Optimized multi-satellite merger of primary production estimates in the California Current using inherent optical properties

CrossMark

Mati Kahru [a,*], Michael G. Jacox [c], Zhongping Lee [b], Raphael M. Kudela [c], Marlenne Manzano-Sarabia [d], B. Greg Mitchell [a]

[a] Scripps Institution of Oceanography, University of California San Diego, La Jolla, CA, USA
[b] Environmental Earth and Ocean Sciences, University of Massachusetts, Boston, MAUSA
[c] Ocean Sciences Department, University of California Santa Cruz, Santa Cruz, CA 95064, USA
[d] Facultad de Ciencias del Mar, Universidad Autónoma de Sinaloa, Mazatlán, Sinaloa, Mexico

ARTICLE INFO

ABSTRACT

Building a multi-decadal time series of large-scale estimates of net primary production (NPP) requires merging data from multiple ocean color satellites. The primary product of ocean color sensors is spectral remote sensing reflectance (Rrs). We found significant differences (13–18% median absolute percent error) between Rrs estimates at 443 nm of different satellite sensors. These differences in Rrs are transferred to inherent optical properties and further on to estimates of NPP. We estimated NPP for the California Current region from three ocean color sensors (SeaWiFS, MODIS-Aqua and MERIS) using a regionally optimized absorption based primary production model (Aph-PP) of Lee et al. (2011). Optimization of the Aph-PP model was required for each individual satellite sensor in order to make NPP estimates from different sensors compatible with each other. While the concept of Aph-PP has advantages over traditional chlorophyll-based NPP models, in practical application even the optimized Aph-PP model explained less than 60% of the total variance in NPP which is similar to other NPP algorithms. Uncertainties in satellite Rrs estimates as well as uncertainties in parameters representing phytoplankton depth distribution and physiology are likely to be limiting our current capability to accurately estimate NPP from space. Introducing a generic vertical profile for phytoplankton improved slightly the skill of the Aph-PP model.

© 2014 Elsevier B.V. All rights reserved.

1. Introduction

Oceanic primary production and phytoplankton biomass (often represented by chlorophyll-a concentration), estimated by satellite ocean color sensors have become important datasets to validate global earth system models (Gregg et al., 2003; Yoder et al., 2011). In fact, estimation of oceanic net primary production (NPP) is considered one of the principal applications of satellite ocean color measurements (McClain, 2009). While the first attempts to use satellite ocean color to estimate oceanic NPP (Eppley et al., 1985) were promising, subsequent progress has been modest. Algorithm testing and comparison campaigns (Campbell et al., 2002; Carr et al., 2006; Friedrichs et al., 2009; Saba et al., 2011) have shown clear limitations of current NPP models. In particular, most of the variability in modeled NPP estimates is directly related to the variability in surface chlorophyll-a concentration whereas including other variables adds little to the skill of the models. It appears that model performance is not directly related with model complexity.

Models underestimate observed NPP variability and fail to capture changes in productivity due to shifts in phytoplankton community composition (Friedrichs et al., 2009).

The primary input to most ocean primary productivity models (Behrenfeld and Falkowski, 1997) has been the concentration of chlorophyll-a (mg m^{-3}, $Chla$). However, operational satellite $Chla$ algorithms (O'Reilly et al., 1998, 2000) are based on the ratio of remote sensing reflectance (Rrs) at blue and green wavelengths and primarily represent changes in the total absorption at blue wavelengths. Total absorption of blue light is affected not just by chlorophyll and other phytoplankton pigments but also by colored dissolved organic matter (CDOM) and non-phytoplankton particulates (Siegel et al., 2005). Errors in estimating $Chla$ are directly transferred to errors in estimated NPP — not just due to the biased phytoplankton biomass but also due to the wide variation of biomass-normalized growth rate (Behrenfeld and Falkowski, 1997). Lee et al. (1996, 2011) proposed primary production algorithms that use phytoplankton absorption, $aph(\lambda)$, instead of $Chla$ as the primary input variable. At least in principle, those algorithms, based on inherent optical properties (IOPs), have some advantages over NPP algorithms using $Chla$. First, when using $aph(\lambda)$ estimated with an inversion model from $Rrs(\lambda)$ instead of a band-ratio derived

* Corresponding author. Tel.: +1 858 534 8947; fax: +1 858 822 0562.
E-mail address: mkahru@ucsd.edu (M. Kahru).

Chla, the influence of absorption by CDOM is minimized. Second, the uncertainty associated with *Chla*-specific absorption coefficient is avoided. The absorption-based productivity algorithm (Aph-PP) proposed by Lee et al. (2011) separates variables determined by phytoplankton physiology (e.g. the photosynthetic quantum yield) from the variables primarily determined by optics (e.g. IOPs). In a case study using a limited in situ dataset from the Southern Ocean, Lee et al. (2011) showed that the Aph-PP algorithm improved the estimates of NPP compared to traditional *Chla*-based methods. Here we apply the Lee et al (2011) Aph-PP method to satellite data from three ocean color sensors (SeaWiFS, MODIS-Aqua, MERIS) and evaluate its performance compared to a large in situ dataset collected by the California Cooperative Oceanic Fisheries Investigations (CalCOFI) and the closely related California Coastal Ecosystem-Long Term Ecological Research (CCE-LTER) program (Ohman and Venrick, 2003).

We start by showing differences between *Rrs* estimates obtained by different sensors. We then show that these differences are transferred to differences in NPP estimates in spite of using a formerly adapted set of algorithms to derive IOPs (*aph*(440)). We emphasize that this study is not aiming to find the best parameters of the NPP model derived from fundamental studies of phytoplankton photophysiology. Rather, we show problems encountered when applying an NPP model to noisy satellite data with sensor-specific anomalies and our empirical attempts to find optimal values for the model parameters.

2. Data and methods

2.1. In situ measurements of NPP

On-deck ^{14}C incubations have been performed on quarterly CalCOFI cruises since 1984. The half-day (local noon to sunset) primary production values, integrated over the euphotic depth, are multiplied by 1.8 to obtain equivalent 24 h productivity (Eppley, 1992). In this study we

used a subset of 861 stations occupied over the time period from 18-Sept-1997 to 12-Feb-2012 when satellite data from at least one of the three sensors of SeaWiFS, MODIS-Aqua (MODISA) and MERIS were available. The CalCOFI study area covers different regions of the California Current from coastal to about 600 km offshore (Fig. 1). In situ NPP measurements have considerable uncertainty with more uncertainty at low values. Saba et al. (2011) assumed uncertainties in NPP decreasing as a linear function of increasing $\log_{10}$(NPP), from 50% for NPP $\leq$50 mg C m^{-2} d^{-1} to 20% for NPP $\geq$ 2000 mg C m^{-2} d^{-1}. The distribution of in situ CalCOFI vertically integrated NPP values is log-normal with the mode between 200 and 400 mg C m^{-2} d^{-1} (Fig. 1 in Kahru et al., 2009).

