Impact of circulation on export production, dissolved organic matter, and dissolved oxygen in the ocean: Results from Phase II of the Ocean Carbon-cycle Model Intercomparison Project (OCMIP-2)


Received 10 October 2006; revised 7 May 2007; accepted 22 May 2007; published 8 August 2007.

[1] Results are presented of export production, dissolved organic matter (DOM) and dissolved oxygen simulated by 12 global ocean models participating in the second phase of the Ocean Carbon-cycle Model Intercomparison Project. A common, simple biogeochemical model is utilized in different coarse-resolution ocean circulation models. The model mean (±1σ) downward flux of organic matter across 75 m depth is 17 ± 6 Pg C yr⁻¹. Model means of globally averaged particle export, the fraction of total export in dissolved form, surface semilabile dissolved organic carbon (DOC), and seasonal net outgassing (SNO) of oxygen are in good agreement with observation-based estimates, but particle export and surface DOC are too high in the tropics. There is a high sensitivity of the results to circulation, as evidenced by (1) the correlation of surface DOC and export with circulation metrics, including chlorofluorocarbon inventory and deep-ocean radiocarbon, (2) very large intermodel differences in Southern Ocean export, and (3) greater export production, fraction of export as DOM, and SNO in models with explicit mixed layer physics. However, deep-ocean oxygen, which varies widely among the models, is poorly correlated with other model indices. Cross-model means of several biogeochemical metrics show better agreement with observation-based estimates when restricted to those models that best simulate deep-ocean radiocarbon. Overall, the results emphasize the importance of physical processes in marine biogeochemical modeling and suggest that the development of circulation models can be accelerated by evaluating them with marine biogeochemical metrics.

1. Introduction

The export of organic matter from surface waters to the deep ocean, known as “export production,” is a central process in marine biogeochemical cycling. This flux, which occurs via gravitational settling, ocean circulation and active transport by organisms, maintains vertical gradients in a myriad of chemical species in the sea. It also helps regulate the abundance of a number of key gases in the atmosphere, including carbon dioxide and nitrous oxide. Export production directly influences the oxygen content of the ocean and indirectly (through the influence on organic matter burial) the oxygen content of the atmosphere, and therefore is key for understanding the Earth’s redox chemistry. Seasonal variations in atmospheric oxygen are also strongly driven by export production [Keeling and Shertz, 1992]. Export production is generally equated with “new production,” which is that fraction of primary production fueled by nutrients external to the euphotic zone [Dugdale and Goering, 1967], though the equality breaks down under nonsteady state conditions and when lateral transport is significant [Plattner et al., 2005]. Finally, export production represents the input of energy and organic matter to aphotic zone ecosystems, including the benthos.

Export production is difficult to quantify on a global scale because of its high spatial and temporal variability and because accurate direct measurements have been elusive. There are many methods for estimating export production on basin-wide scales including: extrapolation of direct measurements of the sinking flux based on sediment trap measurements [Martin et al., 1987] or new production based on $^{15}$N uptake [Eppley and Peterson, 1979], sometimes in combination with satellite measurements [Laws et al., 2000; Gnanadesikan et al., 2004; Dunne et al., 2005]; tracer-based estimates, which utilize mass balances for nutrients [Chavez and Toggweiler, 1995; Louanchi and Najjar, 2000], oxygen [Keeling and Shertz, 1992; Najjar and Keeling, 2000] or dissolved inorganic carbon [Lee, 2001]; and numerical models that simulate or invert the large-scale distributions of biogenic tracers [e.g., Bacastow and Maier-Reimer, 1990; Najjar et al., 1992; Gnanadesikan et al., 2004; Schlitzer, 2002].

Estimates of global export production based on these methods vary considerably, from about 3 to 20 Pg C yr$^{-1}$. Some of this spread reflects different quantities that are being estimated by different methods. For example, export estimates based on sediment traps may be referenced to a fixed depth, such as 150 m, whereas those based on primary production and $f$ ratios may integrate over the euphotic zone, whose depth varies in space and time. Model sensitivity studies by Doney et al. [2003] show a strong dependence on the depth across which export is computed (nearly a factor of 2 difference between 75 and 150 m), owing to the rapid remineralization of organic matter in the upper thermocline. Methods based on seasonal variations in tracers focus on a particular time of year (typically spring and summer) and region (typically middle and high latitudes). However, even within a given method estimates vary considerably. As we will show here, numerical forward model estimates of export across 75 m, using the same biogeochemical module but different circulation models, vary from 9 to 28 Pg C yr$^{-1}$. This is of concern because numerical models are the only means of predicting future variations in the marine carbon cycle and hence atmospheric CO$_2$ and climate.

The purpose of this paper is to explore the sensitivity of export production to the circulation generated by ocean circulation models and to evaluate the degree to which model biases produce erroneous results. These biases are uncovered by evaluating model skill at simulating the distribution of export production and tracers related to it, such as dissolved organic carbon, dissolved oxygen and tracers of ocean ventilation (the process that brings surface water and its properties into the ocean’s interior). By conducting such an evaluation, we are emphasizing that it is important for marine biogeochemical models to be able to simulate the distributions of relevant tracers and fluxes. In essence, we are conducting a joint assessment of ocean circulation models, a simple biogeochemical model, and an incomplete observational record. Though our focus is on the sensitivity of export to physical processes, we recognize the importance of biogeochemical factors that are not considered here, including: micronutrient limitation, silicic acid availability, nitrogen fixation, the microbial loop, and grazing [e.g., Sarmiento and Gruber, 2006, chapter 5 and references therein].

The work presented here was conducted as part of Phase II of the Ocean Carbon-cycle Model Intercomparison Project (OCMIP-2). A variety of simulations were conducted with carefully designed protocols to allow for a straightforward evaluation of the models with observations and comparison among models. In addition to model physical processes [Doney et al., 2004], simulations were generated for chlorofluorocarbons (CFCs) [Dutay et al., 2002], anthropogenic CO$_2$ [Watson and Orr, 2003; Orr et al., 2005], radiocarbon [Matsumoto et al., 2004], mantle helium [Dutay et al., 2004] and the natural marine carbon cycle. Of the latter, the nutrient and oxygen component is presented here. Watson and Orr [2003] also present aspects of the natural marine carbon cycle, with a focus on air-sea CO$_2$ fluxes. An overall summary of OCMIP-2 model results is described in the report of Orr [2002].

2. Methods

OCMIP-2 was designed mainly to evaluate the impact of ocean circulation processes on the marine carbon cycle. To facilitate such a comparison among models, a common biogeochemical model was designed. The model includes five prognostic variables: inorganic phosphate, semilabile dissolved organic phosphorus (DOP), dissolved oxygen (O$_2$), dissolved inorganic carbon and alkalinity. Transport was generated by the individual participating models in OCMIP-2. Of the 13 OCMIP-2 modeling groups, 12 ran the simple biogeochemical model. Table 1 lists the 12 groups and the characteristics of the physical model used by each group. Though all are coarse-resolution, non-eddy-resolving models, they differ considerably in the choice of sub-grid-scale parameterizations, sea-ice components and surface forcing. The coarse resolution of the models and the data sets available for evaluating them precludes any evaluation
of the models’ phosphate and oxygen dynamics in the coastal ocean.

A detailed description of the phosphorus-oxygen model and justification for parameter values has not been previously published but was made available through the internet [Najjar and Orr, 1999]. We present the complete model description in Appendix A and give a brief summary of it here. Production of organic matter is simulated by restoring model phosphate toward an observation-based monthly mean climatology of nutrients in the upper 75 m [Louanchi and Najjar, 2000] with a timescale of 30 days. Two thirds of the production goes to DOP with a remineralization lifetime of 6 months, and the remainder is remineralized instantly in the water column below (between 75 m and the bottom) following a power law function with depth, reflecting the rapid sinking of particulate organic matter (POM). The three model parameters for organic matter cycling (fraction of production that is dissolved, the DOP restoring technique, and the climatology that was used). Thus model errors are not only due to the nutrient-restoring technique, but also the climatology that was used. This evaluation points to a deficiency of the nutrient-restoring technique, particularly on the seasonal timescale [Gnanadesikan et al., 2002; Orr, 2002]. We explored the potential artifacts of nutrient restoring by comparing model surface phosphate fields with the more recent climatological fields of Conkright et al. [2002], which includes many more phosphate measurements and is presumably closer to the true climatology. Most correlation coefficients are greater than 0.9 for the annual mean and less than 0.4 after removing the annual mean, revealing the difficulty in capturing the annual cycle. This is even the case (though less so), when comparing model surface phosphate to the Louanchi and Najjar [2000] climatology, which it is restored to, pointing to a deficiency of the nutrient restoring method, at least with the 30-day restoring timescale used here. Thus model errors are not only due to the nutrient-restoring technique, but also the climatology that was used. This evaluation points to the poor representation of biogeochemical processes on seasonal timescales in the OCMIP-2

3. Results and Discussion

3.1. Effects of Nutrient Restoring

Previous studies using the OCMIP-2 biogeochemical model have suggested that significant errors arise as a result of the nutrient-restoring technique, particularly on the seasonal timescale [Gnanadesikan et al., 2002; Orr, 2002]. We explored the potential artifacts of nutrient restoring by comparing model surface phosphate fields with the more recent climatological fields of Conkright et al. [2002], which includes many more phosphate measurements and is presumably closer to the true climatology. Most correlation coefficients are greater than 0.9 for the annual mean and less than 0.4 after removing the annual mean, revealing the difficulty in capturing the annual cycle. This is even the case (though less so), when comparing model surface phosphate to the Louanchi and Najjar [2000] climatology, which it is restored to, pointing to a deficiency of the nutrient restoring method, at least with the 30-day restoring timescale used here. Thus model errors are not only due to the nutrient-restoring technique, but also the climatology that was used. This evaluation points to the poor representation of biogeochemical processes on seasonal timescales in the OCMIP-2
models. We have therefore chosen to focus most of our attention on the annual mean distribution of tracers, which we believe to be more robust.

