

Assessment of uncertainties of ocean color parameters for the ocean Carbon-based Productivity Model

MA Sheng¹, YANG Xiaofeng, TAO Zui, LI Ziwei, ZHOU Xuan

State Key Laboratory of Remote Sensing Science, Jointly Sponsored by the Institute of Remote Sensing Applications of Chinese Academy of Sciences and Beijing Normal University, Beijing, China

E-mail: masheng1987@gmail.com

Abstract. With the developments of ocean color remote sensing technology, some ocean color parameters can be derived by satellite globally. These terms, including chlorophyll concentration (*Chl*), particulate backscattering coefficients (b_{bp}), photosynthetically available radiation (*PAR*), have been proved to be related to NPP of phytoplankton. Based on these parameters with other auxiliary data, a carbon-based productivity model (CbPM) had been developed. The model derives phytoplankton carbon (*C*) from b_{bp} and utilizes the ratios of *C* and *Chl* to describe the phytoplankton growth rates (μ) which has physiological dependencies on light (through variations in *PAR*), nutrients, and temperature. This paper indicated how the uncertainties in satellite derived parameters (*Chl*, b_{bp} and *PAR*) propagated through the CbPM using Monte Carlo method. Comparisons on the individual contributor to the random uncertainty in NPP between these input items were discussed. The analysis results showed that among the three parameters, the biggest contribution to the uncertainty in the model output came from *Chl*. Therefore, improvements in the accuracy of *Chl* would have the largest potential to improve the ability of CbPM in estimating NPP of phytoplankton.

1. Introduction

Marine net primary production (NPP) is widely used as an indispensable diagnostic of the state and development of ecosystem health and biogeochemical cycles [1, 2]. Traditionally, oceanic NPP has been measured from ships, whose geographic and temporal coverage was restricted by its duration and cost [3]. With the developments in satellite-borne ocean color remote sensing technology, many properties of ocean surface can be derived, such as chlorophyll concentration and particulate backscattering coefficients of water. These properties were employed in mathematical models in combination with other remotely sensed quantities to estimate NPP [4-6]. These models can make for reliable quantification of spatial and temporal variability in phytoplankton productivity [7].

Just as any other models, assessments of satellite-based estimates of NPP are subject to uncertainty. One part of the uncertainty is inherent in the models and indicates the level of success in modeling the physiological process of phytoplankton. A number of studies have assessed this intrinsic uncertainty in NPP models using field measurements data [8, 9]. Another important part of the uncertainty is the one resulting from uncertainties in model input terms. Many studies had documented the sensitivities of

¹ MA sheng, CAS Olympic S&T Park No.20 Da Tun Road, Chao Yang, Beijing (P. O. Box 9718)

E-mail: masheng1987@gmail.com



NPP models to perturbations of a few input terms [3, 7, 10, 11]. All of these studies showed that uncertainties in input variables contributed considerably to NPP model uncertainties. However, information on what level of confidence the uncertainty bounds were related to the model output was not provided by these studies. Milutinović et al. [12] examined how input uncertainties affected NPP estimates from Vertically Generalized Productivity Model (VGPM) of Behrenfeld and Falkowski [5] using a Monte Carlo method. The study provided a wonderful method to quantify the impacts of input uncertainties to model output.

In this study, the input uncertainties through the carbon-based productivity model (CbPM) of Behrenfeld et al. [4] were propagated following the method proposed by Milutinović et al. [12]. CbPM describes the phytoplankton physiological process much more scientifically than VGPM. The uncertainty of three satellite-derived terms (chlorophyll concentrations (Chl), particulate backscattering coefficients at 443 nm ($b_{bp}(443)$) and photosynthetically available radiation (PAR)) was evaluated by performing comparison with reference data sets. Random values were repeatedly drawn from the input uncertainties and introduced into the CbPM using a Monte Carlo method, yielding an output frequency distribution that described the uncertainty in the modeled NPP. At last, the simultaneous impact and individual contributions of the three input uncertainties was quantified, which enabled identification of the greatest introduction in the NPP estimation.