2.2. Match-ups between satellite and in situ data

The validation of satellite products using quasi-simultaneous and spatially collocated measurements (match-ups) of satellite and in situ data followed the general procedures of previous studies (e.g. Bailey and Werdell, 2006; Kahru and Mitchell, 1999; Kahru et al., 2012; Werdell and Bailey, 2005). We acquired coincident Level-2 (i.e. processed to surface quantities but unmapped) data of SeaWiFS (1997–2010, version 2010.0), MODISA (2002–2012, version 2012.0) and MERIS (2002–2012, 3rd reprocessing). Full resolution (~1 km) SeaWiFS and MODISA data were obtained from NASA's Ocean Color web (http://oceancolor.gsfc.nasa.gov/) and the RR data (~1 km) of MERIS were obtained from ESA's MERIS Catalogue and Inventory (http://merci-srv.eo.esa.int/merci/welcome.do). For each Level-2 pixel we used the corresponding Level-2 flags. A pixel was determined valid if none of the following flags were set: ATMFAIL, LAND, HISATZEN, CLDICE, CHLFAIL, SEAICE, NAVFAIL, HIPOL and PRODFAIL (see http://oceancolor.gsfc.nasa.gov/VALIDATION/flags.html for explanation of the flags). For MERIS the following flags made the pixel invalid: LOW SUN, HIGH_GLINT, ICE_HAZE, SUSPECT, COASTLINE, PCD_19,
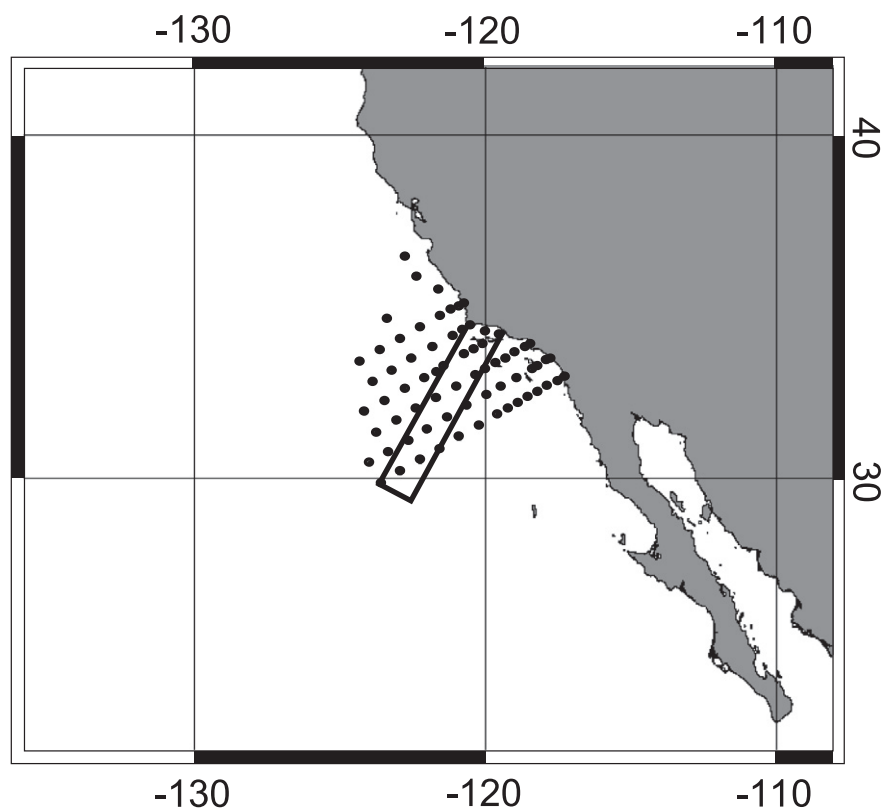

Fig. 1. Locations of the satellite (MODIS-Aqua, MERIS and SeaWiFS) match-ups (filled circles) with in situ primary productivity measurements. The rectangle shows the location of the strip from coast to offshore where satellite-to-satellite *Rrs443* match-ups were assembled.

PCD_18, PCD_17, PCD_16, PCD_15, PCD_14, PCD_1_13, CLOUD and LAND. All variables in Level-2 files were extracted from 3×3 pixel windows centered at the pixel nearest to the in situ sample. We accepted only those match-ups with at least 3 valid pixels (out of 9) and within 24 h of the in situ sample. We used this relaxed matchup criteria (rather than the more common 3 hour window) to increase the number of matchups, particularly at high NPP values. We note that the temporal decorrelation scales for the southern California Current are ~5 days (Frolov et al., 2013), or significantly longer than the 24 hour window. The mean $Rrs(\lambda)$ value of valid pixels within the 3×3 pixel window was used as input to the Lee et al. (2002) quasi-analytic algorithm (QAA) model as optimized by Kahru et al. (2013) and the derived coefficients of total absorption at 490 nm, $a(490)$, phytoplankton absorption at 440 nm, $aph(440)$, and particulate backscattering at 490 nm, $bbp(490)$ were used as input to Aph-PP. The spatial distribution of match-ups with in situ measurements of NPP is shown in Fig. 1.

2.3. Satellite to satellite match-ups

In order to evaluate the errors of the satellite $Rrs(\lambda)$ estimates we used Rrs data from spatially and temporally overlapping satellite sensors. We created satellite to satellite match-ups by finding the same pixel in Level-3 globally binned and mapped 9-km datasets for the same day. Daily Level-3 data are binned using the best quality Level-2 data, excluding lower quality pixel values that are present in Level-2 datasets. Temporally overlapping daily data are available from three sensors (SeaWiFS, MODISA and MERIS). Differences of a few hours in the timing between sensors are unavoidable as SeaWiFS overpass time was nominally at local noon but drifted later towards the afternoon while MODISA overpass time was approximately 1:30 PM and MERIS overpass time approximately 10 AM. SeaWiFS and MODISA Level-3 datasets were downloaded from NASA's Ocean Color website (http://oceancolor.gsfc.nasa.gov/) and MERIS Level-3 data were obtained from ESA's GlobColour Project at ftp://hermes.acri.fr/meris_l3. MERIS data were remapped to the same grid as the NASA data. Comparing Level-3 Rrs data instead of Level-2 data was justified as Level-3 data are binned to a common grid (either 4-km or 9-km grid) and gridded data are more directly comparable. We also compared the approximately 1-km unbinned Level-2 pixels of SeaWiFS to those of MODISA by mapping all 2004 Level-2 datasets to a common 1-km map.

2.4. Applying the Aph-PP model to match-ups with in situ data

The absorption-based productivity (Aph-PP) algorithm (Lee et al., 2011) is based on the relationship PP = $\phi \times$ PAR $\times$ aph where PP is primary production, ϕ is quantum yield of photosynthesis and aph is the coefficient of phytoplankton absorption. A spectrally integrated form of PP at depth z is:

$$PP(z) = \phi_m \frac{K_\phi \exp(-\nu \times E(z))}{K_\phi + E(z)} [aph \times E(z)], \qquad (1)$$

with

$$E(z) = E_0 \exp(-K_{PAR}(z) \times z). \qquad (2)$$

$K_{PAR}(z)$ is the vertical attenuation coefficient of PAR, which is modeled as a function of absorption and backscattering coefficients according to Lee et al. (2005):

$$K_{PAR}(z) = K_1 + K_2/(1 + z)^{0.5} \qquad (3)$$

with

$$K_1 = \left[K_{1-a0} + K_{1-a1}\,(a(490))^{0.5} + K_{1-a2}\,b_b(490) \right] (K_{1-7} + K_{2-7}\sin(\theta_a)) \qquad (4)$$

$$K_2 = [K_{2-a0} + K_{2-a1}\,a(490) + K_{2-a2}\,b_b(490)] (K_{1-6} + K_{2-6}\cos(\theta_a)) \quad (5)$$

and assuming that the solar zenith angle above the surface θ_a is equal to 30°.

The term associated with ϕ_m and K_ϕ on the right side of Eq. (1) represents phytoplankton physiology. Default values of the model parameters used were: $\phi_m = 0.12$ mg C E^{-1}, $K_\phi = 10$ E m^{-2} d^{-1}, and $\nu = 0.01$ (E m^{-2} $d^{-1})^{-1}$. aph on the right side of Eq. (1) is the absorption coefficient of phytoplankton at 443 nm. $E(z)$ in Eqs. (1) and (2) is the PAR value at depth z, with E_0 for surface PAR. Water-column PP (PP$_{eu}$) was calculated by integrating PP(z) between 0 and 100 m with a step of 2 m (or down to 1% of E_0).

Surface PAR (E_0 in Eq. (2), E m^{-2} d^{-1}, Frouin et al., 2003) was obtained from the standard mapped products of SeaWiFS, MODIS-Terra and MODIS-Aqua and merged by averaging valid pixels from different sensors. Due to their different orbits, the relatively small gaps between subsequent satellite orbits fall to different locations and were covered by one or more sensors. Additional errors may be introduced by using daily PAR instead of instantaneous PAR as the estimated $K_{PAR}(z)$ is not exactly the attenuation coefficient of daily PAR. We note that alternative PAR calculation schemes exist (Mobley and Boss, 2012) which could potentially improve the Aph-PP model. For this analysis we used the standard formulation to match the application of Lee et al. (2011) for the Southern Ocean.

