Marine geochemical data assimilation in an efficient Earth System Model of global biogeochemical cycling


1 School of Geographical Sciences, University of Bristol, Bristol, UK
2 Frontier Research Center for Global Change, 3173-25 Showa-machi, Kanazawa-ku, Yokohama, Kanagawa 236-0001, Japan
3 Earth Sciences, The Open University, Walton Hall, Milton Keynes, UK
4 School of Environmental Sciences, University of East Anglia, Norwich, UK
5 National Oceanography Centre, Southampton, European Way, Southampton SO14 3ZH, UK

Received: 17 July 2006 – Published in Biogeosciences Discuss.: 10 August 2006
Revised: 13 December 2006 – Accepted: 15 January 2007 – Published: 25 January 2007

Abstract. We have extended the 3-D ocean based “Grid ENabled Integrated Earth system model” (GENIE-1) to help understand the role of ocean biogeochemistry and marine sediments in the long-term (~100 to 100,000 year) regulation of atmospheric CO2, and the importance of feedbacks between CO2 and climate. Here we describe the ocean carbon cycle, which in its first incarnation is based around a simple single nutrient (phosphate) control on biological productivity. The addition of calcium carbonate preservation in deep-sea sediments and its role in regulating atmospheric CO2 is presented elsewhere (Ridgwell and Hargreaves, 2007).

We have calibrated the model parameters controlling ocean carbon cycling in GENIE-1 by assimilating 3-D observational datasets of phosphate and alkalinity using an ensemble Kalman filter method. The calibrated (mean) model predicts a global export production of particulate organic carbon (POC) of 8.9 Pg C yr\(^{-1}\), and reproduces the main features of dissolved oxygen distributions in the ocean. For estimating biogenic calcium carbonate (CaCO\(_3\)) production, we have devised a parameterization in which the CaCO\(_3\):POC export ratio is related directly to ambient saturation state. Calibrated global CaCO\(_3\) export production (1.2 Pg C yr\(^{-1}\)) is close to recent marine carbonate budget estimates.

The GENIE-1 Earth system model is capable of simulating a wide variety of dissolved and isotopic species of relevance to the study of modern global biogeochemical cycles as well as past global environmental changes recorded in palaeoceanographic proxies. Importantly, even with 12 active biogeochemical tracers in the ocean and including the calculation of feedbacks between atmospheric CO2 and climate, we achieve better than 1000 years per (2.4 GHz) CPU hour on a desktop PC. The GENIE-1 model thus provides a viable alternative to box and zonally-averaged models for studying global biogeochemical cycling over all but the very longest (>1,000,000 year) time-scales.

1 Introduction

The societal importance of better understanding the role of the ocean in regulating atmospheric CO2 (and climate) is unquestionable. Reorganizations of ocean circulation and nutrient cycling as well as changes in biological productivity and surface temperatures all modulate the concentration of CO2 in the atmosphere, and are likely central to explaining the observed variability in CO2 over the glacial-interglacial cycles of the past ~800,000 years (Siegenthaler et al., 2005). These same marine carbon cycle processes will also affect the uptake of fossil fuel CO2 in the future. Interactions between marine biogeochemistry and deep-sea sediments together with imbalances induced between terrestrial weathering and the sedimentary burial of calcium carbonate (CaCO\(_3\)) exert further controls on atmospheric CO2 (Archer et al., 1998). These sedimentation and weathering processes are suspected to have dominated the recovery of the Earth system from catastrophic CO2 release in the geological past (Zachos et al., 2005), and will largely determine the long-term future fate of fossil fuel CO2 (Archer et al., 1998).

Quantifying the behavior of atmospheric CO2 is complicated by the fact that CO2 exerts an influence on marine biogeochemical cycles via the radiative forcing of climate. For instance, anthropogenically-driven increases in atmospheric CO2 are amplified because CO2 solubility decreases at higher temperature and as a result of thermally (and/or fresh water) induced changes in ocean stratification and transport (Sarmiento et al., 1998; Plattner et al., 2001).
Understanding how atmospheric CO\textsubscript{2} is regulated thus necessarily requires feedbacks between CO\textsubscript{2} and climate to be taken into account. Coupled GCM ocean-atmosphere plus carbon cycle models (e.g., “HadCM3LC”; Cox et al., 2000) are important tools for assessing climate change and associated feedbacks over the next few hundred years. However, their computational demands currently make them unsuitable for investigating time-scales beyond 1000 years or for conducting sensitivity studies. So-called “off-line” carbon cycle models in which the ocean circulation has been pre-calculated such as the off-line tracer transport HAMOCC3 model of Maier-Reimer (1993) are much faster. However, the fixed circulation field employed in these models means that the importance of feedbacks with climate cannot easily be explored, except in a highly parameterized manner (Archer et al., 2004; Archer, 2005).

In contrast, marine biogeochemical box models, which consider the global ocean in terms of relatively few (typically between about 3 and about 12) distinct volumes (the “boxes”), are extremely computationally efficient. Models of this type have an illustrious pedigree (e.g., Sarmiento and Toggweiler, 1984; Broecker and Peng, 1986) and continue today to be the tools of choice for many questions involving processes operating on time-scales of 10,000 years or longer such as those involving ocean-sediment interactions and weathering feedbacks (e.g., Zeebe and Westbroek, 2003; Ridgwell, 2005; Yool and Tyrrell, 2005). Box models are also important tools in situations where the inherent uncertainties are large and need to be extensively explored (e.g., Parekh et al., 2004). Box models, as with GCMs, have specific limitations. Validation against marine observations and sediment records is problematic because large volumes of the ocean are homogenized in creating each “box”, whereas biogeochemical processes are extremely heterogeneous both within ocean basins as well as between them. It is also difficult to incorporate a responsive climate or circulation, which is why box models are in effect off-line tracer transport models. The coupled meridional box model of Gildor et al. (2002) is one exception to this. Concerns have also been recently raised regarding whether box models present an inherently biased picture of certain aspects of the oceanic control of atmospheric CO\textsubscript{2} (Archer et al., 2003).

This has led to the development of fast climate models with reduced spatial resolution and/or more highly parameterized “physics” – known variously as Earth System (Climate) Models (ESMs) (Weaver et al., 2001) and Earth System Models of Intermediate Complexity (EMICs) (Claussen et al., 2002). While ESMs have been instrumental in helping address a range of carbon cycling and climate change questions, we believe that they have yet to realize their full potential. In particular, simulations on the time-scale of ocean-sediment interaction (order 10,000 years) are still relatively computationally expensive in many ESMs based around 3-D ocean circulation models, whereas in faster “2.5-D” models (e.g., Marchal et al., 1998) the zonal averaging of the basins complicates comparison between model and paleoceanographic data (Ridgwell, 2001).

In this paper we present a representation of marine biogeochemical cycling within a 3-D ocean based Earth system model, which we calibrate for the modern carbon cycle via a novel assimilation of marine geochemical data. In a further extension to the model we describe the addition of carbonate preservation and burial in deep sea sediments (Ridgwell and Hargreaves, 2007). Together, these developments allow us to explore important questions surrounding future fossil fuel CO\textsubscript{2} uptake by the ocean and including the role of major feedbacks with climate (Ridgwell et al., 2006). However, as we discuss later, the initial biogeochemical modelling framework has important gaps, particularly with respect to the lack of a multi-nutrient description of biological productivity. There is also no representation of organic carbon burial or of suboxic diagenesis and the sedimentary control on ocean phosphate, nitrate, and sulphate concentrations. Continuing developments to this model will progressively address these areas and enable more extreme environmental conditions and events occurring earlier in Earth history to be explicitly assessed.

