainties were reasonable for these quantities: SST

Table 4

As in Table 3, except statistics are computed from the 1990s subset of the ClimPP data

Model # or type	Bias	F_{med}	RMSD_{CP}	Total RMSD	Correlation	Standard deviation
1	-0.11	0.77	0.17	0.20	0.79	0.13
2	0.01	1.02	0.16	0.16	0.78	0.24
3	-0.12	0.74	0.20	0.24	0.67	0.25
4	0.04	1.11	0.17	0.18	0.72	0.18
5	-0.26	0.54	0.17	0.31	0.76	0.23
6	-0.01	0.99	0.18	0.18	0.78	0.27
7	-0.09	0.82	0.21	0.22	0.63	0.23
8	-0.20	0.64	0.16	0.25	0.77	0.22
9	0.04	1.09	0.16	0.17	0.78	0.16
10	-0.36	0.43	0.17	0.40	0.72	0.18
11	-0.14	0.73	0.16	0.21	0.77	0.23
12	-0.19	0.65	0.23	0.30	0.79	0.37
13	-0.15	0.72	0.16	0.22	0.78	0.23
14	0.03	1.06	0.17	0.17	0.76	0.24
15	-0.08	0.82	0.16	0.18	0.78	0.18
16	-0.05	0.89	0.15	0.16	0.79	0.19
17	-0.02	0.97	0.15	0.15	0.79	0.21
18	-0.19	0.65	0.16	0.24	0.77	0.21
19	0.15	1.43	0.18	0.24	0.75	0.11
20	0.02	1.04	0.16	0.16	0.78	0.16
21	0.04	1.10	0.17	0.18	0.75	0.24
22	-0.33	0.47	0.28	0.43	0.59	0.34
23	-0.15	0.71	0.28	0.32	0.52	0.31
24	0.09	1.23	0.20	0.22	0.61	0.17
25	0.24	1.75	0.22	0.33	0.58	0.22
26	0.03	1.07	0.21	0.21	0.63	0.23
27	0.02	1.05	0.20	0.20	0.64	0.22
28	0.17	1.48	0.24	0.29	0.26	0.06
29	-0.06	0.87	0.27	0.27	0.44	0.26
30	0.17	1.48	0.22	0.28	0.46	0.12
DI/WI	0.13±0.03	1.35	0.18±0.01	0.24±0.02	0.75±0.01	0.23±0.02
DR/WI	0.08±0.03	1.22	0.16±0.01	0.19±0.01	0.77±0.01	0.21±0.02
DR/WR	0.08±0.03	1.20	0.16±0.01	0.19±0.02	0.77±0.01	0.19±0.02
SatPPM	0.11±0.02	1.29	0.17±0.01	0.22±0.01	0.76±0.01	0.21±0.01
BOGCM	0.14±0.03	1.38	0.23±0.01	0.28±0.02	0.53±0.04	0.22±0.03
All	0.12±0.02	1.31	0.19±0.01	0.24±0.01	0.69±0.02	0.21±0.01

Standard deviation of this subset of the ClimPP dataset is 0.25.

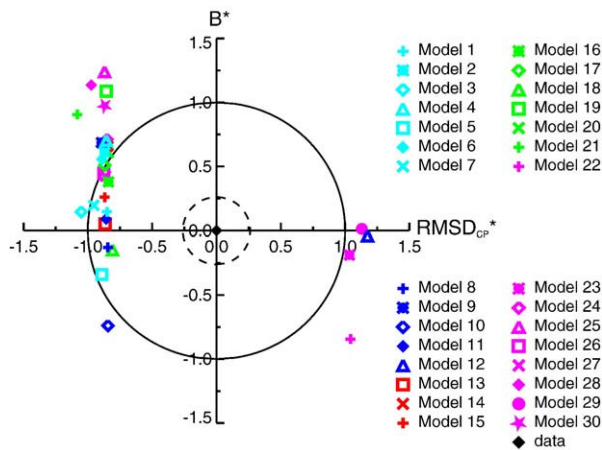


Fig. 3. Target diagram displaying B^* (Eq. (6)) and $\text{RMSD}_{\text{CP}}^*$ (Eq. (7)) for the 30 participating models relative to the ClimPP database. Concentric circles represent RMSD^* isolines: the inner dashed circle represents the normalized observational PP uncertainty, and the outer solid circle represents the normalized deviation of the productivity data. Blue symbols = DI/WI models (darker blue are subset of VGPM variants), red = DR/WI, green = DR/WR, and pink = BOGCMs.

uncertainty = ± 1 °C, PAR uncertainty = ± 10 mol quanta $\text{m}^{-2} \text{d}^{-1}$, and MLD uncertainty = ± 20 m. Uncertainty for chl_0 was estimated (Barber, pers. comm.) to range between $\pm 50\%$ for the minimum chlorophyll value, and $\pm 15\%$ for the maximum chlorophyll value as a linear function of $\log(\text{chl}_0)$. This resulted in chl_0 uncertainties ranging between ± 0.01 and 0.11 mg chl m^{-3} .

Because the four input variables each have three possible values (original value, original value + uncertainty, original value – uncertainty) there are a total of $3^4 = 81$ perturbations. The error perturbation analysis was conducted on these 81 perturbations of each of the ~ 1000 data points. The single perturbation yielding the lowest RMSD for each data point was then selected for an additional ‘best-case’ scenario assessment. This ‘best-case’ scenario RMSD was computed using a range of PP values based on the uncertainty in the ClimPP observations, which was assumed to vary as a linear function of $\log(\text{PP})$, between 30% of PP for the minimum observed productivities, and 10% of PP for PP greater than 0.7 g C $\text{m}^{-2} \text{d}^{-1}$. This corresponds to uncertainties ranging from ± 0.02 to ± 0.07 g C $\text{m}^{-2} \text{d}^{-1}$, with a mean uncertainty of ± 0.05 g C $\text{m}^{-2} \text{d}^{-1}$ (the mean of $\log \text{PP} - \log[\text{PP} - \text{uncertainty}] = 0.073$). These uncertainty levels, which likely underestimate actual measurement errors, are only used in the perturbation analysis to identify the ‘best-case’ scenario total RMSD (Section 4.4), but not in the statistics described in Tables 3 and 4.

In order to further quantify how the observational uncertainties described above affect our assessment of model skill, a principal component analysis (PCA) was performed. The PCA was conducted on the reductions in RMSD obtained from the 21 SatPPMs after (1) considering the uncertainties associated with the ClimPP data, (2) perturbing each input variable individually according to its respective uncertainty and (3) considering the uncertainties in the ClimPP data as well as all the input data fields simultaneously. This PCA ordination reveals how each model responds to the accuracy criteria that reflect the uncertainty in field data due to sampling and analytical errors.

4. Model skill comparison results

Bias, F_{med} , RMSD_{CP} , total RMSD, correlation and standard deviation were computed for each of the 30 participating models and are reported in Table 3. Mean values \pm one standard error are included for all 30 models, as well as for each model type (DI/WI, DR/WI, DR/WR, SatPPM, and BOGCM). These results are discussed below and presented in a variety of graphical representations.

4.1. Target diagram

Following Jolliff et al. (2009–this volume) the bias and RMSD_{CP} of the 30 participating models were normalized by the standard deviation of the log-transformed ClimPP data (σ_d), and plotted on a target diagram (Fig. 3), where, as defined by Eq. (2), the distance from the origin to each symbol represents the total normalized RMSD:

$$(\text{RMSD}^*)^2 = (B^*)^2 + (\text{RMSD}_{\text{CP}}^*)^2 \quad (8)$$

On this plot the inner dashed circle denotes the normalized observational PP uncertainty ($\text{RMSD}^* = 0.073/\sigma_d = 0.26$). Thus the skill of any models yielding points falling within this inner circle (the bulls-eye) would be indistinguishable from each other. This, however, is a difficult goal to achieve for patchy quantities such as ocean productivity; in this analysis no points are located within this circle.

Another useful way to characterize the results of multiple models on the target diagram (Jolliff et al., 2009–this volume) is to examine whether or not the standard deviation of the observations exceeds the total RMSD for each model. If it does not, i.e. if $\text{RMSD}^* > 1$, then the model is showing less skill (in terms of total RMSD) than does the mean of the observations.

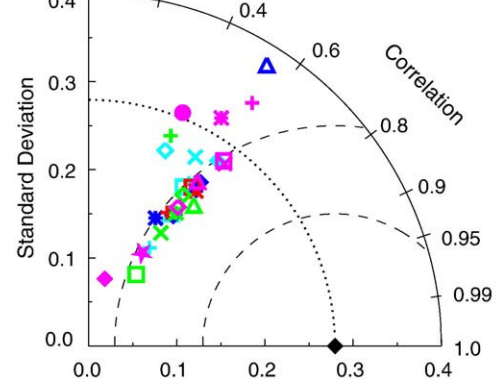
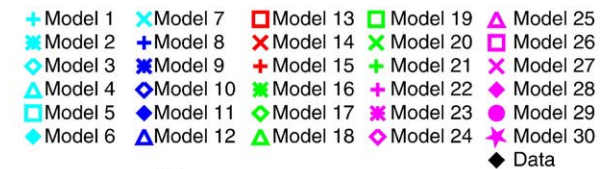


Fig. 4. Taylor diagram of $\log(\text{PP})$. The distance from the origin is the standard deviation of the modeled productivities. The azimuth angle represents the correlation between the observations and the modeled productivities, and the distance between each model symbol and the data (black diamond) is the RMSD_{CP} . Dashed lines are isolines of $\text{RMSD}_{\text{CP}} = 0.25$ and $\text{RMSD}_{\text{CP}} = 0.15$. Dotted line represents the standard deviation of the data. Colors as in Fig. 3.

Model results falling inside the outer circle, i.e. those with total RMSD values that are less than the standard deviation of the observations, tend to provide a better instantaneous estimate of productivity than the mean of the observations. In this analysis, eleven of the 30 models (37%) fell within this outer circle (Fig. 3). These include 42% of the DI/WI models, 67% of the DR/WI models, 33% of the DR/WR models and 22% of the BOGCMs. The model with the lowest normalized total RMSD ($\text{RMSD}^*=0.82$) is Model 18, a DR/WR model. Four models have nearly equally low RMSD: two VGPM variants (Models 8 and 11), a depth-resolved model (Model 13) and the simplistic $\text{sqrt}(\text{chl}_0)$ relationship (Model 1). The performance of the latter model, which does not use PAR, SST or MLD data, illustrates the importance of chl_0 in the estimation of integrated productivity.

The target diagram (Fig. 3) also illustrates that most (23 out of 30) models overestimated observed productivity ($B^*>0$), whereas only 23% (7 out of 30) models underestimated productivity. Three models (one SatPPM #19, and two BOGCMs #25, 28) had positive median biases greater than two ($F_{\text{med}}>2.0$; $B>0.3$; Table 3), i.e. in these cases the median modeled value was more than twice the median observed value. Nearly half of the BOGCMs (four of nine) were associated with absolute biases greater than a factor of 1.7 ($F_{\text{med}}>1.7$ or $F_{\text{med}}<0.59$), whereas only two of the 21 SatPPMs, both DR/WR models, were associated with biases of this magnitude. Interestingly, the overall magnitude of the mean bias was lower for the DI/WI models (0.10 ± 0.02) than for the DR/WR models (0.17 ± 0.04) and the BOGCMs (0.19 ± 0.02).

The centered pattern RMSD (RMSD_{CP}) provides a measure of how well the variability of a certain field is being modeled. Whereas bias for these models varied substantially, the magnitude of RMSD_{CP} diverged much less among models (Table 3), ranging only from $|\text{RMSD}_{\text{CP}}^*|=0.84$ to 1.13 (Fig. 3). The RMSD_{CP} was slightly, but not significantly, higher for the BOGCMs ($|\text{RMSD}_{\text{CP}}^*|=0.95\pm 0.04$) than for the SatPPMs ($|\text{RMSD}_{\text{CP}}^*|=0.90\pm 0.02$). There was also little significant difference in $|\text{RMSD}_{\text{CP}}^*|$ for the different categories of SatPPMs. All but four models (three BOGCMs and one VGPM variant) underestimated the variance of observed productivity ($\text{RMSD}_{\text{CP}}^*<0$).

4.2. Taylor diagram

Taylor diagrams (Taylor, 2001) complement the target diagrams described above by representing the RMSD_{CP} , the correlation, and the standard deviation of each of the 30 PP models with respect to the ClimPP dataset (Fig. 4). Unlike the target diagram, the Taylor diagram presents no information on bias, however it has the advantage of being able to illustrate more detailed information pertaining to the difference in variability associated with the modeled and observed fields (Section 3.4).