The values of the coefficients in the Aph-PP model (Eq. (1)) as well as the coefficients in the light parameterization (Eqs. (4) and (5)) were used as the starting point of the minimization process which optimizes the values of these 13 coefficients for each of the 3 sensors by minimizing the resulting differences with in situ match-ups. For the optimization we used the Trust-Region method, a variant of the Levenberg–Marquardt method as implemented in the NMath numerical libraries (http://www.centerspace.net/). As both Chla and NPP are known to follow lognormal distributions, we used log_{10} transformed data in the optimization. After running the minimization procedure there was still residual bias which was removed with the following empirical adjustment: $YAdjusted = (Y - intercept)/slope$, where Y is the vector of predicted values after applying the coefficients derived in the optimization procedure, $YAdjusted$ is the adjusted vector of output values, $intercept$ and $slope$ are respectively the intercept and slope of the reduced major axis (RMA) regression between Y and the vector of in situ measurements X. RMA regression is more appropriate than the standard ordinary least squares regression as the independent variable (either the measured in situ values or the estimates using another satellite sensor) is measured with error (Sokal and Rohlf, 1995).

2.5. Statistical estimates of model performance

We used four statistical measures to assess the performance of the algorithms in comparisons between satellite products and in situ observations (satellite to in situ match-ups) and three statistical measures between different satellite sensors (i.e. satellite to satellite match-ups). In satellite to in situ match-ups O_i is the ith observation of an in situ variable and P_i is the corresponding predicted satellite variable. In satellite to satellite match-ups the choice of the observed versus predicted variable is arbitrary but we used MODISA as the common variable when comparing with MERIS and SeaWiFS values. For a measure of covariance that shows how much variance in one variable can be predicted from another we used the coefficient of determination (R^2) on log_{10}

transformed variables. As an estimate of general bias (e.g. too high or too low) we used the Median Relative Percent Error (MRPE) that was calculated as MRPE $= 100 *$ median $((P_i - O_i)/O_i)$. As an estimate of scatter we used the Median Absolute Percent Error (MAPE) that was calculated as MAPE $= 100 *$ median $(Abs((P_i - O_i)/O_i))$. For comparison with previous studies (e.g. Kahru et al., 2009) we added a measure of total variance, root–mean–square difference (RMSD) applied to $\log_{10}$ transformed variables:

$$RMSD = \left(\frac{1}{n} \sum_{i=1}^{n} \left[\log P_i - \log O_i \right]^2 \right)^{1/2}$$

3. Results

3.1. Comparison of remote sensing reflectance from different sensors

Temporally and spatially overlapping satellite data from SeaWiFS, MODISA and MERIS were used to evaluate the accuracy of Rrs measurements by comparing the Rrs values of daily, binned 9-km datasets with corresponding Rrs values of another sensor. In order to limit the number of overlapping pixels (which can be overwhelming) but to cover the full range of water masses from coastal to offshore, we selected a test area as a strip of approximately 100 km wide and approximately 560 km long stretching from the coast to offshore in SW direction just south of Pt. Conception along the border between Central and Southern California (Fig. 1). Fig. 2 shows a comparison of $Rrs443$ acquired by different sensors
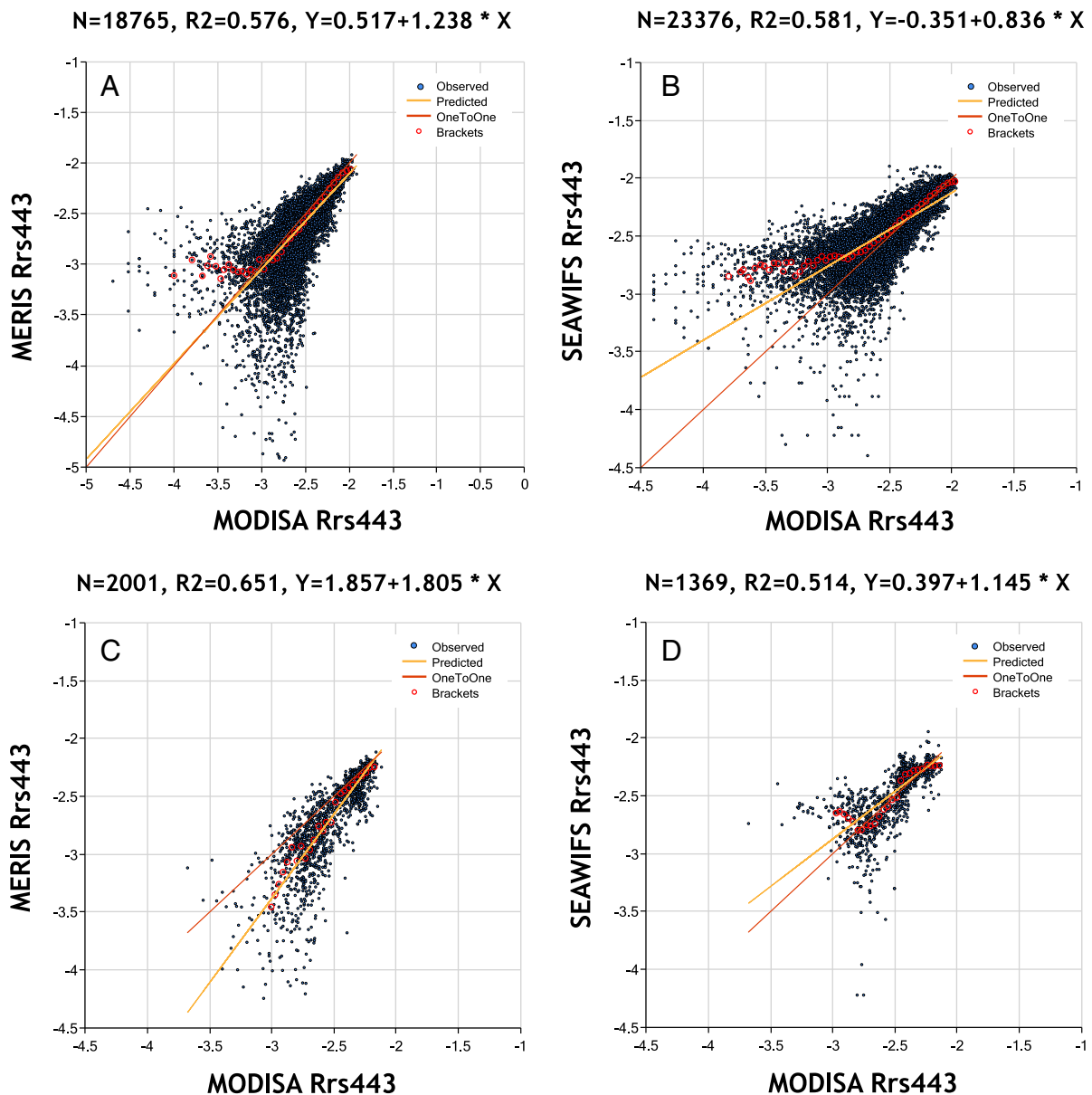

Fig. 2. Inter-sensor comparison of $\log_{10}$-transformed remote sensing reflectance at 443 nm ($Rrs443$). Black dots show same day match-ups between Level-3 global 9-km binned datasets, red line is the one-to-one line, yellow line is the least squares linear regression. Small red circles are the median bracket points that were generated by dividing the full horizontal range (in $\log_{10}$ units) into 100 equal sections and finding median values for the abscissa variable and the corresponding points of the ordinate variable. A, MERIS versus MODISA in 2004; B, SeaWiFS versus MODISA in 2004; C, subset of A of January, 2004; D, subset of B of January, 2004.

