Satellite views of global phytoplankton community distributions using an empirical algorithm and a numerical model

C. S. Rousseaux\textsuperscript{1,2}, T. Hirata\textsuperscript{3}, and W. W. Gregg\textsuperscript{1}

\textsuperscript{1}Global Modeling and Assimilation Office, NASA Goddard Space Flight Center, Greenbelt, Maryland, USA
\textsuperscript{2}Universities Space Research Association, Columbia, Maryland, USA
\textsuperscript{3}Faculty of Environment Earth Science, Hokkaido University, Japan

Received: 13 December 2012 – Accepted: 15 January 2013 – Published: 24 January 2013

Correspondence to: C. S. Rousseaux (cecile.s.rousseaux@nasa.gov)

Published by Copernicus Publications on behalf of the European Geosciences Union.
Abstract

We compared the functional response of a biogeochemical data assimilation model versus an empirical satellite-derived algorithm in describing the variation of four phytoplankton (diatoms, cyanobacteria, coccolithophores and chlorophytes) groups globally and in 12 major oceanographic basins. Global mean differences of all groups were within ~15% of an independent observation data base for both approaches except for satellite-derived chlorophytes. Diatoms and cyanobacteria concentrations were significantly (p < 0.05) correlated with the independent observation data base for both methods. Coccolithophore concentrations were only correlated with the in situ data for the model approach and the chlorophyte concentration was only significantly correlated to the in situ data for the satellite-derived approach. Using monthly means from 1998–2007, the seasonal variation from the satellite-derived approach and model were significantly correlated in 11 regions for diatoms and in 9 for coccolithophores but only in 3 and 2 regions for cyanobacteria and chlorophytes. Most disagreement on the seasonal variation of phytoplankton composition occurred in the North Pacific and Antarctic where, except for diatoms, no significant correlation could be found between the monthly mean concentrations derived from both approaches. In these two regions there was also an overestimate of diatom concentration by the model of ~60% whereas the satellite-derived approach was closer to in situ data (8–26% underestimate). Chlorophytes were the group for which both approaches differed most and that was furthest from the in situ data. These results highlight the strengths and weaknesses of both approaches and allow us to make some suggestions to improve our approaches to understanding phytoplankton dynamics and distribution.

1 Introduction

Phytoplankton composition plays a major role in biogeochemical cycles in the ocean. The intensity of carbon fixation and export is strongly dependent on the phytoplankton
community composition. The approaches to characterize the phytoplankton community composition at a global scale can be roughly classified in two categories: modelling approaches and satellite-derived approaches. Biogeochemical models coupled to physical circulation can describe the complex interactions between the physics and biology in the oceans (e.g. Le Quere et al., 2005; Moore et al., 2004; Dunne et al., 2005; Doney and Ducklow, 2006; Gregg et al., 2003; Dutkiewicz et al., 2009). The phytoplankton composition in the modelling approaches is generally based on biogeochemical functions. Data assimilation techniques can also be used to constrain the model to track observations time series and to optimize certain variables. Satellite approaches can rely on bio-optical algorithms (spectral-based approach), which can then be used to estimate particle size distribution (e.g. Ciotti et al., 2002; Mouw and Yoder, 2006), functional groups (e.g. Alvain et al., 2005) or other parameters of interest. Another way phytoplankton composition can be derived from satellite is by relying on phytoplankton concentration (expressed as chlorophyll concentration or absorption coefficient, e.g. Uitz et al., 2006; Aiken et al., 2007) as an indicator for phytoplankton community composition. Each of these approaches has their own strengths and weaknesses. Some authors for example (Anderson, 2005; Flynn, 2005) have suggested that the amount of observations available may be inadequate to constrain the high number of parameters in some complex biogeochemical models and to evaluate the performance of these models. Weaknesses from the satellite-based approach include uncertainties in the water leaving radiances (e.g. changes in the sensor’s performance, contamination with light from adjacent pixels, etc), from the relationship that links water leaving radiances to chlorophyll and from the inherent uncertainties of in situ dataset used to derive the algorithms.