This paper is laid out as follows: we first describe the ocean biogeochemistry model itself (Sect. 2), then how the model is calibrated by assimilating observed marine geochemical data (Sect. 3). In Sect. 4 we present the predictions of the model for modern marine biogeochemical cycling, and discuss the specific limitations of very low resolution 3-D ocean circulation models such as we use here. We conclude in Sect. 5, summarising the pros and cons of the model as well as outlining the future development path.

2 Description of the “GENIE-1” model

The basis of our work is the fast climate model of Edwards and Marsh (2005) (“C-GOLDSTEIN”), which features a reduced physics (frictional geostrophic) 3-D ocean circulation model coupled to a 2-D energy-moisture balance model (EMBM) of the atmosphere and a dynamic-thermodynamic sea-ice model (see Edwards and Marsh, 2005, for a full description). The ocean model used here is non-seasonally forced and implemented on a 36 × 36 equal-area horizontal grid, comprising 10° increments in longitude but uniform in sine of latitude, giving 3.2° latitudinal increments at the equator increasing to 19.2° in the highest latitude band. The ocean has 8 z-coordinate levels in the vertical. The grid and bathymetry is shown in Fig. 1.

Along with temperature and salinity, the oceanic concentrations of biogeochemical tracers are advected, diffused, and convected on-line by the ocean circulation model. To enable a broad range of questions concerning global biogeochemical cycling to be addressed, the GENIE-1 model currently contains the definitions of 49 dissolved tracers and isotopic properties relevant to future global change and paleoceanography,
We have developed a representation of marine biogeochemical cycling called BIOGEM (for BIOGEOchemical Model) that calculates the redistribution of tracer concentrations occurring other than by transport by the circulation of the ocean. This happens through the removal from solution of nutrients (PO₄) together with DIC and ALK, by biological activity in the sunlit surface ocean layer (the euphotic zone), which has a depth of hₑ=175 m in the 8-level ocean model configuration. The resulting export of constituent species back into inorganic solution, but at greater depth compared to the level at which the tracers were originally removed from solution. Further redistribution of tracers occurs through gas exchange with the atmosphere as well as due to the creation and destruction of dissolved organic matter. BIOGEM has a conceptual relationship with the climate model as shown schematically in Fig. 2. We refer to the overall composite model as GENIE-1.

We carry out the biogeochemical manipulation of tracer concentration fields asynchronously to maximize overall model speed. In this study, BIOGEM is called once every 5 time-steps taken by the ocean circulation model (i.e., 5×0.01 yr=0.05 yr). We find no discernable impact of this decoupling compared to carrying out the same time step
length as the ocean (0.01 yr), even in response to rather extreme (15 000 PgC) releases of (fossil fuel) CO₂ to the atmosphere.

2.1 Ocean biogeochemical cycling

The low vertical resolution of the ocean circulation model and need to maximize computational speed for long simulations and sensitivity analyses dictates that the “biological” part of the marine carbon cycle be relatively abstracted. A more important reason for starting off with as simple a description of surface biological productivity as possible is to reduce the number of free parameters in the model which must be optimized, particularly since this work represents the first attempt to assimilate marine geochemical observations within a prognostic 3-D circulation model of the ocean. Of course, as the nature of the assimilation problem becomes better defined and the capabilities and limitations of available optimization techniques better understood, we expect to subsequently re-calibrate the marine carbon cycle in conjunction with increasingly sophisticated multiple-nutrient biological schemes.

In this paper we estimate new (export) production directly from available surface nutrient concentrations, a robust tactic used in many early ocean carbon cycle models. In other words, what we have is “conceptually not a model of biology in the ocean but rather a model of biogenically induced chemical fluxes (from the surface ocean)” (Maier-Reimer, 1993). Overall, the biological export scheme is functionally similar to that of Parekh et al. (2005), and we adopt their notation where relevant. The main difference is that we currently consider only a single nutrient, phosphate (PO₄), rather than co-limitation with iron (Fe).

The governing equations in BIOGEM for the changes in phosphate and dissolved organic phosphorus (DOP) concentrations occurring in the surface ocean layer (but omitting the extra transport terms that are calculated by the ocean circulation model; Edwards and Marsh, 2005) are:

\[
\frac{\partial \text{PO}_4}{\partial t} = -\Gamma + \lambda \text{DOP} \tag{1}
\]

\[
\frac{\partial \text{DOP}}{\partial t} = \nu \Gamma - \lambda \text{DOP} \tag{2}
\]

\[
\Gamma = u_0 \text{PO}_4 \cdot \frac{\text{PO}_4}{K \text{PO}_4} \cdot (1 - A) \cdot \frac{I}{I_0} \tag{3}
\]

where \( \Gamma \) is the biological uptake of \( \text{PO}_4 \). \( \Gamma \) is calculated from: (i) \( u_0 \text{PO}_4 \), an assumed maximum uptake rate of phosphate (\( \text{mol PO}_4 \text{ kg}^{-1} \text{ yr}^{-1} \)) that would occur in the absence of any limitation on phytoplankton growth, and (ii) a Michaelis-Menten type kinetic limitation of nutrient uptake, of which \( K \text{PO}_4 \) is the half-saturation constant. Because of the degree of abstraction of ecosystem function inherent in our model, the appropriate values for either \( u_0 \text{PO}_4 \) or \( K \text{PO}_4 \) are not obvious, and so are subsequently calibrated (see Sect. 3 and Table 1). We apply two modifiers on productivity representing the effects of sub-optimal ambient light levels and the fractional sea ice coverage of each grid cell (\( A \)) (Edwards and Marsh, 2005). A full treatment of the effects of light limitation on phytoplankton growth is beyond the scope of this current paper. The strength of local insolation (\( I \)) is therefore simply normalized to the solar constant (\( I_0 \)) to give a limitation term that is linear in annual incident insolation. A proportion (\( \nu \)) of \( \text{PO}_4 \) taken up by the biota is partitioned into dissolved organic phosphorus (DOP). The relatively labile dissolved organic molecules are subsequently remineralized with a time constant of \( 1/\lambda \). The values of \( \nu \) and \( \lambda \) are assigned following the assumptions of the OCMIP-2 protocol (Najjar and Orr, 1999): \( \nu=0.66 \) and \( \lambda=0.5 \text{ yr}^{-1} \).

The particulate organic matter fraction is exported vertically and without lateral advection out of the surface ocean layer at the next model time-step. Because there is no explicit standing plankton biomass in the model, the export flux of particulate organic phosphorus (\( F_{z=h_e}^{\text{POC}} \), in units of \( \text{mol PO}_4 \text{ m}^{-2} \text{ yr}^{-1} \)) is equated directly with \( \text{PO}_4 \) uptake (Eq. 1):

\[
F_{z=h_e}^{\text{POC}} = \int_0^{h_e} \rho \cdot (1 - \nu) \cdot \Gamma \, dz \tag{4}
\]

where \( \rho \) is the density of seawater and \( h_e \) the thickness of the euphotic zone (175 m in the 8-level version of this ocean model).

