Biogeosciences

Using satellite-derived backscattering coefficients in addition to chlorophyll data to constrain a simple marine biogeochemical model

H. Kettle

The School of GeoSciences, University of Edinburgh, Edinburgh, UK

Now at: Biomathematics and Statistics Scotland, James Clerk Maxwell Building, The Kings Buildings, Edinburgh, UK

Abstract. Biogeochemical models of the ocean carbon cycle are frequently validated by, or tuned to, satellite chlorophyll data. However, ocean carbon cycle models are required to accurately model the movement of carbon, not chlorophyll, and due to the high variability of the carbon to chlorophyll ratio in phytoplankton, chlorophyll is not a robust proxy for carbon. Using inherent optical property (IOP) inversion algorithms it is now possible to also derive the amount of light backscattered by the upper ocean (b_b) which is related to the amount of particulate organic carbon (POC) present. Using empirical relationships between POC and b_b, a 1-D marine biogeochemical model is used to simulate b_b at 490 nm thereby allowing the model to be compared with both remotely-sensed chlorophyll or b_b data. Here I investigate the possibility of using b_b in conjunction with chlorophyll data to help constrain the parameters in a simple 1-D NPZD model. The parameters of the biogeochemical model are tuned with a genetic algorithm, so that the model is fitted to either chlorophyll data or to both chlorophyll and b_b data at three sites in the Atlantic with very different characteristics. Several inherent optical property (IOP) algorithms are available for estimating b_b, three of which are used here. The effect of the different b_b datasets on the behaviour of the tuned model is examined to ascertain whether the uncertainty in b_b is significant. The results show that the addition of b_b data does not consistently alter the same model parameters at each site and in fact can lead to some parameters becoming less well constrained, implying there is still much work to be done on the mechanisms relating chlorophyll to POC and b_b within the model. However, this study does indicate that including b_b data has the potential to significantly effect the modelled mixed layer detritus and that uncertainties in b_b due to the different IOP algorithms are not particularly significant.

Correspondence to: H. Kettle
(helen@bioss.ac.uk)

1 Introduction

Quantifying the global carbon cycle is crucial for predicting our future climate. The oceans play an important role in the carbon cycle as they absorb CO_2 from the atmosphere enabling the transport of carbon to the deep ocean through physical and biological processes. Physical processes enable CO_2 rich waters from the ocean surface to sink downwards possibly resurfacing up to hundreds of years later. However, biological processes can also cause particulate organic carbon (POC) to sink below the thermocline (a process termed as "export production"). Thus for climate change prediction it is crucial to quantify the amount of CO_2 that is transferred from the atmosphere to the ocean through the air-sea interface (the air-sea CO_2 flux) and the amount of carbon that is subsequently exported to the deep ocean. Ideally, validation of ocean carbon cycle models would involve comparison of the simulated air-sea CO_2 flux and export production with measured data. Unfortunately these data are not available at the time and space scales necessary. However, chlorophyll concentrations inferred from satellite ocean colour data (Chl) are available globally at an adequate time-space resolution. Chl gives an indication of the amount of living phytoplankton in the ocean. This is useful since it is algal photosynthesis that removes CO_2 from the water allowing more CO_2 from the atmosphere to enter the ocean. Ideally, to model the carbon cycle, we need to know the amount and rate of carbon fixed by phytoplankton but these are not directly related to the amount and rate of change of chlorophyll since the carbon to chlorophyll ratio (C:Chl) within phytoplankton is highly dynamic (Geider et al., 1997). Therefore, chlorophyll is not a robust proxy for carbon unless there are also data on phytoplankton physiology (C:Chl). Thus it is perfectly possible to correctly predict chlorophyll concentrations without correctly predicting carbon concentrations.

Satellite ocean colour data are simply measurements of the amount of sunlight (at certain wavelengths) that is scattered

back out of the ocean (known as the water leaving radiance). Water leaving radiance can be used to estimate the inherent optical properties (IOPs) of the ocean surface waters – these include the absorption coefficient, a (m^{-1}), and the backscattering coefficient, b_b (m^{-1}). The total backscattering coeffient is the sum of the particulate backscattering coefficient (b_{bp}) and the backscattering due to seawater (b_{bw}). There are a number of algorithms available to calculate b_b or b_{bp} from the water leaving radiance, e.g., Loisel and Poteau (2006), Smyth et al. (2006), the QAA (Lee et al., 2002), and the GSM (Garver and Siegel, 1997; Maritorena et al., 2002; Siegel et al., 2002). Furthermore, b_{bp} derived from satellite data has been empirically related to the amount of POC in the surface waters (e.g. Loisel et al., 2001, 2002; Stramski et al., 1999). POC is a combination of the amount of carbon contained within living phytoplankton (C_P) and biological detritus (C_D). Other possible products from ocean colour data are the IOP a (absorption coefficient) and the apparent optical property K_d (attenuation coefficient for downwelling light). However, in the open ocean variations in a and K_d will be strongly related to Chl so it is unlikely that there is significantly different information here to constrain the models much beyond simply using Chl.

In this study, the potential to use satellite b_b data in addition to satellite Chl data to calibrate a simple 1-D open ocean biogeochemical model is investigated. The model consists of the Hadley Centre Ocean Carbon Cycle model (HadOCC; Palmer and Totterdell, 2001) coupled to the 1-D General Ocean Turbulence Model (GOTM, Burchard et al., 1999). HadOCC is a simple nutrient-phytoplankton-zooplankton-detritus (NPZD) plus carbonate chemistry model that is used in climate prediction models such as HadCM3. The model is applied to three sites in the Atlantic ocean that have very different physical and biological characteristics. The model parameters are tuned, independently for each station, to co-incident satellite Chl and b_b data using a genetic algorithm to search the parameter space. The hypothesis is that by using satellite b_b to validate the POC in global 3-D climate prediction models, the detrital and phytoplankton components of the model may be better constrained than when simply using Chl. Since there are several IOP algorithms available for estimating b_b the model is tuned to each of these different datasets to assess whether the uncertainty in the b_b data invalidates the hypothesis.

2 Sites

Three sites in the Atlantic are chosen due to the availability of observed data at these locations (see http://www.noc.soton.ac.uk/animate). They are: the Central Irminger Sea (CIS) at 60° N, 40° W, the Estacion Europea de Series Temporales del Oceano, Islas Canarias (ESTOC) at 29° N, 15.5° W and the Porcupine Abyssal Plain (PAP) at 49° N, 16° W. These sites have diverse characteristics, as demonstrated by SeaWiFS satellite measurements of their photosynthetically active ra-
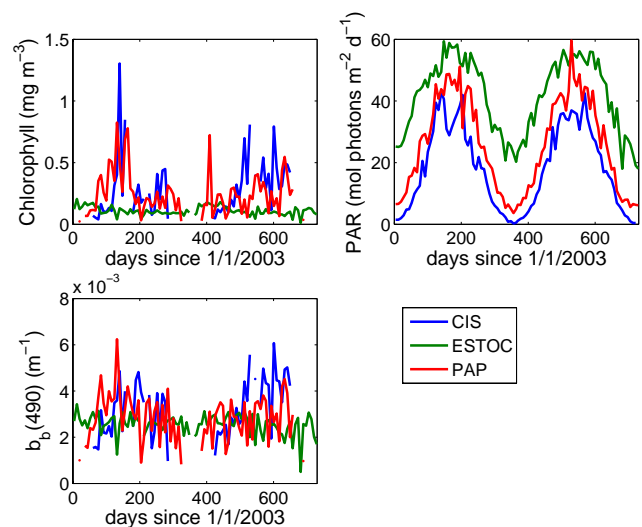

Fig. 1. Comparing PAR (SeaWiFS), chlorophyll (SeaWiFS) and $b_b(490)$ (LP algorithm) at the three different sites in the Atlantic.

diation (PAR), chlorophyll concentrations and backscattering coefficients (Fig. 1) thereby providing a range of conditions over which to test the hypothesis.