2. Primary productivity model and input data

The Carbon-based Productivity Model (CbPM) [4] calculates column integrated net primary productivity (NPP [$\text{mg C m}^{-2} \text{ day}^{-1}$]) as the product of phytoplankton carbon biomass (C [mg m^{-3}]), growth rate (μ [divisions d^{-1}]), depth of the euphotic layer (Z_{eu} [m]), and light-dependent function ($h(I_0)$, unitless):

$$NPP = C \times \mu \times Z_{eu} \times h(I_0). \quad (1)$$

Phytoplankton carbon biomass C is assessed from particulate backscattering coefficients at 443 nm (b_{bp} , m^{-1}), by subtracting a background value 0.00035 m^{-1} and then multiplying a scalar of 13,000 mg C m^{-2} .

$$C = 13000 \times (b_{bp} - 0.00035) \quad (2)$$

The growth rate of a natural phytoplankton community (μ : divisions d^{-1}) is a function of light, nutrients, and temperature and can be described as follows.

$$\mu = \mu_{\max} \times f(N, T) \times g(I_g). \quad (3)$$

Where, $\mu_{\max}$ is the maximum potential community growth rate under optimal conditions and is assigned a value of 2 divisions d^{-1} . $g(I_g)$ accounts for reductions in growth rate due to light limitation and it is expressed as $g(I_g) = 1 - \exp(-3I_g)$. $f(N, T)$ accounts for reductions in growth rate due to nutrient and temperature limitation at a given light level. The function of $f(N, T)$ is quantified by dividing satellite chlorophyll:carbon data ($Chl:C_{sat}$) by a maximum potential community Chl:C value for a given I_g .

$$f(N, T) = Chl : C_{sat} / [0.022 - (0.045 - 0.022)\exp(-3I_g)] \quad (4)$$

Where, I_g is the mixed layer light levels. Cloud-corrected surface photosynthetically available radiation (PAR: Einsteins $\text{d}^{-1} \text{ m}^{-2}$), daylength (DL: h), mean regional mixed layer depths (MLD: m) and mixed layer light attenuation coefficients at 490 nm (k_{490} : m^{-1}) data are used to calculate I_g following:

$$I_g = (PAR / DL) \times \exp(-k_{490} \times MLD / 2) \quad (5)$$

Following Morel and Bethon's model, depth of the euphotic layer Z_{eu} is calculated by chlorophyll concentrations (Chl : mg/L). First, total chlorophyll content within the productive water column (Chl_{tot}) was determined as

$$\begin{cases} Chl_{tot} = 38.0 \times Chl^{0.425}, & \text{if } Chl < 1.0 \text{ mg} \cdot \text{m}^{-3} \\ Chl_{tot} = 40.2 \times Chl^{0.507}, & \text{if } Chl \geq 1.0 \text{ mg} \cdot \text{m}^{-3} \end{cases} \quad (6)$$

Secondly, Z_{eu} was calculated from Chl_{tot} following

$$Z_{eu} = 200.0 \times Chl_{tot}^{-0.293}. \quad (7)$$

At last, the light-dependent function was described as

$$h(I_0) = 0.66125 \times PAR / (PAR + 4.1). \quad (8)$$

In this study, all input quantities to the CbPM, except daylength and mixed layer depths, were produced by, or related to, satellite remote sensing. Four types of coincident global monthly Level-3 binned data products of Moderate Resolution Imaging Spectroradiometer (MODIS) from January 2002 to December 2011 were used. These data were provided by the NASA Ocean Biology Processing Group (<http://oceancolor.gsfc.nasa.gov>), namely Chl, PAR, bbp, k490. MLD data in 2010 was generated by the Hybrid Coordinate Ocean Model (HYCOM) of National Energy Research Scientific Computing Center (NERSC). All the data had the resolution of equal-area grid bin size of 18km×18km. The missing data were not corrected to avoid introducing additional biases to the input field.

3. Evaluation of uncertainties in input quantities

Although the CbPM model is not perfect, for the purpose of this study the model itself was assumed to be flawless, so that uncertainties in the model input were the only source of uncertainty in the output. Following Svetlana Milutinović's approach [12], in this study, for the purpose of quantifying uncertainty in a given input term (Chl , b_{bp} and PAR), three types of discrepancies (δ , δ^{REL} and δ^{LOG}) and their frequency distribution between model results (MOD) and the simultaneous co-located in situ observations (served as reference, REF) was determined. The mean of δ represented the bias (B , B^{REL} and B^{LOG}), while its centered-pattern (or zero-centered) root mean square difference ($RMSD_0$, $RMSD_0^{REL}$ and $RMSD_0^{LOG}$) served as a measure of the remaining uncertainty component. The detailed explanation of this methodological approach with accompanying notation and equations can be found in the work of Svetlana Milutinović [12].