In the production of organic matter, dissolved inorganic carbon (DIC) is taken out of solution in a 106:1 molar ratio with \( \text{PO}_4 \) (Redfield et al., 1963) while \( \text{O}_2 \) takes a -170:1 ratio with \( \text{PO}_4 \) (Anderson and Sarmiento, 1994). The effect on total alkalinity (ALK) of the biological uptake and remineralization of nitrate (NO₃) is accounted for via a modification of ALK in a -1:1 ratio with the quantity of NO₃ transformed. Because we do not model the nitrogen cycle explicitly in this paper, we link ALK directly to \( \text{PO}_4 \) uptake and remineralization through the canonical 16:1 N:P ratio (Redfield et al., 1963). For convenience, we will describe the various transformations involving organic matter in terms of carbon (rather than phosphorus) units, the relationship between organic matter export fluxes being simply:

\[
F_{z=h_e}^{\text{POC}} = 106 \cdot F_{z=h_e}^{\text{POC}} \tag{5}
\]

We represent the remineralization of particulate organic carbon (POC) as a process occurring instantaneously throughout the water column. We partition POC into two distinct fractions with different fates in the water column, following Ridgwell (2001) but adopting an exponential decay as an alternative to a power law. The POC flux at depth \( z \) in the water column is:

\[
F_z^{\text{POC}} = F_{z=h_e}^{\text{POC}} \cdot \left( (1-r^{\text{POC}}) + r^{\text{POC}} \cdot \exp \left( \frac{z}{t^{\text{POC}}} \right) \right) \tag{6}
\]

The parameters: \( r^{\text{POC}} \), the length-scale and \( r^{\text{POC}} \), the initial partitioning of POC into a labile fraction are both calibrated
Table 1. EnKF calibrated biogeochemical parameters in the GENIE-1 model.

<table>
<thead>
<tr>
<th>Name</th>
<th>Prior assumptions (mean and range)</th>
<th>Posterior mean</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>$n_{P,O_4}$</td>
<td>1.65 $\mu$mol kg$^{-1}$ yr$^{-1}$</td>
<td>1.91 $\mu$mol kg$^{-1}$ yr$^{-1}$</td>
<td>maximum PO$_4$ uptake (removal) rate (Eq. 3)</td>
</tr>
<tr>
<td>$K_{P,O_4}$</td>
<td>0.2 $\mu$mol kg$^{-1}$</td>
<td>0.21 $\mu$mol kg$^{-1}$</td>
<td>PO$_4$ Michaelis-Menten half-saturation concentration (Eq. 3)</td>
</tr>
<tr>
<td>$r_{POC}$</td>
<td>0.05 (0.02–0.08)</td>
<td>0.055</td>
<td>initial proportion of POC export as fraction #2 (Eq. 6)</td>
</tr>
<tr>
<td>$r_{POC}$</td>
<td>600 m (200–1000)</td>
<td>556 m</td>
<td>e-folding remineralization depth of POC fraction #1 (Eq. 6)</td>
</tr>
<tr>
<td>$r_{CaCO_3:POC}$</td>
<td>0.036 (0.015–0.088)$^c$</td>
<td>0.022$^d$</td>
<td>CaCO$_3$:POC: export rain ratio scalar (Eq. 8)</td>
</tr>
<tr>
<td>$\eta$</td>
<td>1.5 (1.0–2.0)</td>
<td>1.28</td>
<td>thermodynamic calcification rate power (Eq. 9)</td>
</tr>
<tr>
<td>$r_{CaCO_3}$</td>
<td>0.4 (0.2–0.6)</td>
<td>0.489</td>
<td>initial proportion of CaCO$_3$ export as fraction #2 (Eq. 11)</td>
</tr>
<tr>
<td>$r_{CaCO_3}$</td>
<td>600 m (200–1000)</td>
<td>1055</td>
<td>e-folding remineralization depth of CaCO$_3$ fraction #1 (Eq. 11)</td>
</tr>
</tbody>
</table>

$^a$ the range is quoted as 1 standard deviation either side of the mean  
$^b$ quoted as the mean of the entire EnKF ensemble  
$^c$ assimilation was carried out on a log$_{10}$ scale  
$^d$ Note that the rain ratio scalar parameter is not the same as the CaCO$_3$:POC export rain ratio as measured at the base of the euphotic zone, because $r_{CaCO_3:POC}$ is further multiplied by ($\Omega^{-1}$)$^b$ to calculate the rain ratio, where $\Omega$ is the surface ocean saturation state with respect to calcite (see Sect. 2.1). Pre-industrial mean ocean surface $\Omega$ is $\approx$5.2 in the GENIE-1 model, so that the global CaCO$_3$:POC export rain ratio can be estimated using the 8-parameter assimilation as being equal to $(5.2–1)^{1.28} \times 0.022=0.14$.

(Table 1). Because we explicitly resolve the individual “components” (i.e., C, $^{13}$C, P, . . .) of organic matter, the GENIE-1 model can be used to quantify the effect of fractionation between the components of organic matter during remineralization (e.g., Shaffer et al., 1999) as well as between different carbon isotopes. However, we assume no fractionation during remineralization in this present study. The residual flux of particulate organic material escaping remineralization within the water column is remineralized at the ocean floor, making the ocean-atmosphere system closed with respect to these tracers (i.e., there is no loss or gain to the system).

The modern ocean is oxic everywhere at the resolution of our model (e.g., see Fig. 6). However, O$_2$ availability may be insufficient under different ocean circulation regimes and continental configurations. To broaden the applicability of the GENIE-1 model to past climates and biogeochemical cycling, we limit remineralization according to the total availability of electron acceptors – if dissolved O$_2$ is depleted and NO$_3$ is selected as an active tracer in the model, denitrification occurs to provide the necessary oxidant: 2NO$_3^-$→N$_2$+3O$_2$. If NO$_3^-$ becomes depleted and SO$_4^{2-}$ has been selected as an active tracer, sulphate reduction occurs: 2H$^+$+SO$_4^{2-}$→H$_2$S+2O$_2$. If the total concentration of selected electron-accepting tracer species (O$_2$, NO$_3^-$, SO$_4^{2-}$) is still insufficient, remineralization of POC is restricted. Our strategy thus differs from other modeling approaches in which remineralization always strictly conforms to a predefined profile and the consequences of excess oxidation over O$_2$ availability are resolved either by numerically preventing negative oxygen concentrations occurring (e.g., Zhang et al., 2001, 2003) or by allowing the tracer transport of negative O$_2$ concentrations (e.g., Hotinski et al., 2001). We treat the remineralization of dissolved organic matter in an analogous manner if O$_2$ availability is insufficient.

H$_2$S created through sulphate reduction is oxidized in the presence of O$_2$ at a rate (mM H$_2$S h$^{-1}$):

$$\frac{d[H_2S]}{dt} = k \cdot [H_2S] \cdot [O_2]^2$$  \hspace{1cm} (7)

where [H$_2$S] and [O$_2$] are the dissolved concentrations of hydrogen sulphide and oxygen, respectively, and $k$=0.625 mM$^{-2}$ h$^{-1}$ (Zhang and Millero, 1993).

While we account for remineralisation of carbon and nutrients within the water column as well as the associated depletion of dissolved oxygen, and in more extreme circumstances, of nitrate and sulphate, we do not explicitly account for important long-term sedimentary controls on ocean chemistry, with the exception of carbon and alkalinity associated with CaCO$_3$ – see Ridgwell and Hargreaves (2007). Thus, we are currently unable to assess feedback between bottom-water dysoxia and anoxia on global nutrient cycling, particularly involving phosphate, nitrate, and/or iron. For instance, O$_2$-depleted bottom waters are known to be associated with enhanced phosphorous regeneration from marine sediments (Ingall and Jahne, 1994), leading to the possibility of positive feedback between productivity and anoxia which might be central to the understanding of past Oceanic Anoxic Events (Van Cappellen and Ingall, 1994; Handoh and Lenton, 2003). Future developments to the GENIE-1 model will address the sedimentary control of nutrient cycling, as

www.biogeosciences.net/4/87/2007/  
Biogeosciences, 4, 87–104, 2007
well as explicitly simulating sulphate reduction and sulphide release to the ocean, which currently we treat implicitly as a water-column process in order to avoid problems encountered in previous models when O2 demand due to organic matter remineralisation exceeds dissolved O2 availability (Hotinski et al., 2001; Zhang et al., 2001, 2003).