Models and satellite-derived approaches have been used to assess global changes in phytoplankton biomass and community composition at various time scales. Temporal oscillations of phytoplankton biomass are often very variable. The number, timing and magnitude of annual blooms may differ remarkably among locations. Seasonal cycles at mid to high latitudes (> 45° N or S) are traditionally viewed as one of the most
important sources of biological and biogeochemical variation in the oceans. While sub-
tropical waters (20–40° N or S) have seasonal maximum in local winter and minima
in summer, subpolar waters (> 40° N or S) have local spring or summer maxima and
winter minima (Yoder and Kennelly, 2006). Seasonal cycles have different forcing de-
pending on the region. For example, the spring-summer blooms in subpolar waters are
related to winter mixing that replenishes surface waters with nutrients. Winter mixing
is followed by spring-summer increases of incident solar irradiance and water column
stratification leading to a well lit, initially nutrient-rich mixed layer conducive to phyto-
plankton growth and biomass increase. Winter blooms in subtropical waters are also
responses to winter mixing, although mixing in the subtropics is generally weaker than
that which occurs farther poleward. In the subtropics, comparatively high winter so-
lar irradiance at the lower latitudes and shallow mixed layers leads to an immediate
phytoplankton response (winter blooms) to nutrients.

In this study we compare the functional response of a numerical model (NASA Ocean
Biogeochemical Model, NOBM; Gregg et al., 2003) versus an empirical algorithm (Hi-
rata et al., 2011) in describing the seasonal variation of four phytoplankton groups
globally and in 12 major oceanographic basins. The satellite-derived approach of Hirata
et al. (2011) was chosen because it discriminate between sufficient PFTs to allow for
a comparison with the phytoplankton groups from the model. Elucidating broad-scale
seasonal patterns in phytoplankton biomass is of interest because it can help reveal
general controls of phytoplankton temporal dynamics. The comparison of a model and
a satellite-based approach in characterizing the seasonal patterns of phytoplankton
biomass will highlight strengths and weaknesses in our approach to understand large
scale climate effects.

2 Material and methods

The NOBM is a global biogeochemical model that is coupled with a circulation and
radiative model (Gregg and Casey, 2007). The model contains 4 explicit phytoplankton
taxonomic groups: diatoms, cyanobacteria, chlorophytes, and coccolithophores. The phytoplankton groups differ in maximum growth rates, sinking rates, nutrient requirements, and optical properties. In the model, the diatoms and cyanobacteria represent functional extremes. While the high growth rates of diatoms allow them to flourish in areas of abundant nutrients (high latitude, coastal and equatorial upwelling, see Gregg et al., 2003), their large sinking rate prevent them from dominating in quiescent regions. Cyanobacteria have a slow growth rates, their high nitrogen uptake efficiency, slow sinking rate and ability to fix atmospheric nitrogen allow them to sustain in low nitrogen areas (e.g. mid-ocean gyres). Coccolithophores tolerate lower nutrient conditions and have the property of sinking faster than most phytoplankton (Fritz and Balch, 1996). Finally, chlorophytes occupy the transitional regions between the high nutrients regions dominated by diatoms and the nutrient-scarce regions dominated by cyanobacteria. Total chlorophyll from Sea-viewing Wide Field-of-view Sensor (SeaWiFS) data are assimilated in NOBM for the period from 1998 until 2007 (Gregg, 2008) using a multi-variate assimilation approach (Rousseaux and Gregg, 2012).

The phytoplankton composition derived from the model is compared to the satellite algorithm of Hirata et al. (2011). This algorithm is based on the existence of a relationship between chlorophyll and phytoplankton composition. By quantifying the relationship between satellite derived chlorophyll and HPLC, this algorithm allows the distinction of 3 phytoplankton size classes (micro-, nano- and pico-plankton) and 7 phytoplankton “functional” types (diatoms, prymnesiophytes, green algae, dinoflagellates, prokaryote, picoeukaryote and Prochlorococcus sp.). These phytoplankton groups do not map directly onto the classifications of NOBM. Diatoms and chlorophytes in NOBM are consistent with the classification of diatoms and green algae respectively in Hirata et al. (2011). Cyanobacteria in NOBM do not correspond as well with the prokaryotes of Hirata et al. (2011). Additionally, the classification of coccolithophores from NOBM only represent a portion of the prymnesiophytes of Hirata et al. (2011). With these differences in mind, we proceed to statistical comparisons. For clarity, we refer to these
groups as diatoms, chlorophytes, cyanobacteria and coccolithophores whether we dis-
cuss NOBM or the satellite-derived approach.