3.2. Export Production

[10] Global export production across 11 of the OCMIP-2 models (SOC only reported particle export) varies from 9 to 28 Pg C yr$^{-1}$, with a cross model mean ($\pm$ 1 standard deviation, 1$s$) of 17 $\pm$ 6 Pg C yr$^{-1}$ (Figure 1). This is somewhat higher than the satellite-based estimate of new production by Laws et al. [2000] (12 $\pm$ 0.9 Pg C yr$^{-1}$) and the inverse modeling estimate of export (across 133 m) by Schlitzer [2002] (12.7 Pg C yr$^{-1}$). Similarly, mean ($\pm$1$s$) particle export across 75 m estimated by the models is 13 $\pm$ 3 Pg C yr$^{-1}$, in reasonable agreement with satellite-based estimates (9.7–12 Pg C yr$^{-1}$ [Gnanadesikan et al., 2004]) and inverse estimates (9.8 Pg C yr$^{-1}$ [Schlitzer, 2002]). The slight bias of the models could reflect the depth (75 m) chosen for computing model export production [Doney et al., 2003]. The data-based estimates also have significant uncertainties. For example, error is introduced to the satellite-based export estimates in two steps: first in the primary production algorithm, which shows particularly large regional differences that are masked in the global mean [Gnanadesikan et al., 2002, 2004], and second in the conversion of primary production to export using an export ratio that typically depends on temperature and primary production [Laws et al., 2000; Dunne et al., 2005]. The inverse modeling results of Schlitzer [2002] may have a bias because of the lack of seasonality.

[11] To make the data-based and modeled estimates of export most comparable, they should be brought to a common depth. However, it is not clear how to do this for at least one of two reasons: (1) the depth of the data-based estimate is not known and (2) the appropriate depth correction is not known. To estimate POC export, Gnanadesikan et al. [2004] exploited satellite algorithms for primary production and observation-based estimates of the ratio of export production to primary production [Dunne et al., 2005]. These ratios were not based on a common depth and furthermore were derived from a wide variety of methods, many of which (such as the $^{15}$N technique) cannot be depth-adjusted in any objective way. Schlitzer’s [2002] inverse model gives export at 133 m, the base of the second grid box. We consider two equally defensible ways to bring the Schlitzer [2002] flux and the model-mean flux to the same depth. In the first method, we bring the 75-m model-mean particle flux (13 Pg C yr$^{-1}$) to 133 m by using the assumed particle flux scaling (see Appendix A). This yields a model-mean POC flux of 7.8 Pg C yr$^{-1}$ at 133 m, which is 20% less than that of Schlitzer [2002]. In the second method, we bring the POC flux at 133 m from Schlitzer [2002] to 75 m. In this model, 75% of the export comes from the top box, which is 60 m thick, and 25% comes from the second box. Assuming the flux varies linearly between 60 and 133 m, the flux at 75 m would be 7.9 Pg C yr$^{-1}$; in this case the model-mean flux is more than 60% greater than that of Schlitzer [2002]. The corrections give inconsistent results because the flux increases with depth in the first method and decreases with depth in the second method. Overall, given the difficulties and uncertainties in depth corrections to the observation-based estimates of POC flux, we have chosen not to apply them.

[12] The differences in export production among the models are associated with differences in the parameterizations of lateral diffusive mixing and mixed layer processes (refer to Table 1 and Figure 1). Only four of the 11 models reporting total production have lateral mixing along the horizontal and yet three of these (UL, PIUB and IGCR) are among the five most productive models. Hori-
Horizontal mixing effectively increases cross-isopycnal mixing, which has been shown to be a strong determinant of export production [Gnanadesikan et al., 2002, 2004]. Only three of the 11 models reporting total production have explicit treatments of the mixed layer, and all of them (UL, NCAR and IPSL) are among the four most productive models. Of the four models that have isopycnal mixing using the parameterization of Gent and McWilliams [1990] and no explicit treatment of the mixed layer, three (PRIN, LLNL and MPIM) are the least productive.

Global total, dissolved and (especially) particulate export production are correlated with the radiocarbon content of the deep sea (Figure 2 and Table 2). The three models with the highest radiocarbon content of Circumpolar Deep Water (CDW) [Matsumoto et al., 2004] are also the three most productive (UL, IPSL and PIUB), while the two with the lowest radiocarbon are the least productive (PRIN and LLNL). The correlation coefficient between export production and CDW $^{14}$C is 0.77. Similar correlations are found with other ventilation measures: anthropogenic CO$_2$ uptake, CFC-11 inventory, and the radiocarbon content of North Pacific Deep Water (NPDW) (Figure 2, Table 2), similar to the finding of Gnanadesikan et al. [2002, 2004]. Why is $^{14}$C, a deep-ocean circulation tracer, a good predictor of export production? One could argue that it is the vigor of the upper ocean circulation, which is directly responsible for delivering nutrients to the euphotic zone, that is most relevant. However, export production also depends on the nutrient content of the thermocline. Figure 3 shows that the global mean nutrient profile differs substantially among the models. Models with weak deep-ocean ventilation, such as PRIN, accumulate nutrients in the deep ocean as a result of particle remineralization. This accumulation must come at the cost of nutrients elsewhere, and so the thermocline is drained of nutrients. A model at the other end of the spectrum, with vigorous deep ocean circulation and a relatively high thermocline phosphate content, is PIUB. The correlations between the $^{14}$C content of deep water (either NPDW or CDW) and the phosphate content of the upper km and below 2 km further demonstrate the relationship between deep-ocean ventilation and nutrients (Table 2).

Figure 4 shows the latitudinal variation of total and particulate export derived from the models and data-based methods in five latitude bands: southern high latitudes (90°S to 40°S), southern subtropics (40°S to 10°S), tropics (10°S to 10°N), northern subtropics (10°N to 40°N) and northern high latitudes (40°N to 90°N). Compared to the satellite-based estimates, particle export in the models is generally high in the tropics, low north of 40°N, and extremely...
variable south of 40°S. The OCMIP-2 models tend to agree more with the inverse estimates of Schlitzer [2002] than the satellite-based estimates, probably because the OCMIP-2 and inverse results have a similar basis (nutrient distributions combined with circulation models).

[15] Tropical export production varies from 2.7 to 9.3 Pg C yr⁻¹ among the models, with a mean of 5.1 Pg C yr⁻¹. This is considerably higher than the 2.3 Pg C yr⁻¹ estimated from inverse methods, and the 2.2 Pg C yr⁻¹ estimated by Chavez and Toggweiler [1995] on the basis of nitrate budgets for a larger tropical area (15°S to 15°N). Model-mean particle export in the tropics is 4.5 Pg C yr⁻¹, which exceeds satellite-based export estimates (1.6–2.8 Pg C yr⁻¹). The model estimates of tropical export production are likely too high because coarse-resolution ocean models tend to excessively trap nutrients in upwelling areas [Aumont et al., 1999]. This collection of models also tends to be too cold in equatorial surface waters, suggesting as well that either the upwelling rate is too high or upwelling is coming from too deep in the water column [Doney et al., 2004].