Open ocean dwelling calcifying plankton such as coccolithophors and foraminifera produce calcium carbonate (CaCO3) in addition to organic compounds (see Ridgwell and Zeebe, 2005, for a review of the global carbonate cycle). As the ecological processes which regulate calcifier activity are not well understood, early models incorporating marine carbonate production calculated the export flux of CaCO3 in a fixed production ratio with POC (e.g., Broecker and Peng, 1986; Yamanaka and Tajika, 1996). Derivatives of this approach modify the CaCO3:POC rain ratio as some function of temperature (e.g., Marchal et al., 1998) and/or opal flux (e.g., Heinez et al., 1999; Archer et al., 2000; Heinez, 2004). Current ecosystem models include a measure of competition between calcifying phytoplankton such as coccolithophores and non-calcifying ones such as diatoms to estimate CaCO3 production (e.g., Moore et al., 2002; Bopp et al., 2003). However, there are drawbacks with this strategy because there are doubts as to whether autotrophic coccolithophores are the dominant source of CaCO3 in the open ocean (Schiebel, 2002), or whether Emiliania huxleyi is a sufficiently representative species of global coccolith carbonate production to be chosen as the functional type calcifying species in ecosystem models (Ridgwell et al., 2006).

Recent work has introduced a further dimension as to how CaCO3 production should be represented in models. Experiments have shown that planktic calcifiers such as coccolithophores (Riebesell et al., 2000; Zondervan et al., 2001; Delille et al., 2005) and foraminifera (Bijma et al., 1999) produce less carbonate at lower ambient carbonate ion concentrations (CO32−). A progressive reduction in surface CO32− is an expected consequence of fossil fuel invasion into the ocean (Kleypas et al., 1999; Caldeira and Wickett, 2003; Freely et al., 2004; Orr et al., 2005). Several recent studies have incorporated a response of the CaCO3:POC rain ratio to changes in surface ocean carbonate chemistry, by employing a parameterization based on the deviation from modern surface ocean conditions of either CO2 partial pressure (Heinez, 2004) or CO32− (Barker et al., 2003). We present a new approach here, one which relates the export flux of CaCO3 (FzCaCO3) to the POC flux (FzPOC) via a thermodynamically-based description of carbonate precipitation rate:

\[
F_{z=\gamma}^{CaCO3} = \nu_{\gamma} \cdot F_{z=h}^{CaCO3} \cdot F_{z=h}^{POC}
\]

where \(\nu_{\gamma}\) is a spatially-uniform scalar, and \(\gamma\) is a thermodynamically-based local modifier of the rate of calcification (and thus of the CaCO3:POC rain ratio), defined:

\[
\gamma = (\Omega - 1)^{-\eta} \quad \Omega > 1.0
\]

(9a)

(9b)

where \(\eta\) is a constant and \(\Omega\) is the ambient surface saturation state (solubility ratio) with respect to calcite, defined by the ratio of the product of calcium ion (Ca2+) and carbonate ion (CO32−) concentrations to \(K_{SP}\), the solubility constant (Zeebe and Wolf-Gladrow, 2001):

\[
\Omega = \frac{[Ca^{2+}][CO_3^{2-}]}{K_{SP}}
\]

(10)

In formulating this parameterization we have drawn on descriptions of abiotic carbonate system dynamics in which the experimentally observed precipitation rate can be linked to saturation via an equation with the same form as Eq. (9) (e.g., Burton and Walter, 1987). What this equation says is that the precipitation rate increases with a greater ambient environmental degree of supersaturation with respect to the solid carbonate phase (\(\Omega > 1.0\)), with the power parameter \(\eta\) controlling how non-linear the response of calcification is. At ambient saturation states below the point of thermodynamic equilibrium (\(\Omega = 1.0\)) (“undersaturation”) no carbonate production occurs. It should be noted that although coccolithophorid and foraminiferal calcification rates are observed to respond to changes in saturation (e.g., Bijma et al., 1999; Riebesell et al., 2000; Zondervan et al., 2001; Delille et al., 2005), we do not explicitly capture other important controls. Instead, we have implicitly collapsed the (poorly understood) ecological and physical oceanographic controls on carbonate production onto a single, purely thermodynamic dependence on \(\Omega\).

For this paper we take our prior assumptions regarding the suspected values of \(\eta\) (Table 1) from previous analysis of neritic (shallow water) calcification (Opdyke and Wilkinson, 1993; Zeebe and Westbroek, 2003; Ridgwell, 2004; Langdon and Atkinson, 2005). Elsewhere we collate available observational data on pelagic calcifiers and explore the effect of alternative prior assumptions in \(\eta\) (Ridgwell et al., 2006).

The remineralization (dissolution) of CaCO3 in the water column is treated in a similar manner to particulate organic carbon (the parameter nomenclature being analogous to Eq. 6):

\[
F_{z}^{CaCO3} = F_{z=h}^{CaCO3} \cdot \left(1 - \rho_{CaCO3} \cdot \frac{z_h - z}{\rho_{CaCO3}} \cdot \exp\left(\frac{z_h - z}{\rho_{CaCO3}}\right)\right)
\]

(11)

The values of \(\rho_{CaCO3}\) and \(\rho^{CaCO3}\) are also calibrated (Table 1). While all CaCO3 reaching the ocean floor is assumed dissolved in this ocean-only analysis, we develop the model and data assimilation further in Ridgwell and Hargreaves (2007) by explicitly accounting for the preservation of carbonates in marine sediments.
2.2 Air-sea gas exchange

The flux of gases (mol m⁻² yr⁻¹) across the air-sea interface is given by:

\[ F = k \cdot \rho \cdot (C_w - \alpha \cdot C_a) \cdot (1 - A) \] (12)

where \( k \) is the gas transfer velocity (m yr⁻¹), \( \rho \) the density of sea-water (kg m⁻³), \( C_w \) the concentration of the gas dissolved in the surface ocean (mol kg⁻¹), \( C_a \) the concentration of the gas in the overlying atmosphere (atm), and \( A \) the fractional ice-covered area. The parameter \( \alpha \) is the solubility coefficient (mol kg⁻¹ atm⁻¹) and is calculated from the coefficients listed by Wanninkhof (1992) (and references therein). Gas transfer velocities are calculated as a function of wind speed following Wanninkhof (1992):

\[ k = a \cdot u^2 \cdot (Sc/660)^{-0.5} \] (13)

where \( u \) is the annual mean climatological wind speed and \( Sc \) is the Schmidt Number for the specific gas following Wanninkhof (1992) (and references therein). We use the scalar annual average wind speed field of Trenberth et al. (1989) for calculating air-sea gas exchange and set the scaling constant \( a=0.31 \) (Wanninkhof, 1992), which gives us a global annual mean gas transfer coefficient for CO₂ equal to 0.058 mol m⁻² yr⁻¹ µatm⁻¹ in the calibrated model (Sect. 3.2).

For completeness, we include the parameterization of air-sea gas exchange for all the gases listed by Wanninkhof (1992) although this calculation is not made if the relevant dissolved tracers in the ocean and/or atmosphere are not selected. To facilitate analysis of the biogeochemical consequences of anoxia and the existence of a sulphidic ocean at times in the geologic past, we also allow for the exchange of H₂S, for which we take the solubility following Millero (1986) and Schmidt number from Khalil and Rasmussen (1998).