To estimate uncertainty of the three input terms with Svetlana Milutinović's approach, the field data from 2002 to 2011 was acquired respectively from the website of SeaWiFS Bio-optical Archive and Storage System (SeaBASS). After matching these field data and MODIS modeled daily data using GPS coordinate, many data location was obtained, which was shown in Figure 1.

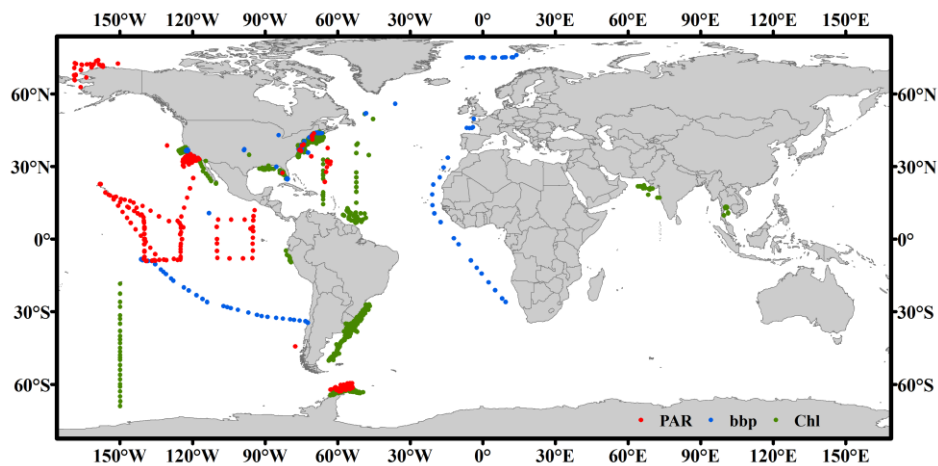


Figure 1. Locations of the in situ data used to determine the respective predictive uncertainties of Chl (green dots), b_{bp} (blue dots) and PAR (red dots).

An approach similar to Mirosław Darecki et. al. [13] was used to evaluate the uncertainty of a standard MODIS Chl product in this study. As Chl values tend to be lognormally distributed [14], normal probability density function is a suitable model for frequency distribution of δ^{LOG} . Therefore,

using the 1639 matched data (between MOD and REF), the δ^{LOG} , B^{LOG} , $\text{RMSD}_0^{\text{LOG}}$ was calculated following the equation (9). The scatterplots and log errors frequency distribution of satellite derived *Chl* data versus in situ data were shown in figure 2. The comparison showed that the *Chl* uncertainty statistics in $\log_{10}$ scale is $B^{\text{LOG}}(\text{Chl})$ of 0.086 and $\text{RMSD}_0^{\text{LOG}}(\text{Chl})$ of 0.248. The uncertainty is similar to that of Gregg et al. [15], which is 0.077 and 0.237.

$$\begin{aligned}\delta^{\text{LOG}} &= \log_{10}(\text{MOD}) - \log_{10}(\text{REF}) \\ B^{\text{LOG}} &= \overline{\delta^{\text{LOG}}} = n^{-1} \sum_{i=1}^n [\log_{10}(\text{MOD}_i) - \log_{10}(\text{REF}_i)] \\ \text{RMSD}_0^{\text{LOG}} &= ((n-1)^{-1} \sum_{i=1}^n (\delta_i^{\text{LOG}} - B^{\text{LOG}})^2)^{-2}\end{aligned}\quad (9)$$

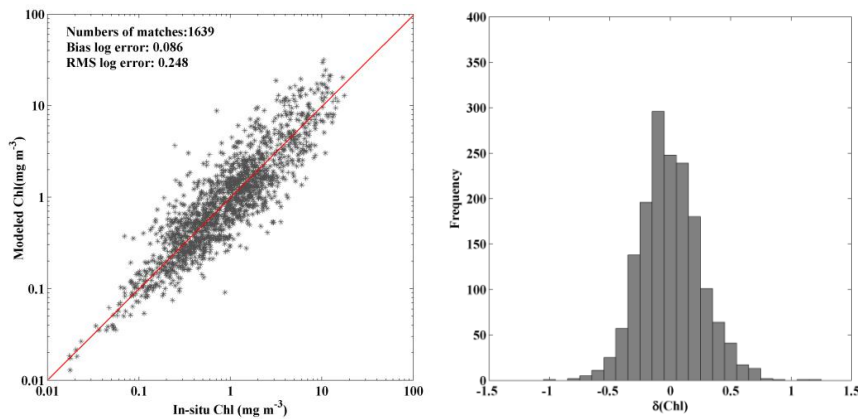


Figure 2. Scatterplots and log errors frequency distribution of satellite derived *Chl* data versus in situ data. The frequency was normally distributed.