3 Satellite data

The ocean colour data were downloaded from ftp://oceans.gsfc.nasa.gov/SeaWiFS/Binned/8Day/. These datafiles contain binned normalised water leaving radiances, nLw (at wavelengths 412, 443, 490, 510, 555 and 670 nm and chlorophyll-a data (from the OC4v4 algorithm), averaged over 8 days intervals. The remote sensing reflectance (R_{rs}) is computed from nLw using

$$R_{rs}(\lambda) = \frac{nLw(\lambda)}{F0} \tag{1}$$

where $F0(\lambda)$ are 173.00, 190.15, 196.47, 188.16, 183.01, 151.14.

4 Methods

4.1 Deriving the particulate backscattering coefficient, $b_{bp}(490)$

b_{bp} can be derived from nLw or R_{rs} using a variety of IOP inversion algorithms. Here b_{bp} at 490 nm (required for relating to POC) is obtained using the following four IOP algorithms:

LP (Loisel and Poteau, 2006; Loisel and Stramski, 2000; Loisel et al., 2001.) $b_b(490)$ obtained from this algorithm is converted to $b_{bp}(490)$ using

$$b_{bp}(490) = b_b(490) - b_{bw}(490) \tag{2}$$

where $b_{bw}(490){=}0.001581378\,\mathrm{m}^{-1}$ is the backscattering coefficient due to water.

GSM (The Garver-Siegel-Maritorena semi-analytical algorithm; Garver and Siegel, 1997; Maritorena et al., 2002; Siegel et al., 2002.) This algorithm gives $b_{bp}(443)$ which is converted to $b_{bp}(490)$ using

$$b_{bp}(490) = b_{bp}(443) \left(\frac{490}{443}\right)^{-\gamma} \tag{3}$$

where

$$\gamma = -0.855 * \log(\mathrm{Chl}) + 1.259 \tag{4}$$

for Chl in $\mathrm{mg\,m}^{-3}$ (Loisel et al., 2006).

PML (Plymouth Marine Laboratory; Smyth et al., 2006.) This is a semi-analytical model using empirically derived spectral slopes between neighboring wavebands in combination with a radiative transfer model. This gives $b_{bp}(490)$ which is used directly.

QAA (The Quasi-Analytical Algorithm, Lee et al., 2002.) This gives $b_{bp}(490)$ which is used directly.

Where the solar zenith angle is a necessary input (e.g., for LP and PML) it is set to $0°$. The code for computing the LP and QAA algorithms was obtained from http://www.ioccg.org/groups/software.html and the algorithms were driven by eight-day averages of nLw or R_{rs} (Eq. 1) from SeaWiFS. The results of the GSM (v4) algorithm for $b_{bp}(443)$ (using SeaWiFS 5.2) were downloaded from http://www.science.oregonstate.edu/ocean.productivity/inputBbpGsmData.php. And finally, the PML algorithm results were kindly provided by Tim Smyth.

The different IOP algorithms can give quite different values of $b_{bp}(490)$ as shown in Figs. 2 and 3. In Fig. 3 it is clear that the QAA and LP results agree very well (Fig. 3c) but there is considerably less agreement between the other algorithms, with the largest discrepancies between PML and GSM (Fig. 3e). However, the intention here is not to compare these algorithms but rather to investigate whether the b_b data produced by the different algorithms are similar enough to each other to constrain the ocean biogeochemical model in a consistent way.

4.2 Biogeochemical model

The biogeochemical model used here is the NPZD Hadley Centre Ocean Carbon Cycle model (HadOCC; Palmer and Totterdell, 2001) coupled to the 1-D General Ocean Turbulence Model (GOTM, Burchard et al., 1999) (henceforth denoted GOTM-HadOCC). For further details and model equations please refer to Kettle and Merchant (2008). The difference in the model used here is the calculation of C:Chl, for which the empirical equation of Cloern et al. (1995) is used
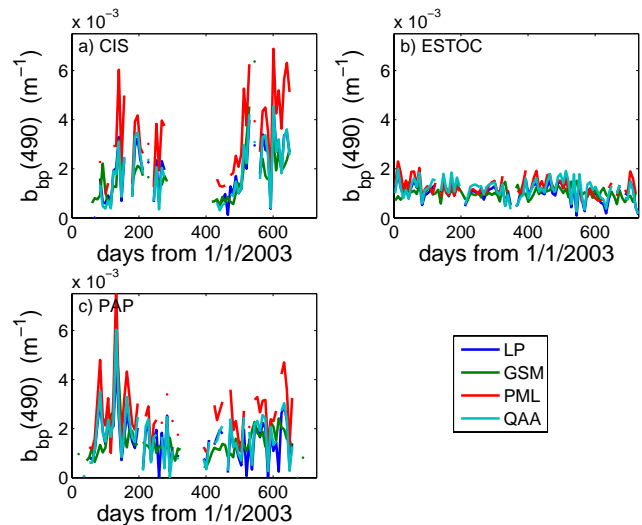

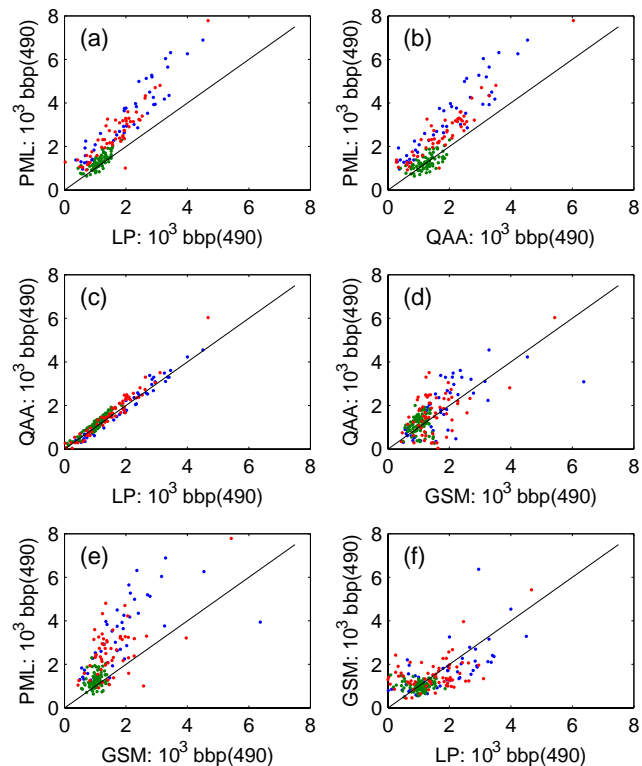**Fig. 2.** Satellite-derived $b_{bp}(490)$ at each site using 4 different IOP inversion algorithms.