2.3 Isotopic tracers and fractionation

Fractionation occurs between \(^{12}\text{C}\) and \(^{13}\text{C}\) during the biological fixation of dissolved carbon (as \( \text{CO}_2(\text{aq}) \)) to form organic and inorganic (carbonate) carbon, as well as during air-sea gas exchange. For the production of organic carbon (both as POC and DOC) we adopt the fractionation scheme of Ridgwell (2001):

\[ \delta^{13}\text{C}_{\text{POC}} = \delta^{13}\text{C}_{\text{CO}_2(\text{aq})} - \varepsilon_f + (\varepsilon_f - \varepsilon_d) \cdot \frac{K_Q}{[\text{CO}_2(\text{aq})]} \] (14)

where \( \delta^{13}\text{C}_{\text{CO}_2(\text{aq})} \) and \([\text{CO}_2(\text{aq})]\) are the isotopic composition and concentration of \( \text{CO}_2(\text{aq}) \), respectively. \( \varepsilon_f \) and \( \varepsilon_d \) are fractionation factors associated with enzymic intercellular carbon fixation and \( \text{CO}_2(\text{aq}) \) diffusion, respectively, and assigned values of \( \varepsilon_f = 25\%e \) and \( \varepsilon_d = 0.7\%e \) (Rau et al., 1996, 1997). \( K_Q \) is an empirical approximation of the model of Rau et al. (1996, 1997) as described by Ridgwell (2001):

\[ K_Q = 2.829 \times 10^{-10} - 1.788 \times 10^{-7} \cdot T + 3.170 \times 10^{-5} \cdot T^2 \] (15)

with \( T \) the ocean temperature in Kelvin.

For CaCO₃, we adopt a temperature-dependent fractionation for calcite following Mook (1986). The air-sea \(^{13}\text{C}/^{12}\text{C}\) fractionation scheme follows that of Marchal et al. (1998), with the individual fractionation factors all taken from Zhang et al. (1995). Solution of the isotopic partitioning of carbon between the different aqueous carbonate species (i.e., \( \text{CO}_2(\text{aq}), \text{HCO}_3^-, \text{CO}_3^{2-} \)) follows Zeebe and Wolf-Gladrow (2001).

For radiocarbon, the \(^{14}\text{C}/^{12}\text{C}\) fractionation factors are simply the square of the factors calculated for \(^{13}\text{C}/^{12}\text{C}\) at every step. Radiocarbon abundance also decays with a half-life of 5730 years (Stuiver and Polach, 1977). We report radiocarbon isotopic properties in the \( \Delta^{14}\text{C} \) notation, which we calculate directly from model-simulated \( \delta^{13}\text{C} \) and \( \delta^{14}\text{C} \) values by:

\[ \Delta^{14}\text{C}=1000 \cdot \left( \frac{\delta^{14}\text{C}}{1000} \cdot \frac{0.975^2}{(1 + \delta^{13}\text{C}/1000)^2} - 1 \right) \] (16)

which is the exact form of the more commonly used approximation: \( \Delta^{14}\text{C}=\delta^{14}\text{C}-2 \cdot (\delta^{13}\text{C} + 25) \cdot (1 + \delta^{14}\text{C}/1000) \) (Stuiver and Polach, 1977). \( \delta^{14}\text{C} \) is calculated using \( \delta^{14}\text{C}=\left( \text{As}/\text{Abs} - 1 \right) \cdot 1000 \), where \( \text{As} \) is the (model-simulated) sample activity and \( \text{Abs} \) is the absolute international standard activity, related to the activity of the oxalic acid standard (\( \text{AOx} \)) by \( \text{Abs}=0.95 \cdot \text{AOx} \). \( \text{AOx} \) is assigned a ratio of 1.176 \times 10⁻¹² (Key et al., 2004).

2.4 Definition of the aqueous carbonate system

The aqueous carbonate chemistry scheme used in the calibration is that of Ridgwell (2001). In this, alkalinity is defined following Dickson (1981) but excluding the effect of NH₃, HS⁻, and S²⁻. The set of carbonate dissociation constants are those of Mehrbach et al. (1973), as refitted by Dickson and Millero (1987), with \( p\text{H} \) calculated on the seawater \( p\text{H} \) scale (pHWS). Numerical solution of the system is via an implicit iterative method, seeded with the hydrogen ion concentration \( ([\text{H}^{+}]) \) calculated from the previous timestep (Ridgwell, 2001) (or with \( 10^{-7.8} \) from a “cold” start). We judge the system to be sufficiently converged when \( ([\text{H}^{+}]) \) changes by less than 0.1% between iterations, which typically occurs after just 2 or 3 iterations. This brings \( p\text{H} \) and \( f\text{CO}_2 \) to within ±0.001 units (pHWS) and ±0.2 µatm, respectively, compared to calculations made using the model of Lewis and Wallace (1998).

Dissolved calcium, total boric acid, sulphate, and fluorine can all be selected as prognostic tracers in the ocean biogeochemistry model and their oceanic distributions simulated...
explicitly. Here, we do not select them, and instead estimate their concentrations from salinity (Millero, 1982, 1995) in order to solve the aqueous carbonate system. Furthermore, because we do not consider the marine cycling of silicic acid (H$_4$SiO$_4$) here we implicitly assume a zero concentration everywhere. The error in atmospheric CO$_2$ induced by this simplification compared to prescribing observed concentrations of H$_4$SiO$_4$ (Conkright et al., 2002) in the model is $<1$ µatm.

Subsequent to the EnKF calibration presented here, refinements have been made to the representation of aqueous carbonate chemistry in the GENIE-1 model including a more comprehensive definition of alkalinity – still following Dickson (1981) but now the only exclusion being the contribution from S$^{2-}$. The set of carbonate dissociation constants remain those of Mehrbach et al. (1973) (refitted by Dickson and Millero, 1987), the apparent ionization constant of water follows Millero (1992), and pH is calculated on the seawater pH scale: otherwise, all other dissociation constants and assumptions follow those adopted in Zeebe and Wolf-Gladrow (2001). The impact of these model parameterization changes has only a very minor impact on the marine carbon cycle: equivalent to a reduction in mean ocean DIC of $<1$ µmol kg$^{-1}$ with atmospheric CO$_2$ held constant at 278 ppm (see Sect. 4.1).

3 Data assimilation and calibration of the marine carbon cycle

The degree of spatial and temporal abstraction inherent in representations of complex global biogeochemical processes inevitably gives rise to important parameters whose values are not well known a priori. Because of the computational cost of most 3-D ocean biogeochemical models, calibration of the poorly known parameters usually proceeds by trial-and-error with the aid of limited sensitivity analysis. The relative speed of the GENIE-1 model allows us to explore a new efficient and optimal approach to this problem by assimilating 3-D fields of marine geochemical data using a version of the ensemble Kalman filter which has been developed to simultaneously estimate multiple parameters in climate models.

In the following sections it should be noted that the ensemble Kalman filter (EnKF) methodology is described in detail in Annan et al. (2005a) where it is applied to identical twin testing of the GENIE-1 climate model as proof-of-concept for the technique. Subsequent application of the EnKF to optimization of the climate component of the GENIE-1 model is described in Hargreaves et al. (2004). Here we simply apply the same EnKF methodology to the ocean biogeochemistry as was performed in this previous calibration of the model climatology.