In the Taylor diagram (Fig. 4), the distance between the black data symbol on the x -axis and each colored model symbol represents RMSD_{CP} . This RMSD_{CP} is composed of a component relating to the variance of the model estimates (the distance from the origin is the standard deviation) and a component relating to the correlation between the observations and model estimates (the azimuth angle). Although the RMSD_{CP} for each model has already been shown on the target diagram (Fig. 3), it

is not possible to determine from that diagram whether a given RMSD_{CP} results from getting the correlation wrong or from getting the magnitude of the variability wrong.

For the 30 participating models, the correlation coefficients (azimuth angles) are very similar for many of the models: all but six models are correlated with the observations with a

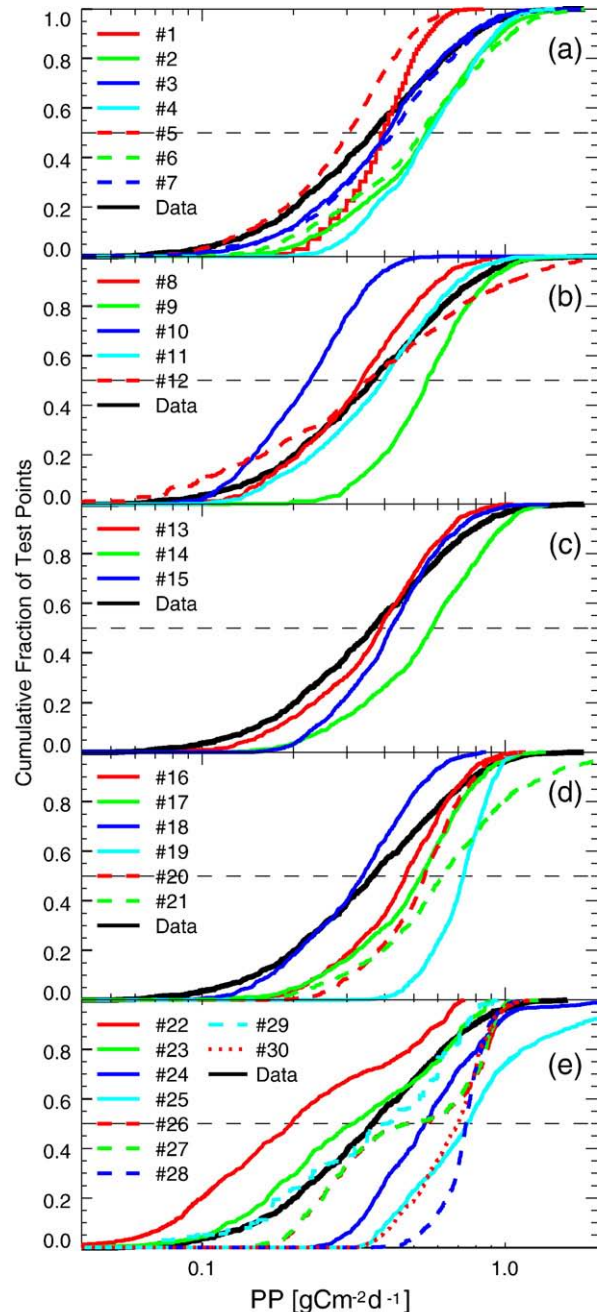


Fig. 5. Cumulative fraction of test points as a function of productivity for the ClimPP observations (black lines) and the 30 PP models (colored lines) for (a) the DI/WI non-VGPMs, (b) the DI/WI VGPMs, (c) the DR/WI models, (d) the DR/WR models, and (e) the BOGCMs. Results of Model #26 are nearly indistinguishable from those of Model #27. Medians are represented by the intersections of the black dashed line and the curves.

coefficient between 0.5 and 0.6 (Fig. 4; Table 3). Of the four models with correlation coefficients less than 0.4, two are BOGCMs, one is a DI/WI model and one is a DR/WR model. Model 18, a DR/WR model, yields the highest correlation.

The distance from the origin to each model point is the standard deviation of the modeled productivities; points closer to the origin than the dotted line underestimate the variance of the data, and those outside the dotted line overestimate this variance. As was evident from the target diagram (Fig. 3) most models underestimate the observed variance, however the Taylor diagram (Fig. 4) illustrates that the amount by which these models underestimate the observed variance varies considerably from model to model, and independently of correlation or RMSD. In particular, the simplistic Model 1, which was shown to produce a very low RMSD value on the target diagram, is shown here to substantially underestimate the observed variance of the data (but less so than Models 19, 28 and 30 which are relatively complex). BOGCMs, with the exception of Models 24, 28 and 30, typically do slightly better than SatPPMs at reproducing the magnitude of the observed variance (i.e. are closer to the dotted line in Fig. 4), although the difference is not statistically significant.

4.3. Cumulative distribution functions

The cumulative distribution of ~1000 test points allows a comparison between the range and median of the PP observations and those obtained from the 30 participating PP models. Most of the DI/WI models (Fig. 5a,b) underestimate the range of the ^{14}C PP data, which extends roughly from 0.1 to $1.0 \text{ g C m}^{-2} \text{ d}^{-1}$. Whereas the lower PP values are substantially overestimated by some of the DI/WI models (e.g. #1, 2, 4, 6, 9), other models tend to underestimate the higher PP values (e.g. #1, 5, 8, 10). Only Model 12 overestimates the range of PP values, and it does so both by underestimating the minimum PP values and overestimating the maximum PP values. Not surprisingly, without any SST, PAR or MLD dependence, Model 1 (i.e. $\text{PP} = \sqrt{\text{chl}}$) produces a range of PP values that is much too small; however, this model reproduces the median of the PP data better than any other DI/WI model, with the exception of Model 12 (which interestingly has the largest range). For this DI/WI subset of models, #11, a VGPM variant with an alternate

definition of the 1% light level, provides the best overall fit to the cumulative fractional distribution of the test points.

In general, the depth-resolved models (DR/WI and DR/WR) do not reproduce the observed range and median of the ^{14}C PP data (Fig. 5c,d) any better than the depth-integrated models (Fig. 5a,b). Eight of the nine DR models overestimate the median PP: #14 by more than 50% and #19 by nearly 100%. This mismatch occurs because these models generally produce very few PP values less than $0.2 \text{ g C m}^{-2} \text{ d}^{-1}$, whereas nearly 20% of the PP observations are less than this value. The WR models (#16–21) also do not reproduce the observed range of PP any better than their WI counterparts (#1–15).

The BOGCM results (Fig. 5e) are more variable than those computed from the SatPPMs, which is to be expected since the BOGCMs do not use the chl_0 input fields provided to the participants and used by the SatPPMs. Moreover, the BOGCMs have different vertical/horizontal resolutions and use different mixing parameterizations, which may further contribute to these deviations. Even so, as was the case for the SatPPMs, most of the BOGCMs produce too few low PP values. The minimum PP for Models 24, 25, 28 and 30 ($0.22, 0.32, 0.35$ and $0.32 \text{ g C m}^{-2} \text{ d}^{-1}$, respectively) are only slightly lower than the median of the observations ($0.37 \text{ g C m}^{-2} \text{ d}^{-1}$). This strong positive bias evident for many of the BOGCMs may result from the tendency of coarse resolution general circulation models to upwell excessive amounts of nutrients in the eastern tropical Pacific (Gnanadesikan et al., 2002; Doney et al., 2004), and thus overestimate these lower productivity observations. The excessive nutrient flux greatly expands the moderately productive HNLC area. On the other hand, the BOGCMs are less biased for the high productivity observations than the low PP observations, because in most of these models excessive nutrient flux will result in a community shift towards diatoms, which will increase export more strongly than primary production.

Not all of the BOGCMs are associated with positive biases. Two BOGCMs over-predict the number of low PP values: Model 22 yields 50% of its results lower than 0.2, whereas only 20% of the observations are below this value PP. Other BOGCMs do quite well: the cumulative distributions of Models 23 and 29 agree with the observations better than many of the SatPPMs. The skill of Model 29 in this respect is particularly interesting, as this model uses climatological rather than interannual forcing

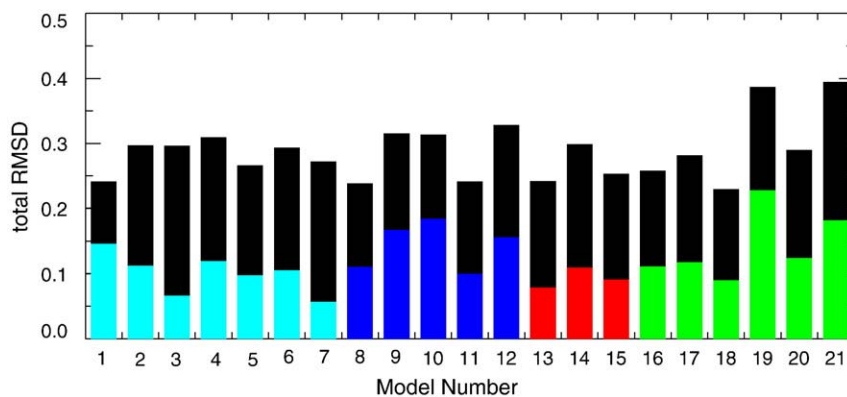


Fig. 6. Total RMSD between each of the satellite-based PP models and the ClimPP database. The portion of the total RMSD that may be attributable to uncertainty in the input variables is black; colors defined in Fig. 3.

fields. Overall, the performance of the BOGCMs varies substantially, and as a group of models they do not match the range and median of the observed PP any better or worse than the SatPPMs.

4.4. Error perturbation analysis

Measurements of the input variables (SST, MLD, PAR, chl₀) used by the SatPPMs are not error-free, and the uncertainty in their values affects the ability of the models to optimally reproduce the ¹⁴C PP measurements, which are also associated with some degree of uncertainty. To examine the effect of these uncertainties, an error perturbation analysis (Section 3.4) was conducted for the SatPPMs (Fig. 6). The maximum reduction in RMSD resulting from perturbations of each of the four input variables individually, PP, and all five quantities simultaneously (the best-case scenario perturbation) is also reported in Table 5.

When the input variables were independently perturbed, not surprisingly most (two-thirds) of the SatPPMs showed larger reductions in model–data misfit with perturbations of chl₀, than with PAR, MLD or SST (Table 5). Although chl₀ was perturbed by less than 50%, which may be less than the uncertainty in satellite chlorophyll algorithms, the decrease in RMSD averaged 25%. This is consistent with the results of Campbell et al. (2002), which demonstrated that errors in surface chlorophyll produced less than proportionate errors in integrated productivity. Model 12 demonstrated the greatest sensitivity to perturbations in chl₀, probably because this VGPM variant computes P_{opt}^B as a function of chl₀ alone, with no input from SST. On the contrary, Model 10, a VGPM variant that computes P_{opt}^B as a function of both SST and chl₀, showed one of the smallest sensitivities to perturbations in chl₀, demonstrating only 13% improvement in RMSD.

With this type of error analysis, it is possible to estimate the maximum fraction of the total RMSD that may be attributed to uncertainties in the input variables and in the PP data themselves. For example, as noted earlier, the model with the lowest total RMSD is Model 18 (RMSD=0.23; Fig. 6). However, when uncertainties in the input variables and PP data are taken into account, the best-case scenario RMSD for this model drops by 60% to the very small RMSD=0.09.