Fig. 3. Comparing IOP inversion algorithms for $b_{bp}(490)$ (m^{-1}): **(a)** Loisel and PML, **(b)** QAA and PML, **(c)** Loisel and QAA, **(d)** GSM and QAA, **(e)** GSM and PML, **(f)** Loisel and GSM for CIS (blue), ESTOC (green) and PAP (red).

rather than that of Geider et al. (1997). The Cloern equation was used because it uses both physical and biological properties (further details are given in Sect. 4.2) and so is not as highly dependent on biological parameter values as the Geider equation. In this application, underwater light

and photosynthesis are modelled using 6 wavebands, with the downwelling attenuation coefficient modelled according to Morel and Maritorena (2001) and the absorption of light by phytoplankton from Bricaud et al. (1998) (see Kettle and Merchant, 2008 for more detail). The remineralisation of detritus (D) back to nutrient (not described by Kettle and Merchant, 2008) is computed as follows: below 100 m the amount of detritus remineralised is $\frac{\gamma}{\text{depth}}D$ (see Table 2 for γ); above 100 m is it fixed at $0.1D$ (mmol N m^{-3} d^{-1}). The model is driven with ERA-40 reanalysis meteorological data from ECMWF at 6 hourly resolution and the physical model (GOTM) is tuned to reproduce the observed temperatures in the ANIMATE dataset (see Kettle and Merchant, 2008). At CIS the shortage of satellite data caused by its high latitude means it was necessary to tune the model to 2 years of data (2003–2004), but at ESTOC and PAP the model is only tuned to 2003 data due to computation time constraints.

4.2.1 Model set-up

The initial profiles of inorganic nutrient are taken from the Levitus climatology (Levitus et al., 1993) and concentrations of phytoplankton and zooplankton are both initialised at 0.01 mmol N m^{-3} and detritus is started at zero. Since for parameter optimisation the model must be runs thousands of times, the vertical grid is reduced to 25 layers with geometrical zooming towards the surface so that the surface layer has thickness 5 cm, there are 10 levels above 10 m and 16 levels above 100 m, eventually going to thicknesses of 900 m at abyssal depths. The vertical grid is constant in time and is the same at each location. GOTM-HadOCC requires some "spin-up" time to remove sensitivity to the initial conditions so the model is run for 12 months before the results are compared with the satellite data. The driving data for the spin-up year is simply a copy of the 2003 data.

4.2.2 Computing Chl within GOTM-HadOCC

Chlorophyll concentration (mg m^{-3}) is computed using

$$\text{Chl} = \frac{M_C(\text{C}:\text{N})_P}{C:Chl}\text{Phyto} \tag{5}$$

where Phyto is the concentration of nitrogen in the phytoplankton compartment (mmol N m^{-3}), M_C is the molar mass of carbon (12.01 g mol^{-1}), and $(\text{C}:\text{N})_P$ is the molar C:N ratio in phytoplankton (6.625 mol C (mol N)$^{-1}$). The C:Chl (mass) ratio, is computed using physical variables and the nutrient limitation on growth (N_{lim}) according to Cloern et al. (1995):

$$\text{C}:\text{Chl} = \left(0.003 + 0.0154e^{(0.05T_M - 0.059(I_M))}N_{\text{lim}}\right)^{-1} \tag{6}$$

where T_M is the temperature (°C) of the mixed layer and I_M is the mean daily irradiance in the mixed layer (mol photons m^{-2} d^{-1}) (the expression is inverted as Cloern

et al. (1995) use Chl:C). Below the mixed layer C:Chl is set to 40. The nutrient limitation on growth is given by:

$$N_{\text{lim}} = \frac{N}{K_{\text{nit}} + N} \tag{7}$$

where N is inorganic nutrient (nitrogen) and K_N is the half saturation parameter (see Table 2).

4.2.3 Computing $b_{bp}(490)$ within GOTM-HadOCC

The scattering of underwater light is affected by the amount of POC in the water. In GOTM-HadOCC, POC (mg m^{-3}) is calculated by adding together the carbon in the phytoplankton and detrital compartments:

$$\text{POC} = M_C((\text{C}:\text{N})_P\text{Phyto} + (\text{C}:\text{N})_D D) \tag{8}$$

where D is the nitrogen concentration (mmol N m^{-3}) in the detritus compartment and $(\text{C}:\text{N})_D$ is the C:N ratio in detritus (set to 7.5). Note here that POC is assumed not to include zooplankton however, it is possible that small zooplankton ($< \sim 0.5$ cm) may be included in field measurements of POC. Since the zooplankton compartment in GOTM-HadOCC contains all sizes of zooplankton, it was omitted here but future work could be to look at the difference in the results if some fraction of the zooplankton compartment was also included as POC.

POC is then used to simulate $b_{bp}(490)$ through the following series of equations.

The scattering coefficient due to POC, $b_{\text{POC}}(490)$ (m^{-1}), is estimated from POC using a simple linear relationship:

$$b_{\text{POC}}(490) = \frac{\text{POC}}{400} \tag{9}$$

(Loisel et al., 2002; Claustre et al., 1999). This is converted to the backscattering coefficient due to POC by

$$b_{b\text{POC}}(490) = 0.0096b_{\text{POC}}(490)\text{Chl}^{-0.253} \tag{10}$$

for Chl in mg m^{-3} (Twardowski et al., 2001). The total particulate backscattering coefficient is then assumed to be the sum of $b_{b\text{POC}}$ and a background value ($b_{b\text{BG}}$) which is unrelated to biological activity and set to 0.17×10^{-3} m^{-1} (see Fujii et al., 2007 for more details), such that:

$$b_{bp}(490) = b_{b\text{POC}}(490) + b_{b\text{BG}}. \tag{11}$$

Thus the minimum value for $b_{bp}(490)$ predicted by GOTM-HadOCC is 0.17×10^{-3} m^{-1} which is just slightly higher than the minimums shown by the satellite data for the three sites (Fig. 2).

4.2.4 Optical weighting

In order to compare depth-resolved model variables with the equivalent remote sensing variable, the vertical profile must be weighted according to its contribution to the light leaving the water. Since the light must travel down through the water

and back out again, the diminishing factor is $\exp(-2\,K_d(z))$ (assuming the attenuation of upwelling and downwelling irradiance is the same) where K_d is the attenuation coefficient for downwelling light. Thus for a variable X its remotely sensed equivalent, X_{rs} is computed here using:

$$X_{rs} = \frac{\int_0^{z_{90}} X(z)g(z)dz}{\int_0^{z_{90}} g(z)dz} \qquad (12)$$

where

$$g(z) = \exp\left(-2\int_0^{z_{90}} K_d(z')dz'\right)$$

(Gordon and Clark, 1980), where K_d is spectrally averaged and z_{90} is the first attenuation depth (the depth above which 90% of the light received by the satellite originates – this varies with changing Chl concentration).