[16] To remove overall model biases and focus on the spatial variability of export production, Figure 5 shows the fraction of global carbon export (total and particulate) in the five latitude bands. As might be expected from the large-scale wind-driven circulation and distribution of mixed layer depth, the subtropics have lower-than-average export production (per unit area) and the tropics and northern high latitudes are higher than average in this regard. Satellite-based and inverse estimates of particle export show differences with this pattern: the tropics and northern subtropics have export per unit area roughly equal to the global mean. Surprisingly, for most models, the southern high latitudes have lower-than-average export production per unit area; the satellite-based and inverse estimates here are close to the global mean. Most striking, however, are the large differences among the models in this region, where the fraction of global export production varies from 2 to 33% of the global average. The coefficient of variation (standard deviation divided by the mean) of this fraction is 0.58 in southern high latitudes, whereas it varies between 0.10 and 0.21 elsewhere. This is consistent with the OCMIP-2 simulations of the uptake of CFCs and anthropogenic CO₂, which show the largest intermodel variation in the Southern Ocean [Dutay et al., 2002; Watson and Orr, 2003]. Inverse modeling studies of natural and anthropogenic air-sea CO₂

### Table 2. Pearson Correlation Coefficient Between Various Model Metrics

<table>
<thead>
<tr>
<th></th>
<th>&gt;2 km AOU</th>
<th>&gt;2 km Phosphate</th>
<th>0–1 km Phosphate</th>
<th>1994 CFC-11 Inventory</th>
<th>1990s Anthro. CO₂ Uptake</th>
<th>CDW ¹⁴C</th>
<th>NPDW ¹⁴C</th>
<th>0–50 m DOC</th>
<th>DOM Export Fraction</th>
<th>DOM Export</th>
<th>Particle Export</th>
</tr>
</thead>
<tbody>
<tr>
<td>Total export</td>
<td>0.36</td>
<td>−0.08</td>
<td>0.13</td>
<td>0.67</td>
<td>0.78</td>
<td>0.77</td>
<td>0.59</td>
<td>0.75</td>
<td>0.64</td>
<td>0.89</td>
<td>0.92</td>
</tr>
<tr>
<td>Particle export</td>
<td>0.42</td>
<td>−0.19</td>
<td>0.29</td>
<td>0.47</td>
<td>0.66</td>
<td>0.83</td>
<td>0.63</td>
<td>0.86</td>
<td>0.30</td>
<td>0.63</td>
<td></td>
</tr>
<tr>
<td>DOM export</td>
<td>0.27</td>
<td>0.10</td>
<td>−0.08</td>
<td>0.73</td>
<td>0.74</td>
<td>0.54</td>
<td>0.45</td>
<td>0.44</td>
<td>0.90</td>
<td></td>
<td></td>
</tr>
<tr>
<td>DOM fraction</td>
<td>0.32</td>
<td>0.37</td>
<td>−0.33</td>
<td>0.55</td>
<td>0.48</td>
<td>0.21</td>
<td>0.20</td>
<td>0.10</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0–50 m DOC</td>
<td>0.16</td>
<td>−0.35</td>
<td>0.56</td>
<td>0.52</td>
<td>0.73</td>
<td>0.91</td>
<td>0.71</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>NPDW ¹⁴C</td>
<td>0.02</td>
<td>−0.50</td>
<td>0.53</td>
<td>0.61</td>
<td>0.69</td>
<td>0.91</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CDW ¹⁴C</td>
<td>0.14</td>
<td>−0.36</td>
<td>0.57</td>
<td>0.65</td>
<td>0.83</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Anthro. CO₂</td>
<td>0.06</td>
<td>−0.44</td>
<td>0.32</td>
<td>0.87</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CFC inventory</td>
<td>−0.03</td>
<td>0.20</td>
<td>0.02</td>
<td>0.93</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0–1 km PO₄</td>
<td>−0.35</td>
<td>−0.93</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>&gt;2 km PO₄</td>
<td>0.46</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

*Correlations with absolute values between 0.5 and 0.7 are italicized, and those greater than 0.7 are boldface.

Figure 3. Global mean phosphate profiles from the models and OBS (observations) [Conkright et al., 2002].
fluxes using a variety of ocean models, some of which participated in OCMIP-2, also show large intermodel differences in the Southern Ocean [Mikaloff Fletcher et al., 2006, 2007].

The two models with the highest export fractions in the northern high latitudes, IGCR and PIUB, also have some of the deepest mixed layers in the North Atlantic [Doney et al., 2004]. IGCR, however, has deep Southern Ocean mixed layers and yet has a small fraction of export there. PIUB, IGCR and LLNL have very low export fractions in the Southern Ocean. Two of these models (PIUB and IGCR) have lateral mixing along horizontal surfaces. This may diffuse nutrients northward across steeply sloping isopyncals in the Southern Ocean, which would decrease productivity in the Southern Ocean and increase it where those nutrients are upwelled, in the tropics. This hypothesis is supported by the

Figure 4. Total (gray) and particle (black) carbon export in five latitude bands for the 12 OCMIP models (Table 1) and their mean, the inverse method of SO2 [Schlitzer, 2002], and satellite-based estimates (particle only) of GO4 [Gnanadesikan et al., 2004]. DOC export results were not available from SOC. For GO4, the three bars refer to usage of the primary production algorithms of Behrenfeld and Falkowski [1997] (left), Carr [2002] (middle), and Marra et al. [2003] (right).
large inverse correlation between export fractions in the southern high latitudes and the tropics (Table 3).

A second hypothesis for the low export in the southern high latitudes is strong tropical upwelling. For example, the PIUB model is known to have too much upwelling around the equator [Marchal et al., 1998], which, when combined with the rapid remineralization of particulate and dissolved organic matter, leads to strong nutrient trapping in the tropics. Phosphate is then too low elsewhere, including the southern high latitudes, which, in conjunction with the nutrient-restoring approach, gives rise to low export production in the southern high latitudes. This hypothesis is also consistent with the correlation of export fraction between the southern high latitudes and the tropics (Table 3).

There is also a substantial anticorrelation between export fractions in southern high latitudes and northern...
subtropics. This may be due to the same northward nutrient transport in the thermocline hypothesized above, followed by an upward nutrient flux due to vertical diffusion and wintertime convection. Using similar reasoning, one would expect an inverse relationship between export fractions in the southern high latitudes and the southern subtropics, but this is not observed. The southern subtropics are distinguished by their relatively constant export fractions among the models (Figure 5), with a mean (±1σ) of 23 ± 2.3%. The coefficient of variation of the export fraction in this region is 0.10, the lowest of the five regions considered, and may reflect a nutrient source that is similar in all of the models. A candidate is lateral advection due to Ekman transport from subpolar waters, which has been suggested for the subtropical North Atlantic [Williams and Follows, 1998]. The effect would be expected to be even larger in the Southern Hemisphere, where subpolar surface nutrient concentrations and wind stresses are higher. The constancy across models could be explained by the similarity in wind stress climatologies used by the models as well as the similar surface phosphate fields.

### 3.3. Dissolved Organic Matter

[29] Comparison of model DOP with observations is not straightforward because of the assumption of a C:P ratio close to the standard Redfield ratio in the production and decomposition of dissolved organic matter (DOM) in the OCMIP-2 models, which is in clear violation of observations that show strongly and systematically non-Redfield behavior of DOM. Karl and Björkman [2002] synthesized historical DOP measurements and found decreases in open-ocean DOP from the surface to 300 m of less than 0.1 μmol L⁻¹, when averaged over large areas or long time periods. Corresponding DOC decreases are typically 30 μmol L⁻¹ or more [Hansell, 2002]. Thus semilabile DOM has a C:P ratio of at least 300, far exceeding the standard Redfield ratio. Because the OCMIP-2 models were designed to capture export of carbon, and used fairly standard Redfield ratios (C:P = 117), model DOP is actually more of an analogue for bulk DOM (see Appendix A) as opposed to the DOP pool itself. We have therefore chosen to make comparisons with DOC observations. This also has the advantage of comparing with a larger database of higher-quality observations. The choice of conventional Redfield ratios in model DOM cycling undoubtedly produces unrealistic artifacts, which are discussed in section 4.

[21] We evaluate the models with the DOC data synthesis presented by Siegel et al. [2002], who developed correlations of surface (0–50 m) DOC concentration with winter sea surface temperature (SST). This algorithm is based on nearly 500 DOC measurements and has an RMS error of 2.6 to 5.1 μmol L⁻¹, depending on the region. Their algorithm applies to all surface waters except the Atlantic Ocean north of about 40°N. We used their algorithm in combination with winter SST from Stephens et al. [2002] to estimate surface ocean DOC and subtracted off an estimate of refractory DOC to arrive at the semilabile component. Hansell and Carlson [1998a] found deep-ocean DOC equatorward of 40° latitude (away from extremely old and extremely young deep waters) to vary between 39 and 43 μmol L⁻¹, so we chose 40 μmol L⁻¹ as the estimated refractory component [see also Anderson and Williams, 1999]. The global mean (excluding the North Atlantic north of about 40°N) surface semilabile DOC is estimated to be 29 ± 5 μmol L⁻¹, where the uncertainty is qualitatively based on uncertainties in the algorithm and in the refractory component. We converted model concentrations of DOP to semilabile DOC by multiplying by 117 [Anderson and Sarmiento, 1994], the elemental ratio used in the model.

[22] Figure 6 shows global mean profiles of modeled semilabile DOC along with an estimate of surface DOC based on the observations. Model surface (0–50 m) values vary from 18 to 36 μmol L⁻¹, with a mean of 29 μmol L⁻¹, equal to the observation-based estimate. Surface DOC in the models is highly correlated with particle export (Table 2 and Figure 7), which reflects the specification in the models that DOC production scales with particle export. At 300 m, the models have very low semilabile DOC, typically less than 1 μmol L⁻¹. The only observational synthesis at depth we are aware of is from the modeling study of Anderson and Williams [1999], who present observed (total) DOC values of about 50 μmol L⁻¹ at 300 m, though their model estimate is about 45 μmol L⁻¹. The variety of observations presented by Hansell [2002] show that the DOC vertical gradient all but disappears by about 300 m and typical values are between 40 and 50 μmol L⁻¹, suggesting a semilabile component of 0–10 μmol L⁻¹. Thus semilabile DOC in the models may be too low in thermocline. This could reflect a remineralization timescale for semilabile DOC that is too short in the models. Kwon and Primeau [2006] recently optimized the model parameters used here to fit the phosphate distribution, and found estimates of the DOC lifetime between 1 and 3 years, several times the value used here (0.5 years). Schlitzer [2002] also found, through trial and error, that model DOC was closest to observations with a DOC lifetime of 1 year. Our model misfit may also reflect the lack of multiple DOM pools with multiple lifetimes, or a dependence of the DOC lifetime on the physical and biological environment.