The phytoplankton groups from the model and satellite-derived approach are vali-
This data set includes 469 surface-layer observations of phytoplankton group con-
centrations (Fig. 1). The co-located, coincident match ups are assembled over ocean
basins and over the months for a year. The spatial distribution of the four phytoplank-
ton groups from both approaches are compared by calculating a climatology using the
period 1998–2007. We then proceed to compare the seasonal variation of the four phy-
toplankton groups using both approaches in the 12 major oceanographic basins (Fig. 1)
using monthly averages for the period from 1998 until 2007. The difference in relative
abundance is expressed as model/satellite algorithm minus in situ, with resulting units
of percent. Comparisons between the two approaches are expressed as differences in
absolute abundances with resulting units of mg m\(^{-3}\) chlorophyll.

3 Results

3.1 Comparison of satellite-derived algorithm and model representations of the
global phytoplankton community with in situ data

Global mean differences of all groups from the model and the satellite-derived algorithm
were within ~15% of observations except for satellite-derived chlorophytes (globally
an underestimate of \(-21\%\), Table 1). Global diatom and cyanobacteria concentrations
were significantly correlated with observations. Coccolithophore concentrations were
only significantly correlated to the observations when using the model. Chlorophyte
concentrations were only significantly correlated to the observations when using the
satellite-derived approach. At a regional scale, the satellite-derived diatom concentra-
tions were always within 10% of in situ data except in the Antarctic (underestimate
of 26%). Except for high latitudes (> 40° N or S), satellite-derived cyanobacteria were
underestimated in all regions with an underestimate > 14% in the North Central Pacific, North and Equatorial Indian and Equatorial Atlantic where the satellite-derived approach underestimated cyanobacteria by 23–30%. The diatoms and cyanobacteria concentrations from the model also mostly agreed (< 25% difference) with in situ data except for Antarctic and North Pacific where diatoms were overestimated by ~ 60% and in the North and Equatorial Indian where the model underestimated cyanobacteria by ~ 30–50%.

Using either approach, the coccolithophore concentrations were always in good agreement (within ~ 30%) with in situ data. Chlorophytes were the group for which both approaches differed most from in situ data. Using the model, chlorophyte concentrations were underestimated in the North Pacific (~ 73%) and overestimated in the Equatorial Indian (76%). The chlorophytes from the model were overall closer to in situ data than those from the satellite-derived approach. Although differences of ~ 57% were found in the North Pacific, in all the other regions the satellite-derived chlorophyte concentrations were within ~ 34% of in situ data.

3.2 Comparison of global phytoplankton community distributions between the satellite-derived algorithm and the model

3.2.1 Diatoms

Using the model, high diatom concentrations were spatially more widespread than using the satellite-derived approach (Fig. 3). This was reflected in the overall higher diatom concentrations from the model compared to those from the satellite-derived approach. For example, the North Pacific was the region where the model identified the highest diatom concentration (0.33 mg m\(^{-3}\)) whereas the satellite-derived average concentration was 0.15 mg m\(^{-3}\). Although not as productive (in terms of chlorophyll concentration) as the northernmost latitudes, Antarctic was also a region of abundant diatoms and where both approaches differed. Here, diatom concentration from the model was 0.19 mg m\(^{-3}\) compared to 0.03 mg m\(^{-3}\) using the satellite-derived approach.
3.2.2 Cyanobacteria

Both approaches agreed on the overall distribution of cyanobacteria, except for high latitude and upwelling regions. For example, the satellite-derived approach identified the North Atlantic and Pacific as the regions with highest cyanobacteria concentrations (\( \sim 0.05 \text{ mg m}^{-3} \), Figs. 2, 3) whereas the model detected low cyanobacteria concentrations (< 0.01 mg m\(^{-3}\)) in these two regions. This was also the case for the Antarctic where satellite-derived cyanobacteria concentrations were of \( \sim 0.04 \text{ mg m}^{-3} \), while the model did not identify the presence of any cyanobacteria in this region. Using the model, cyanobacteria were most abundant in the North Central Atlantic and Equatorial Indian although their distribution was relatively homogeneous among regions compared to the other phytoplankton groups.