3.1 EnKF methodology

The model parameters were optimized with respect to 3-D data fields that are available for the present day climatological distributions of phosphate (PO$_4$) (Conkright et al., 2002) and alkalinity (ALK) (Key et al., 2004) in the ocean. The data were assimilated into the model using an iterative application of the ensemble Kalman filter (EnKF), which is described more fully in Annan et al. (2005a, b). The EnKF was originally introduced as a state estimation algorithm (Evensen, 1994). We introduce the parameters into the analysis simply by augmenting the model state with them. The EnKF solves the Kalman equation for optimal linear estimation by using the ensemble statistics to define the mean and covariance of the model’s probability distribution function. In other words, the resulting ensemble members are random samples from this probability distribution function. Although this method is only formally optimal in the case of a linear model and an infinite ensemble size, it has been shown to work well in cases similar to ours (Hargreaves et al., 2004; Annan et al., 2005a).

The parameter estimation problem studied here is a steady state problem, somewhat different in detail (and in principle simpler) than the more conventional time-varying implementations of the EnKF. However, the prior assumption of substantial ignorance, combined with the nonlinearity of the model and high dimensionality of the parameter space being explored, means that a direct solution of the steady state problem does not work well. Therefore, an iterative scheme has been developed which repeats a cycle of ensemble inflation, data assimilation and model integration over a specified time interval, in order to converge to the final solution. (See the references mentioned above for the technical details.) As demonstrated in Annan et al. (2005a) and Annan and Hargreaves (2004), this iterative method converges robustly to the correct solution in identical twin testing. Further applications using real data have also been successful with a range of different models (e.g., Hargreaves et al., 2004; Annan et al., 2005b, c). The method is relatively efficient, requiring a total of approximately 100 times the equilibrium time of the model to converge, and it is exact in the case of a linear model and a large ensemble size.

This method has already been used to calibrate the physical part of this model with long-term average climatological data to produce a reasonable representation of the preindustrial climate (Hargreaves et al., 2004). For this work we use the ensemble mean parameter set derived using the same method but with a slightly more recent version of the model following the correction of an error in the equation of state for seawater. The result of this is a $\sim$2 Sv stronger Atlantic meridional overturning compared to Hargreaves et al. (2004) and closer agreement with observational estimates. The revised parameter values that define the physical climate model are detailed in Table 2 and are kept fixed during the calibration of the biogeochemical model.
As mentioned above, the EnKF ensemble randomly samples the probability distribution function defined by the model, data, and prior assumptions. Therefore, the ensemble members do not themselves converge to the optimum but instead sample the region around it, with the ensemble mean being a good estimate of the optimum. Where all uncertainties are well defined, the spread of the ensemble members indicates the uncertainty surrounding this optimum. However, this is not the case here, since many model parameterizations are poorly understood and may be inadequate in various ways. A good example is the uncertainty surrounding the role of dust and the marine iron cycle (Jickells et al., 2005). The formulation of the biogeochemical model is thus inherently more uncertain than that of a physical model, and, at this stage, there is no clear way to estimate the true uncertainty of the calibrated model. We therefore use the EnKF to produce a single calibrated model version, taking the mean of this ensemble as an estimate of the best parameters.

The marine carbon cycle is calibrated against long-term average observations of PO$_4$ (Conkright et al., 2002) and ALK (Key et al., 2004) distributions in the ocean. We chose these data targets on the basis that PO$_4$ will help constrain the cycling of organic matter within the ocean, while ALK will (primarily) help constrain the cycling of CaCO$_3$ in the ocean. We assume that their observed distributions are relatively unaffected by anthropogenic change (Orr et al., 2005) (although see Feely et al., 2004). To create the PO$_4$ and ALK assimilation targets, we transformed the 3-D data-sets of Conkright et al. (2002) and Key et al. (2004), respectively, to the GENIE-1 model grid, and salinity-normalized the data. For the surface data target, although the model surface layer is 175 m thick, the observed data is integrated only over the uppermost 75 m for calculating surface boundary conditions – the depth assumed in the OCMIP-2 protocol as the nominal consumption depth separating the production zone (above) from the consumption zone (below) (Najjar and Orr, 1999).

The calibration of the biogeochemistry used an ensemble size of 54, which was chosen primarily for computational convenience. Ocean chemistry was initialized with uniform concentrations of: 2244 µmol kg$^{-1}$ DIC (estimated pre-Industrial) (Key et al., 2000), 2363 µmol kg$^{-1}$ ALK (Key et al., 2000), 2.159 µmol kg$^{-1}$ PO$_4^{3-}$ (Conkright et al., 2001), and 169.6 µmol kg$^{-1}$ O$_2$ (Conkright et al., 2002). Initial concentrations of dissolved organic matter are zero, as are $\delta^{13}$C and $\delta^{14}$C. Atmospheric $pO_2$ was initially set at 0.2095 atm and allowed to evolve freely in response to net air-sea gas exchange thereafter. Atmospheric CO$_2$ was continually restored to a value of 278 ppm throughout the assimilation. The parameters we considered in the EnKF assimilation as well as our prior assumptions regarding their likely values are listed in Table 1.

### Table 2. EnKF calibrated climate parameters in the GENIE-1 model.

<table>
<thead>
<tr>
<th>Parameter$^a$</th>
<th>Original posterior mean$^b$</th>
<th>Revised posterior mean$^c$</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Ocean</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Wind-scale</td>
<td>1.67</td>
<td>1.93</td>
</tr>
<tr>
<td>Isopycnal diffusion (m$^2$ s$^{-1}$)</td>
<td>4,126</td>
<td>4,489</td>
</tr>
<tr>
<td>Diapycnal diffusion (m$^2$ s$^{-1}$)</td>
<td>1.81 × 10$^{-5}$</td>
<td>2.72 × 10$^{-5}$</td>
</tr>
<tr>
<td>$1/\eta$ friction (days)</td>
<td>3.43</td>
<td>2.94</td>
</tr>
<tr>
<td><strong>Atmosphere</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>T diffusion amplitude (m$^2$ s$^{-1}$)</td>
<td>3.75 × 10$^6$</td>
<td>4.67 × 10$^6$</td>
</tr>
<tr>
<td>T diffusion width (radians)</td>
<td>1.31</td>
<td>1.08</td>
</tr>
<tr>
<td>T diffusion slope</td>
<td>0.07</td>
<td>0.06</td>
</tr>
<tr>
<td>T advection coefficient</td>
<td>0.06</td>
<td>0.11</td>
</tr>
<tr>
<td>$q$ diffusion (m$^2$ s$^{-1}$)</td>
<td>1.75 × 10$^6$</td>
<td>1.10 × 10$^6$</td>
</tr>
<tr>
<td>$q$ advection coefficient</td>
<td>0.14</td>
<td>0.23</td>
</tr>
<tr>
<td>FWF adj (Sv)</td>
<td>0.28</td>
<td>0.23</td>
</tr>
<tr>
<td><strong>Sea-ice</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sea ice diffusion (m$^2$ s$^{-1}$)</td>
<td>6,249</td>
<td>6,200</td>
</tr>
</tbody>
</table>

$^a$ see Edwards and Marsh (2005) and Lenton et al. (2006)
$^b$ Hargreaves et al. (2004)
$^c$ parameter values employed here (re-calibrated as per Hargreaves et al., 2004, but using a corrected equation of state – see Sect. 3.1)

3.2 Results of the EnKF assimilation

The mean and standard deviation of the ensemble values of the controlling biogeochemistry parameters are listed in Table 1. Most of the parameters showed only weak correlations in the posterior ensemble, with the striking exception of $r_0$ CaCO$_3$:POC$^{-1}$ and $\eta_1$ as shown in Fig. 4. We can trace this relationship back to Eqs. (8) and (9), where it implies that $r_0$ CaCO$_3$:POC$^{-1}$; $(\Omega-1)^\eta_1$ is close to being constant. We interpret this to mean that although total global production of carbonate is relatively tightly constrained by the data, the spatial variation, which is a function of local saturation state (\Omega) in the model, is not as well constrained.