[23] Model DOM export across 75 m is highly variable, ranging from 2 to 12 Pg C yr⁻¹, with a mean (±1σ) of 4.5 ± 3 Pg C yr⁻¹; the coefficient of variation is 0.68, more than twice that of particle export. DOM export in the models depends on how much DOM is produced in the upper 75 m and how rapidly the DOM is removed from surface waters. The first is twice the particle export and the second depends on the vigor of the upper ocean circulation, which is why DOM export is positively correlated with particle export.

---

**Table 3. Pearson Correlation Coefficient Between Fraction of Global Export Production in Various Regions for the 11 Models Reporting Total Export**

<table>
<thead>
<tr>
<th>Region</th>
<th>90°S–40°S</th>
<th>40°S–10°S</th>
<th>10°S–10°N</th>
<th>10°N–40°N</th>
</tr>
</thead>
<tbody>
<tr>
<td>40°N–90°N</td>
<td>−0.23</td>
<td>−0.20</td>
<td>0.01</td>
<td>0.08</td>
</tr>
<tr>
<td>10°N–90°N</td>
<td>−0.81</td>
<td>0.34</td>
<td>0.44</td>
<td></td>
</tr>
<tr>
<td>10°S–10°N</td>
<td>−0.82</td>
<td>−0.28</td>
<td></td>
<td></td>
</tr>
<tr>
<td>40°S–10°S</td>
<td>−0.14</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

*See Figure 1.*
and indices of upper ocean circulation, such as CFC and anthropogenic CO$_2$ uptake (Table 2). The fraction of export production in dissolved form varies considerably among the models, from 8 to 43%, with a mean $\pm 1\sigma$ of 25 $\pm$ 8%. Observation-based estimates of the fraction of export production in dissolved form were made by Hansell and Carlson [1998b] from the ratio of DOC accumulation to the drawdown of dissolved inorganic carbon (either directly measured or estimated from nitrate). The global estimate of the DOC export fraction, based on extrapolation from three diverse sites, was 17%. Hansell [2002] revised this estimate to 20 $\pm$ 10% using DOC remineralization ratios based on DOC-oxygen relationships in the aphotic zone. The revised estimate is in reasonable agreement with the OCMIP-2 models, as only two models are outside of the observed range. Models that have higher export production tend to have a greater fraction of it in dissolved form, as indicated by the positive correlation between the two quantities (Table 2). This means that DOM export in the models is more sensitive to circulation than particle export. We suggest that this is because DOM export in the models is doubly sensitive to circulation: both DOM production (which depends on the rate of nutrient delivery to the surface ocean) and DOM removal (i.e., its lifetime in the surface ocean) depend on circulation. In contrast, particle export equals particle production in the models, so the lifetime of particles in the upper 75 m has no dependence on circulation. Three of the 11 models that report DOM export have explicit mixed layer physics (IPSL, NCAR and UL; Table 1), and these three have the highest DOM export fractions (Figure 7). This highlights the importance of seasonal convective overturning in DOM export in the models, a phenomenon that has been shown to be an

![Figure 6](image_url)

**Figure 6.** Global mean semilabile DOC profiles for the 12 models. Model DOC is computed from the model DOP concentration and a C:P ratio of 117. Also shown (black circle) is an estimate of global mean surface (0–50 m, excluding the North Atlantic, north of about 40°N) semilabile DOC based on the algorithm of Siegel et al. [2002], with a refractory component of 40 $\mu$mol L$^{-1}$ removed. See text for details.

![Figure 7](image_url)

**Figure 7.** (a) Global mean surface (0–50 m) semilabile DOC and (b) fraction of global export as DOM as a function of global particle export for the models (symbols) and observation-based estimates (bars). The observation-based estimate for DOC is the same as in Figure 6. The observation-based estimate of the DOC export fraction is from Hansell [2002].

![Figure 8](image_url)

**Figure 8.**
important mechanism for DOC export in the real ocean [e.g., Carlson et al., 1994; Hansell, 2002].

[24] Zonal average semilabile DOC is presented in Figure 8 for the models and the observation-based estimate. The abrupt change in the observation-based estimate at about 40°S, and to a lesser extent at 40°N, reflects the change in algorithm used to estimate DOC from SST. Observation-based semilabile DOC is highest in tropical waters and decreases with latitude. The model latitudinal pattern is different, with secondary maxima typically between 40° and 50° latitude in both hemispheres. This is only hinted at in the observation-based estimate and perhaps cannot be captured in simple algorithms based on SST. Pacific transects, in fact, do show modest minima at about 35°S [Hansell, 2002] and 35°N [Abell et al., 2000]. The model DOC distribution clearly reflects the latitudinal pattern of production, wherein DOC is produced in tropical and high-latitude regions and laterally transported toward the subtropics where it is consumed and subducted. The most glaring difference between the models and the observation-based estimate, however, is in the tropics, where the models are generally higher by about 50%. This may reflect export production that is too high in tropical regions (as discussed earlier in section 3.2), a DOM production fraction that is too high, or a DOM lifetime that is too long. The latter is consistent with the Equatorial Pacific modeling study of Archer et al. [1997], who found DOC lifetimes of 0.1–0.3 years, compared to the value of 0.5 years used here. The model differences along the equator seem to be related to the fraction of export production that occurs in the tropics. For example, the model with the highest fraction (PIUB) also has the highest DOC, whereas the model with the lowest fraction (AWI) is the third lowest in DOC. The models are also consistently too low in the

Figure 8. Zonal average semilabile DOC from the 12 models and that based on observations. The former are the model DOP fields (0–50 m mean) multiplied by a C:P ratio of 117 [Anderson and Sarmiento, 1994]. The latter does not include the North Atlantic north of about 40°N, and is based on the DOC algorithms of Siegel et al. [2002], with a refractory component of 40 μmol L⁻¹ subtracted off.
3.4. Oxygen

3.4.1. Seasonal Net Outgassing

The annual cycle in the air-sea flux of oxygen has often been used to estimate export production [Jenkins and Goldman, 1985; Emerson, 1987; Keeling and Shertz, 1992; Bender et al., 1996; Najjar and Keeling, 2000]. Oxygen is outgassed owing to net community production and heating in extratropical regions during the summer, when surface waters are stratified. The thermal portion of seasonal net outgassing (SNO) is due to the annual cycle in surface heat flux combined with the temperature dependence of oxygen solubility (decrease with warming). The biological portion of SNO is due to the annual cycle in vertical mixing combined with the seasonality of net community production. During the winter, oxygen-depleted water is entrained into the mixed layer, creating an oxygen demand on the atmosphere. The nutrients that are also entrained are subsequently consumed during the following spring and summer, resulting in the photosynthetic production of oxygen, some of which is outgassed. An important test for marine biogeochemical models is their ability to simulate the seasonality in the air-sea 02 flux. Here we present model results for SNO, which is defined for a given region as follows. First the spatially integrated flux is computed for each month of the year over the region. Then the annual mean flux for the region is subtracted from each month. Finally, SNO is computed as the sum of the fluxes over the months in which the flux is out of the ocean. 

Jin et al. [2007] have computed this diagnostic in a similar way. Figure 9 shows SNO in five regions for the 12 models and from the observation-based climatology of GK [Garcia and Keeling, 2001]. The two quantities are positively correlated (r = 0.83), supporting the contention that SNO is an export production proxy. Jin et al. [2007] recently used the SNO-export relationship in a global ocean ecosystem model along with models do not have seasonality (AWI and PIUB) and hence have zero SNO. The IGCR model has no seasonality in its physics, but imposed a seasonal variation in surface nutrients, and so has modest SNO. Consistent with the observation-based estimate, model SNO is smallest in the tropics and greatest at high southern latitudes. Model SNO for the fully seasonal models (i.e., all but AWI, PIUB and IGCR), when summed globally over the five regions, varies from 5.4 to 10.6 \times 10^{14} \text{mol O}_2, with a mean of 7.5 \times 10^{14} \text{mol O}_2. This is slightly less than the 8.5 \pm 0.7 \times 10^{14} \text{mol O}_2 estimated by Garcia and Keeling [2001]. One possible reason for this bias is the underestimation of seasonality in the model physics. Among the nine models with seasonal physics, the two with the largest SNO (SOC and UL) have explicit mixed layer dynamics while the three with the smallest SNO (PRIN, LLNL and CSIR) do not. The mean of the four models with explicit mixed layer dynamics is 8.6 \times 10^{14} \text{mol O}_2. The models also likely underestimate SNO because of the nutrient restoring model. As noted above (section 3.1), the model surface phosphate fields severely underestimate the true seasonality because of the climatology used and because of the restoring technique itself. This most likely introduces errors in the amplitude and phase of the annual cycle that lead to an underestimation of SNO.