3.2.3 Chlorophytes and coccolithophores

The satellite-derived approach identified the North Pacific and North Atlantic as the regions with most, and equally abundant, chlorophyte and coccolithophore concentrations. Although the coccolithophores and chlorophytes from the model were abundant in the North Atlantic, they were almost absent in the North Pacific. Despite this, both approaches were always within 0.11 mg m\(^{-3}\) of each other for both coccolithophores and chlorophytes. The largest difference in coccolithophores was observed in the North Pacific (0.11 mg m\(^{-3}\), Figs. 2, 3) and the North Indian (0.11 mg m\(^{-3}\)) where the coccolithophores from the satellite-derived approach was higher than that of the model.

3.3 Seasonal phytoplankton community variation

3.3.1 Diatoms

Seasonal variation of diatoms from the model was significantly (\( p < 0.05 \)) correlated with those derived from satellite in 11 out of the 12 regions (Fig. 4). Both approaches identified the presence of a spring bloom in the North Atlantic and Pacific followed by
a fall bloom in the North Pacific. In the North Central Pacific and Atlantic, there was also a spring bloom although the chlorophyll maximum occurred one month earlier and persisted for longer than in the North Atlantic and Pacific. In the North Indian region, both approaches detected an increase in diatoms during the monsoon season (~ August). In the Equatorial regions, both approaches were significantly correlated in the Equatorial Atlantic and Indian. In both these regions, diatom concentrations reached a maximum in boreal summer that was followed by a smaller winter increase in diatom concentrations. In the southern temperate regions (10° S–40° S), the diatom concentrations from the model all presented the same pattern: increase of diatom concentrations from austral autumn until spring when they reached a maximum. This seasonal cycle could also be distinguished using the satellite-derived approach although the magnitude over which diatom concentrations varied was smaller than in the model. In the South Atlantic, for example, the satellite-derived diatom concentrations reached in spring were 2–5 times lower than that from the model. In the Antarctic, both approaches identified the existence of an increase in diatom concentrations during the austral summer in the Antarctic. Here too, the model suggested diatom concentrations that were 3–4 times greater than that of the satellite-derived approach.

### 3.3.2 Cyanobacteria

The cyanobacteria concentrations derived from the two approaches were only significantly correlated in 3 out of the 12 regions: Equatorial Atlantic, North and South Indian (Fig. 5). In the subpolar regions (> 40° N or < 40° S), the seasonal variation in satellite-derived cyanobacteria resembled that of diatoms, but cyanobacteria concentrations from the model was very low all year round (< 0.01 mg m⁻³). In the northern temperate and equatorial regions (10° N–40° N), monthly cyanobacteria concentrations from both approaches were significantly correlated. Although the seasonal variation of cyanobacteria from both approaches was significantly (p < 0.05) correlated in the North Indian, the satellite-derived approach identified maximum concentration in March whereas the concentration of cyanobacteria from the model peaked in August. In the Equatorial
Atlantic, cyanobacteria concentrations increased in summer. In the Southern Hemisphere, both approaches were only significantly correlated in the South Indian Ocean where cyanobacteria reached a maximum during the austral winter and a minimum in summer. The absence of cyanobacteria in Antarctic in the model did not allow for a correlation analysis.

### 3.3.3 Coccolithophores

The seasonal variation of coccolithophore concentrations from the model correlated with those from the satellite-derived approach in 9 out of the 12 regions (Fig. 6). The 3 regions where there was no significant correlation were: Antarctic, North Pacific and Equatorial Atlantic.