Certain models are inherently more sensitive to changes in input variables than others (Fig. 6), and therefore their PP estimates are more affected by uncertainties in these variables. Since Model 1 uses only one of the provided input fields (chl₀), it is affected only by uncertainties in that field and PP, and not surprisingly, less of its total RMSD can be attributed to these uncertainties than most other models, although even in this case the fraction is quite large (39%). The relatively low original RMSD may partially be explained by the fact that, by only using one of the input fields provided, it is unaffected by uncertainties in the other fields. These data uncertainties inherently have a greater effect on the skill of the more comprehensive models that incorporate information from all the provided fields: for example 78% of the total RMSD of Model 3 can be attributed to uncertainties in the data. On average, uncertainties in the input variables and the PP data can account for more than half (mean=58%; range=39–79%) of the total RMSD computed for the 21 SatPPMs. If the uncertainties associated with measurements of chl₀, SST, PAR and MLD are reduced in the future as

Table 5

Maximum possible reductions in total RMSD (and maximum percentage reduction in total RMSD) resulting from perturbing chl₀, SST, PAR, MLD, PP and all five variables simultaneously

Model #	chl ₀ (%)	SST (%)	PAR (%)	MLD (%)	PP (%)	All (%)
1	0.051 (21)	0.000 (0)	0.000 (0)	0.000 (0)	0.050 (21)	0.094 (39)
2	0.078 (26)	0.006 (2)	0.086 (29)	0.000 (0)	0.053 (18)	0.183 (61)
3	0.086 (29)	0.008 (3)	0.075 (25)	0.092 (31)	0.052 (17)	0.227 (76)
4	0.062 (20)	0.000 (0)	0.067 (22)	0.037 (12)	0.054 (17)	0.188 (61)
5	0.070 (26)	0.008 (3)	0.043 (16)	0.036 (13)	0.049 (18)	0.167 (62)
6	0.082 (28)	0.014 (5)	0.082 (28)	0.000 (0)	0.053 (18)	0.187 (64)
7	0.045 (17)	0.000 (0)	0.065 (24)	0.108 (40)	0.051 (19)	0.212 (79)
8	0.066 (28)	0.017 (7)	0.012 (5)	0.000 (0)	0.049 (20)	0.126 (53)
9	0.073 (23)	0.021 (7)	0.015 (5)	0.000 (0)	0.054 (17)	0.146 (46)
10	0.040 (13)	0.049 (16)	0.010 (3)	0.000 (0)	0.048 (15)	0.128 (41)
11	0.065 (27)	0.038 (16)	0.013 (5)	0.000 (0)	0.051 (21)	0.140 (58)
12	0.122 (37)	0.000 (0)	0.012 (4)	0.000 (0)	0.052 (16)	0.171 (52)
13	0.072 (30)	0.006 (3)	0.069 (29)	0.000 (0)	0.051 (21)	0.161 (67)
14	0.068 (23)	0.019 (6)	0.088 (29)	0.000 (0)	0.054 (18)	0.188 (63)
15	0.068 (27)	0.018 (7)	0.059 (24)	0.000 (0)	0.051 (20)	0.161 (64)
16	0.065 (25)	0.001 (0)	0.052 (20)	0.000 (0)	0.052 (20)	0.144 (55)
17	0.076 (27)	0.010 (4)	0.058 (21)	0.000 (0)	0.053 (19)	0.163 (58)
18	0.059 (26)	0.000 (0)	0.058 (25)	0.000 (0)	0.049 (21)	0.138 (60)
19	0.044 (11)	6E-06 (0)	0.065 (17)	1E-6 (0)	0.059 (15)	0.158 (41)
20	0.061 (21)	0.009 (3)	0.066 (23)	0.004 (1)	0.054 (19)	0.165 (57)
21	0.111 (28)	0.000 (0)	0.066 (17)	0.014 (4)	0.053 (14)	0.211 (54)

This data matrix is the basis for the principal component analysis (Fig. 7).

observations of these quantities become more accurate, the skill of Models 2–21 would be expected to increase relative to that of the simplistic Model 1.

4.5. Principal component analysis of reductions in RMSD

An R-mode principal component analysis (PCA) was performed to further examine how individual and simultaneous perturbations of the input variables and ClimPP observations reduce the total RMSD between observed and predicted PP for each SatPPM. The PCA was computed on the reductions in RMSD that resulted from individual perturbations of chl₀, SST, PAR, MLD, PP and simultaneous perturbations of all five quantities (Table 5). Eighty-one percent of the variance of reduction in RMSD is explained by the first two principal components arising from such an analysis, with 60% of the variance explained by PC1 and 21% explained by PC2. In Fig. 7 the original total RMSD associated with each model, assuming the provided input variables and ClimPP data have no associated uncertainty (Table 3), is displayed as the symbol size (better fitting models are represented by smaller circles); vectors represent variables and they point towards the largest decrease in total RMSD associated with uncertainties in that specific variable. Thus models whose projection onto a specific vector fall near or beyond the head of that vector tend to be those that are most affected by uncertainties in that variable. As several variables are simultaneously considered, the order of the models' projections onto vectors cannot exactly match the reductions in RMSD, but it can be regarded as the best overall approximation.

The simultaneous perturbation of 'all' variables (Fig. 7) is strongly correlated with the first principal component (PC1), as is illustrated by the fact that the angle between the 'all' vector and PC1 is very small. Therefore, models with the

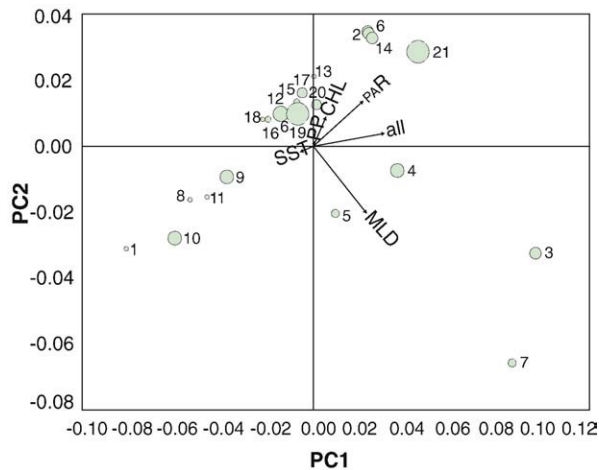


Fig. 7. Principal component analysis of reduction in total RMSD obtained from optimal perturbation of observed input and output variables, where perturbations represent hypothetical levels of uncertainty in each variable. Bubble size is proportional to the original total RMSD (i.e. assuming perfectly known input variables). Vectors point towards the largest reductions in total RMSD obtained assuming different uncertainty criteria (all = uncertainty in all variables; PP = uncertainty in PP only; SST = uncertainty in SST only; etc.).

highest values of PC1 (e.g. #3, 4, 7, 21) are those for which the performance is most improved when uncertainties in all the input and output variables are considered, i.e. these models are generally the most sensitive to sampling and analytical errors. On the contrary, models that are located in the far left hand quadrants (negative values of PC1; e.g. #1, 8, 10) tend to be less affected by the data perturbation and are also generally the best fitting models (smallest original RMSD assuming no data perturbations).

The vectors describing perturbations of MLD and PAR are the longest vectors, indicating that perturbations in these input variables produce the most diverse responses among the various models. Perturbations in PAR and MLD yield substantial reductions in RMSD for some models, and very little such reductions for other models. For example, the skill of Models 3 and 7 are strongly affected by inaccuracies in MLD, whereas the skills of Models 2, 6, 14, and 21 are more strongly affected by inaccuracies in PAR and chl_0 . It is interesting to note that the vectors that represent PAR and MLD are orthogonal, i.e. independent of each other in the PCA. This implies that improvements in RMSD due to PAR data perturbations tend to be independent of those obtained by perturbing MLD data. Both MLD and PAR strongly contribute (by combining their effects in a way that is similar to the combination of two orthogonal forces) to the reduction in RMSD obtained by perturbing all variables.

Vectors describing perturbations of chl_0 and SST are much shorter, reflecting that the effects of their perturbation on the differences between models as shown in the ordination are minor, or in other words these variables affect most models in a similar way. Similarly, PP data perturbation has a virtually negligible role in discriminating models according to their sensitivity. All vectors, except the one representing SST, point towards the positive end of PC1. This implies that their effects, although not necessarily correlated, contribute to the reduction of the RMSD of all the models that are located in the right

part of the ordination. SST, on the contrary, points towards the left part of the ordination, nearly in opposition to PAR. This means that PAR and SST perturbations tend to have opposite effects on the reduction of RMSD: models that are improved by perturbing SST are generally insensitive to perturbations in PAR, and models that are improved by perturbing PAR are relatively insensitive to perturbations in SST. However, the effect of SST perturbations on RMSD reduction is considerably smaller than that of the PAR perturbations (Table 5).

5. Correlations of model–data misfit with environmental variables

Although documentation of the relative skill of these 30 productivity models in the tropical Pacific is useful, model improvement requires not only an assessment of model skill, but also an understanding of the particular environmental conditions that affect performance. To address this issue, a correlation analysis was undertaken to better characterize model performance in the context of environmental forcing. Specifically, the correlation was estimated between model–data misfit (Δ) and ten variables (Fig. 8), including the input variables (SST, chl_0 , PAR and MLD) as well as longitude, degrees from the equator, year day, year, the depth of the chlorophyll maximum, and biomass-normalized PP integrated over the euphotic zone (P_b^B).

As expected, correlations with surface chlorophyll are relatively high for the 21 SatPPMs. Model–data misfits in all but one of these models are positively correlated with chl_0 , indicating that modeled PP overestimates measured PP when surface chlorophyll concentration is high, and modeled PP underestimates measured PP when surface chlorophyll is low. Because the SatPPMs by definition only utilize chlorophyll data from the surface layer, one might also expect an equally high negative correlation with the depth of the chlorophyll maximum, i.e. that the models would underestimate productivity

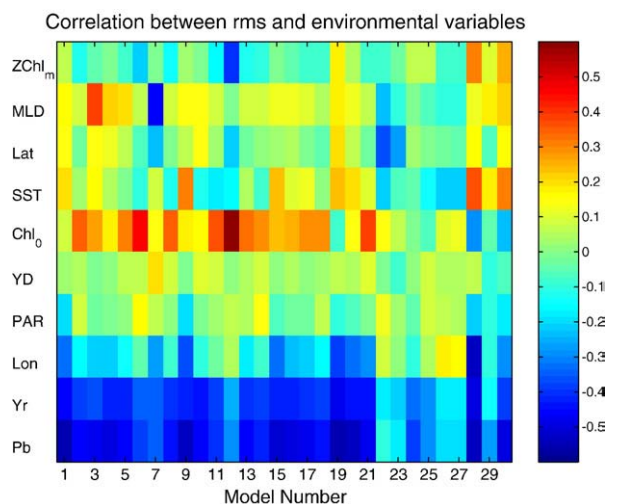


Fig. 8. For each of the 30 participating models, correlation between model–data misfit and various environmental variables: depth of the chlorophyll maximum ($ZChl_m$), MLD, degrees from equator (Lat), SST, chl_0 , year day (YD), PAR, longitude (Lon), year (Yr), and biomass-normalized PP integrated over the euphotic zone (P_b).

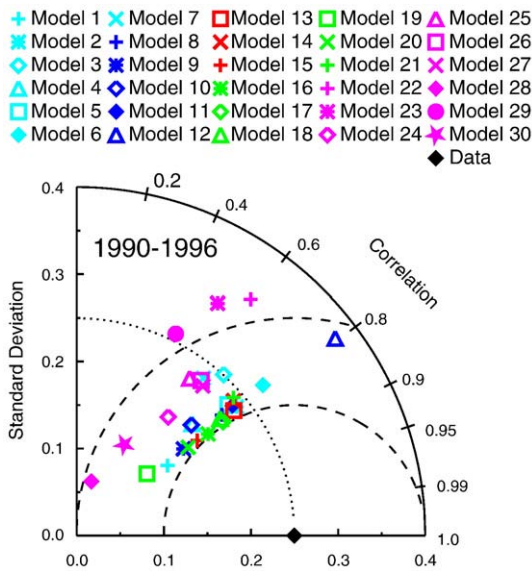


Fig. 9. As in Fig. 4, except statistics are computed for the 1990s subset of the ClimPP data.

when the maximum chlorophyll concentration occurred below one optical depth. Because large parts of this region are characterized by deep chlorophyll maxima (Barber et al., 1996; Le Borgne et al., 2002), the ClimPP dataset provides a good test of the impact of chlorophyll vertical structure on model misfit. However, Fig. 8 illustrates that the correlation between misfit and the depth of the chlorophyll maximum is generally less than 0.1. Only Model 12 produced misfits that were strongly correlated to the depth of the chlorophyll maximum, overestimating the ClimPP data when the maximum chlorophyll concentration was at the surface and underestimating the ClimPP data when the maximum chlorophyll concentration was deep. This latter behavior is what would be expected from models that use only a surface value, and has been a concern for the use of ocean color. The low correlation between misfit and depth of the chlorophyll maximum for the bulk of the models suggests that model assumptions typically used to characterize the depth profile of chlorophyll are fairly sound, and/or that the variations of PP in the deep chlorophyll maximum do not contribute strongly to vertically integrated PP.