4.3 Tuning model parameters

Eleven of the model parameters are tuned, based on a recent sensitivity analysis of the model by Scott et al. (2008). Table 2 shows the tuning parameters and their range of possible values. Given the similarity between the LP and QAA $b_b(490)$ datasets it was deemed unnecessary to tune the model to both datasets and so the QAA dataset was omitted from the following work. Thus GOTM-HadOCC is tuned independently at each site to the following four datasets:

C Chlorophyll

L Chlorophyll and $b_b(490)$ from LP

G Chlorophyll and $b_b(490)$ from GSM

P Chlorophyll and $b_b(490)$ from PML

For a given parameter set, GOTM-HadOCC outputs Chl_{rs} and b_{brs} which are then compared to the observed dataset using the cost function:

$$\text{cost} = \frac{1}{J}\Sigma_{j=1}^{J}\frac{1}{N_j\sigma_j^2}\Sigma_{n=1}^{N_j}(\text{obs}_{jn} - \text{mod}_{jn})^2 \qquad (13)$$

where j is the data type (i.e., Chl or b_b), J is the number of datatypes (i.e., 1 or 2), N_j is the number of observations of type j, σ_j^2 is the variance of the observed data for the given site over the entire time period (see Table 1) and mod_j is an 8 day moving average of the 6-hourly model output. When there is more than one type of observation, i.e., both Chl and b_b, they are adjusted to each have the same number of data points.

4.3.1 Optimisation using a genetic algorithm

To search through the parameter space to find the optimum parameter set (i.e., the one which minimizes the cost function (Eq. 13)), a genetic algorithm (GA) is employed (following

Table 1. Variances of observed data.

Site	Chl mg m^{-3}	$b_b(490) \times 10^{-7}$ m^{-1}		
		LP	GSM	PML
CIS	0.0819	9.47	3.82	17.9
EST	9.5 x 10^{-4}	0.935	0.502	1.24
PAP	0.0364	9.36	8.54	18.9

Schartau and Oschlies, 2003). This approach has also been recently tested by Ward et al. (2006) for a 1-D marine ecosystem model and was found to give results comparable to those for optimisation using the variational adjoint method. A GA is simply a search/optimisation technique based on Darwin's theory of natural selection (for an good introduction to GAs see Goldberg, 1989). The basic idea is that a set of model parameters is viewed as an "individual" whose fitness, F, is determined by

$$F = \frac{1}{\text{cost}}. \qquad (14)$$

A number of individuals make up a "population" within which the individuals "reproduce" to make a population of new individuals (the next generation). Whether or not attributes from a certain individual are transferred to the next generation depends on whether the individual is allowed to reproduce, which in turn depends on its fitness. Individuals with high fitness levels are more likely to reproduce than those with low fitness (survival of the fittest). The offspring from two individuals is a new individual (parameter set) with characteristics from each parent. To bring in new information, mutation may be used to make small modifications to some of the children. Thus, the procedure produces successive generations of parameter sets that give models whose output is a better match to the observed data, eventually converging to the parameter set that gives the best possible model fit.

Here a micro-genetic algorithm (μGA) coded and published by Carroll (1996) and made freely available at http://cuaerospace.com/carroll/ga.html is used. Details of the μGA are given by Krishnakumar (1989). The μGA is based on the same operations as a general GA but it does not contain mutation and gives greater influence on elitism principles thereby assuring that the best individual (parameter set) is transferred to the next generation. As soon as all the individuals of one generation show less than 5% difference between each other, a new random population is generated (although the best individual is retained).

Here, each parameter in the set of 11 tuneable parameters (Table 2) is represented by a 6 digit binary string thus enabling each parameter to take 64 possible values (the resulting parameter increments are given in Table 2). The string for each individual (i.e. parameter set) is thus 66 (11×6)

digits long and the population is chosen to comprise of 11 individuals. The optimisation is run for 2000 generations (11×2000 model runs for each dataset and site) to ensure convergence.

5 Results

5.1 Model fit to data

At each site GOTM-HadOCC is tuned in turn to the C, L, G and P datasets (defined in Sect. 4.3). In most cases the GA has achieved the optimum fitness after ∼1000 generations (see Fig. 4). Figures 5 and 6 show the model outputs of Chl_{rs} and b_{brs} (from here on simply referred to as Chl and b_b) when run with the optimised parameter sets, compared with the appropriate satellite data. The fits are remarkably good given GOTM-HadOCC is only 1-D with no accounting for advection. Broadly speaking, tuning also to b_b data increases the magnitude of both modelled Chl and b_b (Figs. 5 and 6). Table 3 gives the RMSEs for the different model fits. Note the number of observed data points varies between IOP algorithms as some of the algorithms produce more invalid values than others (e.g. negative numbers). In general, and as expected, adding in b_b data causes an increase in Chl error. Somewhat surprisingly Table 3 shows that in some cases the b_b RMSE is actually smaller when the model is not tuned to b_b (e.g., at CIS and PAP tuning to dataset P gives a larger b_b RMSE than tuning to dataset C; at ESTOC tuning to L and P are worse than tuning to C). This has to do with the way the errors are calculated whereby the models tuned to dataset C are compared with the mean of the observed b_b datasets, whereas the models tuned to a particular b_b dataset are then compared with that dataset. The combined RMSE results show that GOTM-HadOCC produces the best fit to dataset G which implies that the SeaWiFS chlorophyll data and the GSM b_b data are related in a way that is the most consistent with the mechanisms contained within GOTM-HadOCC.

5.2 Optimised parameter values

Table 4 shows the optimised parameter values for each dataset at each of the three sites. Using a student t-test it is possible to test whether the parameters in the model tuned to dataset C are significantly different (i.e., outside the 95th percentile) to those obtained when b_b data are also used (datasets L, G and P) – this is shown by the line labelled "significant" in Table 4. Also, to see how well constrained the optimised parameters are, it is useful to look at the sensitivity of the model fitness to variations in each of the parameters when the other parameters are at their optimum values (Fennel et al., 1995) – this is shown in Fig. 7. Well constrained parameters are evident by a well defined curve whose gradient changes from positive to negative at the optimum parameter value, e.g., P_m for dataset C at CIS (Fig. 7). Note that optimised parameters occuring at the bounds of their predefined range are
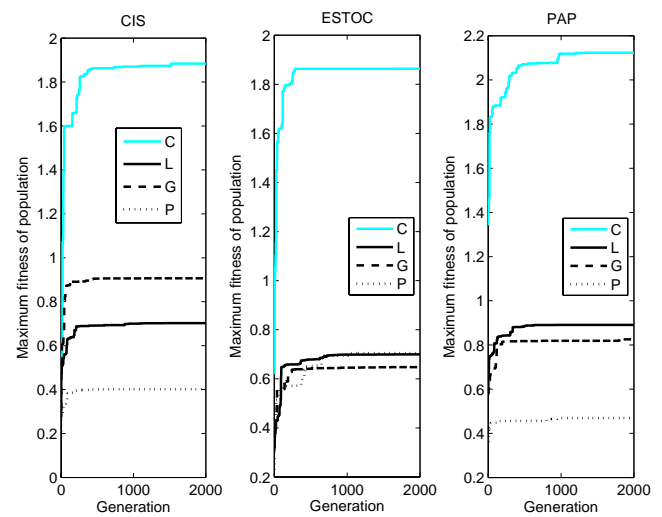

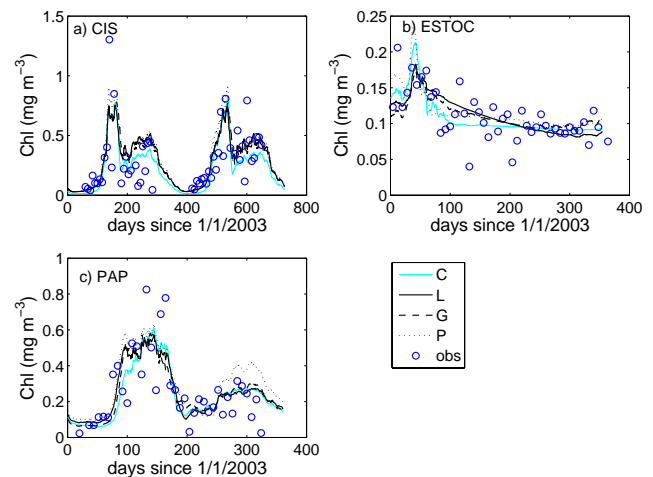**Fig. 4.** Convergence of GA at each site.