In the Antarctic and North Pacific, the lack of correlation was a result of the low coccolithophore concentrations in the model (< 0.02 mg m\(^{-3}\)) contrasted with higher satellite-derived concentrations 0.07–0.15 mg m\(^{-3}\). In the North Atlantic both approaches agreed that coccolithophore concentrations reached a maximum in spring. The seasonal variation of coccolithophore in the North Central Pacific and the North Central Atlantic resembled each other: higher concentrations between winter and spring than those during the second half of the year. In the Equatorial regions, the coccolithophore concentrations from the model were significantly correlated with the satellite-derived approach in the Equatorial Indian and Pacific. Similarly to diatoms, the southern temperate regions were all dominated by a clear seasonal cycle in coccolithophore concentrations for both approaches with a maximum in the austral winter-spring in the South Indian, South Pacific and South Atlantic. In the Antarctic, no correlation between the coccolithophore concentrations from the model could be found with those derived from satellite data.
3.3.4 Chlorophytes

The chlorophyte concentrations derived from the two approaches were only significantly correlated in two out of the 12 regions: South Indian and Equatorial Atlantic (Fig. 7) and the correlation in the South Indian was negative. In the Equatorial Atlantic, both approaches identified an increase in chlorophyte concentrations during the boreal summer followed by a weaker peak in winter. Here the chlorophyte concentrations from the model were always 0.05–0.07 mg m\(^{-3}\) greater than that from the satellite-derived approach.

4 Discussion

Global mean differences of all groups are within \(\sim 15\%\) of in situ data for both approaches except for satellite-derived chlorophytes. At a global scale, diatom and cyanobacteria concentrations from both approaches are significantly correlated with in situ data. These results reflect the good knowledge we have on the functions fulfilled by diatoms and cyanobacteria, the functional extremes. The lack of significance between satellite-derived coccolithophores and in situ data is most likely because the classification of prymnesiophytes in the satellite-derived approach is much broader than the coccolithophore group in the model. The lack of understanding on how chlorophytes functionally relate to other phytoplankton groups leads to the misrepresentation of the function of chlorophytes in the model.

Seasonally, both approaches agree on the variation of diatoms and coccolithophores but there are some discrepancies for chlorophytes and cyanobacteria. Although most of the difference is due to mischaracterization of these groups by model and satellite-derived methods, some of this difference may be due to model weighting during the assimilation and to data gap filling (Gregg and Casey, 2009). The satellite-derived approach is also directly linked to the seasonal variation in the satellite data coverage which may explain some of the differences observed between both approaches.
While the lack of agreement on the seasonal variation in phytoplankton composition derived from both approaches can indicate some weaknesses in the methods, the presence of significant correlation must be taken carefully. For example, in the North and Equatorial Indian, both approaches indicate that most phytoplankton groups reach a maximum in August. This summer maximum in the Equatorial Indian is most likely related to the monsoon cycle as observed in the subtropical waters of this region. In the Equatorial Indian, most of the increase in chlorophyll in summer occurs on the western side of the basin. In this region, satellites can be contaminated by the atmosphere (Gregg, 2002). For example, dust plumes accompanying the high winds of the southwest monsoon are known to influence the chlorophyll concentration in the Indian Ocean (Wang et al., 2005).

The group for which both approaches agree most on the seasonal variation and are close to in situ data in most regions is diatoms. Both in the model and the satellite-derived approach, diatoms tend to dominate in regions where nutrients are abundant (high latitudes, coastal and equatorial upwelling areas, and regions of strong monsoonal influences). While both approaches agree on the global distribution of diatoms and are significantly correlated to in situ data, high diatom concentrations in the satellite-derived approach are spatially less widespread than in the model. Most often the high satellite-derived diatom concentrations are located along the coasts whereas the model allows them to grow in a more widespread area (i.e. North Pacific, Antarctic, see Fig. 3). This may be a result of the excess of iron in the model in these High Nutrient Low Chlorophyll (HNLC) regions. This suggests that the parameters in the model may need to be adjusted, probably as a function of nutrient, since there is a relatively large overestimate of diatoms when compared to that of the in situ data in the North Pacific and Antarctic. Despite this difference both approaches agree on the overall spatial distribution of diatoms and indicate the presence of high diatom concentration in regions that have been previously characterized by the presence of diatom-rich blooms. Both approaches indicate that the previously, well accepted and reported spring bloom in the North Atlantic and Pacific and austral summer bloom in the Antarctic is made
of a relatively high proportion of diatoms. The mechanism driving the intense spring-summer bloom in the temperate and subpolar latitudes is well known (Sverdrup, 1953). The deepening of the mixed layer depth in winter allows for surface waters to be replenished with nutrients which in turns allows the phytoplankton to flourish in spring-summer. During these events, the dominance of diatoms in these regions has been previously reported. For example, Marañón et al. (2000) found that diatoms make up to 80% of the total phytoplankton carbon in the North Atlantic in May and was reduced in September–October which supports both the model and satellite-derived data.