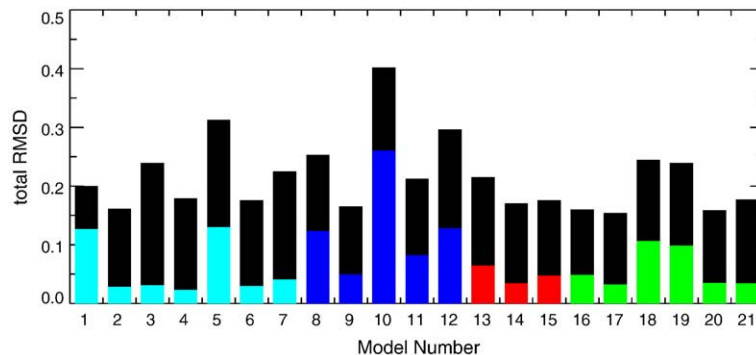


Fig. 10. As in Fig. 6, except RMSD is computed for the 1990s subset of the ClimPP data.

Somewhat surprisingly, for all of the participating models but especially the SatPPMs, model–data misfit shows a strong inverse correlation with year, i.e. the models overestimated the ClimPP data in the earlier years. A comparison of the RMSD statistics computed for the 1990–1996 time period ($N=340$; Table 4) versus the entire time period (1983–1996; $N=948$; Table 3) reveals that although the bias is similar for these two time periods, the RMSD_{CP} is substantially less for the SatPPMs in the 1990s. Specifically, the correlation between the SatPPM estimates of PP and the ClimPP data (Fig. 9) is higher for the 1990s data (mean for all SatPPMs = 0.76 ± 0.01) than it is for the entire dataset (0.52 ± 0.01). This is true even for the simplistic Model 1, implying that the correlation between surface chlorophyll and integrated productivity was stronger in the 1990s than it was in the 1980s.

Not only is total RMSD reduced for the SatPPMs (0.29 ± 0.01 to 0.22 ± 0.01), but the average fraction of this total RMSD that may be related to uncertainties in input variables and PP data also increased from 58% to 68% after 1990 (Fig. 10). For seven SatPPMs, in the best-case scenario up to 80% of the RMSD for the 1990s may be attributed to uncertainties in the input variables and ClimPP data. The improved performance of the models since 1990 could reflect the fact that some of the participating SatPPMs and many of the BOGCMs were developed using data acquired after 1990. However, this was not true for many of the SatPPMs; the VGPM variants, for example, were based primarily on in situ data collected prior to the 1990s (Behrenfeld and Falkowski, 1997b).

Model–data misfit was not only highly correlated with year, but was similarly highly correlated with observations of P_1^{B} (Fig. 8). To further investigate these correlations, model–data misfit is plotted as a function of P_1^{B} data and time (Fig. 11). Consistent with the correlation results, all thirty models yield a clear trend of overestimating PP for low P_1^{B} values and underestimating PP when P_1^{B} is high; absolute misfit ($|\Delta|$) typically is greatest for the lowest P_1^{B} values. Furthermore, the oldest data (from the early and mid-1980s) are generally characterized by relatively low P_1^{B} values ($P_1^{\text{B}} < 2.5 \text{ mmol C mg chl}^{-1} \text{ d}^{-1}$), whereas P_1^{B} measured from 1995 and 1996 exceed $2.5 \text{ mmol C mg chl}^{-1} \text{ d}^{-1}$ (Fig. 12a). This change in P_1^{B} holds over the entire ClimPP domain, not just the equatorial zone (Fig. 12b,c). Thus, the surprisingly high correlations between year and misfit occur because: (1) all models, both the SatPPMs and the BOGCMs, perform poorly when P_1^{B} is low ($< 1 \text{ mmol C mg chl}^{-1} \text{ d}^{-1}$), and (2) 95% of these low P_1^{B} values

occur in the 1980s (although there are more data points poleward of 10° in the 1980s than the 1990s (Fig. 12c), even when only points between 10°N and 10°S are included, 95% of the $P_1^B < 1 \text{ mmol C mg chl}^{-1} \text{ d}^{-1}$ values still occur in the 1980s portion of the ClimPP dataset). It is possible that the improved

performance of the models since 1990 is a reflection of inherent problems associated with modeling low photosynthesis rates for a given biomass.

The interdecadal variability in productivity per unit biomass directly results from an increase in measured

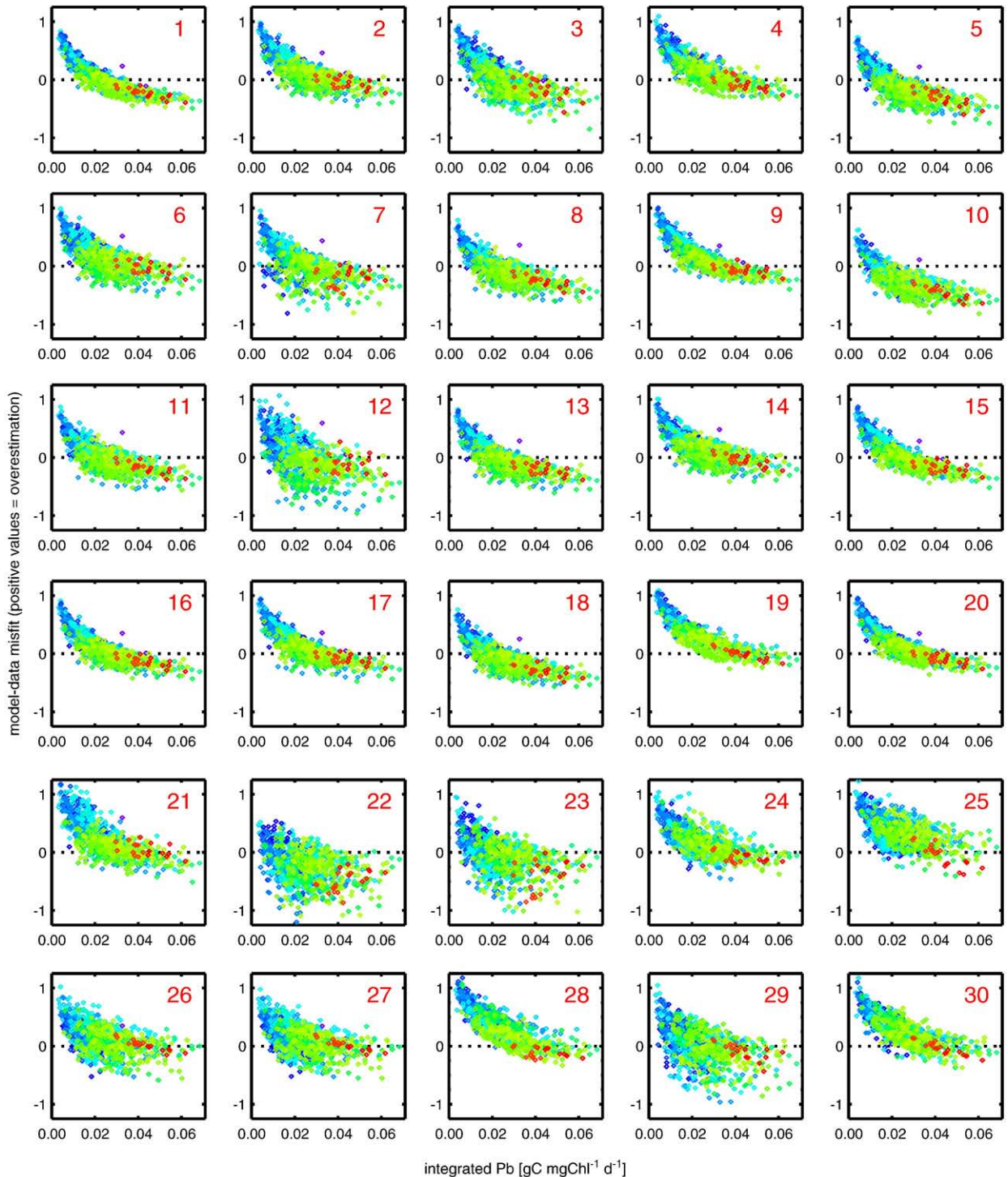


Fig. 11. Model–data misfit as a function of observations of P_1^B , color-coded to demonstrate that low P_1^B values occurred in the 1980s. Model numbers in the upper right corner of each plot correspond to definitions in Table 2.

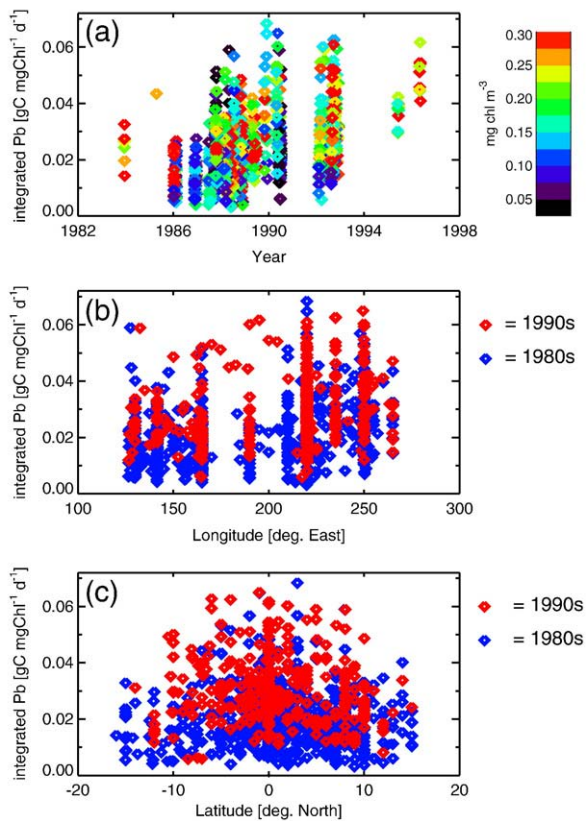


Fig. 12. P^B plotted as a function of (a) time, color-coded to demonstrate that low P^B values are not correlated with chlorophyll, (b) longitude, color-coded according to decade, and (c) latitude, again color-coded to decade.

productivity: mean productivity increased by as much as 62% between these two decades. Biomass-normalized productivity, on the other hand, increased by a smaller percentage (47%) since mean integrated chlorophyll for the 1990s was also slightly higher (8%) than that observed in the 1980s. Interestingly, although average integrated chlorophyll was lower in the 1980s, mean surface chlorophyll, which is what SatPPMs use to estimate PP, was slightly (10%) higher in the 1980s than it was in the 1990s. As a result, nearly all the SatPPMs overestimated PP in the 1980s. As an aside we note that over the entire ClimPP domain there was no significant difference in SST between pre-1990 and post-1990; however, in the eastern equatorial Pacific, the 1990s time period was associated with colder temperatures.

These trends and interdecadal differences do not reflect any change in observational methodology. As discussed above, the entire ClimPP database was collected using the same trace-metal clean procedures (Chavez and Barber, 1991; Sanderson et al., 1995). In no instances did the Go-Flo bottles contain black rubber closing springs, which were found to be the major source of trace metal contamination (Fitzwater et al., 1982). In addition, the entire database was reprocessed using the same model for estimating PAR attenuation (Morel, 1988) and the same method for normalizing light extinction estimates in all of the ClimPP observations.

The interdecadal changes documented above are consistent with findings from other recent studies of interdecadal

change in this region. The Pacific Decadal Oscillation regime shift between 1988 and 1992 has been associated with major physical, chemical and biological changes in this ocean basin (Hare and Mantua, 2000). For example, Chavez et al. (2003) proposed multidecadal regime shifts in the eastern tropical Pacific Ocean, consisting of a warm phase with high landings of sardines followed by a cool phase with high landings of anchovies. Chavez et al. (2003) hypothesized that the regime shift occurred around 1990 when the landings of sardines began to decrease and a cooling of SST was observed.

In another study, Takahashi et al. (2003) present equatorial (5°N – 5°S) observations that demonstrate a decrease in $p\text{CO}_2$ from 1979 to 1990 followed by an increase in seawater $p\text{CO}_2$ between 1990 and 2001. Similarly Feely et al. (2006) document a decrease in SST in the central equatorial Pacific and a simultaneous increase in the rate of the 28°C $f\text{CO}_{2\text{sw}}$ during the 1990s, again suggestive of an increasing influx of colder and CO_2 -rich waters into this region. They note that these trends could result from either increased lateral flow of colder subtropical waters driven by equatorward winds or increased upwelling of colder deep waters. The increase in P^B documented here is more consistent with the latter scenario, since an increased rate of upwelling would bring high-nutrient iron-rich water to the surface, possibly resulting in a change in community structure and phytoplankton species composition, and generating the higher rates of biomass-normalized productivity documented here. The absence of lower mean SST after 1990 is attributable to the large areal extent over which the data extend (30° of latitude and nearly 150° of longitude) and thus is not inconsistent with an increased upwelling scenario.