Fig. 5. Comparing Chl from optimised models (8-day running means) with observed satellite data.

considered to be unconstrained by the optimisation. Looking at Table 4 and Fig. 7 together allows the real difference in the model parameter values caused by adding in b_b data to be determined.

At CIS, K_N, P_m, ϕ, R and V are found to be significantly altered by adding in b_b data (Table 4). However, Fig. 7 shows that the model is virtually insensitive to changes in K_N, and for the b_b datasets, R and V are at the lower bounds of their predefined ranges. Thus of the parameters that show a significant and consistent shift in value due to b_b data, only P_m and ϕ are well constrained. Alternatively, the parameters μ_1, K_f and g_{max} are reasonably well constrained but their values are not significantly and consistently altered by adding in b_b data. Thus, it can be concluded that at CIS, adding in b_b data causes only P_m and ϕ to each change towards a different and

Table 2. Parameters of the NPZD model. Note the fixed parameters are shown in Table 2 in Kettle and Merchant (2008) with the exception that F_{th} is set to $0.01 \, \mathrm{mmol \, N \, m^{-3}}$ at all sites.

Parameter for variation	Symbol	Units	Lower bound	Upper bound	Interval
Half saturation constant of N uptake	K_N	$\mathrm{mmol \, N \, m^{-3}}$	0.01	0.85	0.0133
Max. photosynthetic rate	P_m	$\mathrm{d^{-1}}$	0.1	5.1	0.0808
Yield	ϕ	$\mathrm{mol \, C \, mol \, quanta^{-1}}$	0.01	0.12	0.0018
Respiration rate	R	$\mathrm{d^{-1}}$	0.005	0.095	0.0014
Conc. dependent specific mortality	m_0	$\mathrm{d^{-1}(mmol \, N \, m^{-3})^{-1}}$	0.008	0.25	0.0038
Constant specific mortality (zoo.)	μ_1	$\mathrm{d^{-1}}$	0.03	0.2	0.0027
Zoo-dependent specific mortality	μ_2	$\mathrm{d^{-1}(mmol \, N \, m^{-3})^{-1}}$	0.03	0.57	0.0086
Half saturation constant for Z grazing	K_f	$\mathrm{mmol \, N \, m^{-3}}$	0.4	1.0	0.0095
Max. Z grazing rate	g_{max}	$\mathrm{d^{-1}}$	0.06	2.0	0.0308
Detrital sinking rate	V	$\mathrm{m \, d^{-1}}$	3.0	32.0	0.4603
Deep remineralisation rate	γ	$\mathrm{m \, d^{-1}}$	3.8	13.36	0.1518

well constrained value. At ESTOC, P_m, μ_1 and K_f are significantly different when b_b data are added (Table 4) but only K_f is well constrained (the others have values at the bounds of their ranges). Figure 7 shows that, at ESTOC, very few parameters are well constrained indicating that there are large uncertainties associated with the model simulations at this location. At PAP, K_N, m_0, and V are significantly different when b_b data are added but only m_0 is within its bounds and is well constrained. Thus the information added by including b_b data does not consistently constrain the same parameters at each site. In fact, at all of the sites there are some parameters that are much better constrained when only Chl data are used (e.g., at CIS: ϕ, μ_1, K_f; at ESTOC: P_m, K_f, and at PAP: K_N, R, m_0, μ_1 and K_f). This implies that the way $b_{bp}(490)$ is generated by GOTM-HadOCC needs significant refinement (i.e. Eqs. 9–11) in order that the mechanisms relating Chl and $b_{bp}(490)$ are consistent with the satellite data.

5.3 Model behaviour

Given that the optimised parameter values may show a range of values for the different tuning datasets, it is perhaps more informative to examine the actual model outputs to ascertain the real effect of including b_b data. Figure 8 shows the modelled nutrient, phytoplankton, zooplankton and detritus within the mixed layer for each site for each optimised parameter set. Results vary considerably between sites with all of the differently optimised models having very similar behaviour at CIS, very different behaviour at ESTOC and differences which depend on the model compartment examined at PAP. At ESTOC, it is clear that the satellite data do not have sufficient dynamical range to constrain the model behaviour. At all of the sites, however, the models tuned to datsets G and L show very similar behaviour implying that the b_b data from these 2 algorithms are not significantly different. At CIS and PAP the models tuned to dataset P (dotted

Table 3. RMS errors between the tuned-model output and the satellite data. Note b_b output from models tuned to dataset C are compared with the mean of the 3 b_b datasets whereas the others are compared with the b_b datset they were tuned to. The combined error is computed by scaling the two RMSEs by the mean for the 4 tuning datasets and then adding the two scaled error values. The number of observed data points is shown in brackets (note CIS covers 2 years).

Dataset	C	L	G	P
	RMSE with observed Chl ($\mathrm{mg \, m^{-3}}$)			
CIS	0.18	0.19	0.19	0.21
EST	0.025	0.026	0.026	0.026
PAP	0.13	0.14	0.15	0.16
	RMSE with observed b_b ($10^{-4} \, \mathrm{m^{-1}}$)			
CIS	13.4	10.0	8.11	21.0
EST	2.41	2.88	2.11	3.21
PAP	11.5	9.24	7.14	16.32
	Combined error (dimensionless)			
CIS	1.94 (54)	1.77 (48)	1.61 (51)	2.67 (44)
EST	1.86 (45)	2.10 (43)	1.80 (45)	2.24 (38)
PAP	1.96 (38)	1.77 (33)	1.66 (36)	2.60 (28)

lines) have outputs which are generally similar to those for L and G with the notable exception of the complete extinction of zooplankton at PAP! At CIS the model tuned to dataset C (the chlorophyll only dataset) shows slightly less phytoplankton (and subsequently less nutrient uptake), and zooplankton and detritus than the models that are also tuned to b_b. At PAP, the model tuned to dataset C shows similar concentrations of nutrient and phytoplankton but significantly less detritus (the zooplankton compartment is clearly not well constrained by the satellite data).