The MODIS b_{bp} products at 443nm was estimated with the Garver-Siegel-Maritorena (GSM) semi-analytical algorithm[16-18]. Using the 558 in situ derived values as reference data, the performance of the modeled $b_{bp}(443)$ was evaluated. After three types of discrepancies (δ , δ^{REL} and δ^{LOG}) were calculated, only the frequency distribution of the δ^{LOG} could be modeled as normal probability density function. The scatterplots and log errors frequency distribution of satellite derived $b_{bp}(443)$ data versus in situ data were shown in figure 3. The comparison showed that the $b_{bp}(443)$ uncertainty statistics in $\log_{10}$ scale is $B^{\text{LOG}}(b_{bp})$ of 0.046 and $\text{RMSD}_0^{\text{LOG}}(b_{bp})$ of 0.255, which is consistent with the work of Maritorena et al. [16].

To estimate the uncertainty of MODIS derived PAR data, 342 in situ measurements of PAR was applied. After three types of discrepancies were calculated, only the frequency distribution of the δ could be modeled as normal probability density function. The δ , B , RMSD_0 was calculated following the equation (10). The scatterplots and errors frequency distribution of satellite derived *PAR* data versus in situ data were shown in figure 4. The comparison showed that the *PAR* uncertainty statistics is $B(\text{PAR})$ of 2.14 E m⁻² day⁻¹ and $\text{RMSD}_0(\text{PAR})$ of 5.27 E m⁻² day⁻¹, which is consistent with the work of Frouin et al.[19].

$$\begin{aligned}\delta &= \text{MOD} - \text{REF} \\ B &= \bar{\delta} = n^{-1} \sum_{i=1}^n (\text{MOD}_i - \text{REF}_i) \\ \text{RMSD}_0 &= ((n-1)^{-1} \sum_{i=1}^n (\delta_i - B)^2)^{-2}\end{aligned}\quad (10)$$

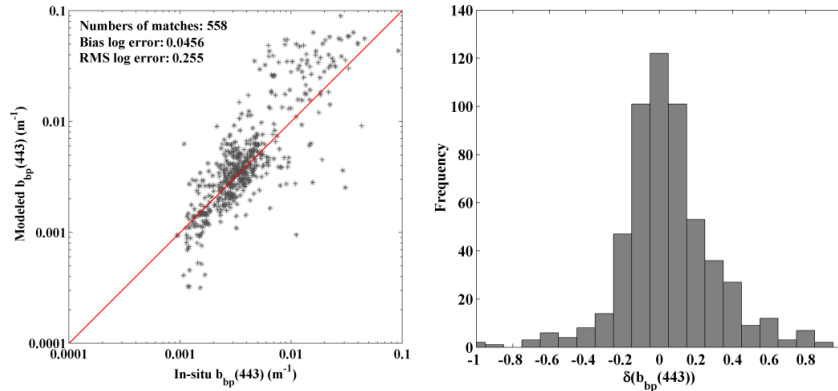


Figure 3. Scatterplots and log errors frequency distribution of satellite derived $bbp(443)$ data versus in situ data. The frequency was normally distributed.