Although cyanobacteria concentrations from both approaches are close to in situ data, there are some differences in the spatial distribution and seasonal variation. In the model cyanobacteria exhibit nearly opposite behavior to diatoms: presence in the central ocean basins and absence in the high latitudes and upwelling regions. This is the opposite of the satellite-derived approach: here the higher the diatom concentrations, the higher the cyanobacteria concentrations. This suggests that a temperature dependence in the algorithm may improve the estimate of cyanobacteria in the satellite-derived approach.

Similarly to diatoms, the seasonal variation of coccolithophores is well represented by both the model and satellite-derived approach. The vast blooms of coccolithophores observed in the North Atlantic for example (e.g. Okada and McIntyre, 1979; Robertson et al., 1994; Boyd et al., 1997; Balestra et al., 2004) are well represented by both the model and the satellite-derived approach. Using calcite as a representation of coccolithophore abundance, Gregg and Casey (2007) qualitatively compare the model-derived coccolithophores distribution to that from Balch et al. (2005). In both datasets, coccolithophores were abundant in the Southern Ocean transition region, around 40° S in the Atlantic and Indian basins. Further south in the extremes of the Southern Ocean however, coccolithophores are abundant using the satellite-derived approach whereas they are absent in the model. The data on the existence of coccolithophores south of 50° S are contradictory. While some have reported the existence of coccolithophores in this region (i.e. Wright and van den Enden, 2000; Winter et al., 1999), other do not
find any coccolithophores south of 53° S (Wright et al., 1996). Another region that indicates some deficiencies in the model is the North Pacific where the model detects very low coccolithophore concentrations. Coccolithophores in the North Pacific can be found in substantial amount (Okada and Honjo, 1973; Lam et al., 2001; Crawford et al., 2003) and satellite algorithms have also estimated high calcite production in this region (Balch et al., 2005). The region where the model is furthest from the in situ data is in the Equatorial Pacific. Although it has been reported that coccolithophores are present in the western Equatorial Pacific up to 35%, (Okada and Honjo, 1973; Hagino et al., 2000), some other investigators (DiTullio et al., 2003; Ishizaka et al., 1997) reported low to negligible relative abundance of coccolithophores in this area. The abundance of coccolithophores has also been reported to be dependent on climate variability with lower abundance during El Niño events compared to non El Niño events in the western Pacific (Ishizaka et al., 1997).

For both the model and satellite-derived approach, chlorophyte is the group where there is the largest difference with in situ data. Particularly striking is the underestimate by the model of chlorophytes in the North Pacific (by −72%) and the overestimate in the Equatorial Indian (by ~75%). These two relative differences are each based on only 1 or 2 months when data are available which may explain the discrepancies. Furthermore the comparison of model and satellite-derived phytoplankton composition with in situ data can be complicated because in situ data contain themselves some level of uncertainties that is generally not reported (e.g. sky and sea conditions, standard error among replicates, interannual variation, local oceanographic phenomena, etc). Another uncertainty arises because in situ data are collected at specific points whereas models describe the average over an area of several kilometres (i.e. representation error).

Another reason for the difference in chlorophyte concentrations with in situ data and between approaches could be related to the variety of phytoplankton included as chlorophytes in the model. In the model, chlorophytes are a transitional group and represent a wide range of phytoplankton, such as flagellates, *Phaeocystis* spp., etc.
This expectation is probably unrealistic, and probably accounts for the lack of statistical significance in their relative abundances. An improvement could be achieved by adding an additional group that would represent specifically *Phaeocystis* spp. A good example of a region where the inclusion of *Phaeocystis* spp. as chlorophytes causes a problem is the Antarctic. Here, the model underestimates chlorophytes because of the low temperatures that don’t allow chlorophytes to grow. This is a deficiency in the model since it is reported that some species such as *Phaeocystis antarctica* are able to grow in low temperature areas (e.g. Smith et al., 1998).