Despite the shortcomings of the model SNO simulations, the models offer the opportunity to investigate the basic premise that SNO is related to export production [Keeling and Shertz, 1992; Bender et al., 1996; Najjar and Keeling, 2000; Garcia and Keeling, 2001]. We expect little correlation between SNO and tropical export production owing to the lack of seasonality in the tropics, so we regressed global SNO versus extratropical export production for the models with seasonal physics (except SOC). The two quantities are positively correlated (r = 0.83), supporting the contention that SNO is an export production proxy. Jin et al. [2007] recently used the SNO-export relationship in a global ocean ecosystem model along with...
observation-based estimates of SNO to arrive at a global ocean export at 76 m of 14.9 ± 2.5 Pg C yr⁻¹.

3.4.2. Deep-Ocean Oxygen Content

[28] We now shift our attention to the cycling of oxygen in deep waters. The models show widely different global mean profiles of oxygen (Figure 10a). The fact that the apparent oxygen utilization (AOU, Figure 10b) mirrors these profiles indicates that the major differences among the models are not due to solubility effects. Model AOU is, in general, much greater than observed. The mean AOU for waters deeper than 2 km varies by more than a factor of two among the models (106 to 242 μmol L⁻¹) but is generally too high, similar to Gnanadesikan et al. [2002, 2004]; the cross-model mean is 183 μmol L⁻¹, which can be compared with the observed value of 156 μmol L⁻¹ (computed from the atlas of Locarnini et al. [2002]). The large range of deep-ocean AOU is surprising in light of box model studies [e.g., Sarmiento et al., 1988] that suggest that the oxygen content of the deep ocean is regulated by the concentration of surface phosphate in regions of deep water formation (as well as the Redfield ratio and the phosphate inventory of the ocean), regardless of the rate of ocean circulation and export production. Because all of the OCMIP-2 models have essentially the same surface nutrient distribution, the box model concept fails to explain the GCMs, similar to the results of earlier GCM studies [Sarmiento and Orr, 1991; Gnanadesikan et al., 2004]. The box model could, however, provide an explanation for why the models, in general, have too much deep-ocean AOU. We suspect, owing to the few wintertime surface nutrient data available in high latitudes, that the climatology used for phosphate restoring [Louanchi and Najjar, 2000] is systematically too low during the winter. This could reduce deep ocean oxygen significantly. If, for example, the wintertime value of surface phosphate is underestimated by 0.2 μmol L⁻¹, this would translate into an oxygen underestimate of about 30 μmol L⁻¹, roughly equal to the mean underestimate of the models. This speculation is tempered by the fact that the nutrient-

Figure 10. Global mean profiles of (a) oxygen and (b) apparent oxygen utilization (AOU) from the models and observations. Observations are from Locarnini et al. [2002].
restoring technique produces a global mean bias of the opposite sign of about 0.1 μmol L⁻¹.

[29] Is there too much deep-ocean AOU in the models because the sinking particle flux to the deep ocean is too great? From an average of 24 sediment trap observations made between 2 and 3 km depth, Berelson et al. [2007] estimated a mean (±1σ) sinking particle flux of 0.60 ± 0.49 Pg C yr⁻¹. Measurements at these depths should not be biased significantly by trapping artifacts [Yu et al., 2001]. Using the expression for the particle sinking flux in the models (see Appendix A), the amount reaching 2.5 km is 4.3% of the amount exported from 75 m. With a mean particle export of 13 Pg C yr⁻¹, we get a mean 2.5-km sinking flux of 0.56 Pg C yr⁻¹, in good agreement with the observations. This result suggests that it is not the restoring scheme, remineralization scheme, or surface nutrient supply but rather the deep circulation that is the main cause of low deep-ocean oxygen in the models. To be conclusive, however, more sediment trap observations are needed, as most of the long-term deployments have been in the Northern Hemisphere [Francois et al., 2002].

[30] The correlations of deep-ocean AOU with other metrics are generally low (Table 2), the largest being with particle export and deep ocean phosphate. As particle export increases, AOU generally increases (Figure 11a and Table 2), which means that the increased oxygen demand caused by the increased export is larger than the oxygen increase associated with increased circulation, in contrast to findings by Gnanadesikan et al. [2004].

[31] The PRIN model is fundamentally distinct among the OCMIP-2 models in terms of its deep-ocean oxygen dynamics as shown in Figures 11a and 11b. In fact, if the PRIN model is eliminated, then correlations of AOU with export and deep ocean ¹⁴C increase dramatically (from 0.42 to 0.77 for particle export and from 0.14 to 0.71 for CDW ¹⁴C). The PRIN model may be an outlier because it is one of only two models (the other being SOC) among the OCMIP-2 model suite with low diapycnal diffusivity (0.15 cm² s⁻¹).

[32] Sarmiento et al. [1988] found no sensitivity of deep-ocean oxygen to ventilation because as ventilation increases in their box model, so does export production, which exactly counteracts the ventilation. In the OCMIP-2 model suite, there is some independence of deep-ocean ventilation and export because other processes influence the latter, like near-surface and mixed layer dynamics. Considering the deep ocean as a well-mixed box of volume $V$, the mean AOU below 2 km should be related to the POC sinking flux at 2 km ($F_{2km}$) and the ventilation time of the water below 2 km ($\tau$) through the following:

$$[\text{AOU}] = \frac{F_{2km}}{r_{C:O2}} \tau,$$

where $r_{C:O2}$ is the Redfield ratio of carbon to dissolved oxygen during remineralization, taken to be the same value
used in the models (0.69). This assumes that water entering the deep ocean is saturated with dissolved oxygen. We estimate $\tau$ from the difference in radiocarbon age between the deep (>2 km) North Atlantic and North Pacific Ocean (based on the corresponding mean radiocarbon values). As a check on the approach, we apply it to observations. We use $F_{2km} = 0.6$ Pg C yr$^{-1}$ [Berelson et al., 2007], and $\tau = 1600$ years based on $\Delta^{14}C$ for the North Atlantic and Pacific of $-70$ and $-230^{\circ}$, respectively [Matsumoto et al., 2004]. We then estimate [AOU] to be 160 $\mu$mol L$^{-1}$, essentially equal to the observed value.

[31] Figure 11c shows that the GCM AOU can be predicted reasonably well in this way, except that the GCMs tend to have more AOU than expected. Also, the PRIN model, which was an outlier on plots of AOU versus particle export and NDPW $^{14}C$, fits the trend line better. Other models (like AWI and IPSL), however, seem more anomalous from this perspective. A possible explanation for the higher-than-expected AOU in the GCMs is the assumption of gas-phase equilibrium of oxygen with the atmosphere during deep water formation in our simple box model. Newly ventilated deep waters may be significantly undersaturated. In a modeling study carrying explicit preformed tracers (and diagnosing True Oxygen Utilization, as well as AOU), Ito et al. [2004] demonstrated that southern source deep and bottom waters are undersaturated by as much as 60 $\mu$mol L$^{-1}$. In their model, this is due to the short exposure time and strong cooling of the deep winter mixed layers where the deep and bottom waters are forming. Increasing the AOU of the box model by roughly this amount would improve the agreement in Figure 11c.

### 3.5. Restricting the Models

[34] The models have a disconcertingly large range of export production, 9–28 Pg C yr$^{-1}$, which we doubt is a true reflection of the uncertainty in global export production. We therefore would like to restrict or weight the models when computing statistics of model global export production using some objective criteria. A logical criterion is some circulation metric, though we acknowledge that weighting the models in this way is not guaranteed to reduce the model spread, nor would the mean be more accurate. We tried weighting the models by using natural radiocarbon skill scores, as in work by Mikaloff Fletcher et al. [2006, 2007] but found that this had a very modest effect on mean export. We then attempted to restrict the models as in work by Matsumoto et al. [2004] on the basis of the models’ ability to simulate CFCs and $^{14}C$. The 11 models reporting total export have a mean ($\pm 1$σ) of 17 ± 6 Pg C yr$^{-1}$. All but two models are within the observed range of CFC-11 inventory. The remaining nine models have a mean export of 16 ± 5 Pg C yr$^{-1}$. If we select only the seven models within the range of observation-based estimates of anthropogenic CO$_2$ uptake, we get a mean export production of 15 ± 5 Pg C yr$^{-1}$. If we further restrict models to those five that also have a $^{14}C$ content of CDW within three times the observational error, the mean decreases further to 14 ± 5 Pg C yr$^{-1}$. Going through the same exercise for particle export yields a decrease as well, from 13 ± 3 Pg C yr$^{-1}$ for all 12 models to 11 ± 3 Pg C yr$^{-1}$ for the five satisfying the $^{14}C$ criterion. Thus restricting models on the basis of circulation metrics tends to modestly, though systematically, decrease export production. This is consistent with the results of Matsumoto et al. [2004]: restricting the models reduced the mean anthropogenic CO$_2$ uptake. To evaluate the significance of the export decrease, we computed the mean total export of all possible combinations of five of the 11 models (i.e., $11 \times 10 \times 9 \times 8 \times 7$ combinations). Only 7% of these have total export less than 14 Pg C yr$^{-1}$, which suggests a significant decrease. Other model-mean metrics changed when restricted to the five models satisfying the $^{14}C$ criteria: surface DOC decreased modestly from 29 ± 6 to 27 ± 5 $\mu$mol L$^{-1}$, the fraction of export as DOM decreased from 25 ± 8 to 21 ± 8%, and deep ocean (>2 km) AOU decreased from 183 ± 41 to 149 ± 34 $\mu$mol L$^{-1}$. Overall, restricting the models on the basis of the $^{14}C$ distribution brings them into better agreement with observation-based estimates, as shown in Table 4.