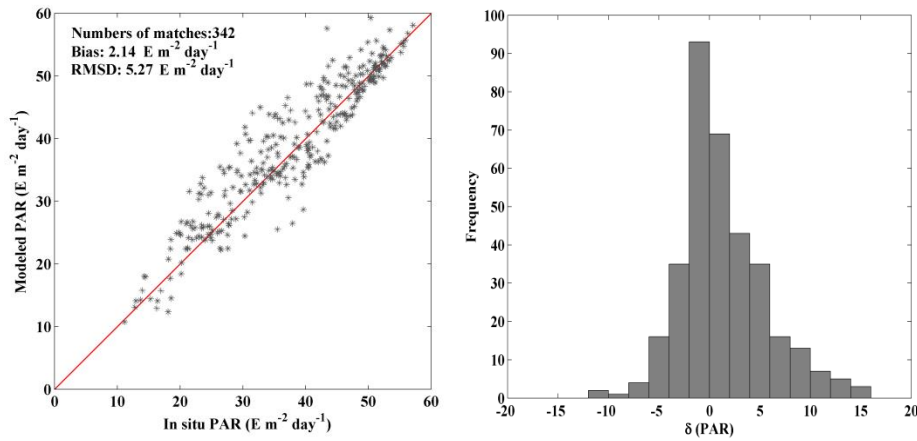


Figure 4. Scatterplots and errors frequency distribution of satellite derived PAR data versus in situ data. The frequency was normally distributed.

4. Propagation of uncertainties through the CbPM

Monte Carlo method is applied in this study to validly propagate uncertainties through the CbPM model. To achieve a compromise between sufficient quality and required computer processing time, 1500 random sampling was produced from uncertainty distributions previously assigned to the input quantities for every pixel. In any given bin (i.e. grid cell), bias was subtracted from the nominal value of each input quantity and the random samplings was produced following equation (11).

$$\begin{aligned} \log_{10}(\mathbf{Chl}) &\sim N[\log_{10}(Chl) - B^{\text{LOG}}(Chl), \text{RMSD}_0^{\text{LOG}}(Chl)] \\ \log_{10}(\mathbf{b}_{bp}) &\sim N[\log_{10}(b_{bp}) - B^{\text{LOG}}(b_{bp}), \text{RMSD}_0^{\text{LOG}}(b_{bp})] \\ \mathbf{PAR} &\sim N[PAR - B(PAR), \text{RMSD}_0(PAR)] \end{aligned} \quad (10)$$

Where random values are indicated in bold type, nominal values in light type and N symbolizes a particular normal probability distribution, whose mean and standard deviation are given in square brackets. The summary of uncertainties in input terms was shown in table 1. For each individual set of randomly generated input values, NPP was computed by the CbPM, resulting in an ensemble of up to 1500 NPP values in every bin. The propagation of these uncertainties through the CbPM resulted in a frequency distribution that represented uncertainty in the model NPP estimate in any given bin in

January and July of 2010. The properties of distribution in each bin were summarized using the mean and the coefficient of variation (CV).

Table 1. Summary of uncertainties in input quantities that were propagated through the CbPM.

Input quantity	Scale	Component of input uncertainty	
		Bias	Zero-centered RMSD
<i>Chl</i>	Logarithmic	0.086	0.248
$b_{bp}(443)$	Logarithmic	0.046	0.255
<i>PAR</i>	Linear	2.14	5.27

The mean values for uncertainty distributions of NPP ($\overline{NPP_{MC}}$, where MC signifies the Monte Carlo method) was generated. Differences between nominal NPP values (presented in figure 5a and b respectively) and $\overline{NPP_{MC}}$ (Δ , $\text{mg C m}^{-2} \text{ day}^{-1}$) were mapped in figure 5c and d for January and July 2010. Figure 5c and d suggests that, overall, NPP is overestimated and the magnitude of this generally positive bias in NPP is proportional to NPP itself. In order to complement the information contained in figures 5a-d, the value of Δ was divided by the related values of $\overline{NPP_{MC}}$. Both in January and July, the bulk of percentage bias values spanned between -2% and $+18\%$ and the mean of which is 13% and 15% .