Table 4. Optimised model parameter sets for each tuning dataset at each site. Numbers in bold indicate the parameter value is at the bounds of its predefined range. Significant: "yes" indicates that the parameter value for dataset C is outside of the 95th percentile for the other 3 datasets using the student t-test (i.e. adding in b_b data changes the parameter value in a consistent way).

|	K_N	P_m	ϕ	R	m_0	μ_1	μ_2	K_f	g_{max}	V	γ	fitness
CIS												
C	0.72	1.22	0.040	0.011	**0.250**	0.079	**0.570**	0.657	1.415	3.92	7.14	1.88
L	**0.85**	0.58	0.089	**0.005**	**0.250**	0.044	0.544	0.924	1.107	**3.00**	**3.80**	0.70
G	**0.85**	0.58	0.092	**0.005**	**0.250**	0.054	**0.570**	0.700	0.922	**3.00**	13.36	0.91
P	**0.85**	0.74	0.090	**0.005**	**0.250**	0.081	**0.570**	**1.000**	1.538	**3.00**	**3.80**	0.40
significant	yes	yes	yes	yes	no	no	no	no	no	yes	no	
EST												
C	**0.85**	0.66	0.113	**0.005**	**0.008**	**0.200**	0.201	0.533	1.784	4.84	**3.80**	1.86
L	**0.01**	4.53	**0.120**	0.015	0.012	0.143	0.450	0.905	0.676	19.57	**3.80**	0.70
G	**0.01**	4.62	0.113	0.028	**0.008**	0.116	**0.570**	0.838	1.230	31.54	4.86	0.65
P	**0.85**	**5.10**	0.064	**0.005**	0.012	0.116	0.073	0.981	1.938	**3.00**	**3.80**	0.70
significant	no	yes	no	no	no	yes	no	yes	no	no	no	
PAP												
C	0.69	3.16	**0.010**	0.019	0.158	0.060	**0.030**	0.552	0.645	7.60	**13.36**	2.12
L	0.82	4.53	0.014	0.016	0.242	0.087	0.090	**0.400**	0.553	**3.00**	**13.36**	0.89
G	**0.85**	0.66	0.057	0.062	0.223	0.041	0.253	0.533	0.399	**3.00**	12.90	0.83
P	**0.85**	2.43	0.015	0.016	0.219	0.184	0.364	0.486	1.969	**3.00**	4.10	0.47
significant	yes	no	no	no	yes	no	no	no	no	yes	no	
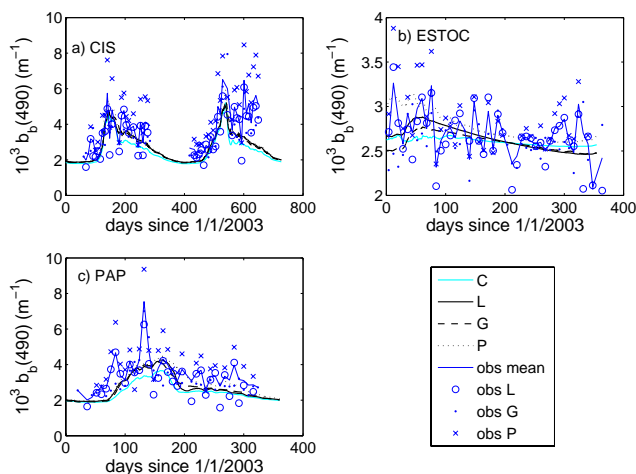
Fig. 6. Comparing $b_b(490)$ from optimised models (8-day running means) with observed satellite data (averge of all IOP algorithm results).

Interestingly at ESTOC the phytoplankton compartment shows a wide range of values which implies the model is simulating very different phytoplankton physiology (C:Chl ratios) since the chlorophyll concentrations are similar (Fig. 5). Figure 9 shows C:Chl ratios in the mixed layer for each site. At CIS they are virtually identical for the different models; at PAP there are large differences when the model is tuned to dataset P and small differences between the other models in the latter half of the year. At ESTOC, however, there are

very large differences in C:Chl between the models. This indicates that correctly simulating Chl at this location with GOTM-HadOCC will not necessarily result in the correct estimation of carbon fixation. Furthermore, the fairly large differences between the models tuned to datasets with b_b included indicates that at this location (where there is no strong seasonal cycle), the differences in the b_b values from different IOP algorithms is significant. This is supported by the values for mean column primary production predicted by the models which are tuned to the datasets that include b_b data (Table 5) – at ESTOC the values vary by up to 191%, at CIS by up to 24% and at PAP by up to 54%.

The other important variable for climate prediction is the export of detrital material below the thermocline (export production). This is a balance between the rate at which the material sinks and the rate at which it is remineralised back to nutrient. In order to compare the effect of the different parameter sets on export production, the amount of detritus below 200 m on the last day of the simulation is used. The choice of depth level is arbitrary since the physics of the models at each site are the same. Table 5 shows that the amount of detrital carbon below 200 m is highly variable even for models that have very similar surface behaviour (e.g., models tuned to datasets L and G at CIS). This is because it is highly sensitive to the remineralisation rate, γ which is not well constrained since its effects are seen beneath the depth to which the satellite sensor can retrieve information.
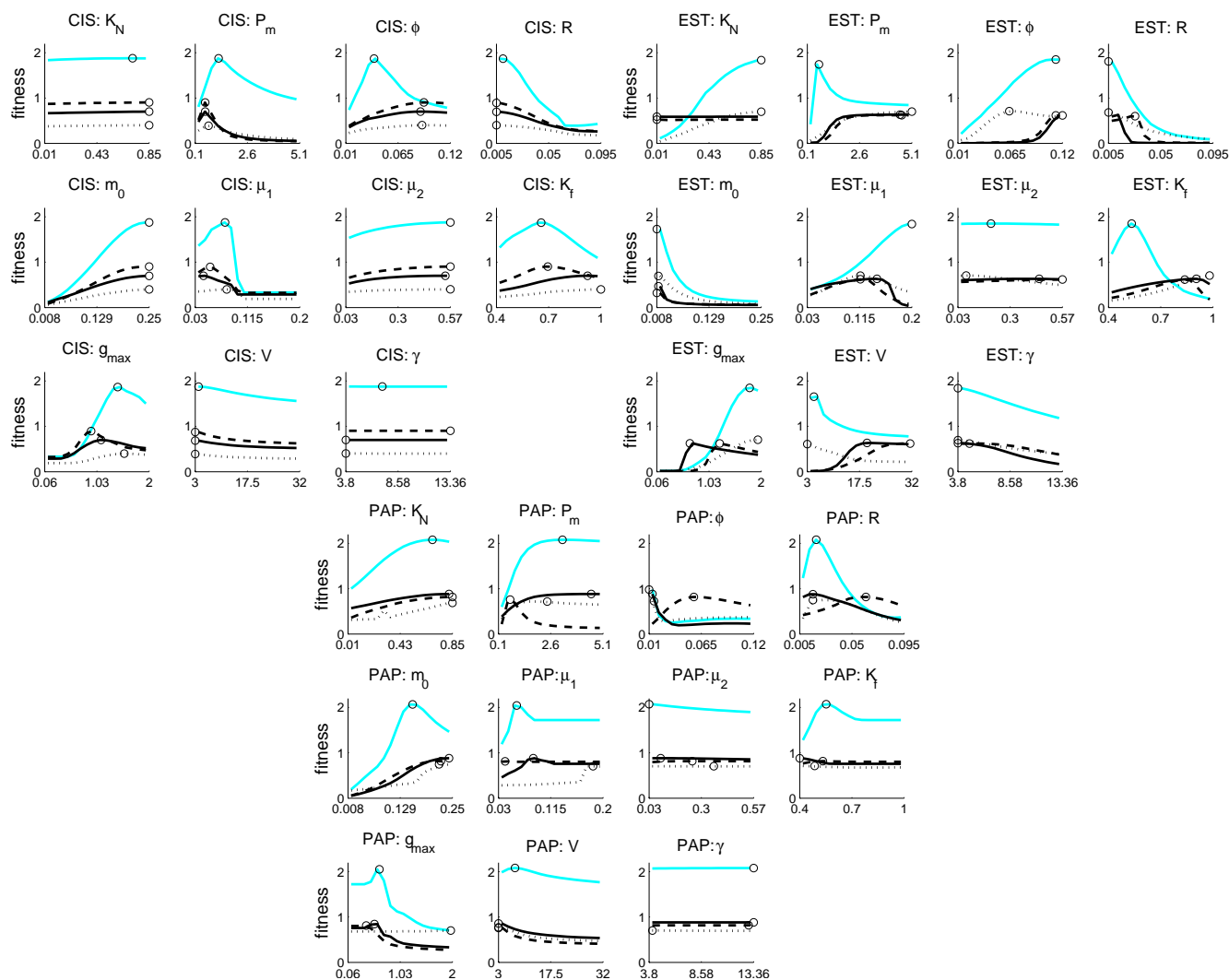
Fig. 7. Sensitivity of model fitness to variations in each parameter when the other parameters are at their optimised values for datasets C (pale blue), L (black), G (dashed) and P (dotted). The optimum parameter values are marked by circles.