### 4. Summary and Conclusions

[35] We have examined export production, DOC concentration, seasonal net outgassing of oxygen and deep ocean AOU in 12 ocean circulation models with the same, simple marine biogeochemical model. We find that cross-model means and standard deviations (if interpreted as an error estimate) compare favorably with observation-based estimates, though the overall spread among the models is considerable (between a factor of 2 and 3). The models tend to be in better agreement with observation-based estimates when they also satisfy circulation-based metrics. On the other hand, regional distributions of some quantities

---

Table 4. Summary of Globally Averaged and Integrated Quantities From the Models and Observation-Based Estimates

<table>
<thead>
<tr>
<th></th>
<th>Model Range</th>
<th>All-Model Mean ±1σ</th>
<th>Five-Model Mean ±1σ*</th>
<th>Observation-Based Estimate</th>
<th>Reference for Observation-Based Estimate</th>
</tr>
</thead>
<tbody>
<tr>
<td>Total export, Pg C yr$^{-1}$</td>
<td>9–28</td>
<td>17 ± 6</td>
<td>14 ± 5</td>
<td>12 ± 1</td>
<td>Laws et al. [2000], Schlitzer [2002]</td>
</tr>
<tr>
<td>Particle export, Pg C yr$^{-1}$</td>
<td>7–17</td>
<td>13 ± 3</td>
<td>11 ± 3</td>
<td>10–12</td>
<td>Gnanadesikan et al. [2004], Schlitzer [2002]</td>
</tr>
<tr>
<td>0–50 m [DOC], $\mu$mol L$^{-1}$</td>
<td>18–36</td>
<td>29 ± 6</td>
<td>27 ± 5</td>
<td>29 ± 5</td>
<td>Hays [2002]</td>
</tr>
<tr>
<td>Fraction of export as DOC, %</td>
<td>14–43</td>
<td>25 ± 8</td>
<td>21 ± 8</td>
<td>20 ± 10</td>
<td>Locarnini et al. [2002]</td>
</tr>
<tr>
<td>&gt;2 km [AOU], $\mu$mol L$^{-1}$</td>
<td>121–242</td>
<td>183 ± 41</td>
<td>149 ± 34</td>
<td>156</td>
<td>Garcia and Keeling [2001]</td>
</tr>
<tr>
<td>Seasonal net outgassing, $\times 10^{14}$ mol</td>
<td>5.4–10.6</td>
<td>7.5 ± 1.1$^b$</td>
<td>7.0 ± 0.7</td>
<td>8.5 ± 0.7</td>
<td>Fasham et al. [2002]</td>
</tr>
</tbody>
</table>

$^a$Refers to five models that are within 3 σ of observed $^{14}C$ of Circumpolar Deep Water.

$^b$Excludes IGCR, AWI, and PIUB.
show considerable differences with observation-based estimates. Notably, model particle export and semilabile DOC are considerably higher than the observation-based estimates in the tropics, possibly reflecting well-known problems of coarse-resolution marine biogeochemical models in this region.

[16] However, our main message about the OCMIP-2 marine biogeochemical model results is not about model skill but the high sensitivity of marine biogeochemical fluxes and tracer distributions to ocean circulation. For example, the coefficients of variation in total export production (0.35), DOM export production (0.68), DOM export fraction (0.34), and even deep-ocean AOU (0.20), are considerably larger than those for ventilation metrics, like CFC inventory (0.16) and anthropogenic CO$_2$ uptake (0.12). Thus small changes in model ocean circulation, as measured by the uptake of transient tracers, may lead to large changes in marine biogeochemical stocks and rates (e.g., as discussed in section 3.3). This emphasizes two points made by Doney [1999]: that marine biogeochemical models require good circulation models and that biogeochemical metrics offer a new opportunity for evaluating ocean circulation models. Such metrics should provide a useful complement to purely physical tracers of ocean circulation (such as temperature, salinity, radiocarbon, and chlorofluorocarbons). Specific examples identified here are export production and the semilabile DOC distribution, both of which are highly sensitive to ocean circulation, particularly the exchange of water between the mixed layer and the thermocline. It could be argued that the sensitivity of marine biogeochemical metrics to ocean circulation is overstated here because of the nutrient-restoring approach used. However, we believe that a similar conclusion would have emerged if a common prognostic model was used because the surface nutrient fields would have varied dramatically among the models.

[17] We conclude by making several recommendations for marine biogeochemical modeling strategies and the observations needed to pursue them. We recognize that advances in modeling ocean circulation and (particularly) marine biogeochemical processes have been dramatic since the OCMIP-2 model was designed [Najjar and Orr, 1999] and the simulations generated. See, for example, the reviews of Christian and Anderson [2002] and Hood et al. [2006]. Nevertheless, simple marine biogeochemical models are now routinely used in Earth-system models, and have an important role to play in gaining insight into the processes regulating the large-scale cycling of carbon, oxygen and nutrients in the ocean, as others have shown [e.g., Gnanadesikan et al., 2002, 2004] and we have argued here.

[18] Our first recommendation is to incorporate idealized tracers, such as tracers of preformed water properties, into marine biogeochemical models [e.g., Ito et al., 2004; Friis et al., 2006]. Carrying preformed nutrient tracers in the OCMIP-2 models, perhaps “colored” differently in different source regions, would go a long way toward explaining the wide differences among the models and the correlations of export fraction among various regions (Table 3). Combinations of nutrients and oxygen (i.e., PO and NO) have been used to estimate sources of deep water, and so carrying preformed nutrient and oxygen tracers in models will also provide insight into ocean ventilation [e.g., Broecker et al., 1985]. Other idealized tracers include “solubility” and “biological” forms of dissolved gases, like oxygen [e.g., Jin et al., 2007] would be helpful for understanding the wide range of SNO results from the models.

[19] We also recommend that methods for estimating biogeochemical rates in the ocean using tracer data be evaluated by applying them to model data. For example, Friis et al. [2006] evaluated a method of estimating CaCO$_3$ dissolution from the distribution of alkalinity and CFCs [Feely et al., 2004], and found that it produced substantially inflated rates compared to the actual rates in the model. Thus we recommend that similar approaches be applied to methods of estimating various other rates and ratios, including remineralization rates, Redfield ratios, the fraction of oxygen utilization due to DOC remineralization, and export production [e.g., Jenkins, 1987; Anderson and Sarmiento, 1994; Doval and Hansell, 2000; Lee, 2001].

[20] Our model results and observations [e.g., Michaels et al., 1994; Carlson et al., 1994] show the importance of the seasonality of vertical mixing to upper ocean biogeochemical fluxes. We therefore recommend that global marine biogeochemical models routinely include seasonal forcing and mixed layer submodels, and that they are evaluated with observed climatologies of mixed layer depth [e.g., de Boyer Montégut et al., 2004].

[21] With regard to observations, we would like to highlight the importance of quantifying the DOM pool and POM fluxes. Specifically, we suggest that a global climatology of DOC, much like we now have for other tracers [e.g., Key et al., 2004], be developed; the surface synthesis of Siegel et al. [2002] is a step in the right direction. With respect to deep-ocean particle fluxes, there is a need for greater spatial coverage and for methods to extrapolate measurements to basin-wide scales with error estimates. We also join the chorus of marine biogeochemists who would value reliable sinking flux estimates in the shallow portion of the aphotic zone.

[22] We note the shortcomings of the formulation of organic matter cycling in the OCMIP-2 models. The low C:P ratios of the OCMIP-2 models means that they have too much recycling of phosphorus, which may substantially inflate the rate of export production in the models. Approaches are therefore needed to constrain C:N and C:P ratios of semilabile DOM, which will require more simultaneous measurements of the marine DON and DOP pools [e.g., Abell et al., 2000]. Combining comprehensive DOM measurements with models, especially inverse models, will lead to better constraints on the rates of DOM production and remineralization, including their elemental (C, N and P) ratios. This may help resolve conflicting errors regarding DOM lifetime (section 3.3), perhaps leading to parameterizations that capture its spatial and temporal variability. With respect to the parameterization of particle fluxes, we recommend that more sophisticated approaches [e.g., Armstrong et al., 2002] be evaluated with tracer data, perhaps in the context of inverse modeling.
[43] In summary, a full hierarchy of models is required to address the complexity of marine biogeochemical cycles and their interaction with the climate system. The large numbers of parameters and tracers involved require one to balance the complexity of the applied biogeochemical model and the spatiotemporal resolution of the underlying physical model with computational efficiency. Cost-efficient, coarse resolution models tuned to observation-based metrics [e.g., Müller et al., 2006] are well suited for sensitivity analyses, multiscenario evaluations, or for Monte Carlo approaches. On the other hand, high-resolution, eddy resolving models [e.g., Plattner et al., 2005] allow one to study biogeochemical processes in a highly detailed spatiotemporal setting.