during the full year of 2004 as well as monthly subsets of January, 2004. Year 2004 was chosen as data from SeaWiFS, MODISA and MERIS were available and all sensors were in good health with minimal calibration problems. Each point in Fig. 2 corresponds to the same pixel value obtained on the same day but estimated at a slightly different overpass time by a different satellite sensor. In spite of the vicarious calibration (Franz et al., 2007) where satellite measurements of the different sensors are fitted to the same high-quality field measurements of ocean remote sensing reflectance, we note systematic differences. In addition to systematic differences, there is considerable random scatter, particularly at low values. The coefficient of determination is slightly lower between MODISA and MERIS but both relative and absolute errors are slightly higher between MODISA and SeaWiFS (MRPE −9.7% versus 11% and MAPE 15.4% versus 18.0%, Table 1). During any particular month the correspondence between $Rrs443$ of different sensors can be even lower, e.g. $R^2 = 0.51$ in January, 2004 (Fig. 2D) or R^2 can be higher but the relationship can deviate more from the one-to-one relationship (Fig. 2C). The ratio of the median bracket points (Fig. 3) shows that SeaWiFS $Rrs443$ values tend to be higher than the corresponding MODISA $Rrs443$ values and MERIS $Rrs443$ tends to be lower than the corresponding MODISA $Rrs443$. At high $Rrs443$ the ratios are close to 1, indicating better agreement offshore, but at low $Rrs443$ the scatter increases drastically and the median ratios diverge, indicating significant differences between sensors in the coastal zone where $Rrs443$ is typically low. When Rrs values approach zero, the scatter in log_{10} transformed values increases dramatically and we get spurious estimates. Small differences between sensors are expected as different satellite sensors have different optical characteristics (even if they have the same nominal waveband), different sensitivities, signal to noise ratios, calibration and degradation histories. Some of the observed variability is due to true changes in surface properties during different satellite overpass times while the rest must be due to errors caused by different observation geometries, variable influences by sun glint, solar zenith angle, and clouds.

In this paper, we do not intend to make a comprehensive analysis of the differences and errors between the various Rrs bands of the different sensors and just note the significant differences. We use the $Rrs443$ band as an example as it is important for the detection of $Chla$ and phytoplankton. Both systematic and random variations in $Rrs443$ (and other bands) affect our estimates of IOPs (or $Chla$) and limit our ability to accurately estimate NPP. We use the existence of systematic differences between Rrs estimates by different satellite sensors as justification for designing empirical algorithms for IOPs and NPP optimized for each sensor and not for ideal, model generated data.

3.2. Optimizing the Aph-PP model for individual satellite sensors

We used the Lee et al. (2002) quasi-analytic algorithm (QAA, http://www.ioccg.org/groups/Software_OCA/QAA_v5.pdf) that was optimized to our IOP match-ups (Kahru et al., 2013). The optimized QAA was applied to Level-2 satellite Rrs match-ups with in situ NPP data and the following products were used as input to the Aph-PP model: the total absorption coefficient at 490 nm ($a490$), coefficient of particulate backscatter at 490 nm ($bbp490$), and coefficient of phytoplankton absorption at 440 nm ($aph440$). For PAR, we used a merged
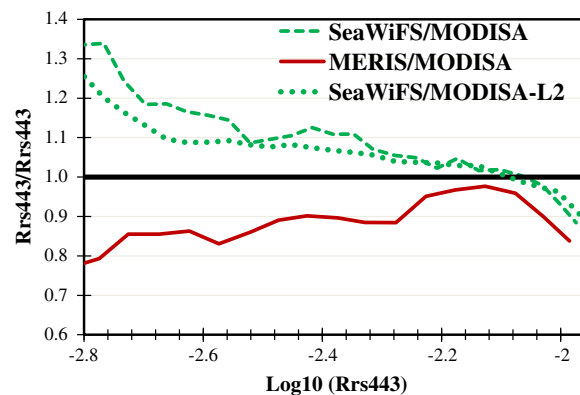

Fig. 3. Ratio of SeaWiFS $Rrs443$ to MODISA $Rrs443$ (green dashed, green dotted) and MERIS $Rrs443$ to MODISA $Rrs443$ (red). The plots show the median bracket points of the satellite to satellite pixel match-ups over the year of 2004. SeaWiFS/MODISA and MERIS/MODISA are using daily Level-3 binned 9-km datasets whereas SeaWiFS/MODISA-L2 is using 1-km Level-2 data mapped to a common 1-km projection. The heavy black line corresponds to the ideal case of no inter-sensor bias.

daily PAR estimate derived from standard PAR products of SeaWiFS, MODIS-Aqua and MODIS-Terra, matched spatially to the pixel nearest to the in situ NPP measurement. Although the output of the Aph-PP model using standard coefficients has the same order of magnitude with in situ NPP, the estimates are considerably biased (Fig. 4, top row; Table 2). After tuning the 13 coefficients included in Eqs. (1), (4) and (5) by minimizing the differences with in situ data we were able to remove the systematic bias and somewhat reduce the scatter. When models with the same coefficients are applied to Rrs data from different satellite sensors, the pooled datasets have typically lower R^2 than the respective R^2 of the individual sensors (Fig. 4 top row and Figs. 6 to Fig. 7 top row). This is due to the pooling of data with somewhat different relationships (as illustrated in Fig. 4A, B, C). For example, NPP estimates assuming standard Aph-PP model produced R^2 for match-ups with individual sensors as 0.318, 0.373 and 0.497 but only 0.282 for the pooled match-ups (Table 2). The differences in the Rrs values of different sensors increase the divergence in the pooled results, decrease R^2 and make it error-prone to apply models with standard coefficients to Rrs of different sensors. On the other hand, fitting separate models to match-ups from individual sensors makes the output more compatible and produces higher R^2 of the pooled datasets (Table 2). However, even after optimization the coefficient of determination (R^2) remained below 0.6 (Fig. 4, bottom row). NPP values derived with SeaWiFS data tend to have higher R^2, particularly for the standard model (Fig. 4C). The standard Aph-PP model assumes vertically uniform absorption and backscattering properties in the surface mixed layer. Vertical profiles of phytoplankton (measured as $Chla$) in the California Current often show subsurface chlorophyll maxima (Cullen and Eppley, 1981; Millan-Nunez et al., 1997). When the subsurface maximum is above the depth of the euphotic zone then it may add substantially to the production of the water column (Platt and Sathyendranath, 1993). We therefore modeled the vertical $Chla$ distribution with the Gaussian distribution of Platt et al. (1988) using the mean parameter values for Central California inshore region from Millan-Nunez et al. (1997) (Fig. 5). As shown by Frolov et al. (2012), the vertical $Chla$ distribution off Central California is variable and consistent parameterizations allowing accurate prediction of the vertical profile from surface value are not available. We therefore consider our vertical profile a "generic" vertical profile that predicts increasing $Chla$ in the subsurface layer which is consistent with observations but which does not pretend to be an accurate parameterization for any particular match-up. We then assumed that the relative vertical distributions of the IOPs ($a490$, $bbp490$ and $aph440$) were similar to the relative vertical profile of $Chla$. While this assumption is probably not very accurate even in Case

Table 1
Statistical relationships between $Rrs443$ estimated for the same pixel, same day by different satellite sensors. MRPE = median relative percent error, MAPE = median absolute percent error.