Because the interdecadal change in productivity was not associated with a significant and coincident change in any of the input variables (chl_a , SST, PAR or MLD), it is not surprising that the SatPPMs failed to reflect these changes in PP. This reflects an important characteristic of SatPPMs in general. Regardless of complexity, i.e. depth/wavelength resolved or integrated, these models all incorporate empirical information regarding the relationship of physiological state with environmental conditions. If changes in the rate of upwelling at depth alter deep nutrient concentrations without effecting a coincident change in SST or MLD, a scenario which is likely to occur for regions of shallow MLD as is frequently the case in the tropical Pacific, the skill of SatPPMs is likely to be less. Although BOGCMs in principle have the capacity to model such interdecadal changes in upwelling and the resultant effects on nutrient concentrations at depth, the models submitted to this study have not demonstrated the ability to do this in this case. It is possible that complexities in community structure missing in these BOGCMs may be required in order to understand the relative biomass invariance. These results demonstrate that temporal variability in the form of interdecadal regime shifts pose a significant challenge for both SatPPMs and BOGCMs alike, and further suggest that other changes associated with scenarios of global warming are likely to be equally problematic.

6. Discussion and summary

Thirty models, including both SatPPMs and BOGCMs, participated in this PPARR comparison exercise. PP model skill was assessed by computing root mean square differences

(RMSD) and by using a variety of skill assessment tools to determine how well each model reproduced \log_{10} PP from an extensive in situ tropical Pacific database (ClimPP). This region of the ocean was chosen specifically because earlier PP modeling efforts demonstrated that the tropical Pacific represents a significant challenge for satellite-based PP models.

A primary result of this model intercomparison effort is that although skill varied substantially among the participating models, model skill was not associated with model complexity or model type. Overall, there was little significant difference in the performance of the three different classes of SatPPMs represented here (based on depth and wavelength resolution). Also, neither SatPPMs nor BOGCMs performed significantly better based on the full suite of skill assessment methodologies, revealing that using output from SatPPMs to evaluate BOGCM simulations of PP is not an adequate test of BOGCM skill.

Most models substantially underestimated PP variability, often by more than a factor of two. Specifically many of the models, including all 9 depth-resolved satellite-based models (though Model 18 to a much lesser degree), produced too few low PP values ($<0.2 \text{ g C m}^{-2} \text{ d}^{-1}$). PP model developers need to be aware of this model shortcoming, and should note that an improvement in the ability of productivity models to estimate PP at low rates of productivity will have a significant impact on PP model skill in the tropical Pacific and other oligotrophic regions across the globe.

Whereas nearly all models underestimated PP variability, model performance differed substantially in terms of how well the models reproduced the mean and median PP. Several models overestimated the median productivity by a factor of two, others underestimated productivity by nearly a factor of two, and others presented almost no bias. Thus, the models with the greatest skill were generally those with the lowest bias. PPARR2 concluded that removing bias was an immediate goal to improve model performance. The present study confirms that several participating models have accomplished this goal. Although, as noted above, in general there was little difference in the performance of different classes of models, one exception is in terms of bias: the simplest models were characterized by significantly less bias than those that resolved wavelength and depth.

In a best-case scenario, more than half of the total RMSD associated with the SatPPMs could be attributed to uncertainties in input variables and the PP data. Whereas the skill of certain models could increase dramatically if the input variables were known more accurately (e.g. many of the models that use the MLD fields), the skill of other models, including the simple square root of chlorophyll relationship that uses only the chl_0 input field, would not undergo the same level of improvement. In general, the subset of models based on the Vertically Generalized Productivity Model (Behrenfeld and Falkowski, 1997b) show a relatively small improvement in skill when uncertainties in input variables and PP data are considered, primarily because of their inherent insensitivity to PAR. On the contrary, the skills of PP models that utilize MLD in their computations of productivity are severely limited by uncertainties in MLD estimates, suggesting that further development of models that require MLD input fields is not advised, especially since high quality MLD fields are not readily available.

The specific models that demonstrate the greatest skill vary depending on which skill assessment methods are used. For instance, if total root mean square difference is used as the single criteria for skill, Model 18 (Mélín, 2003), which uses biogeographical provinces to define the necessary model parameters, performs best. Model 1, the simplistic $\sqrt{\text{chl}_0}$ model, performs nearly as well according to this criterion, however this is partially due to the fact that models that underestimate variance result in lower total RMSD scores (Jolliff et al., 2009-this volume). Ten of the 30 models are associated with biases that are smaller than the levels of PP uncertainty we assume here. One of the BOGCMs (Model 29: Tjiputra et al., 2007) produces the lowest bias, however one of the VGPMs (Model 12) and the depth-resolved Model 13 (Armstrong, 2006) also have negligible biases. In terms of reproducing the observed PP variability, two depth and wavelength-resolved models outperform the others (Model 16: Smyth et al., 2005 and Model 18: Mélín, 2003), Model 18, as well as Model 27 (Gregg, 2003) produce PP fields that are most highly correlated with the data, whereas Model 29 (Tjiputra et al., 2007) produces PP fields that have nearly the same variance as the ClimPP data. Cumulative fractional distributions indicate that Model 11 (Morel and Maritorena, 2001) produces PP distributions that most closely reproduce the observations. It is clearly not possible to identify one single model that is most skilled according to all the criteria listed above, however, the specific models highlighted here are examples of models that perform particularly well in these analyses.

In order to better understand why certain models have greater skill than others and what environmental conditions control performance, correlations between model–data misfit and a number of relevant variables were computed. Surprisingly, model–data misfit was highly correlated to year for almost all of the SatPPMs as well as for some of the BOGCMs. Both primary productivity and biomass-normalized productivity in the ClimPP database were substantially higher in the 1990s than in the 1980s. This interdecadal variability may be associated with a change in upwelling rate within the tropical Pacific, a hypothesis that is consistent with previous studies (Takahashi et al., 2003; Chavez et al., 2003; Feely et al., 2006). The observed interdecadal change in integrated productivity per unit biomass within the ClimPP database was not associated with a coincident change in the surface expression of chlorophyll, temperature, or PAR, probably because of the typically shallow MLD in the tropical Pacific. As a result, the SatPPMs were unable to correctly estimate this interdecadal variability in productivity. Although BOGCMs should theoretically have the ability to replicate the increased upwelling of nutrients and the effect of this on productivity, this study has shown that this remains a challenge for the current generation of BOGCMs. These results suggest that interdecadal and global changes will be a significant challenge for both SatPPMs and BOGCMs alike.

The success of these types of comparison exercises is illustrated by the continual modification and improvement of the participating models and the resulting improvement in model skill. The previous PPARR2 exercise (Campbell et al., 2002) specifically highlighted bias as a major contributor to total RMSD. As demonstrated here, bias has been effectively reduced in several participating models. PPARR2 used the Joint Global Ocean Flux Study equatorial Pacific process study

data collected on and off the equator at 140°W in 1992 to compute the average total RMSD of \log_{10} PP for their twelve participating models. Using the identical dataset in an analogous computation, the average total RMSD is computed from 20 of the SatPPMs participating in this exercise (Models 2–21). The resulting RMSDs (on equator: 0.21; off equator: 0.23) are 58% and 35% lower, respectively, than those computed 6 years ago in PPARR2 (on equator: 0.50; off equator: 0.31). If this rate of model improvement continues, within a little more than a decade total RMSD can be expected to fall to within the level of observational uncertainty (0.07).

Acknowledgements

This research was supported by a grant from the National Aeronautics and Space Agency Ocean Biology and Biogeochemistry program (NNG06GA03G), as well as by numerous other grants to the various participating investigators. Jason Jolliff, John Cullen and an anonymous reviewer provided many useful comments, which improved the final version of the manuscript. The AVHRR Oceans Pathfinder SST data were obtained from the Physical Oceanography Distributed Active Archive Center (PO.DAAC) at the NASA Jet Propulsion Laboratory, Pasadena, CA. Shortwave radiation data were provided by the NOAA-ESRL Physical Sciences Division, Boulder Colorado from their website at <http://www.cdc.noaa.gov>. We thank Raghu Murtugudde for providing the mixed layer depths for the ClimPP database. We are also grateful for the hard work of the many scientists who helped acquire the ClimPP database. This is VIMS contribution #2904.

Appendix A

A.1. DI/WI SatPPMs

Model 1: This model estimates PP as:

$$PP \left[\text{g C m}^{-2} \text{ d}^{-1} \right] = \left(\text{chl}_0 \left[\text{mg chl m}^{-3} \right] \right)^{1/2}$$

(Eppley et al., 1985). It ignores any external forcing or changes in physiological state. While other models incorporate information regarding geography or forcing fields, this model assumes that the standing stock is sole determinant of photosynthetic rate. All biomass performs identically. This simplicity is inherently elegant because biomass is, for most of the ocean, an excellent indicator of nutrient supply and presence of light.

Model 2: This is a variant of the original Howard, Yoder, Ryan model (HYR, Howard and Yoder, 1997; Model 3) which integrates photosynthesis to the euphotic depth as defined in Behrenfeld and Falkowski (1997b) rather than to the MLD (Carr, 2002).

Model 3: This is the original HYR model, which for many years was a standard MODIS algorithm. Maximum growth rate is parameterized as a function of SST according to Eppley (1972). Primary productivity is integrated to the MLD rather than to the euphotic depth.

Model 4: This HYR variant uses MLD information to quantify the photoadaptive variables within the euphotic zone, as well as to address water column partitioning of primary productivity relative to euphotic depth.

Model 5: This model uses an artificial neural network to perform a generalized nonlinear regression of PP on several predictive variables, including latitude, longitude, day length, MLD, SST, $P_{\text{opt}}^{\text{B}}$ (computed according to Behrenfeld and Falkowski (1997b)), PAR, and chl_0 (Scardi, 2000; Scardi, 2001). Since there are insufficient data to calibrate the neural network throughout the tropical Pacific, PP values from other models (VGPM, HYR, and the MOD-27 formulation (Esaias, 1996)) were considered measurements where there were none.

Model 6: This model is based on the formulation obtained through dimensional analysis by Platt and Sathyendranath (1993). The photosynthetic parameter ($P_{\text{max}}^{\text{B}}$) is assigned by combining a temperature-dependent relationship for the maximum growth rate (Eppley, 1972) with a variable carbon to chlorophyll ratio following the statistical relationship of Cloern et al. (1995). Primary production is integrated to the 1% light level.

Model 7: The Carbon-based Production Model (CbPM) represents a new approach to NPP assessment that utilizes satellite information on both surface chlorophyll and particulate backscatter (bbp) at 443 nm, which is converted into phytoplankton carbon biomass (C) (Behrenfeld et al., 2005). This expanded information set allows the ratio of chl:C to be calculated and related to phytoplankton growth rates (m), such that NPP is then calculated as the product of C and m . In the current application of the CbPM, no field data were provided on bbp, so phytoplankton C was estimated using 'typical values' derived from SeaWiFS data for a given sample location. This lack of coincident field bbp data compromises the performance of the CbPM in the current model-field data comparison. Vertical structure in this version of the CbPM is as assumed in Model 8.

Model 8: The Vertically Generalized Production Model (VGPM) developed by Behrenfeld and Falkowski (1997b) is one of the most widely known and used WIDI PP models and is one of two standard MODIS algorithms. The maximum observed photosynthetic rate within the water column, $P_{\text{opt}}^{\text{B}}$, is obtained as a 7th-order polynomial of SST.

Model 9: This model only differs from Model 8 in that $P_{\text{opt}}^{\text{B}}$ is estimated as an exponential function of temperature following Eppley (1972).

Model 10: This VGPM variant formulates $P_{\text{opt}}^{\text{B}}$ as a function of SST and chl_0 (Yamada et al., 2005; Kameda and Ishizaka, 2005). The model is based on the assumptions that changes in chlorophyll concentration depend on the abundance of large phytoplankton and that chlorophyll-specific productivity is inversely proportional to phytoplankton size.

Model 11: This VGPM variant uses the continuous function of Morel and Berthon (1989) to estimate total integrated chlorophyll, which in turn is used to estimate the euphotic depth with the equations proposed by Morel and Maritorea (2001).