We explored the sensitivity of the model calibration to the assimilated data by using the ALK data of Goyet et al. (2000) as an alternative to Key et al. (2004), although we omitted assimilating surface model layer ALK in this case because the Goyet et al. (2000) data-set is valid only below the mixed layer depth in the ocean. We found that the marine carbon cycle model and GLODAP PO$_4$ data set (Key et al., 2004) were less consistent with the Goyet et al. (2000) ALK distributions compared to the GLODAP ALK data (Key et al., 2004). Assimilation of the latter data rather than the former reduced the root mean square difference between model ALK and PO$_4$ fields taken together by about 10%. Most of the improvement occurred in the fit of the model to the ALK data set but there was also a slight improvement (1%) in the fit of the PO$_4$
Fig. 3. Data assimilation of PO$_4$ and ALK. The top panel shows the basin averaged meridional-depth distribution of phosphate (PO$_4$), with the model simulation immediately below the respective observations (Conkright et al., 2002). The bottom panel shows the basin averaged meridional-depth distributions of alkalinity (ALK) (Key et al., 2004). Note that we plot this data as actual concentrations, whereas the assimilation is carried out on salinity-normalized values.

data set to the model output. We interpret this in terms of the interpolation procedure used by Key et al. (2004) producing a more self-consistent data-set compared to the empirically based reconstruction of Goyet et al. (2000).

Analysis of ocean respiration patterns led Andersson et al. (2004) to propose a double exponential as a useful description for the profile of POC flux with depth. We tested a parameterization for the remineralization of POC in which each of the two fractions was assigned a characteristic (exponential decay) length scale. The mean EnKF calibrated remineralization length scales were 96 m and 994 m, compared to values of 55 m and ~2200 m determined by Andersson et al. (2004). However, the coarse vertical resolution of the model employed here means that we cannot place any confidence in length scales shorter than a few hundred meters. The improved fit to upper ocean PO$_4$ gradients using a double exponential in the EnKF comes at the expense of a rather negligible POC flux to the ocean floor because of the poor data constraint provided by the relatively weak gradients of PO$_4$ in the deep ocean (e.g., see Fig. 3).

Achieving an appropriate particulate flux to the ocean floor is critical to the determination of sedimentary carbonate content and organic carbon burial (see: Ridgwell and Hargreaves, 2007). We therefore retain the Eq. (6) parameterization in which one of the fractions has a fixed and effectively infinite length scale. An identical assumption of exponential decay combined with a recalcitrant fraction was employed in the sediment trap analysis of Lutz et al. (2002). Although they found considerable regional variations in vertical transport, the mean fraction of POC export reaching the abyssal ocean in their analysis was 5.7%, which is almost identical to our EnKF calibrated value of 0.055 (5.5%) for parameter $l_{POC}$ (Table 1). Furthermore, the mean depth scale of decay of the remaining 94.3% of exported POC was found to be 317 m (Lutz et al., 2002), which while somewhat shorter than our calibrated value of 556 m (parameter $l_{POC}$) is of the same order. Thus, independent analysis of annual POC sinking fluxes is largely consistent with our inverse modelling of global PO$_4$ distributions, particularly at depth. We believe that further improvement, particularly to the shallow length-scale will be possible within the context of an ocean circulation model with higher vertical resolution (e.g., 16 levels).

4 Discussion

4.1 The calibrated baseline state of the model

The calibrated model achieves appropriate zonally-averaged nutrient (PO$_4$) distributions for many regions of the ocean, particularly the Equatorial and North Atlantic, the Indian Ocean, and Equatorial Pacific and deep Pacific (Fig. 3). The predicted surface distribution of PO$_4$ (Fig. 5A, B) also agrees
with the observations to a first-order (Conkright et al., 2002), with relatively high (>0.5 μmol kg$^{-1}$) concentrations in the Southern Ocean, North Pacific, North Atlantic, and Eastern Equatorial Pacific, and nutrient depletion (<0.5 μmol kg$^{-1}$) in the mid latitude gyres. However, over-estimated low latitude upwelling results in excess (>0.5 μmol kg$^{-1}$) PO$_4$ in the Western Equatorial Pacific and Equatorial Indian Ocean. In addition, there is no representation of iron limitation, critical in the modern ocean in restricting nutrient depletion in the “High Nutrient Low Chlorophyll” (HNLC) regions of the ocean such as the North and Eastern Equatorial Pacific, and Southern Ocean (Jickells et al., 2005). The calibration must then strike a compromise – PO$_4$ uptake (scaled by the parameter $u_0^{PO_4}$ in Eq. 3) must be sufficiently low that nutrients remain unused in the HNLC regions, yet at the same time, high enough to deplete nutrients elsewhere. The result is that PO$_4$ is generally slightly lower than observed in the HNLC regions but slightly too high elsewhere. However, overall, global export production of particulate organic carbon (POC) is 8.91 GtC (PgC) yr$^{-1}$, consistent with recent estimates (e.g., Amout et al., 2003; Schmittner et al., 2005; Jin et al., 2006).

The lack of an explicit representation of iron limitation in conjunction with the simplistic treatment of light limitation is an obvious deficiency in this initial incarnation of the marine biogeochemical model. Despite this, the general structure of PO$_4$ is still reasonably reproduced. Thus, we would argue that the single nutrient (PO$_4$) limitation scheme is adequate for developing and exploring geochemo-assimilation techniques such as the EnKF.

For alkalinity (ALK), we achieve a generally reasonable simile of the zonally-averaged distributions in each ocean basin (Fig. 3). The main areas of model-data mismatch concern surface concentrations. These are primarily caused by deficiencies in the climate model simulation of surface ocean salinities (see Hargreaves et al., 2004). This is because we assimilate salinity-normalized alkalinity, which means that deviations of model-predicted salinity from observations degrades the fidelity of simulation of the (non salinity-normalized) ALK field. (The importance of salinity-related variations is much less for PO$_4$ because the biologically-induced range in concentrations is an order of magnitude greater.) Global pre-industrial CaCO$_3$ production is 1.21 PgC yr$^{-1}$, which falls towards the centre of the production budget uncertainty of 0.8–1.4 PgC yr$^{-1}$ proposed by Feely et al. (2004). It is also very close to the preferred estimate of 1.14 PgC yr$^{-1}$ diagnosed by Jin et al. (2006) from global nutrient and alkalinity distributions. The mean global CaCO$_3$:POC export ratio (Fig. 5C) is a little less than 0.14, supporting suggestions (Yamanaka and Tajika, 1996; Sarmiento et al., 2002; Jin et al., 2006) that the ratio is rather lower than the value of around 0.2–0.25 characterizing many models (e.g., Broecker and Peng, 1986; Archer et al., 1998; Heinze et al., 1999; Maier-Reimer, 1993). Of the predicted 1.21 PgC of biogenic carbonate exported annually from the surface ocean, approximately 0.62 PgC yr$^{-1}$ reaches 2000 m depth, slightly higher than estimates of 0.4 PgC yr$^{-1}$ based on sediment trap measurements (Feely et al., 2004).
Fig. 6. Basin averaged meridional-depth distributions of dissolved oxygen in the ocean: observed (top) (Conkright et al., 2002) and model-simulated for the year 1994 (bottom).