6 Conclusions

It has been shown here that satellite Chl and b_b data are not sufficient to fully constrain the parameters in a simple 1-D biogeochemical ocean model. In fact including b_b data increases the errors in the simulated Chl in the mixed layer and causes a number of parameters to be less well constrained. This implies that the mechanisms in the model that connect Chl and b_b are not consistent with the observations. This implies one or more of the following:

1. Modelled phytoplankton carbon (which is a function of phytoplankton physiology (C:Chl)) and detrital carbon (and hence POC) are incorrect.

2. Equations 9–12 need refining.

3. b_{BG} is incorrect and should have been included in the optimisation procedure.

Table 5. Comparing results from the differently optimised models for the mean primary production in the water column ($mg\,C\,m^{-2}\,d^{-1}$) and the amount of detrital carbon below 200 m ($g\,C\,m^{-2}$).

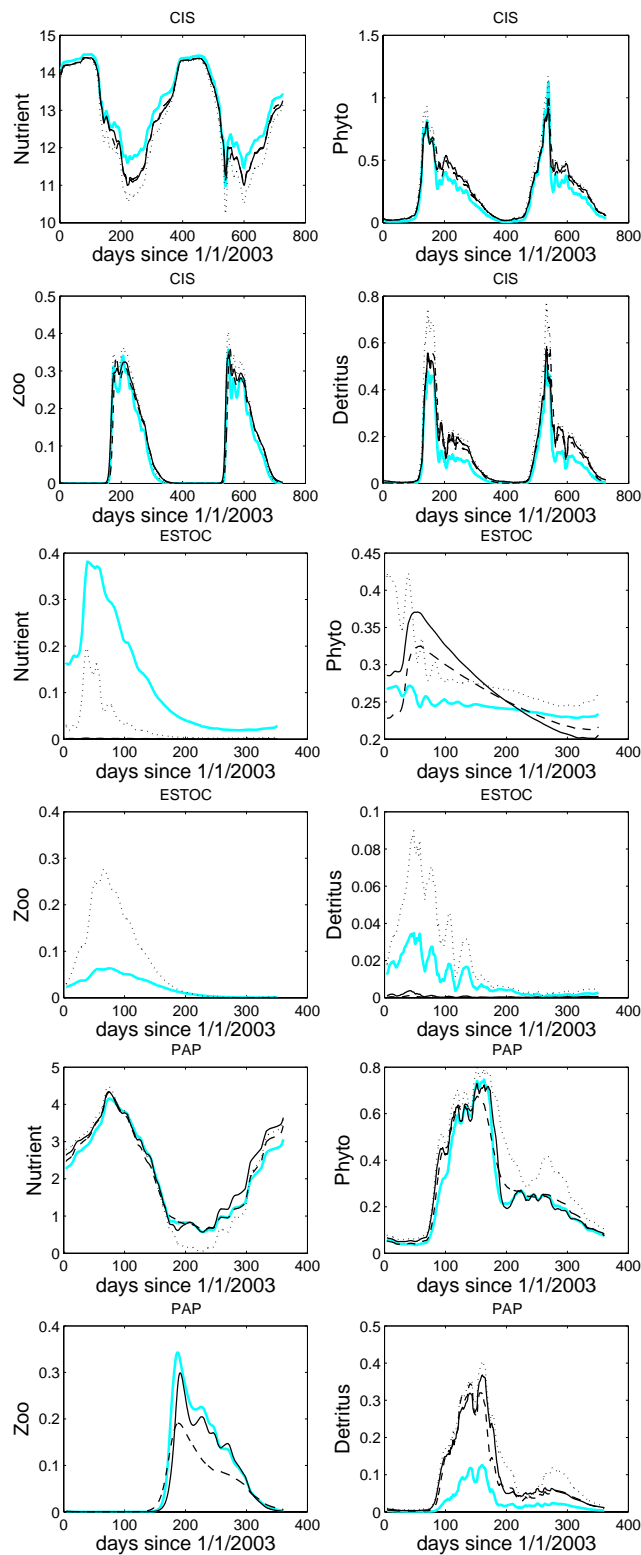| Tuning data:	C	L	G	P
Mean column primary production				
CIS	152.4	181.0	182.5	224.5
EST	92.4	40.9	65.8	119.4
PAP	62.3	86.7	134.2	95.7
Detrital carbon below 200 m				
CIS	0.42	2.12	0.09	2.62
EST	1.95	2.55	1.56	1.06
PAP	0.08	0.01	0.01	0.47

Fig. 8. Modelled 8-day running means of mixed-layer nutrient, phytoplankton, zooplankton and detritus (all in mmol N m^{-3}) produced by the optimised models at each site. Thick pale blue lines represent the model optimised to dataset C, black lines are for datsets L (solid), G (dashed) and P (dotted).
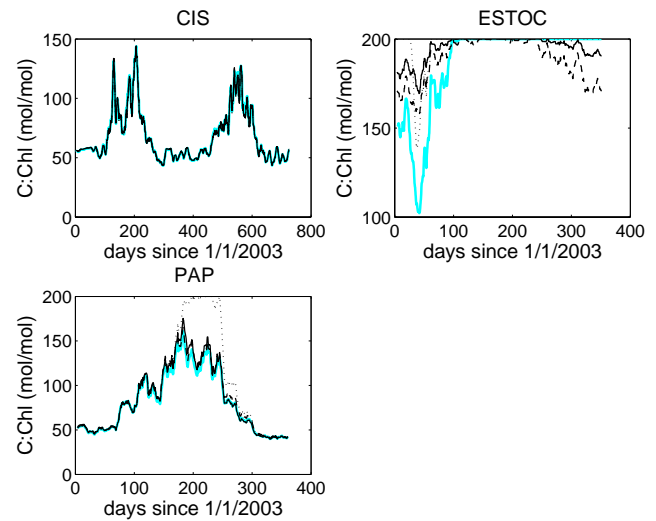

Fig. 9. Modelled 8-day running means of mixed-layer C:Chl produced by the optimised models at each site. Thick pale blue lines represent the model optimised to dataset C, black lines are for datsets L (solid), G (dashed) and P (dotted).

4. Some fraction of zooplankton should be added to POC.

In addition to this, the possibility that some other model parameters, that were not included in the optimisation procedure, are important should also be mentioned.