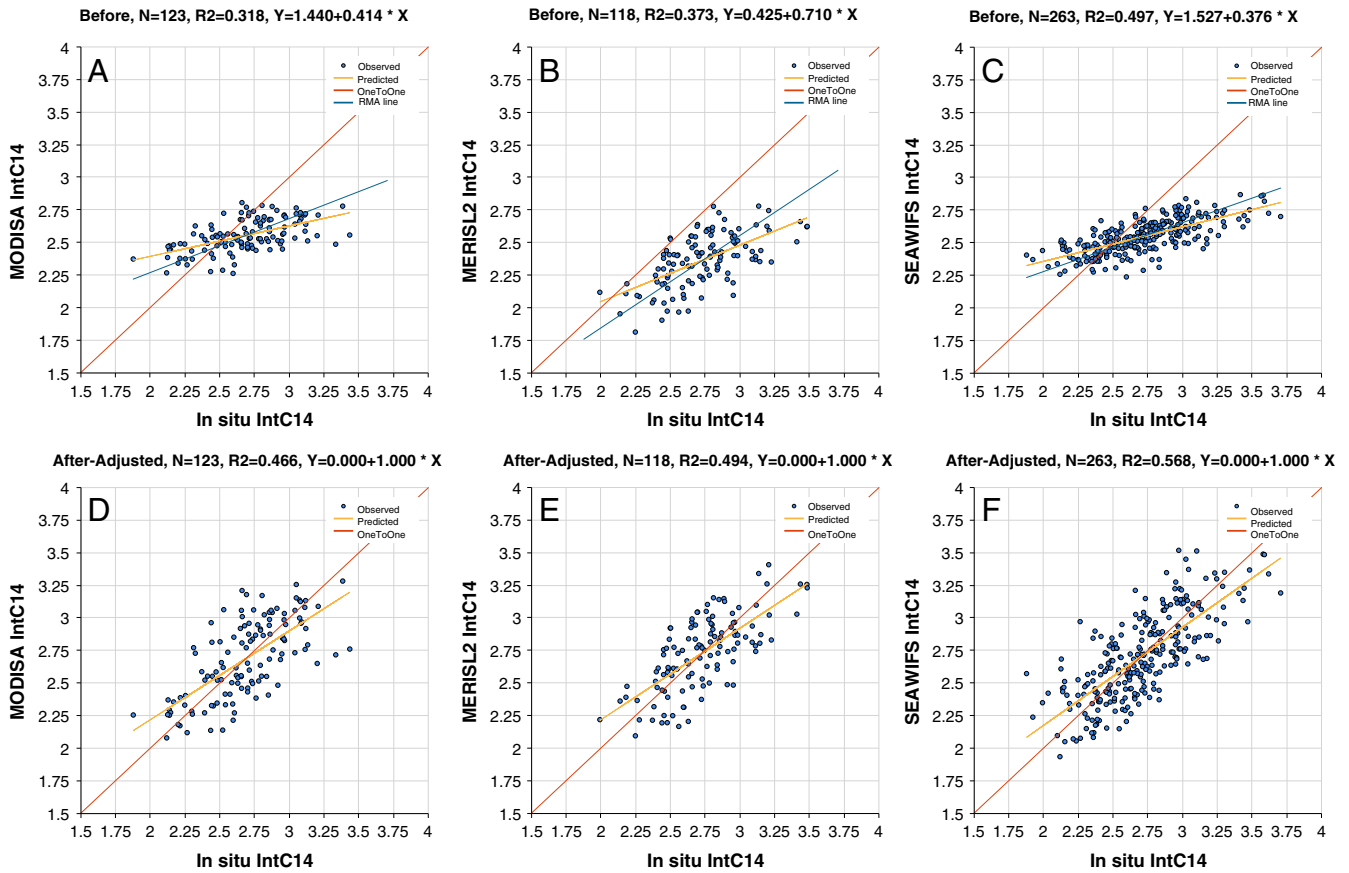|	R^2	MRPE	MAPE
MERIS/MODISA, 2004	0.576	−9.7	15.4
MERIS/MODISA, January−2004	0.651	−14.0	15.9
SeaWiFS/MODISA, 2004	0.581	11.0	18.0
SeaWiFS/MODISA, January−2004	0.513	11.7	19.5
SeaWiFS/MODISA L2, 2004	0.631	7.1	13.4

Fig. 4. Results of applying the Aph-PP model to Level-2 data from MODIS-Aqua (left column), MERIS (middle column) and SeaWiFS (right column) compared to match-ups of in situ daily vertically integrated NPP. The Aph-PP model assumes mixed surface layer and is using the default coefficients (top row) or optimized coefficients (bottom row).

1 waters where IOPs are positively well correlated with *Chla*, it is a first approximation to the vertical structure having a subsurface *Chla* maximum. It appears that by assuming the generic vertical distribution we can obtain a slight increase in the predictive skill of the Aph-PP model (Fig. 6, Table 2).

Table 2
A comparison of the performance of the standard Aph-PP model (mixed layer) versus the optimized model for mixed layer and the optimized model with generalized vertical profile as applied to match-ups with in situ NPP from MODIS-Aqua (MODISA/in situ), MERIS (MERIS/In situ), SeaWiFS (SeaWiFS/in situ) and from all sensors combined (all sensors/in situ). R^2 = coefficient of determination, MRPE = median relative percent error, MAPE = median absolute percent error, RMSD = root–mean–square deviation. Bold numbers show improvement from standard to optimized and to optimized with a generic vertical profile. R^2 and RMSD are calculated for $\log_{10}$-transformed variables.

Match-up type	Measure	Standard Aph-PP	Optimized Aph-PP for mixed layer	Optimized Aph-PP for generalized vertical profile
MODISA/in situ	R^2	**0.318**	**0.466**	**0.532**
	MRPE	**−29.5**	**−2.5**	**−0.9**
	MAPE	**41.7**	**31.6**	32.4
	RMSD	**0.286**	**0.221**	**0.206**
MERIS/in situ	R^2	**0.373**	**0.494**	**0.540**
	MRPE	**−30.1**	**−4.1**	**−2.1**
	MAPE	**43.2**	**32.9**	33.4
	RMSD	**0.291**	**0.225**	**0.210**
SeaWiFS/In situ	R^2	**0.497**	**0.568**	**0.579**
	MRPE	**−41.4**	**1.1**	3.1
	MAPE	**46.6**	**35.9**	32.4
	RMSD	**0.360**	**0.218**	**0.202**
All sensors/in situ	R^2	**0.282**	**0.526**	**0.562**
	MRPE	**−35.2**	**−0.7**	1.2
	MAPE	**42.1**	**34.5**	**31.6**
	RMSD	**0.337**	**0.218**	**0.206**

4. Discussion

Satellite ocean color products are critical for building Climate Data Records (CDR; National Research Council, 2004). However, merging data from multiple satellite sensors in order to extend the length of the time series beyond the limited lifetime of a single sensor is still problematic. Applying models that are sensitive to sensor-specific differences in satellite *Rrs* values may magnify these differences and produce incompatible time series, particularly when switching from one sensor to another. This leads to significant errors in trend estimates. In order to make output from different sensors compatible with each other, we tune the model for each individual sensor by using match-ups with in situ data. Ideally, a more straightforward approach would be to remove the anomalies from satellite *Rrs* instead of fitting in situ data to different estimates of *Rrs*. However, the inter-sensor anomalies in *Rrs* are much more complex than has been acknowledged to date (Mélin, 2013). In our analysis we explicitly acknowledge the existence of differences between the *Rrs* products of different sensors and try to create optimized algorithms for each sensor in order to produce compatible data products that can then be merged in space and time. Our analysis is restricted by availability of well-distributed in situ data. Our method is essentially a "regional optimization" as it depends on a regional set of in situ data and it assumes that the *Rrs* anomalies are temporally consistent. To retrieve IOPs we used the QAA model adapted previously in Kahru et al. (2013). Ideally, that should have removed the effect of inter-sensor differences in *Rrs*. However, the calculated NPP products clearly showed large differences and inaccuracies (Fig. 4) which was probably due to a combination of random errors in satellite retrievals and poorly distributed in situ datasets used in the optimization of the IOP algorithms. Here we have used empirical optimization of the absorption-based productivity model for three different satellite
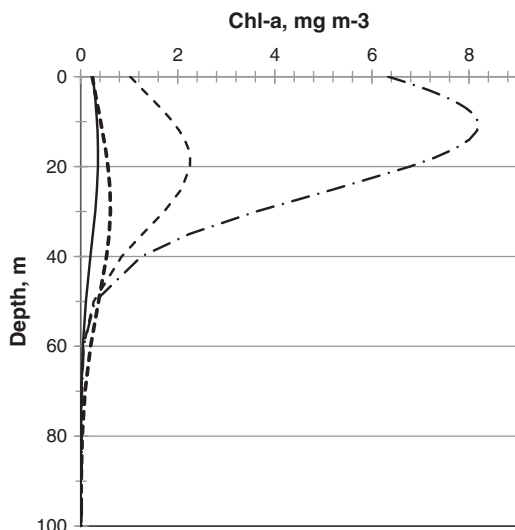
Fig. 5. Generalized vertical profile of *Chla* for surface concentrations of 0.05, 0.3, 1 and 5 mg m^{-3}.

sensors with the goal to produce results that are compatible with in situ measurements.