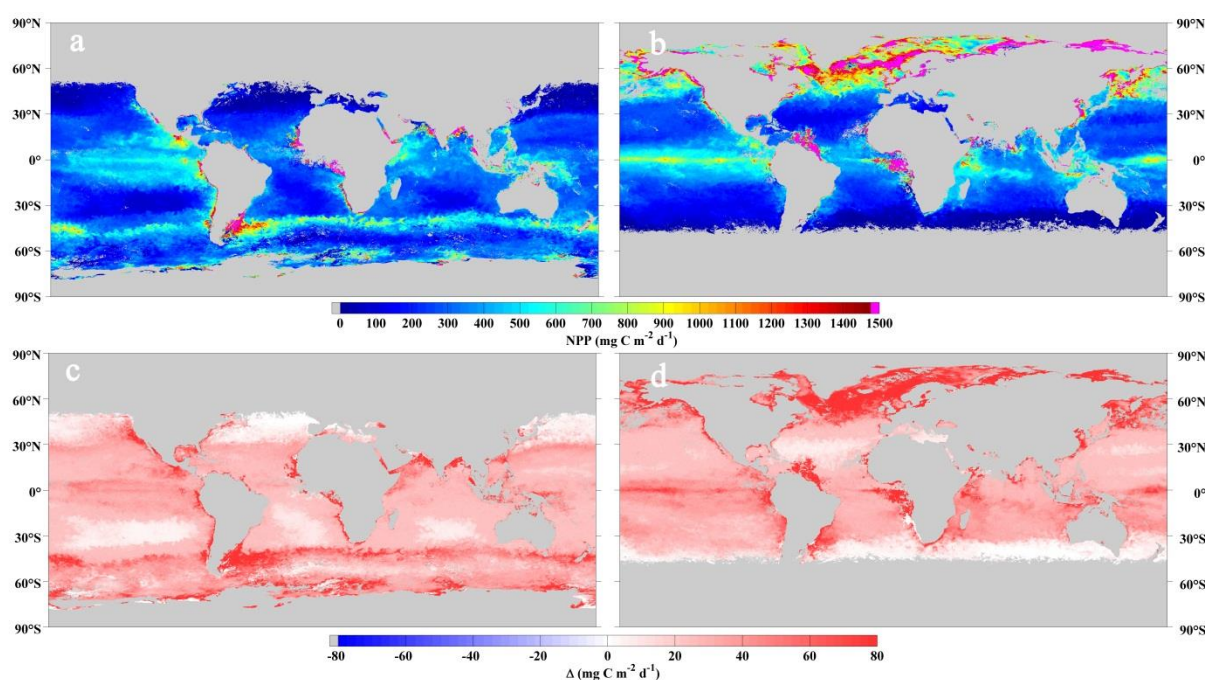


Figure 5. Top: Net primary productivity (NPP [$\text{mg C m}^{-2} \text{ day}^{-1}$]) estimated using the Carbon-based Productivity Model (CbPM) for (a) January and (b) July 2010. Bottom: Difference between nominal NPP values and the mean values of corresponding NPP uncertainty distributions (Δ [$\text{mg C m}^{-2} \text{ day}^{-1}$]) for (c) January and (d) July 2010. Gray color represents locations with no available remote sensing observations, continental shelf (<200 m) and bins with unreliable statistics.

To assess the respective individual contributions of *Chl*, b_{bp} and *PAR* to the overall uncertainty in NPP estimates, the Monte Carlo method was used in a manner that allowed the input quantities to vary randomly one at a time, following the predetermined probability density functions. This way, for a given bin, a frequency distribution that illustrated the component of the overall uncertainty in NPP

stemming from the uncertainty in that input quantity only was derived. This frequency distribution was described by the mean percentage bias and CV in NPP estimates that was solely due to the bias in the particular input quantity. Table 2 lists the averages of NPP uncertainty components (partial percentage bias and partial CV) originating from uncertainties in *Chl*, *b_{bp}* and *PAR* for January and July 2010. From the results in table 2, a conclusion can be draw that the uncertainty of *Chl* contributes the greatest part to the CbPM output uncertainty.

Table 2. Overall and individual contributions of uncertainties in input quantities to the uncertainty in NPP.

Input quantity	Component of NPP uncertainty			
	percentage bias		coefficient of variation	
	January 2010	July 2010	January 2010	July 2010
All	11%	13%	107%	114%
<i>Chl</i>	7%	8%	68%	74%
<i>b_{bp}</i> (443)	1%	2%	31%	42%
<i>PAR</i>	3%	3%	26%	23%

5. Conclusion

The uncertainty of input terms (*Chl*, *b_{bp}*, *PAR*) propagation through the CbPM in one arbitrarily chosen calendar year was analysed in this paper using Monte Carlo method. The analysis showed that an output value of the CbPM was typically overestimated by 12%, whereas the representative measure of its random uncertainty was 110%. The individual contribution of *Chl*, *b_{bp}* and *PAR* to the overall uncertainty in NPP estimates was also assessed respectively. The results showed that the biggest contribution to the systematic uncertainty in the model output came from the *Chl*. Therefore, improvements in the accuracy of this term would have the largest potential to decrease the input-related uncertainty in the model NPP estimates.