### 5 Concluding remarks

A comprehensive global comparison of a data assimilating biogeochemical model and a satellite-derived approach with extensive in situ archives indicate an overall good agreement of the spatial distribution and seasonal variation in phytoplankton community composition. Some regions such as the Antarctic and North Pacific suggest further improvements of both approaches to improve our estimates especially for chlorophytes. The departure from in situ data and the disagreement between both approaches for chlorophytes is most likely due to the inclusion of *Phaeocystis* spp. as chlorophytes in the model and observations and the lack of understanding on the specifics of how they relate to other transitional groups. We suggest adding a group in the model that would represent *Phaeocystis* sp. specifically to improve these estimates. The satellite-derived approach could be improved by adding a temperature dependency to the estimate of cyanobacteria to reduce their abundance in regions of high chlorophyll such as the Antarctic where observations suggest that this group is only present in small abundance. Finally, assessment of global phytoplankton composition could be improved by assimilating the satellite-derived phytoplankton composition in the model and by using the model to further improve the satellite-derived approach. Future satellite ocean color missions may enhance our capacity to distinguish phytoplankton composition at a global scale by increasing the number of wavelengths to allow the differentiation of...
an increasing number of phytoplankton groups. This is especially true for the intermediate, transitional groups between the functional extremes (diatoms and cyanobacteria), such as the chlorophytes and non-coccolithophore prymnesiophytes, which are the most problematic for both approaches.

**Acknowledgements.** We thank the NASA SeaWiFS project for providing the satellite chlorophyll data and the NASA Center for Climate Simulation for computational support. Acknowledgments also go to MARine Ecosystem Model Intercomparison Project (MAREMIP) for promoting the collaboration that resulted in this paper and bringing scientists interested in phytoplankton biodiversity together. This paper was funded by the NASA EOS and MAP Programs.

**References**


Smith, W. O., Carlson, C. A., Ducklow, H. W., and Hansell, D. A.: Growth dynamics of phaeo-
cystis antarctica-dominated plankton assemblages from the ross sea, Mar. Ecol. Prog. Ser.,
168, 229–244, 1998.
Sverdrup, H. U.: On conditions for the vernal blooming of phytoplankton, ICES J. Mar. Sci., 18,
Uitz, J., Claustre, H., Morel, A., and Hooker, S. B.: Vertical distribution of phytoplankton com-
munities in open ocean: an assessment based on surface chlorophyll, J. Geophys. Res, 111,
Winter, A., Elbrachter, M., and Krause, G.: Subtropical coccolithophores in the weddell sea,
Wright, S. W. and van den Enden, R. L.: Phytoplankton community structure and stocks in
the east antarctic marginal ice zone (broke survey, January–March 1996) determined by
Wright, S. W., Thomas, D. P., Marchant, H. J., Higgins, H. W., Mackey, M. D., and Mackey, D. J.: 
Analysis of phytoplankton of the australian sector of the southern ocean: comparisons of mi-
croscopy and size frequency data with interpretations of pigment hplc data using the “chem-
Yoder, J. A. and Kennelly, M. A.: What have we learned about ocean variability from satellite
Table 1. Percent difference in relative abundance between the model (NOBM) or the satellite-derived approach and in situ data. Differences greater than 20% are in bold. The column denominated # represents the number of months when observations were present. The South Indian region is not included in this table since there was no in situ data available for comparison in this region.