Model 12: This VGPM variant differs from Model 8 in that $P_{\text{opt}}^{\text{B}}$ ($\text{mg C mg chl}^{-1} \text{ h}^{-1}$) is estimated as a function of chl_0 (mg chl m^{-3}) only. This empirical relationship was derived from EqPac data ($P_{\text{opt}}^{\text{B}} = 24 \times \text{chl}_0 + 1$).

A.2. DR/WI SatPPMs

Model 13: Photosynthesis per unit chlorophyll was determined using an optimality-based model of nitrogen allocation

and photoacclimation (Armstrong, 2006), which is in turn based on the model of Geider et al. (1998). Photoacclimation and nitrogen allocation are determined as a function of light and temperature; therefore both PAR and SST are used in the productivity algorithm. Maximum photosynthetic rates were taken to be an average of maximum NPP and GPP estimated for *T. weissflogii* in Armstrong (2006), and were assumed to have Eppley temperature dependence. Through the photoadaptation algorithm, chlorophyll reflects nitrogen status, so that no additional assumptions about nutrient limitation are needed. Chlorophyll concentration was assumed constant over the photic zone and equal to surface chlorophyll, so that light decreases exponentially with depth. Photic zone depth (1% light) was determined from chlorophyll concentration and assumed extinction coefficients. The photic zone was assumed to be well-mixed and cells were assumed to be photoacclimated to average (10%) light. Column productivity is the integral over the photic zone of (photosynthesis/chlorophyll)*chlorophyll.

Model 14: In this model the depth distribution of PAR is determined by chl_0 while the depth distribution of chlorophyll is given by the PAR profile and chl_0 . Total productivity is empirically estimated as a function of surface temperature and depth-dependent PAR and chlorophyll (Asanuma, 2006).

Model 15: This model is based on chlorophyll-specific phytoplankton absorption, which is parameterized empirically as a function of SST (Marra et al., 2003). Absorption by photosynthetic pigments is distinguished from total absorption; the former is used to calculate productivity and the latter is used to estimate light attenuation in the water column. The rate of photosynthesis is quantum yield times absorbed irradiance, obtained from a hyperbolic tangent and a constant maximum quantum yield, ϕ_{max} . The depth profile of chlorophyll is estimated assuming a Gaussian shape with parameters determined by the surface value. The attenuation coefficient of chlorophyll is also SST-dependent.

A.3. DR/WR SatPPMs

Model 16: This model is an implementation of the Morel (1991) model in which the depth distribution of chlorophyll is assumed constant throughout the water column. The broadband incident PAR is spectrally resolved using a lookup table generated from a single run of the Gregg and Carder (1990) marine irradiance model where the effects of clouds and aerosols are essentially linearly scaled. The model uses 60-min time and 10-m depth steps at 5-nm wavelength resolution when run using the global datasets (Smyth et al., 2005).

Model 17: This model is the lookup table described in Smyth et al. (2005) which uses a radiative transfer model with parameterizations for Colored Dissolved Organic Matter (CDOM). The model is based on Morel (1991) with incident irradiance parameterized in the same manner as Model 16.

Model 18: This model follows that of Platt and Sathyendranath (1988) as implemented at global scale by Longhurst et al. (1995). It uses biogeographical provinces to define the values of the parameters to describe the light–photosynthesis curve and the chlorophyll depth profile. Photosynthetic parameters were updated using an extended dataset provided by the Bedford Institute of Oceanography and an extensive literature review. Spectral surface irradiance is first estimated independently with the model of Gregg and Carder (1990)

combined with a correction for cloud cover and then scaled to match the PAR values provided for the exercise. Spectral light is subsequently propagated in the water column with a bio-optical model with updated parameterizations of the inherent optical properties. All changes to the original implementation of Longhurst et al. (1995) are detailed by Mélin (2003).

Model 19: This model derives spectral irradiance from PAR using Tanre et al. (1990), and assumes a vertically uniform chlorophyll profile. Quantum yield is parameterized as a maximum value times both a light dependent term (Waters et al., 1994; Bidigare et al., 1992) and a temperature-dependent term. Temperature dependence was assumed to be sigmoidal, and was based on a vertical profile of temperature derived from SST and MLD.

Model 20: This is a spectral light–photosynthesis model published by Morel (1991). It is formulated using chlorophyll-specific wavelength-resolved absorption and quantum yield. Temperature dependence is given by the parameterization of P_{max}^B , which follows Eppley (1972). The chlorophyll profile is determined to be well-mixed or stratified according to the ratio of MLD and the euphotic depth, and if stratified, assigned a Gaussian profile as in Morel and Berthon (1989). Mean photo-physiological parameters are from Morel et al. (1996). The model is run in its ‘satellite’ version Antoine and Morel (1996), where PP is the product of integral biomass, the daily irradiance, and Ψ^* (the cross-section of algae for photosynthesis per unit of areal chlorophyll biomass). Lookup tables for Ψ^* were previously generated using the full DR/WR model, and are used to increase computational efficiency.

Model 21: This model represents an expansion in both the physiological attributes of the original CbPM (Model 7) and the space- and time-resolution of the model, and is described in detail in Westberry et al. (2008). The model requires surface chlorophyll concentration data and particulate backscatter coefficients (bbp) at 443 nm. As with Model #7, execution of Model #21 for the current study required the estimation of bbp from the SeaWiFS record, as no coincident field data for bbp were available. Vertical structure in chlorophyll concentration in this model is driven primarily by photoacclimation of cellular chlorophyll, with additional adjustments made for changes in nutrient stress through the water column.

A.4. BOGCMs

Model 22: This BOGCM (Christian et al., 2002) uses a 20-layer reduced-gravity σ -coordinate model, with the first layer corresponding to the ocean’s mixed layer. The grid employed here stretches from 124°E to 76°W with uniform 1° resolution in longitude, and from 30°S to 30°N with maximal latitudinal resolution in the equatorial band (1/3° at latitudes <10°). The model was run using NCEP weekly (6 day) mean wind stress and monthly interannual precipitation. The nitrogen and iron based ecosystem model has two size classes each of phytoplankton, zooplankton, and detritus, and the nutrients nitrate, ammonium, and iron. Phytoplankton growth is simultaneously limited by irradiance as well as nitrogen and iron availability, with fixed Fe/N ratios so that iron in all compartments except the dissolved pool is not modelled explicitly.

Model 23: This BOGCM is a modification of Model 22 (Wang et al., 2005, 2006) and differs primarily by having a saturating

(Ivlev) rather than a nonsaturating (multiplicative) grazer functional response. In addition it utilizes an additional quadratic mortality term that confers stability and suppresses oscillatory solutions in the case of saturable grazing (Wang et al., 2005). Additional minor modifications from the published versions are discussed in Christian et al. (in press).

Model 24: The Biogeochemical Elemental Cycling (BEC) ocean model simulates the biogeochemical cycling of carbon, oxygen, nitrate, phosphate, iron, silicate, and alkalinity. The ecosystem has multiple, potentially growth limiting nutrients (nitrogen, phosphorus, silicon, and iron) and four phytoplankton groups (diatoms, diazotrophs, coccolithophores, and picophytoplankton) (Moore et al., 2002, 2004; Doney et al., 2009–this issue). Growth rates can be limited by available nutrients and/or light levels. Full carbonate chemistry, as well as sinking particulate and semi-labile dissolved organic pools are included in a global, 3D context without nutrient restoring (Moore et al., 2004). The BEC runs within the coarse resolution ocean component of the Community Climate System Model (Yeager et al., 2006), driven with time-varying atmospheric forcing from NCEP reanalysis and satellite data products (Doney et al., 2007).

Model 25: This prognostic ocean biogeochemistry/ecology model considers 25 tracers including three phytoplankton groups, two forms of dissolved organic matter, heterotrophic biomass, and dissolved inorganic species for coupled C, N, P, Si, Fe, CaCO₃, O₂ and lithogenic cycling with flexible N:P:Fe stoichiometry. The model includes such processes as gas exchange, atmospheric deposition, scavenging, N₂ fixation and denitrification. Loss of phytoplankton is parameterized through the size-based relationship of Dunne et al. (2005), river inputs, and sediment processes. This biogeochemistry was run in the Modular Ocean Model version 4 (Griffies et al., 2005) with 50 vertical z-coordinate layers and a nominally 1° global resolution horizontal B-grid with tri-polar coordinates to resolve the arctic and finer detail of 1/3° near the equator. The model was forced with 6-hourly and interannually varying forcing from the atmospheric reanalysis of Large and Yeager (2004).

Model 26: The NASA Ocean Biogeochemical Model (NOBM) simulates four phytoplankton groups (diatoms, chlorophytes, cyanobacteria, and coccolithophores) and four nutrients (nitrate, ammonium, silica, and iron) (Gregg and Casey, 2007). The model is approximately 0.8° resolution with 14 vertical layers in quasi-isopycnal configuration. The model was forced by monthly mean winds and shortwave radiation from NCEP for 1979–2004.

Model 27: In this NOBM variant (Gregg, 2008), NODC data for 1979–2004 were assimilated using the Conditional Relaxation Analysis Method. The assimilation affected the model representations of total chlorophyll (sum of the 4 phytoplankton groups), but not the individual community distributions directly. Primary production was affected by the change in total chlorophyll, as well as by indirect effects such as subsurface irradiance resulting from absorption and scattering by the changed chlorophyll field.

Model 28: The PISCES-T model is a two phytoplankton, two zooplankton, three detritus, three nutrient Dynamic Green Ocean Model. It has variable Fe:C, Chl:C and Si:C ratios and fixed N:C:O. It is embedded in the OPA8 OGCM, initialized with observations, and forced by daily NCEP reanalysis

(Buitenhuis et al., 2006). No nutrient restoring was used. The model was run from 1948 to 2006 (Le Quere et al., 2007).

Model 29: The Hamburg Ocean Carbon Cycle (HAMOCC5) model uses monthly climatology forcing with approximately 3.5° × 3.5° horizontal degree resolution and 22 vertical layers with thickness varying from 50 m at the surface to 700 m in the deepest layer (Maier-Reimer, 1993). The model employs an NPZD-type ecosystem model with two phytoplankton functional groups (diatoms and calcifiers) with multi-nutrient (nitrate, phosphate, silicate and iron) limitation. Productivity is simulated with a 3-day time step (Six and Maier-Reimer, 1996; Howard et al., 2006).

Model 30: The physical model is based on the Regional Ocean Modeling System (ROMS), which has been configured for the Pacific Ocean at 50-km horizontal resolution and 20 vertical layers (Wang and Chao, 2004). The biogeochemical model is based on the Carbon, Silicate, Nitrogen Ecosystem (CoSINE) model developed by Chai et al. (2002, 2003). The CoSINE model includes silicate, nitrate and ammonium, two phytoplankton groups, two zooplankton grazers, and two detrital pools. The model is initialized with climatological, temperature, salinity, nutrients and TCO₂. When forced with daily air–sea fluxes (wind stress, heat flux, and freshwater exchange) derived from NCEP/NCAR reanalysis (Kalnay et al., 1996) for the period of 1984 to 2005, the Pacific ROMS-CoSINE model is able to realistically reproduce the evolution of ENSO events (1992–93; 1997–98) as well as a sharp change of the circulation and thermal structure for the Pacific Ocean in 1998–99 (Liu and Chai, 2008).

Appendix B. Supplementary data

Supplementary data associated with this article can be found, in the online version, at doi:10.1016/j.jmarsys.2008.05.010.