References

- [1] Behrenfeld M J, Randerson J T, McClain C R, Feldman G C, Los S O, Tucker C J, Falkowski P G, Field C B, Frouin R, Esaias W E, Kolber D D and Pollack N H 2001 Biospheric primary production during an ENSO transition *Science* **291** 2594-7
- [2] Falkowski P G, Barber R T and Smetacek V 1998 Biogeochemical controls and feedbacks on ocean primary production *Science* **281** 200-6
- [3] 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 Quere C, Lohrenz S, Marra J, Melin F, Moore K, Morel A, Reddy T E, Ryan J, Scardi M, Smyth T, Turpie K, Tilstone G, Waters K and Yamanaka Y 2006 A comparison of global estimates of marine primary production from ocean color *Deep-Sea Res Pt II* **53** 741-70
- [4] Behrenfeld M J, Boss E, Siegel D A and Shea D M 2005 Carbon-based ocean productivity and phytoplankton physiology from space *Global Biogeochem. Cycles* **19**
- [5] Behrenfeld M J and Falkowski P G 1997 Photosynthetic rates derived from satellite-based chlorophyll concentration *Limnol. Oceanogr.* **42** 1-20
- [6] Westberry T, Behrenfeld M J, Siegel D A and Boss E 2008 Carbon-based primary productivity modeling with vertically resolved photoacclimation *Global Biogeochem. Cycles* **22**
- [7] 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, Melin 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 and Winguth A 2009 Assessing the uncertainties of model estimates of primary productivity in the tropical Pacific Ocean *J. Mar. Sys.* **76** 113-33
- [8] 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 and Yoder J 2002 Comparison of algorithms for estimating ocean primary production from surface chlorophyll, temperature, and irradiance *Global Biogeochem. Cycles* **16**
- [9] Saba V S, Friedrichs M A M, Carr M E, Antoine D, Armstrong R A, Asanuma I, Aumont O, Bates N R, Behrenfeld M J, Bennington V, Bopp L, Bruggeman J, Buitenhuis E T, Church M J, Ciotti A M, Doney S C, Dowell M, Dunne J, Dutkiewicz S, Gregg W, Hoepffner N, Hyde K J W, Ishizaka J, Kameda T, Karl D M, Lima I, Lomas M W, Marra J, McKinley G A, Melin F, Moore J K, Morel A, O'Reilly J, Salihoglu B, Scardi M, Smyth T J, Tang S L, Tjiputra J, Uitz J, Vichi M, Waters K, Westberry T K and Yool A 2010 Challenges of modeling depth-integrated marine primary productivity over multiple decades: A case study at BATS and HOT *Global Biogeochem. Cycles* **24**
- [10] Milutinovic S, Behrenfeld M J, Johannessen J A and Johannessen T 2009 Sensitivity of remote sensing-derived phytoplankton productivity to mixed layer depth: Lessons from the carbon-based productivity model *Global Biogeochem. Cycles* **23**
- [11] 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, Melin F, Morel A, O'Reilly J, Scardi M, Smith W O, Smyth T J, Tang S, Uitz J, Waters K and 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
- [12] Milutinovic S and Bertino L 2011 Assessment and propagation of uncertainties in input terms through an ocean-color-based model of primary productivity *Remote Sens. Environ.* **115** 1906-17
- [13] Darecki M and Stramski D 2004 An evaluation of MODIS and SeaWiFS bio-optical algorithms in the Baltic Sea *Remote Sens. Environ.* **89** 326-50
- [14] Campbell J W 1995 The Lognormal-Distribution as a Model for Biooptical Variability in the Sea *J Geophys Res-Oceans* **100** 13237-54
- [15] Gregg W W and 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-Oceans* **106** 2499-515
- [16] Frouin Maritorea S, Siegel D A and Peterson A R 2002 Optimization of a semianalytical ocean color model for global-scale applications *Appl. Opt.* **41** 2705-14
- [17] Garver S A and Siegel D A 1997 Inherent optical property inversion of ocean color spectra and its biogeochemical interpretation .1. Time series from the Sargasso Sea *J Geophys Res-Oceans* **102** 18607-25
- [18] 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, Maritorea S, Garver S A, Brody E A, Ubante J and Hammer M A 2001 Bio-optical modeling of primary production on regional scales: the Bermuda BioOptics project *Deep-Sea Res Pt II* **48** 1865-96
- [19] Frouin R and Murakami H 2007 Estimating photosynthetically available radiation at the ocean surface from ADEOS-II Global Imager data *J. Oceanogr.* **63** 493-503