However, at two of the locations chosen (CIS and PAP) using the Chl and b_b datasets together *is* sufficient to constrain modelled outputs of nutrient, phytoplankton and detritus (but not always zooplankton) within the mixed layer. Moreover, including b_b data rather than simply using Chl significantly and consistently alters the modelled detritus at PAP.

The differences in b_b resulting from different IOP inversion algorithms do not appear to be hugely significant as the models tuned to the different datasets show fairly consistent behaviour. However, using the b_b data from the GSM algorithm gives the best simultaneous fit of modelled Chl and modelled b_b to the satellite data at each site.

To conclude, incorporating b_b data into ocean carbon cycle models has great potential to improve simulations of phytoplankton and detritus in the mixed layer but this is a relatively new field of research and as this preliminary investigation shows, there is still much work to be done.

Edited by: L. Bopp

References

Bricaud, A., Morel, A., Babin, M., Allali, K., and Claustre, H.: Variations of light absorption by suspended particles with chlorophyll-*a* concentration in oceanic (case 1) waters. Analysis and implications for bio-optical models, J. Geophys. Res., 103, 31 033–31 044, 1998.

Burchard, H., Bolding, K., and Villarreal, M. R.: GOTM – a general ocean turbulence model. theory, applications and test cases, Tech. Rep. EUR 18745 EN, European Commission, 1999.

Carroll, D. L.: Chemical Laser Modeling with Genetic Algorithms, AIAA, 34(2), 338–346, 1996.

Claustre, H., Morel, A., Babin, M., Cailliau, C., Marie, D., Marty, J.-C., Tailliez, D., and Vaulot, D.: Variability in particle attenuation and chlorophyll fluorescence in the tropical Pacific: Scales, patterns, and biogeochemical implications, J. Geophys. Res., 104(C2), 3401–3422, 1999.

Cloern, J. E., Grenz, C., and Vidergar-Lucas, L.: An empirical model of the phytoplankton chlorophyll:carbon ratio – the conversion factor between productivity and growth rate, Limnol. Oceanogr., 40(7), 1313–1321, 1995.

Fennel, K., Losch, M., Schröter, J., and Wenzel, M.: Testing a marine ecosystem model: sensitivity analysis and parameter optimization, J. Marine Syst., 28, 45-63, 2001.

Fujii, M., Boss, E., and Chai, F.: The value of adding optics to ecosystem models: a case study, Biogeosciences, 4, 817–835, 2007,
http://www.biogeosciences.net/4/817/2007/.

Garver, S. A. and Siegel, D. A.: Inherent optical property inversion of ocean color spectra and its biogeochemical interpretation: I. Time series from the Sargasso Sea, J. Geopys. Res., 102, 18607–18625, 1997.

Geider, R. J., MacIntyre, H. L., and Kana, T.: Dynamic model of phytoplankton growth and acclimation: responses of the balanced growth rate and the chlorophyll-*a*: carbon ratio to light, nutrient-limitation and temperature, Ecol. Prog. Ser., 148, 187–200, 1997.

Goldberg, D.: Genetic Algorithms in Search, Optimization and Machine Learning, Addison-Wesley, 1989.

Kettle, H. and Merchant, C. J.: Modeling ocean primary production: sensitivity to spectral resolution of attenuation and absorption of light, Progr. Oceanogr., 78(2), 135, doi:10.1016/j.pocean.2008.04.002, 2008.

Krishnakumar, K.: Micro-genetic algorithms for stationary and non-stationary function optimization, Intel. Control and Adaptive Syst., 1196, 289–296, 1989.

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

Levitus, S., Conkright, M., Reid, J., Najjar, R., and Mantyla, A.: Distributon of nitrate, phosphate and silicate in the world oceans, Prog. Oceangr., 31, 245–273, 1993.

Loisel, H. and Poteau, A.: Inversion of IOP based on R_{rs} and remotely retrieved K_d, IOCCG. Remote Sensing of Inherent Optical Properties: Fundamentals, Tests of Algorithms, and Applications. Lee, Z-P (ed), Reports of the International Ocean-Colour Coordinating Group, No. 5, IOCCG, Dartmouth, Canada, 35–41, 2006.

Loisel, H. and Stramski, D.: Estimation of the inherent optical properties of natural waters from the irradiance attenuation coefficient and reflectance in the presence of Raman scattering, Appl. Opt,, 39(18), 3001–3011, 2000.

Loisel, H., Stramski, D., Mitchell, B., Fell, F., Fournier-Sicre, V., Lemasle, B., and Babin, M.: Comparison of the ocean inherent optical properties obtained from measurements and inverse modelling, Appl. Optics, 40(15), 2384–2397, 2001.

Loisel, H., Nicolas, J., Deschamps, P., and Frouin, R.: Seasonal and inter-annual variability of particulate organic matter in the global ocean, Geophys. Res. Lett., 29(24), 2196, doi:10.1029/2002GL015948, 2002.

Loisel, H., Nicholas, J., Sciandra, A., Stramski, D., and Poteau, A.: Spectral dependency of optical backscattering by marine particles from satellite remote sensing of the global ocean, J. Geophys. Res., 111, C09024, doi:10.1029/2005JC003367, 2006.

Maritorena, S., Siegel, D. A., and Peterson, A. R.: Optimization of a semianalytical ocean color model for global-scale applications, Appl. Optics, 41, 2705–2714, 2002.

Morel, A. and Maritorena, S.: Bio-optical properties of oceanic waters: a reappraisal, J. Geophys. Res., 106, 7163–7180, 2001.

Palmer, J. R. and Totterdell, I. J.: Production and export in a global ocean ecosystem model, Deep-Sea Res., 48(5), 1169–1198, 2001.

Schartau, M. and Oschlies, A.: Simultaneous data-based optimization of a 1D-ecosystem model at three locations in the North Atlantic: Part 1 – Method and parameter estimates, J. Marine Res., 61, 765–793, 2003.

Scott, V., Kettle, H., Hankin, R., and Merchant, C. J.: Sensitivity analysis of an ocean carbon cycle model: an investigation of parameters affecting CO_2 flux, carbon export and primary production, J. Mar. Res., in review, 2008.

Siegel, D., Maritorena, S., Nelson, N., Hansell, D., and Lorenzi-Kayser, M.: Global distribution and dynamics of colored dissolved and detrital organic materials, J. Geophys. Res., 107, 3228, doi:doi:10.1029/2001JC000965, 2002.

Smyth, T. J., Moore, G. F., Hirata, T., and Aiken, J.: Semianalytical model for the derivation of ocean color inherent optical properties: description, implementation, and performance assessment, Appl. Optics, 45(31), 8116–8131, 2006.

Stramski, D., Reynolds, R. A., Kahru, M., and Mitchell, B. G.: Estimation of particulate organic carbon in the ocean from satellite remote sensing, Science, 285, 239–242, 1999.

Twardowski, M., Boss, E., Macdonald, J., Pegau, W., Barnard, A., and Zaneveld, J.: A model for estimating bulk refractive index from the optical backscattering ratio and the implications for understanding particle composition in case I and case II waters, J. Geophys. Res., 106, 14 129–14 142, 2001.

Ward, B., Friedrichs, M. A. M., Anderson, T., and Osclies, A.: Parameter optimisation techniques and the problem of underdetermination in marine biogeochemical models, J. Marine Syst., in press, 2009.