References

- Antoine, D., Morel, A., 1996. Oceanic primary production: I. Adaptation of a spectral light–photosynthesis model in view of application to satellite chlorophyll observations. *Global Biogeochemical Cycles* 10, 43–55.
- Armstrong, R.A., 2006. Optimality-based modeling of nitrogen allocation and photo acclimation in photosynthesis. *Deep-Sea Research II* 53, 513–531.
- Asanuma, I., 2006. Depth and time resolved primary productivity model examined for optical properties of water, global climate change and response of carbon cycle in the equatorial Pacific and Indian Oceans and adjacent landmasses. *Elsevier Oceanography Series* 73, 89–106 pp.
- Barber, R.T., Borden, L., Johnson, Z., Marra, J., Knudson, C., Trees, C., 1997. Ground truthing modeled k_{PAR} and on deck primary productivity incubations with *in situ* observations. *Ocean Optics XIII* 2963, 834–839.
- Barber, R.T., Chavez, F.P., 1991. Regulation of primary productivity rate in the equatorial Pacific Ocean. *Limnology and Oceanography* 36, 1803–1815.
- Barber, R.T., Marra, J., Bidigare, R.C., Codispoti, L.A., Halpern, D., Johnson, Z., Latasa, M., Goericke, R., Smith, S.L., 2001. Primary productivity and its regulation in the Arabian Sea during 1995. *Deep-Sea Research II* 48, 1127–1172.
- Barber, R.T., Sanderson, M.P., Lindley, S.T., Chai, F., Newton, J., Trees, C.C., Foley, D.G., Chavez, F.P., 1996. Primary productivity and its regulation in the equatorial Pacific during and following the 1991–1992 El Niño. *Deep Sea Research I* 43, 933–969.
- Behrenfeld, M.J., Falkowski, P.G., 1997a. A consumer's guide to phytoplankton primary productivity models. *Limnology and Oceanography* 42, 1479–1491.
- Behrenfeld, M.J., Falkowski, P.G., 1997b. Photosynthetic rates derived from satellite-based chlorophyll concentration. *Limnology and Oceanography* 42, 1–20.
- Behrenfeld, M.J., Boss, E., Siegel, D.A., Shea, D.M., 2005. Carbon-based ocean productivity and phytoplankton physiology from space. *Global Biogeochemical Cycles* 19, GB1006. doi:10.1029/2004GB002299.
- Behrenfeld, M.J., O' Malley, R.T., Siegel, D.A., McClain, C.R., Sarmiento, J.L., Feldman, G.C., Milligan, A.J., Falkowski, P.G., Letelier, R.M., Boss, E.S., 2006.

- Climate-driven trends in contemporary ocean productivity. *Nature* 444, 752–755.
- Berger, W.H., 1989. Global maps of primary productivity. In: Berger, W.H., Smetacek, V.S., Wefer, G. (Eds.), *Productivity of the Ocean: Present and Past*. Wiley, Berlin, pp. 429–455.
- Bianchi, A.A., Bianucci, L., Piola, A.R., Pino, D.R., Schloss, I., Poisson, A., Balestrini, C.F., 2005. Vertical stratification and air–sea CO₂ fluxes in the Patagonian shelf. *Journal of Geophysical Research* 110, C07003. doi:10.1029/2004JC002488.
- Bidigare, R.R., Prézelin, B.B., Smith, R.C., 1992. Bio-optical models and the problems of scaling, primary productivity and biogeochemical cycles in the sea. Plenum Press, New York, pp. 175–212.
- Boyd, P.W., Trull, T.W., 2007. Understanding the export of biogenic particles in oceanic waters: is there consensus? *Progress in Oceanography* 72, 276–312.
- Buitenhuis, E., Le Quééré, C., Olivier, A., Grégory, B., Bunker, A., Hirst, A., Ikeda, T., O'Brien, T., Piontkovski, S., Straile, D., 2006. Biogeochemical fluxes through mesozooplankton. *Global Biogeochemical Cycles* 20, GB2003. doi:10.1029/2005GB002511.
- Campbell, J., Antoine, D., Armstrong, R., Arrigo, K., Balch, W., Barber, R., Behrenfeld, M., Bidigare, R., Bishop, J., Carr, M.-E., Esaias, W., Falkowski, P., Hoepffner, N., Iverson, R., Kiefer, D., Lohrenz, S., Marra, J., Morel, A., Ryan, J., Vedernikov, V., Waters, K., Yentsch, C., Yoder, J., 2002. Comparison of algorithms for estimating ocean primary production from surface chlorophyll, temperature, and irradiance. *Global Biogeochemical Cycles* 16, 1029. doi:10.1029/2001GB001444.
- Carr, M.E., 2002. Estimation of potential productivity in Eastern Boundary Currents using remote sensing. *Deep-Sea Research II* 49, 59–80.
- Carr, M.-E., Friedrichs, M.A.M., Schmeltz, M., Aita, M.N., Antoine, D., Arrigo, K.R., Asanuma, I., Aumont, O., Barber, R., Behrenfeld, M., Bidigare, R., Buitenhuis, E.T., Campbell, J., Ciotti, A., Dierssen, H., Dowell, M., Dunne, J., Esaias, W., Gentili, B., Gregg, W., Groom, S., Hoepffner, N., Ishizaka, J., Kameda, T., Le Quééré, C., Lohrenz, S., Marra, J., Mélin, F., Moore, J., Morel, A., Reddy, T.E., Ryan, J., Scardi, M., Smyth, T., Turpie, K., Tilstone, G., Waters, K., Yamanaka, Y., 2006. A comparison of global estimates of marine primary production from ocean color. *Deep-Sea Research II* 53, 741–770.
- Chai, F., Dugdale, R.C., Peng, T.H., Wilkerson, F.P., Barber, R.T., 2002. One-dimensional ecosystem model of the equatorial Pacific upwelling system. Part I: model development and silicon and nitrogen cycle. *Deep-Sea Research II* 49, 2713–2745.
- Chai, F., Jiang, M.S., Barber, R.T., Dugdale, R.C., Chao, Y., 2003. Interdecadal variation of the transition zone chlorophyll front: a physical–biological model simulation between 1960 and 1990. *Journal of Oceanography* 59, 461–475.
- Chavez, F.P., Barber, R.T., 1991. An estimate of new production in the equatorial Pacific. *Deep-Sea Research* 34, 1229–1243.
- Chavez, F.P., Ryan, J., Lluch-Cota, S.E., Niquen, M.C., 2003. From anchovies to sardines and back: multidecadal change in the Pacific Ocean. *Science* 299, 217–221.
- Chen, D., Rothstein, L.M., Busalacchi, A.J., 1994. A hybrid vertical mixing scheme and its application to tropical ocean models. *Journal of Physical Oceanography* 24, 2156–2179.
- Christian, J.R., Verschell, M.A., Murtugudde, R., Busalacchi, A.J., McClain, C.R., 2002. Biogeochemical modelling of the tropical Pacific Ocean. I: Seasonal and interannual variability. *Deep-Sea Research II* 49, 509–543.
- Christian, J.R., Feely, R.A., Ishii, M., Murtugudde, R., Wang, X., in press. Testing an ocean carbon model with observed sea surface pCO₂ and DIC in the tropical Pacific Ocean. *Journal of Geophysical Research*. doi:10.1029/2007JC004428.
- Cloern, J.E., Grenz, C., VidergarLucas, L., 1995. An empirical model of the phytoplankton chlorophyll:carbon ratio – the conversion factor between productivity and growth rate. *Limnology and Oceanography* 40, 1313–1321.
- Doney, S.C., Lindsay, K., Caldeira, K., Campin, J.-M., Drange, H., Dutay, J.-C., Follows, M., Gao, Y., Gnanadesikan, A., Gruber, N., Ishida, A., Joos, F., Madec, G., Maier-Reimer, E., Marshall, J.C., Matear, R.J., Monfray, P., Mouchet, A., Najjar, R., Orr, J.C., Plattner, G.-K., Sarmiento, J., Schlitzer, R., Slater, R., Totterdell, I.J., Weirig, M.-F., Yamanaka, Y., Yool, A., 2004. Evaluating global ocean carbon models: the importance of realistic physics. *Global Biogeochemical Cycles* 18, GB3017. doi:10.1029/2003GB002150.
- Doney, S.C., Yeager, S., Danabasoglu, G., Large, W.G., McWilliams, J.C., 2007. Mechanisms governing interannual variability of upper-ocean temperature in a global ocean hindcast simulation. *Journal of Physical Oceanography* 37, 1918–1938.
- Doney, S.C., Lima, I., Moore, J.K., Lindsay, K., Behrenfeld, M., Westberry, T.K., Mahowald, N., Glover, D.M., McGillicuddy Jr., D.J., Takahashi, T., 2009. Skill metrics for confronting global upper ocean ecosystem–biogeochemistry models against field and remote sensing data. *Journal of Marine Systems* 76, 95–112 (this issue). doi:10.1016/j.jmarsys.2008.05.015.
- Dunne, J.P., Armstrong, R.A., Gnanadesikan, A., Sarmiento, J.L., 2005. Empirical and mechanistic models for the particle export ratio. *Global Biogeochemical Cycles* 19, GB02390. doi:10.1029/2004GB002390.
- Eppley, R., Steward, E., Abbott, M., Heyman, U., 1985. Estimating ocean primary production from satellite chlorophyll: introduction to regional differences and statistics for the southern California Bight. *Journal of Plankton Research* 7, 57–70.
- Eppley, R.W., 1972. Temperature and phytoplankton growth in the sea. *Fisheries Bulletin* 70, 1063–1085.
- Esaias, W., 1996. MODIS Algorithm Theoretical Basis Document for Product MOD-27: Ocean Primary Productivity. <http://modis.gsfc.nasa.gov/data/atbd/atbd-mod24.pdf>.
- Feely, R.A., Takahashi, T., Wanninkhof, R., McPhaden, M., Cosca, C.E., Sutherland, S.C., Carr, M.-E., 2006. Decadal variability of the air–sea CO₂ fluxes in the equatorial Pacific Ocean. *Journal of Geophysical Research* 111, C08S90. doi:10.1029/2005JC003129.
- Fitzwater, W.E., Knauer, G.A., Martin, J.H., 1982. Metal contamination and its effect on primary production measurement. *Limnology and Oceanography* 27, 544–551.
- Friedrichs, M.A.M., 2002. The assimilation of SeaWiFS and JGOFS EqPac data into a marine ecosystem model of the central equatorial Pacific. *Deep-Sea Research II* 49, 289–319.
- Friedrichs, M.A.M., Hood, R.R., Wiggert, J.D., 2006. Ecosystem model complexity versus physical forcing: quantification of their relative impact with assimilated Arabian Sea data. *Deep-Sea Research II* 53, 576–600.
- Friedrichs, M.A.M., Dusenberry, J.A., Anderson, L.A., Armstrong, R.A., Chai, F., Christian, J.R., Doney, S.C., Dunne, J., Fujii, M., Hood, R., McGillicuddy, D.J., Moore, J.K., Schartau, M., Spitz, Y.H., Wiggert, J.D., 2007. Assessment of skill and portability in regional marine biogeochemical models: role of multiple planktonic groups. *Journal of Geophysical Research* 112, C08001. doi:10.1029/2006JC003852.
- Geider, R.J., MacIntyre, H.L., Kana, T.M., 1998. A dynamic regulatory model of phytoplankton acclimation to light, nutrients, and temperature. *Limnology and Oceanography* 43, 679–694.
- Gnanadesikan, A., Slater, R.D., Gruber, N., Sarmiento, J.L., 2002. Oceanic vertical exchange and new production: a comparison between models and observations. *Deep-Sea Research II* 49, 363–401.
- Gregg, W.W., Carder, K.L., 1990. A simple spectral solar irradiance model for cloudless maritime atmospheres. *Limnology and Oceanography* 35, 1657–1675.
- Gregg, W.W., Casey, N.W., 2007. Modeling coccolithophores in the global oceans. *Deep-Sea Research II* 54, 447–477.
- Gregg, W.W., 2008. Assimilation of SeaWiFS ocean chlorophyll data into a three-dimensional global ocean model. *Journal of Marine Systems* 69, 205–255.
- Gregg, W.W., Friedrichs, M.A.M., Robinson, A.R., Rose, K.A., Schlitzer, R., Thompson, K.R., Doney, S.C., 2009. Skill assessment in ocean biological data assimilation. *Journal of Marine Systems* 76, 16–33 (this volume). doi:10.1016/j.jmarsys.2008.05.006.
- Griffies, S.M., Gnanadesikan, A., Dixon, K.W., Dunne, J.P., Gerdes, R., Harrison, M.J., Rosati, A., Russell, J.L., Samuels, B.L., Spelman, M.J., Winton, M., Zhang, R., 2005. Formulation of an ocean model for global climate simulations. *Ocean Science* 1, 45–79.
- Hare, S.R., Mantua, N.J., 2000. Empirical evidence for North Pacific regime shifts in 1977 and 1989. *Progress in Oceanography* 47, 103–145.
- Howard, E.C., Henriksen, J.R., Buchan, A., Reisch, C.R., Bürgmann, H., Welsh, R., Ye, W., González, J.M., Mace, K., Joye, S.B., Kiene, R.P., Whitman, W.B., Moran, M.A., 2006. Bacterial taxa that limit sulfur flux from the ocean. *Science* 314, 649–652.
- Howard, K.L., Yoder, J.A., 1997. Contribution of the sub-tropical oceans to global primary production. In: Liu, C.-T. (Ed.), *Proceedings of COSPAR Colloquium on Space Remote Sensing of Subtropical Oceans*. Pergamon, New York, pp. 157–168.
- Jolliff, J., Kindle, J.C., Shulman, I., Penta, B., Friedrichs, M.A.M., Helber, R., Arnore, R.A., 2009. Summary diagrams for coupled hydrodynamic–ecosystem model skill assessment. *Journal of Marine Systems* 76, 64–82 (this volume). doi:10.1016/j.jmarsys.2008.05.014.
- Kalnay, E., Kanamitsu, M., Kistler, R., Collins, W., Deaven, D., Gandin, L., Iredell, M., Saha, S., White, G., Woollen, J., Zhu, Y., Leetmaa, A., Reynolds, B., Chelliah, M., Ebisuzaki, W., Higgins, W., Janowiak, J., Mo, K.C., Ropelewski, C., Wang, J., Jenne, R., Joseph, D., 1996. The NCEP/NCAR 40-year reanalysis project. *Bulletin of the American Meteorological Society* 77, 437–471.
- Kameda, T., Ishizaka, J., 2005. Size-fractionated primary production estimated by a two-phytoplankton community model applicable to ocean color remote sensing. *Journal of Oceanography* 61, 663–672.
- Koblentz-Mishke, O.I., Volkovinsky, V.V., Kabanova, Y.G., 1970. Plankton primary production of the world ocean. In: Wooster, W. (Ed.), *Scientific Exploration of the South Pacific*. National Academy of Sciences, Washington, D.C., pp. 183–193.
- Large, W.G., Yeager, S.G., 2004. Diurnal to decadal global forcing for ocean and sea-ice models: the data sets and flux climatologies. TN-460+STR, NCAR, p. 105.