Appendix A

[44] The time evolution equation for phosphate (PO$_4$), DOP and O$_2$ is given by

\[
\frac{\partial C}{\partial t} = L(C) + J_C, \tag{A1}
\]

where $C$ is the concentration of the tracer, $L$ is the linear transport operator defining advection and diffusion and $J_C$ is the biogeochemical source-sink distribution for $C$. Boundary fluxes are zero except for oxygen, which has an air-sea flux described below. Transport was generated by the individual participating models in OCMIP-2.

A1. Phosphorus

[45] The model describing the cycling of phosphorus is similar to the “nutrient restoring” approach adopted by Najjar et al. [1992] and Anderson and Sarmiento [1995]. Phosphate is chosen instead of nitrate as the basic currency of the model so as to avoid the complexities of nitrogen fixation and denitrification. Phosphate also has the advantage of having a greater database of observations for forcing and evaluating the model. We use the conceptual framework of three dissolved organic matter (DOM) pools (labile, semilabile and refractory) with relatively constant lifetimes, though this is a simplification because there is a wide spectrum of DOM lifetimes. Only the semilabile DOM pool is considered in the model because it is thought to be the main contributor to carbon export from surface waters [e.g., Hansell and Carlson, 2001]. This pool is thought to have a lifetime of months, and we approximate its decomposition with a constant lifetime.

[46] To compute $J_{PO_4}$, two regions are defined, separated by the compensation depth, $Z_c$, the depth at which photosynthesis is equal to respiration of the whole biological community, as discussed by Smertacek and Passow [1990]. We prefer not to use the term “euphotic zone depth,” which is often defined on the basis of an arbitrary percentage (usually 0.1 or 1) of the surface irradiance, because it means something quite different from compensation depth. We will refer to the region above the compensation depth as the production zone and the region below the compensation depth as the consumption zone. In the production zone, [PO$_4$] is nudged toward observations, [PO$_4$]*. On a timescale $\tau$, but only if [PO$_4$] > [PO$_4$]*. Otherwise, there is no

nudging. We define this term as $J_{Prod}$, representing the production of organic phosphorus:

\[
J_{Prod} = \frac{1}{\tau} \left( [PO_4] - [PO_4]^* \right), \quad [PO_4] > [PO_4]^* \quad Z < Z_c, \tag{A2}
\]

\[
J_{Prod} = 0, \quad [PO_4] \leq [PO_4]^*. \tag{A2b}
\]

where $Z$ is the depth.

[47] It is assumed that a fixed fraction, $\sigma$, of the phosphate uptake in the production zone is converted to DOP, which is allowed to be advected and diffused by the circulation field and is consumed everywhere following first-order kinetics. Thus we have the source/sink function for DOP,

\[
J_{DOP} = \sigma J_{Prod} - \kappa [DOP], \quad Z < Z_c \tag{A3a}
\]

\[
J_{DOP} = -\kappa [DOP], \quad Z > Z_c. \tag{A3b}
\]

[48] The phosphate not converted to DOP results in an instantaneous downward flux of particulate organic phosphorus at the compensation depth,

\[
F_Z = (1 - \sigma) \int_0^{Z_c} J_{Prod} dZ. \tag{A4}
\]

This flux decreases with depth owing to remineralization following a power law relationship,

\[
F(Z) = F_Z \left( \frac{Z}{Z_c} \right)^{-a}, \quad Z > Z_c. \tag{A5}
\]

The power law form has the advantage over exponential forms in that it captures the observed increase in remineralization length scale with depth. The source/sink term for phosphate is then

\[
J_{PO_4} = -\sigma J_{Prod} + \kappa [DOP], \quad Z < Z_c \tag{A6a}
\]

\[
J_{PO_4} = -\frac{\partial F}{\partial Z} + \kappa [DOP], \quad Z > Z_c. \tag{A6b}
\]

It is assumed that any flux reaching the sediments is remineralized there and diffused instantaneously back into the water column. Therefore the bottom box of the model will have an additional source of phosphate beyond that given by equation (A6b).

[49] The assumption that particulate organic phosphorus produced in the production zone is remineralized instantaneously in the consumption zone directly below is reasonable, and is based on the fact that the particle sinking timescale ($\sim$1 month) is much shorter than basin-wide advective timescales ($\sim$1–100 years).

[50] Note that remineralization of organic phosphorus is assumed to be independent of the amount of dissolved oxygen present. This may not be strictly true, but there is clear evidence of organic matter decomposition in anoxic regions as a result of denitrification. We simply assume that nitrate is as effective at oxidizing organic matter as oxygen.
[51] We now must choose the parameters $Z_c$, $\tau$, $\sigma$, $\kappa$ and $\alpha$. All of these parameters are likely to have spatial and temporal variability, but in the interest of keeping this model as simple as possible, we will use constant values. For $Z_c$, 75 m is used, on the basis of the analysis of seasonal oxygen variations by Najjar and Keeling [1997]. They inferred that the summertime compensation depth varies between about 50 and 100 m, depending on the latitude. The choice of $\tau$ is somewhat arbitrary. We use a value of 30 days, which should be long enough, in most cases, to accommodate differences between observed and modeled circulation, yet short enough to prevent too much of a lag of surface phosphate behind the observations. For $\alpha$, we use a value of 0.9, which is close to what sediment traps yield [Martin et al., 1987; Berelson, 2001]. Even though there is concern that sediment traps do not properly measure flux in the upper few hundred meters of the water column, other lines of evidence support the rapid drop-off in flux measured by sediment traps. A number of tracer-based estimates of remineralization in the upper few hundred meters of the water column yield remineralization length scales of a few hundred meters [Sarmiento et al., 1990, and references therein], similar to what equation (A5) with $\alpha = 0.9$ would yield. This value of $\alpha$ also yields good results for phosphate distributions in ocean general circulation models [Yamanaka and Tajika, 1996, 1997].

[52] A variety of studies suggest that $\kappa$, the semilabile DOP consumption rate constant, is between about (0.2 year)$^{-1}$ and (0.7 year)$^{-1}$. DOC in the Sargasso Sea at 200 m depth is observed to decrease from about 62 $\mu$mol kg$^{-1}$ in February–March to 52 $\mu$mol kg$^{-1}$ in July–August [Carlson et al., 1994], about a 5-month time period. Estimating the refractory component of DOC to be the concentration at 1000 m, 46 $\mu$mol kg$^{-1}$ [Hansell et al., 1995], the semilabile pool decreases from 16 to 6 $\mu$mol kg$^{-1}$. (The labile pool is assumed to be negligibly small.) We can then estimate $\kappa$ by rearranging (A3b) and substituting DOC for DOP,

$$\kappa = \frac{J_{\text{DOC}}}{[\text{DOC}]} = \frac{(-10 \text{ $\mu$mol kg}^{-1}/\text{5 mon})}{11 \text{ $\mu$mol kg}^{-1}} \approx 1 \text{ yr}^{-1}. \tag{A7}$$

Mineralization experiments by Hansell et al. [1995] give reasonable agreement with this value of $\kappa$. They found that DOC in water taken from 200 m depth in the Sargasso Sea decreases at a rate of 0.044 $\mu$mol kg$^{-1}$ day$^{-1}$, for a 101-day dark incubation, somewhat less than inferred from the observed decrease at 200 m depth, 0.067 $\mu$mol kg$^{-1}$ day$^{-1}$ from the Carlson et al. [1994] data. This would suggest a semilabile DOC lifetime of about (0.7 year)$^{-1}$. A 0.5-year lifetime of semilabile DOC was also estimated by Yamanaka and Tajika [1997] by fitting a three-dimensional model to observed vertical and horizontal distributions of DOC in the equatorial Pacific. A similar exercise by Archer et al. [1997] with a higher-resolution GCM found that the semilabile DOC lifetime is somewhat shorter, from 0.1 to 0.3 years, in the equatorial Pacific. On the basis of these four studies we choose an intermediate value of $\kappa$ equal to (0.5 year)$^{-1}$.

[53] We use a value of 0.67 for $\sigma$, which was determined by Yamanaka and Tajika [1997] in the data-fitting exercise mentioned above. It is important to realize that the fraction of the net downward flux of organic matter across the compensation depth that is in dissolved form is less than $\sigma$. This is because DOM is also remineralized above the compensation depth (equation (A3a)). For example, with $\sigma = 0.67$, Yamanaka and Tajika [1997] found that, on a global scale, DOC accounted for less than 30% of the downward flux of organic matter across 100 m depth.