Gregg and Conkright (2001) and Gregg et al. (2009) pioneered the approach of blending satellite *Chla* data with in situ data. Kahru et al. (2012) optimized band-ratio *Chla* algorithms in the California Current for multiple sensors and built a merged time series. Kahru et al. (2013) used the same idea in a more complex task to produce compatible IOP estimates from multiple sensors. Our previous analysis of various NPP algorithms as applied to remotely sensed SeaWiFS data and compared to CalCOFI in situ productivity measurements (Kahru et al., 2009) showed that all tested algorithms were somewhat biased but explained approximately 60% of the total variance. The absorption-based Lee et al. (2011) algorithm has theoretical advantages over traditional productivity algorithms, and is much improved overly purely empirical algorithms since it is based on mechanistic, physiological parameters. This should lead the way to improved estimates of NPP if these underlying parameters can be adequately estimated or derived from remote sensing data. However in practice this approach also did not explain more than about 60% of the total variance and had absolute error over 30% which is similar to other algorithms.

There are multiple reasons why most commonly used NPP models, including the Aph-PP model, have similar errors. First, satellite measurements of *Rrs* have significant errors (~15–18%) which may be magnified when sensitive inversion algorithms are applied to produce IOPs.

In high-chlorophyll waters *Rrs443* approaches zero and relative errors are particularly high. Over the full range of variability, estimates of *a490* are more robust than estimates of other IOPs. Second, parameters dependent on phytoplankton physiology, community composition and light adaptation are poorly known and their variability is not adequately captured in current productivity models. Third, seasonal variability in PAR related to the solar zenith angle is ignored. We intend to work on this in the future. Better characterization of the subsurface light field is probably needed. Fourth, the depth variation of phytoplankton biomass (or absorption) is poorly parameterized. Including a simple generalized depth profile resulted in slightly increased coefficient of determination. Jacox et al. (2013) using in large part the same CalCOFI in situ dataset, showed that adding details of the subsurface chlorophyll and light profiles provided significant improvement in the skill of the models. Including vertical profiles of measured light and chlorophyll concentration improved the predictions of NPP considerably ($R^2 = 0.77$). Mean *Chla* depth profiles can be parameterized but the accuracy of these parameterizations has been shown to be low, at least off Central California (Frolov et al., 2012). As shown by Jacox et al. (2013), using autonomous underwater gliders, capable of measuring subsurface properties on long-term, long-range deployments, can provide significantly improved depth profiles of light and phytoplankton biomass and improve the skill of productivity algorithms. While gliders are increasingly routinely deployed, coverage remains insufficient at this time to dramatically improve satellite-based NPP estimates.

The remaining ~25% variability between modeled and measured NPP includes both the poorly constrained variability in physiology, and the errors associated with in situ productivity estimates (e.g. Marra, 2008). A recent comparison of the CalCOFI NPP data, VGPM, and oxygen isotope estimates suggest that methodological issues could consistently lower the in situ NPP rates (Munro et al., 2013). However, this would probably not account for the model performance reported here, which generally underestimates the in situ NPP values, but it does highlight the difficulty of assessing the accuracy of the in situ data. The absorption-based approach also requires parameterization of the photosynthetic quantum yield. Here we used the default values (see Section 2.4), but a previous analysis of CalCOFI data (Sosik, 1996) showed 10-fold variation in this value, strongly related to environmental factors including optical depth, distance to the nutricline, and temperature, with less variability in the upper mixed layer compared to values deeper in the water column. Similarly, Kudela and Chavez (1997) showed 6-fold variation in the quantum yield of photosynthesis from data collected in central California. Other analyses have shown a strong relationship between quantum yield and iron stress (e.g. Hiscock et al., 2008), and evidence for potential regulation of phytoplankton growth (NPP) by iron in the southern California Current (King and Barbeau, 2007). This suggests that there
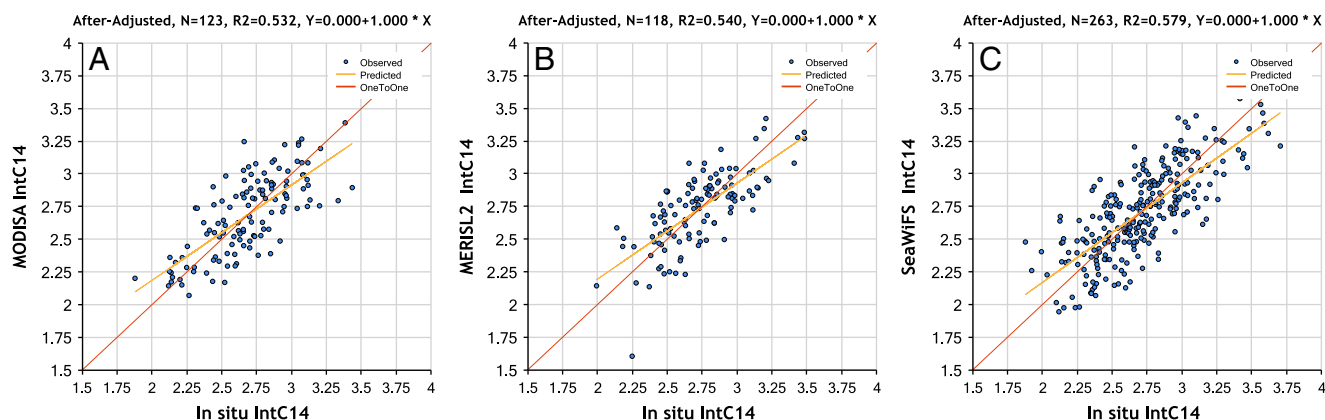

Fig. 6. Results of applying the Aph-PP model to Level-2 data from MODIS-Aqua (A), MERIS (B) and SeaWiFS (C) compared to match-ups of in situ daily vertically integrated NPP. The Aph-PP model assumes generalized vertical profile and optimized coefficients.
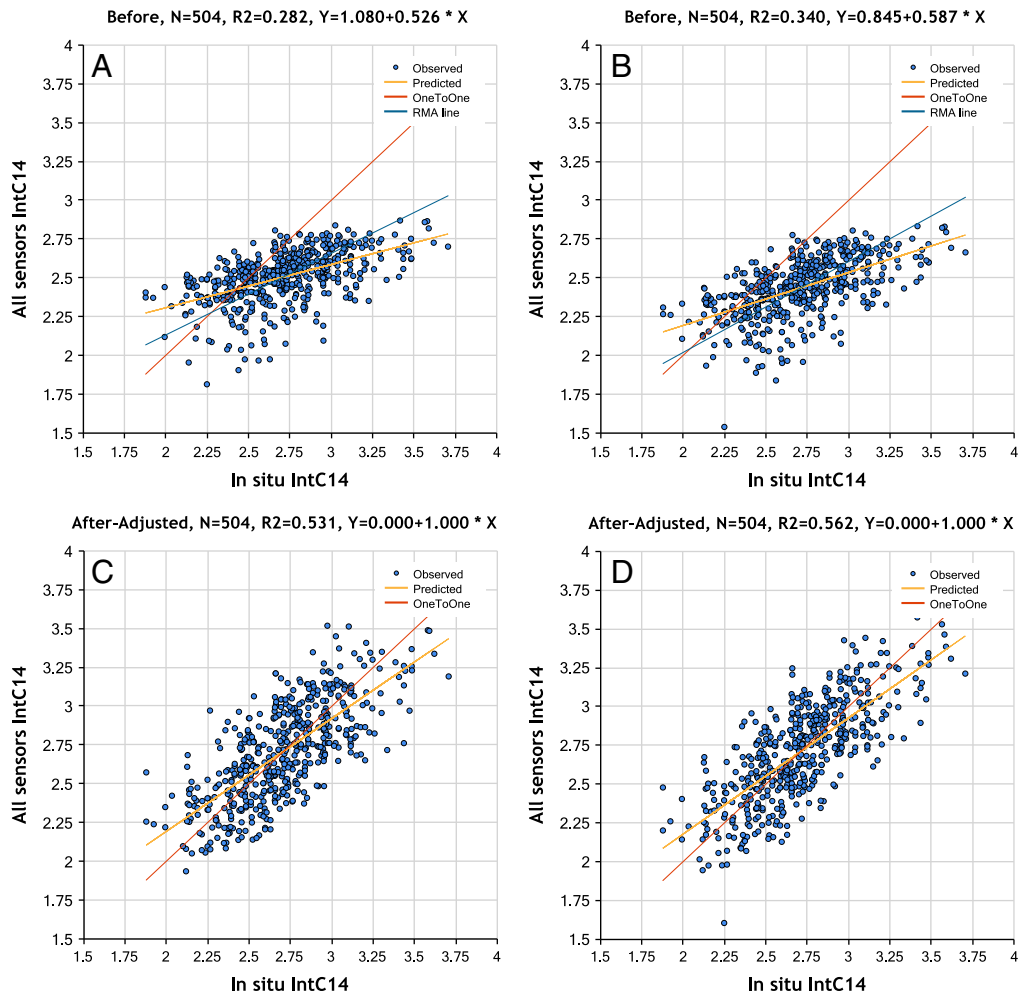
Fig. 7. Results of applying the Aph-PP model to pooled Level-2 satellite data from MODIS-Aqua, MERIS and SeaWiFS compared to in situ measurements of daily vertically integrated NPP. The Aph-PP model assumes mixed top layer (left columns) and generalized vertical profile (right column) while using the default coefficients (top row) and the optimized coefficients (bottom row).