- Larsen, S.H., 2005. Solar variability, dimethyl sulphide, clouds, and climate. *Global Biogeochemical Cycles* 19, GB1014. doi:10.1029/2004GB002333.
- Le Borgne, R., Barber, R.T., Delcroix, T., Inoue, H.Y., Mackey, D.J., Rodier, M., 2002. Pacific warm pool and divergence: temporal and zonal variations on the equator and their effects on the biological pump, 2002. *Deep-Sea Research II* 49, 2471–2512.
- Le Quééré, C., Rödenbeck, C., Buitenhuis, E.T., Conway, T.J., Langenfelds, R., Gomez, A., Labuschagne, C., Ramonet, M., Nakazawa, T., Metzl, N., Gillett, N., Heimann, M., 2007. Saturation of the Southern Ocean CO₂ sink due to recent climate change. *Science* 316, 1735–1738. doi:10.1126/science.1136188.
- Longhurst, A., Sathyendranath, S., Platt, T., Caverhill, C., 1995. An estimate of global primary production in the ocean from satellite radiometer data. *Journal of Plankton Research* 17, 1245–1271.
- Maier-Reimer, E., 1993. Geochemical cycles in an ocean general circulation model: preindustrial tracer distributions. *Global Biogeochemical Cycles* 7, 645–677.
- Maritorena, S., Siegel, D.A., Peterson, A.R., 2002. Optimization of a semi-analytical ocean color model for global-scale applications. *Applied Optics* 41, 2705–2714.
- Marra, J., Ho, C., Trees, C., 2003. An Alternative Algorithm for the Calculation of Primary Productivity from Remote Sensing Data. LDEO Technical Report #LDEO-2003-1.
- Matsumoto, K., Sarmiento, J.L., Key, R.M., Bullister, J.L., Caldeira, K., Campin, J.-M., Doney, S.C., Drange, H., Dutay, J.-C., Follows, M., Gao, Y., Gnanadesikan, A., Gruber, N., Ishida, A., Joos, F., Lindsay, K., Maier-Reimer, E., Marshall, J.C., Matear, R.J., Monfray, P., Najjar, R., Plattner, G.-K., Schlitzer, R., Slater, R., Swathi, P.S., Totterdell, I.J., Weirig, M.-F., Yamanaka, Y., Yool, A., Orr, J.C., 2004. Evaluation of ocean carbon cycle models with data-based metrics. *Geophysical Research Letters* 31, L07303. doi:10.1029/2003GL018970.
- McClain, C.R., Christian, J.R., Signorini, S.R., Lewis, M.R., Asanuma, I., Turk, D., Dupouy-Douchement, C., 2002. Satellite ocean-color observations of the tropical Pacific Ocean. *Deep-Sea Research II* 49, 2533–2560.
- McPhaden, M.J., Zhang, D., 1999. Slowdown of the meridional overturning circulation in the upper Pacific Ocean. *Nature* 415, 603–608.
- Mélin, F., 2003. Potentiel de la télédétection pour l'analyse des propriétés optiques du système océan-atmosphère et application à l'estimation de la photosynthèse phytoplantonique. Ph.D., Université Paul Sabatier, Toulouse, France.
- Moore, J.K., Doney, S.C., Kleypas, J.A., Glover, D.M., Fung, I.Y., 2002. An intermediate complexity marine ecosystem model for the global domain. *Deep-Sea Research Part II* 49, 403–462.
- Moore, J.K., Doney, S.C., Lindsay, K., 2004. Upper ocean ecosystem dynamics and iron cycling in a global three-dimensional model. *Global Biogeochemical Cycles* 18, GB4028. doi:10.1029/2004GB002220.
- Morel, A., 1988. Optical modeling of the upper ocean in relation to its biogenous matter content (case 1 water). *Journal of Geophysical Research* 93 (10), 749–10,768.
- Morel, A., 1991. Light and marine photosynthesis: a spectral model with geochemical and climatological implications. *Progress in Oceanography* 26, 263–306.
- Morel, A., Antoine, D., Babin, M., Dandonneau, Y., 1996. Measured and modeled primary production in the northeast Atlantic (EUMELI JGOFS program): the impact of natural variations in photosynthetic parameters on model predictive skill. *Deep-Sea Research I* 43, 1273–1304.
- 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.
- Morel, A., Maritorena, S., 2001. Bio-optical properties of oceanic waters: a reappraisal. *Journal of Geophysical Research* 106, 7163–7180.
- Murtugudde, R., Seager, R., Busalacchi, A., 1996. Simulation of the tropical oceans with an ocean GCM coupled to an atmospheric mixed-layer model. *Journal of Climate* 9, 1795–1815.
- Murtugudde, R., Busalacchi, A.J., 1998. Salinity effects in a tropical ocean model. *Journal of Geophysical Research* 103, 3283–3300.
- Mélin, F., 2003. Potentiel de la télédétection pour l'analyse des propriétés optiques du système océan-atmosphère et application à l'estimation de la photosynthèse phytoplantonique. Université Paul Sabatier, Toulouse, France.
- Najjar, R.G., Jin, X., Louanchi, F., Aumont, O., Caldeira, K., Doney, S.C., Dutay, J.-C., Follows, M., Gruber, N., Joos, F., Lindsay, K., Maier-Reimer, E., Matear, R.J., Matsumoto, K., Monfray, P., Mouchet, A., Orr, J.C., Plattner, G.K., Sarmiento, J.L., Schlitzer, R., Weirig, M.F., Yamanaka, Y., Yool, A., 2007. Impact of circulation on export production, dissolved organic matter and dissolved oxygen in the ocean: results from OCMIP-2. *Global Biogeochemical Cycles* 21, GB3007. doi:10.1029/2006GB002857.
- Ondrusek, M.E., Bidigare, R.R., Waters, K., Karl, D.M., 2001. A predictive model for estimating rates of primary production in the subtropical North Pacific Ocean. *Deep-Sea Research II* 48, 1837–1863.
- Pauly, D., Christensen, V., 1995. Primary production required to sustain global fisheries. *Nature* 374, 255–257.
- Platt, T., Sathyendranath, S., 1988. Oceanic primary production: estimation by remote sensing at local and regional scales. *Science* 241, 1613–1620.
- Platt, T., Sathyendranath, S., 1993. Estimators of primary production for interpretation of remotely-sensed data on ocean color. *Journal of Geophysical Research* 98, 14561–14576.
- Sanderson, M.P., Hunter, C.N., Fitzwater, S.E., Gordon, R.M., Barber, R.T., 1995. Primary productivity and trace metal contamination measurements from a clean rosette system versus ultra clean Go-Flo bottles. *Deep-Sea Research II* 42, 431–440.
- Scardi, M., 2000. Neuronal network models of phytoplankton primary production. In: Lek, S.G.J.-F. (Ed.), *Artificial Neuronal Networks: Application to Ecology and Evolution*. Springer, Berlin/Heidelberg, pp. 115–129.
- Scardi, M., 2001. Advances in neural network modeling of phytoplankton primary production. *Ecological Modelling* 146, 33–45.
- Siegel, D.A., Westberry, T.K., O'Brien, M.C., Nelson, N.B., Michaels, A.F., Morrison, J.R., Scott, A., Caporelli, E.A., Sorensen, J.C., Maritorena, S., Garver, S.A., Brody, E.A., Ubante, J., Hammer, M.A., 2001. Bio-optical modeling of primary production on regional scales: the Bermuda BioOptics project. *Deep-Sea Research II* 48, 1865–1896.
- Six, K.D., Maier-Reimer, E., 1996. Effects of plankton dynamics on seasonal carbon fluxes in an ocean general circulation model. *Global Biogeochemical Cycles* 10, 559–583.
- Smyth, T.J., Tilstone, G.H., Groom, S.B., 2005. Integration of radiative transfer into satellite models of ocean primary production. *Journal of Geophysical Research* 110, C10014. doi:10.1029/2004JC002784.
- Stow, C.A., Jolliff, J., McGillicuddy, D.J., Doney, S.C., Allen, J.I., Friedrichs, M.A.M., Rose, K.A., Wallhead, P., 2009. Skill assessment for coupled biological/physical models of marine systems. *Journal of Marine Systems* 76, 4–15 (this volume). doi:10.1016/j.jmarsys.2008.03.011.
- Takahashi, T., Sutherland, S.C., Feely, R.A., Cosca, C.E., 2003. Decadal variation of the surface water PCO₂ in the western and central equatorial Pacific. *Science* 302, 852–856.
- Tanré, D., Deroo, C., Duhaut, P., Herman, M., Morcrett, J., Perbos, J., Deschamps, P., 1990. Description of a computer code to simulate the satellite signal in the solar spectrum. *International Journal of Remote Sensing* 11, 659–668.
- Taylor, K.E., 2001. Summarizing multiple aspects of model performance in a single diagram. *Journal of Geophysical Research* 106, 7183–7192.
- Tjiputra, J.F., Polzin, D., Winguth, A.M.E., 2007. Assimilation of seasonal chlorophyll and nutrient data into an adjoint three-dimensional ocean carbon cycle model: sensitivity analysis and ecosystem parameter optimization. *Global Biogeochemical Cycles* 21, GB1001. doi:10.1029/2006GB002745.
- Wang, X.C., Chao, Y., 2004. Simulated sea surface salinity variability in the tropical Pacific. *Geophysical Research Letters* 31, L02302. doi:10.1029/2003GL018146.
- Wang, X.J., Christian, J.R., Murtugudde, R., Busalacchi, A.J., 2005. Ecosystem dynamics and export production in the central and eastern equatorial Pacific: a modeling study of impact of ENSO. *Geophysical Research Letters* 32. doi:10.1029/2204GL021538.
- Wang, X.J., Christian, J.R., Murtugudde, R., Busalacchi, A.J., 2006. Spatial and temporal variability of the surface water pCO₂ and air-sea CO₂ flux in the equatorial Pacific during 1980–2003: a basin-scale carbon cycle model. *Journal of Geophysical Research* 111. doi:10.1029/2005JC002972.
- Ware, D.W., Thomson, R.E., 2005. Bottom-up ecosystem trophic dynamics determine fish production in the northeast Pacific. *Science* 308, 1280–1284.
- Waters, K.J., Smith, R.C., Marra, J., 1994. Phytoplankton production in the Sargasso Sea as determined using optical mooring data. *Journal of Geophysical Research* 99, 18385–18402.
- Westberry, T.K., Behrenfeld, M.J., Siegel, E., Boss, E., 2008. Carbon-based primary productivity modeling with vertically resolved photosynthesis. *Global Biogeochemical Cycles* 22, GB2024. doi:10.1029/2007GB003078.
- Yamada, K., Ishizaka, J., Nagata, H., 2005. Spatial and temporal variability of satellite primary production in the Japan Sea from 1998 to 2002. *Journal of Oceanography* 61, 857–869.
- Yeager, S.G., Shields, C.A., Large, W.G., Hack, J.J., 2006. The low-resolution CCSM3. *Journal of Climate* 19, 2545–2566.