<table>
<thead>
<tr>
<th></th>
<th>Diatoms</th>
<th>Cyanobacteria</th>
<th>Coccolithophores</th>
<th>Chlorophytes</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Satellite</td>
<td>NOBM</td>
<td>#</td>
<td>Satellite</td>
</tr>
<tr>
<td>Global</td>
<td>−2.1</td>
<td>15.0</td>
<td>64</td>
<td>−12.1</td>
</tr>
<tr>
<td>r</td>
<td>0.44</td>
<td>0.77</td>
<td>0.75</td>
<td>0.81</td>
</tr>
<tr>
<td>p</td>
<td>&lt; 0.001</td>
<td>&lt; 0.001</td>
<td>&lt; 0.001</td>
<td>&lt; 0.001</td>
</tr>
<tr>
<td>NAtl</td>
<td>3.2</td>
<td>−5.3</td>
<td>5</td>
<td>0.6</td>
</tr>
<tr>
<td>NPac</td>
<td>−7.7</td>
<td>59.7</td>
<td>12</td>
<td>8.1</td>
</tr>
<tr>
<td>N CAtl</td>
<td>−0.4</td>
<td>−2.6</td>
<td>12</td>
<td>−12.0</td>
</tr>
<tr>
<td>NCPac</td>
<td>−2.9</td>
<td>−3.4</td>
<td>3</td>
<td>−25.4</td>
</tr>
<tr>
<td>NInd</td>
<td>8.3</td>
<td>20.0</td>
<td>5</td>
<td>−22.8</td>
</tr>
<tr>
<td>EAtl</td>
<td>−0.9</td>
<td>−2.0</td>
<td>3</td>
<td>−31.5</td>
</tr>
<tr>
<td>EPac</td>
<td>−7.8</td>
<td>1.6</td>
<td>6</td>
<td>−13.8</td>
</tr>
<tr>
<td>Elnd</td>
<td>0.2</td>
<td>−3.8</td>
<td>2</td>
<td>−30.9</td>
</tr>
<tr>
<td>SATl</td>
<td>9.8</td>
<td>13.4</td>
<td>3</td>
<td>−14.1</td>
</tr>
<tr>
<td>SPac</td>
<td>1.6</td>
<td>25.5</td>
<td>3</td>
<td>−5.4</td>
</tr>
<tr>
<td>Ant</td>
<td>−26.2</td>
<td>61.6</td>
<td>10</td>
<td>13.4</td>
</tr>
</tbody>
</table>
Fig. 1. Location of observations of phytoplankton group relative abundances with basin definitions superimposed. The annotated and referenced data set is available on the NASA Global Modeling and Assimilation Office web site.
Fig. 2. Global and regional area-weighted climatological (1998–2007) concentration for (a) diatom, (b) cyanobacteria, (c) coccolithophores and (d) chlorophytes expressed in mg chl a m$^{-3}$. Regions: North Atlantic (NATL), North Pacific (NPAC), North central Pacific (NCPAC), North Indian (NIIND), Equatorial Atlantic (EATL), Equatorial Indian (EIND), South Atlantic (SATL), South Pacific (SPAC), South Indian (SIND) and Antarctic (Ant).
Fig. 3. Climatology (1998–2007) of the spatial distribution of phytoplankton community using the satellite-derived approach and using the NASA Ocean Biogeochemical Model for cyanobacteria, diatoms, coccolithophores and chlorophytes. Note that the phytoplankton groups of Hirata et al. (2011) do not map directly onto the classifications of NOBM. Satellite-derived prokaryotes are compared to cyanobacteria from the model, similarly for prymnesiophytes (satellite) and coccolithophores (model) and for green algae (satellite) and chlorophytes (model).
Fig. 4. Area-weighted seasonal variation of diatoms using monthly average from 1998–2007 (in units of mg chl $a m^{-3}$) for 12 regions as defined in Fig. 1. Below each subplot are the results of the correlation analysis between the model and the satellite-based approach. The monthly standard errors for 1998–2007 are represented. Regions abbreviations as in Fig. 2.
Fig. 5. Area-weighted seasonal variation of cyanobacteria using monthly average from 1998–2007 (in units of mg chl a m⁻³) for 12 regions as defined in Fig. 1. Below each subplot are the results of the correlation analysis between the model and the satellite-based approach. The monthly standard errors for 1998–2007 are represented. Regions abbreviations as in Fig. 2.
Fig. 6. Area-weighted seasonal variation of coccolithophores using monthly average from 1998–2007 (in units of mg chl a m⁻³) for 12 regions as defined in Fig. 1. Below each subplot are the results of the correlation analysis between the model and the satellite-based approach. The monthly standard errors for 1998–2007 are represented. Regions abbreviations as in Fig. 2.
Fig. 7. Area-weighted seasonal variation of chlorophytes using monthly average from 1998–2007 (in units of mg chl a m$^{-3}$) for 12 regions as defined in Fig. 1. Below each subplot are the results of the correlation analysis between the model and the satellite-based approach. The monthly standard errors for 1998–2007 are represented. Regions abbreviations as in Fig. 2.