is room for improvement in the absorption-based model by parameterizing some of these environmental variables and allowing the physiological parameters to vary accordingly. Optical depth, temperature, and potentially, iron stress (McGaraghan and Kudela, 2012) can be parameterized from remote sensing. The link between nutricline depth and improvement in NPP estimates by incorporating depth profiles (Jacox et al., 2013) also points to potential improvements of the absorption-based models with the inclusion of vertical profiles from gliders.

5. Conclusions

We show that in spite of recent efforts in vicarious calibration of satellite sensors, differences between *Rrs* products (approximately 15–18% for *Rrs443*) of different sensors (MODIS-Aqua, MERIS, SeaWiFS) are large enough to cause significant differences in the derived products, such as IOPs and NPP. Even when using IOP products that have been produced by regional algorithms (tuned to in situ IOP measurements) as inputs to the absorption-based productivity model, the estimates of NPP were biased and systematically different between sensors. Optimization by means of adapting coefficients for each sensor reduced the bias and somewhat reduced the scatter but still failed to explain more than 60% of the total variation in measured NPP. Uncertainties in satellite estimates of *Rrs* as well as uncertainties in parameters representing phytoplankton physiology, light adaptation and depth distribution are likely limiting our current capability to accurately estimate NPP from space in the California Current. Moving towards IOP-based models of NPP is clearly advantageous in that it provides additional variables and is based on a theoretical, rather than empirical, perspective. However, given the similar predictive performance for NPP regardless of algorithm (Kahru et al., 2009), it is unlikely that additional significant advances will be made by developing new algorithms. The most promising near-term improvement will likely come from a combination of reduced error in *Rrs*, merging satellite and in situ observations, and perhaps introduction of more refined estimates of biomass and physiology such as through the development of phytoplankton functional type models (Friedrichs et al., 2009; Nair et al., 2008) and parameterization of varying quantum yields (Sosik, 1996).

Appendix A. List of notations and acronyms.

Symbol	Explanation	Units
a	Coefficient of absorption	m^{-1}
$a490$	Coefficient of absorption at 490 nm	m^{-1}
aph	Coefficient of absorption by phytoplankton	m^{-1}
$aph440$	aph at 440 nm	m^{-1}
bbp	Coefficient of backscattering by particles	m^{-1}
$bbp490$	bbp at 490 nm	m^{-1}
MAPE	Median Absolute Percent Error	%
MRPE	Median Relative Percent Error	%
MERIS	Medium resolution imaging instrument on ESA Envisat satellite	
MODISA	Moderate Resolution Imaging Spectroradiometer on NASA Aqua satellite	
PAR	Photosynthetically available radiation	$E\ m^{-2}\ s^{-1}$
SeaWiFS	Sea-viewing Wide Field-of-view Sensor	
R^2	Coefficient of determination	
RMA	Reduced Major Axis (regression)	
Rrs	Remote sensing reflectance	sr^{-1}
$Rrs443$	Remote sensing reflectance at 443 nm	sr^{-1}

References

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

Behrenfeld, M.J., Falkowski, P.G., 1997. Photosynthetic rates derived from satellite based chlorophyll concentration. Limnol. Oceanogr. 42, 1–20.

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. Glob. Biogeochem. Cycles 16, 1035. http://dx.doi.org/10.1029/2001GB001444.

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 Quéré, C., Lohrenz, S., Marra, J., Mélin, F., Moore, K., 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 Res. II 53, 741–770.

Cullen, J.J., Eppley, R.W., 1981. Chlorophyll maximum layers of the Southern California Bight and possible mechanisms of their formation and maintenance. Oceanol. Acta 4 (1), 23–32.

Eppley, R.W., 1992. Chlorophyll, photosynthesis and new production in the Southern California Bight. Prog. Oceanogr. 30, 117–150. http://dx.doi.org/10.1016/0079-6611(92)90010-W.

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. J. Plankton Res. 7, 57–70.

Franz, B.A., Bailey, S.W., Werdell, P.J., McClain, C.R., 2007. Sensor-independent approach to the vicarious calibration of satellite ocean color radiometry. Appl. Opt. 46, 5068–5082.

Friedrichs, M.A.M., Carr, M.-E., Barber, R.T., Scardi, M., Antoine, D., Armstrong, R.A., Asanuma, I., Behrenfeld, M.J., Buitenhuis, E.T., Chai, F., Christian, J.R., Ciotti, A.M., Doney, S.C., Dowell, M., Dunne, J., Gentili, B., Gregg, W., Hoepffner, N., Ishizaka, J., Kameda, T., Lima, I., Marra, J., Mélin, F., Moore, J.K., Morel, A., O'Malley, R.T., O'Reilly, J., Saba, V.S., Schmeltz, M., Smyth, T.J., Tjiputra, J., Waters, K., Westberry, T.K., Winguth, A., 2009. Assessing the uncertainties of model estimates of primary productivity in the tropical Pacific Ocean. J. Mar. Syst. 76, 113–133.

Frolov, S., Ryan, J.P., Chavez, F.P., 2012. Predicting euphotic-depth-integrated chlorophyll-a from discrete-depth and satellite-observable chlorophyll-a off central California. J. Geophys. Res. 117, C05042. http://dx.doi.org/10.1029/2011JC007322.

Frolov, S., Kudela, R.M., Bellingham, J.G., 2013. Monitoring of harmful algal blooms in the era of diminishing resources: a case study of the U.S. West Coast. Harmful Algae 21–22, 1–12.

Frouin, R., Franz, B.A., Werdell, P.J., 2003. The SeaWiFS PAR product. In: Hooker, S.B., Firestone, E.R. (Eds.), Algorithm Updates for the Fourth SeaWiFS Data Reprocessing, p. 22 (NASA/TM-2003-206892).

Gregg, W.W., Conkright, M.E., 2001. Global seasonal climatologies of ocean chlorophyll: blending in situ and satellite data for the coastal zone color scanner era. J. Geophys. Res. 106 (C2), 2499–2516.

Gregg, W.W., Conkright, M.E., Ginoux, P., O'Reilly, J.E., Casey, N.W., 2003. Ocean primary production and climate: global decadal changes. Geophys. Res. Lett. 30 (15), 1809. http://dx.doi.org/10.1029/2003GL016889.

Gregg, W.W., Casey, N.W., O'Reilly, J.E., Esaias, W.E., 2009. An empirical approach to ocean color data: reducing bias and the need for post-launch radiometric re-calibration. Remote Sens. Environ. 113, 1598–1612. http://dx.doi.org/10.1016/j.rse.2009.03.005.