[54] The 30% figure from Yamanaka and Tajika [1997] can be compared directly with regional observation-based estimates of the fraction of export production that occurs in dissolved form. This should be done with caution, however, because it is likely that this fraction varies spatially. Carlson et al. [1994] estimated that downward mixing of DOC accounts for 23–42% of the export production in the Sargasso Sea. In the equatorial Pacific, a variety of studies estimate DOC export to be roughly one quarter to one half of the export production [Murray et al., 1996; Archer et al., 1997; Hansell et al., 1997; Zhang and Quay, 1997; Quay, 1997; Loukos et al., 1997]. Thomas et al. [1995] find, using DOC and nitrate measurements with a simple box model, that 20 ± 15% of the organic matter export in the tropical Atlantic is in dissolved form. Guo et al. [1994] used AOU-DOC relationships in the Gulf of Mexico, as well as a simple 1-D model, to estimate that DOC accounts for about 20–30% of the downward flux of organic carbon. Finally, Borsheim and Myklestad [1997] show that the seasonal accumulation of DOC in the mixed layer is 20–60% of the export production in the Norwegian Sea. This fraction is an upper limit on the amount of organic matter exported in dissolved form. All of the observation-based estimates appear to be in reasonable agreement with the Yamanaka and Tajika [1997] estimate based on $\sigma = 0.67$.

[55] The final component to be specified for the phosphorus model is $[\text{PO}_4^3-]^*$, the observed distribution of phosphate above the compensation depth. The climatological monthly maps of Louanchi and Najjar [2000] were used and vertically interpolated to model gridboxes with upper boundaries above 75 m. In the grid box containing the compensation depth, the nudging term was multiplied by the fraction of the gridbox above 75 m. The maps were linearly interpolated to the model timestep, except for the PIUB and AWI models, which are nonseasonal, and in which case the annual mean was used. Note that even though the IGCR model is not seasonal, seasonal variations in nutrients were nevertheless imposed. The phosphate maps were created from the quality-controlled phosphate data of the 1994 World Ocean Atlas [Conkright et al., 1994] using the interpolation/smoothing technique of Najjar and Keeling [1997]. Phosphate values in regions of sparse data coverage were determined from phosphate-temperature relationships. Full details of the maps are given by Louanchi and Najjar [2000].

A2. Oxygen

[56] The oxygen model closely follows that of phosphate, being linked by the Redfield ratio, $r_{\text{O}_2:P}$. The main differences are that: (1) oxygen consumption is assumed to be halted below some critical oxygen level $[\text{O}_2]_{\text{crit}}$ and (2) oxygen crosses the air-sea interface. Oxygen is likely to
go below the critical level below the compensation depth. The ocean will then be a net source of oxygen to the atmosphere because oxygen production above the compensation depth will be greater than oxygen consumption below the compensation depth. This imbalance is probably made up for by nitrogen fixation in surface waters. An estimate of the impact on air-sea oxygen fluxes can be made by converting estimates of global denitrification to a mean air-sea oxygen flux. Unfortunately, global denitrification estimates vary by more than an order of magnitude [Middelburg et al., 1996]. Using a value of 300 Tg N yr\(^{-1}\) for denitrification, at the high end of the estimated range, and a \(O_2: N\) Redfield ratio of 10 yields an outgassing of oxygen equal to about 0.6 mol m\(^{-2}\) yr\(^{-1}\), equivalent to an oxygen supersaturation of only 0.5 \(\mu\)mol kg\(^{-1}\) given a gas transfer velocity of 3 m d\(^{-1}\). Since the impact on the dissolved oxygen distribution in surface waters is so small, we do not include nitrogen fixation in this model. The source/sink term for oxygen is then

\[
J_{O_2} = r_{O_2:P} J_{P_{PO_4}}, \quad [O_2] > [O_2]_{crit} \tag{A8a}
\]

\[
J_{O_2} = 0, \quad [O_2] < [O_2]_{crit}. \tag{A8b}
\]

For \(r_{O_2:P}\) we use the value of -170 based on the work of Anderson and Sarmiento [1994]. For \([O_2]_{crit}\) we use a value of 4 \(\mu\)mol kg\(^{-1}\), the oxygen concentration below which zooplankton abundance drops rapidly [Saltzman and Wisshner, 1997, and references therein].

A standard gas transfer formulation is adopted for the air-sea exchange of oxygen,

\[
F_{O_2} = k_{w} \left( [O_2] - [O_2]_{sat} \right), \tag{A9}
\]

where \(F_{O_2}\) is the upward flux of oxygen, \(k_{w}\) is the gas transfer velocity for oxygen, \([O_2]\) is the surface ocean concentration of oxygen, \([O_2]_{sat}\) is the concentration in equilibrium with a water-vapor-saturated atmosphere at a total atmospheric pressure \(P_t = 1\) atm, and \(P\) is the total atmospheric pressure. \([O_2]_{sat}\) is taken from Garcia and Gordon [1992] and computed from model surface temperature and salinity. The monthly mean climatology of Eshensen and Kashin [1981] is used to compute \(P\), and this is interpolated linearly to the model timestep. Variations in total pressure are taken into account when computing the oxygen saturation concentration because these may be important in some regions. Total pressure, for example, decreases by about 3% from 30°S to 60°S, a change that will affect \([O_2]_{sat}\) by about 10 \(\mu\)mol kg\(^{-1}\).

Wanninkhof [1992] has parameterized the instantaneous gas transfer velocity as a function of wind speed at 10 m (\(u\)) and Schmidt number (Sc). Adding the effect of sea ice fraction (\(f_i\)) to his formulation yields

\[
k_w = b (1 - f_i) \left( \frac{Sc}{660} \right)^{-1/2} u^*, \tag{A10}
\]

where \(b\) is a constant adjusted to give the correct global mean gas transfer velocity as deduced from the distribution of natural and bomb radiocarbon and an idealized wind speed distribution. This estimate of the global mean gas transfer velocity has also been found to be consistent with the large-scale seasonal variations of oxygen in the surface ocean and the atmosphere [Keeling et al., 1998]. Sc is computed from the model temperature using the formula of Keeling et al. [1998].

The mean annual cycle at monthly resolution of \(b(1 - f_i)u^*\) was provided to OCMIP-2 modelers, and interpolated linearly to the model timestep. The mean annual cycle at monthly resolution of \(u^*\) is computed from satellite measurements, as described by Dutay et al. [2002]. Note that this monthly climatology includes the effect of short-term (submonthly) variations in the wind speed. The value of \(b\) was determined to be 0.336 so that the global and annual mean gas transfer coefficient for carbon dioxide \((k_wk_0, \text{ where } k_0 \text{ is the CO}_2 \text{ solubility})\) is equal to 0.061 mol m\(^{-2}\) yr\(^{-1}\) \(\mu\)atm\(^{-1}\) for preindustrial times, which is based on \(k_wk_0\text{CO}_2 = 17 \pm 4 \text{ mol m}^{-2} \text{ yr}^{-1}\) estimated by Broecker et al. [1986] on the basis of the global bomb radiocarbon budget, divided by the preindustrial \(\mu\)atm of \(\text{CO}_2\) of 280 \(\mu\)atm.

The value of 0.336 is slightly different from the value of 0.31 derived by Wanninkhof [1992] for “short-term” winds because of the different wind fields used. We note that our gas transfer velocity may be an overestimate because more recent estimates of the bomb radiocarbon inventory are about 25% lower [Peacock, 2004].

### A3. Initial Conditions and Spinup

For \(O_2\), the initial value is not critical because exchange with the atmosphere will ultimately determine its global steady state inventory. Phosphate and DOP, however, do not exchange with the atmosphere, and therefore the sum of their inventories will remain fixed at its initial value. For simplicity, global mean values are chosen for initial conditions: 2.17 \(\mu\)mol kg\(^{-1}\) for phosphate (computed from the atlas of Levitus et al. [1993]) and 170 \(\mu\)mol kg\(^{-1}\) for oxygen [Levitus and Boyer, 1994]. The mean semilabile DOP concentration can be estimated from the semilabile DOC distribution, which decreases from about 40 \(\mu\)mol kg\(^{-1}\) at the surface to close to zero around 400 m. Assuming a linear decrease with depth, the above value of \(r_{C:D}\) and a mean oxygen depth of 4000 m yields a mean semilabile DOP concentration of 0.02 \(\mu\)mol kg\(^{-1}\). The model is run from these initial conditions until a steady state is reached.

Levitus et al. [1993] reported phosphate concentrations in units of \(\mu\)mol L\(^{-1}\), which we incorrectly interpreted as \(\mu\)mol kg\(^{-1}\). All tracer simulations were run assuming volume conservation, and so a fixed density of 1.0245 kg L\(^{-1}\) (the mean surface value) was used to convert all tracer concentrations from \(\mu\)mol kg\(^{-1}\) to \(\mu\)mol L\(^{-1}\). Because of our erroneous interpretation of the Levitus et al. [1993] phosphate, we have overestimated the phosphate inventory by 2–3%. This is a small error, possibly smaller than the actual error in the global mean concentration.
Acknowledgments. R. G. N. and J. L. S. acknowledge the support of the NASA grants NAG5-6451 and NAG5-6591, respectively, as part of the JGOFS Synthesis and Modeling Program. Dork Sahagian and the IGBP/GAIM office provided logistical support. G. K. P. and F. J. acknowledgment, the Swiss National Science Foundation. European contributions were supported by the EU GOSAC Project (ENV4-CT97-0495). We thank the two reviewers of this manuscript and C. LeQuéré for their critical comments and helpful suggestions.

References