Hiscock, M.R., Lance, V.P., Apprill, A.M., Bidigare, R.R., Johnson, Z.I., Mitchell, B.G., Smith Jr., W.O., Barber, R.T., 2008. Photosynthetic maximum quantum yield increases are an essential component of the Southern Ocean phytoplankton response to iron. Proc. Natl. Acad. Sci. U. S. A. 105, 477–480.

Jacox, M.G., Edwards, C.A., Kahru, M., Rudnick, D.L., Kudela, R.M., 2013. The potential for improving remote primary productivity estimates through subsurface chlorophyll and irradiance measurement. Deep-Sea Res. II. http://dx.doi.org/10.1016/j.dsr2.2013.12.008.

Kahru, M., Mitchell, B.G., 1999. Empirical chlorophyll algorithm and preliminary SeaWiFS validation for the California Current. Int. J. Remote Sens. 20, 3423–3429.

Kahru, M., Kudela, R., Manzano-Sarabia, M., Mitchell, B.G., 2009. Trends in primary production in the California Current detected with satellite data. J. Geophys. Res. 114, C02004. http://dx.doi.org/10.1029/2008JC004979.

Kahru, M., Kudela, R.M., Manzano-Sarabia, M., Mitchell, B.G., 2012. Trends in the surface chlorophyll of the California Current: merging data from multiple ocean color satellites. Deep-Sea Res. II 77–80, 89–98.

Kahru, M., Lee, Z., Kudela, R.M., Manzano-Sarabia, M., Mitchell, B.G., 2013. Multi-satellite time series of inherent optical properties in the California Current. Deep-Sea Res. II. http://dx.doi.org/10.1016/j.dsr2.2013.07.023.

King, A.L., Barbeau, K., 2007. Evidence for phytoplankton iron limitation in the southern California Current System. Mar. Ecol. Prog. Ser. 342, 91–103.

Kudela, R.M., Chavez, F.P., 1997. Estimating new production from the quantum yield of nitrate uptake. Proc. SPIE, Ocean Opt. XIII 2963, 471–476.

Lee, Z.P., Carder, K.L., Marra, J., Steward, R.G., Perry, M.J., 1996. Estimating primary production at depth from remote sensing. Appl. Opt. 35 (3), 463–474. http://dx.doi.org/10.1364/AO.35.000463.

Lee, Z.P., Carder, K.L., Arnone, R., 2002. Deriving inherent optical properties from water color: a multi-band quasi-analytical algorithm for optically deep waters. Appl. Opt. 41, 5755–5772.

Lee, Z.P., Du, K.P., Arnone, R., Liew, S.C., Penta, B., 2005. Penetration of solar radiation in the upper ocean — a numerical model for oceanic and coastal waters. J. Geophys. Res. 110, C09019. http://dx.doi.org/10.1029/2004JC002780.

Lee, Z., Lance, V.P., Shang, S., Vaillancourt, R., Freeman, S., Lubac, B., Hargreaves, B.R., Del Castillo, C., Miller, R., Twardowski, M., Wei, G., 2011. An assessment of optical properties and primary production derived from remote sensing in the Southern Ocean (SO GasEx). J. Geophys. Res. 116, C00F03. http://dx.doi.org/10.1029/2010JC006747.

Marra, J., 2008. Approaches to the measurement of plankton production. In: Williams, P.J. Le B., Thomas, D.N., Reynolds, C.S. (Eds.), Phytoplankton productivity: Carbon assimilation in marine and freshwater ecology, pp. 78–108.

McClain, C.R., 2009. A decade of satellite ocean color observations. Annu. Rev. Mar. Sci. 1, 19–42.

McGaraghan, A.R., Kudela, R.M., 2012. Estimating labile particulate iron concentrations in coastal waters from remote sensing. J. Geophys. Res. 117, C02004.

Mélin, F., 2013. In search of long-term trends in ocean colour record. International Ocean Colour Science Meeting 2013, pp. 162–163 (Submitted abstracts).

Millan-Nunez, R., Alvarez-Borrego, S., Trees, C.C., 1997. Modeling the vertical distribution of chlorophyll in the California Current System. J. Geophys. Res. 102 (C4), 8587–8595.

Mobley, C.D., Boss, E.S., 2012. Improved irradiances for use in ocean heating, primary production, and photo-oxidation calculations. Appl. Opt. 51 (27), 6549–6560.

Munro, D.R., Quay, P.D., Juranek, L.W., Goericke, R., 2013. Biological production rates off the Southern California coast estimated from triple O_2 isotopes and O_2:Ar gas ratios. Limnol. Oceanogr. 58 (44), 1312–1328.

Nair, A., Sathyendranath, S., Platt, T., Morales, J., Stuart, V., Forget, M.-H., Devred, E., Bouman, H., 2008. Remote sensing of phytoplankton functional types. Remote Sens. Environ. 112 (8), 3366–3375.

National Research Council, 2004. Climate data records from environmental satellites. Natl. Acad. Press, Washington, D.C.,

O'Reilly, J.E., Maritorena, S., Mitchell, B.G., Siegel, D.A., Carder, K.L., Garver, S.A., Kahru, M., McClain, C.R., 1998. Ocean color chlorophyll algorithms for SeaWiFS. J. Geophys. Res. 103, 24937–24953.

Ohman, M.D., Venrick, E.L., 2003. CalCOFI in a changing ocean. Oceanography 16, 76–85.

O'Reilly, J.E., et al., 2000. SeaWiFS Postlaunch Calibration and Validation Analyses, Part 3. NASA Tech. Memo. 2000-206892, Vol. 11. In: Hooker, S.B., Firestone, E.R. (Eds.), NASA Goddard Space Flight Center (49 pp.).

Platt, T., Sathyendranath, S., 1993. Estimators of primary production for interpretation of remotely sensed data on ocean color. J. Geophys. Res. 98, 14,561–14,576.

Platt, T., Sathyendranath, S., Caverhill, M., Lewis, M.R., 1988. Ocean primary production and available light: further algorithms for remote sensing. Deep-Sea Res. A 35, 855–879.

Saba, V.S., Friedrichs, M.A.M., Antoine, D., Armstrong, R.A., Asanuma, I., Behrenfeld, M.J., Ciotti, A.M., Dowell, M., Hoepffner, N., Hyde, K.J.W., Ishizaka, J., Kameda, T., Marra, J., Mélin, F., Morel, A., O'Reilly, J., Scardi, M., Smith Jr., W.O., Smyth, T.J., Tang, S., Uitz, J., Waters, K., Westberry, T.K., 2011. An evaluation of ocean color model estimates of marine primary productivity in coastal and pelagic regions across the globe. Biogeosciences 8, 489–503. http://dx.doi.org/10.5194/bg-8-489-2011.

Siegel, D.A., Maritorena, S., Nelson, N.B., Behrenfeld, M.J., 2005. Independence and interdependencies among global ocean color properties: reassessing the bio-optical assumption. J. Geophys. Res. 110, C07011. http://dx.doi.org/10.1029/2004JC002527.

Sokal, R.R., Rohlf, F.J., 1995. Biometry: the principles and practice of statistics in biological research, 3rd edition. W.H. Freeman, New York.

Sosik, H.M., 1996. Bio-optical modeling of primary production: consequences of variability in quantum yield and specific absorption. Mar. Ecol. Prog. Ser. 143, 225–238.

Werdell, P.J., Bailey, S.W., 2005. An improved in-situ bio-optical data set for ocean color algorithm development and satellite data product validation. Remote Sens. Environ. 98, 122–140. http://dx.doi.org/10.1016/j.rse.2005.07.001.

Yoder, J.A., Yoder, J.A., Antoine, D., del Castillo, C.E., Evans, R.H., Mobley, C., et al., 2011. Assessing requirements for sustained ocean color research and operations. Committee on Assessing Requirements for Sustained Ocean Color Research and Operations, National Research Council. The National Academies Press, (http://www.nap.edu/openbook.php?record_id=13127).