With a mean ocean ALK of 2363 µmol kg⁻¹ (Key et al., 2004), a (pre-Industrial) atmospheric CO₂ concentration of 278 ppm requires a mean ocean DIC of 2214 µmol kg⁻¹ compared to an observationally-based estimate of 2244 µmol kg⁻¹ (Key et al., 2004). Some 14 µmol kg⁻¹ of this apparent DIC difference is explained by mean surface alkalinity being slightly lower than observed (2301 µmol kg⁻¹) compared to an average of 2310 µmol kg⁻¹ over the uppermost 75 m of the modern ocean (Conkright et al., 2002), itself mainly a consequence of the low surface salinities simulated by the climate model (particularly in the Atlantic) (Hargreaves et al., 2004). Use of a revised “Redfield ratio” value of 1:117 linking phosphorus and carbon in organic matter (Anderson and Sarmiento, 1994) increases DIC by nearly 7 µmol kg⁻¹, while using observed ocean surface temperatures (in place of model simulated temperatures) in the calculation of air-sea gas exchange accounts for another 2 µmol kg⁻¹ DIC. The residual model-data difference (9 µmol kg⁻¹ DIC) is comparable to the uncertainty in the GLODAP data-sets of 5–10 µmol kg⁻¹ (Key et al., 2004). However, we cannot rule out convective ventilation or insufficient sea-ice extent in the Southern Ocean, lack of seasonality, and/or the thickness of the model surface ocean layer, contributing to driving the DIC inventory slightly too low (see discussion in Sect. 4.3).

4.2 The marine biogeochemical cycling of O₂

The model reproduces the main features of the observed distribution of dissolved oxygen (O₂) in the ocean (Fig. 6), with elevated concentrations associated with the sinking and subduction of cold and highly oxygenated waters in the North Atlantic, as well as the occurrence of intermediate depth minima at low latitudes in all three ocean basins. Deep ocean O₂ concentrations are almost everywhere correct to within ~40 µmol kg⁻¹ of observations (Conkright et al., 2002). The area of greatest model-data mismatch concerns the ventilation of intermediate waters in the Southern Ocean and North Pacific. In addition, while the magnitude of O₂ depletion of the oxygen minimum zone in the northern Indian Ocean is correctly predicted, it lies at too shallow a depth and is too restricted in vertical extent. Overall, however, the quality of our O₂ simulation compares favorably with predictions made by considerably more computationally expensive ocean circulation models (e.g., Bopp et al., 2002; Meissner et al., 2005).

4.3 Ocean circulation and biogeochemical cycling in very coarse resolution models

While large-scale heat transports in the ocean are well captured (Hargreaves et al., 2004), the steady-state distribution of dissolved oxygen (Fig. 6) highlights the excess ventilation of intermediate depths above ca. 1500 m in parts of the Southern Ocean and North Pacific in the GENIE-1 model. We can demonstrate that ocean transport rather than processes associated with organic production and/or remineralization is primarily responsible by simulating the transient uptake of CFC-11 and CFC-12 from the atmosphere. We use a similar methodology to OCMIP-2 (Dutay et al., 2002), in which the partial pressure of CFCs in the atmosphere follow observations for the years 1932 to 1998 (Walker et al., 2000) and with air-sea gas exchange calculated explicitly according to Eqs. (12) and (13). The model-predicted 1994 global ocean CFC-11 inventory is 0.88×10⁹ mol, compared to the observationally-based estimate of 0.55±0.08×10⁹ mol (Key et al., 2004; Willey et al., 2004), with Southern Ocean intermediate depths accounting for more than half of the model-data mismatch.

Anthropogenic CO₂ uptake by the ocean can be similarly assessed by forcing atmospheric CO₂ to follow the historical trajectory between 1765 and 2000: constructed from the data of Enting et al. (1990) up until 1994, and Keeling and Whorf (2004) thereafter, and allowing climate to respond. We find a year 1994 anthropogenic CO₂ inventory of 171 PgC compared to a recent data based estimate of 118±19 GtC (Sabine et al., 2004). The areas of model-data mismatch are illustrated in Fig. 7 by the water column integrals of anthropogenic CO₂ – there is excessive ventilation.
of intermediate waters in the North (East) Pacific and particularly in the Southern Ocean where the excess (compared to observations) South of 41° S is 21 PgC. The main areas of anthropogenic CO₂ model-data mismatch are similar to those of CFC-11 and are also consistent with the locations of excess O₂ concentrations in the model – see Fig. 6. However, subduction of anthropogenically CO₂ enriched water to depth in the North Atlantic is well reproduced in addition to there being a reasonable match to the observations at low and mid latitudes.

There may be a fundamental limitation as to how coarse a resolution a 3-D ocean circulation model may have and still be able to simulate decadal-scale uptake processes accurately. This is because the stability of the water column at high latitudes appears to be very sensitive to the vertical resolution. For instance, it has recently been demonstrated (Müller et al., 2006) that a significant improvement in transient tracer uptake is obtained in a very similar ocean GCM to that used here by increasing the number of vertical levels from 8 to 32, although the use of observed climatological surface boundary conditions and separation of eddy-induced and isopycnal mixing effects may also have been critical in this. We have investigated this further by repeating the OCMIP-2 transient tests with 16 rather than 8 levels in the ocean (Edwards and Marsh, 2005) and find that the radiocarbon properties of the deep ocean indeed become much closer to the observations, with model-predicted Δ¹⁴C values below 2000 m in the Southern Ocean of ca. -140‰ and ca. -200‰ in the North Pacific.

Thus, simply increasing the vertical resolution appears to be an effective strategy in creating a more stable water column at high latitudes. However, this comes at a computational price because there is an approximate doubling of the number of cells in the ocean. To address century to millennial-scale (and longer) questions, particularly those involving interaction with deep-sea sediments we retain the 8-level ocean by default, but recognize the limitations we have identified in 8-level vs. 16-levels in representing some decadal (and sub-decadal) processes.

We have no grounds for suspecting that the excessive rates of high latitude intermediate water ventilation are likely to introduce a fundamental bias to the biological (PO₄) cycle. For instance, if the supply of PO₄ and alkalinity into the euphotic zone was generally substantially overestimated, then for the EnKF to be able to achieve a reasonable fit to the observed

---

**Fig. 7.** Water column integrated anthropogenic CO₂ inventory in the ocean for the year 1994: observed (top) (Key et al., 2004) and model-simulated (bottom).

**Fig. 8.** Ocean CO₂ invasion behavior. Predicted evolution of atmospheric CO₂ following an instantaneous doubling to 556 ppm. The atmospheric CO₂ behavior of GENIE-1 is consistent with the four ocean carbon cycle models which ran this experiment out as part of OCMIP-2 (Doney et al., 2004). The OCMIP-2 models are those of: Stocker et al. (1992), Yamanaka and Tajika (1996), Gordon et al. (2000), and Schlitzer (2002). We deliberately do not differentiate between the different OCMIP-2 models here to highlight the comparison between GENIE-1 and models having much greater horizontal and/or vertical resolution, rather than discuss the reasons for the differences amongst the OCMIP-2 models. All OCMIP-2 trajectories are therefore plotted as identical grey lines. The 8-level version of GENIE-1 is shown as a thin solid black line, while the improvement in decadal-scale CO₂ uptake resulting from doubling the vertical resolution is illustrated by the thin dashed black line. The bottom plot shows the same data, but plotted on a logarithmic time axis to help visually separate out the different time-scales of CO₂ invasion.
3-D structure of PO₄ and ALK in the ocean, global export would have to be similarly overestimated. In contrast, as noted in Sect. 4.1, the diagnosed export of POC lies towards the centre of the spread of current (high resolution) GCM estimates while the estimated global export flux of CaCO₃ is near identical to the recent GCM data-assimilation study of Jin et al. (2006), itself one of the lowest values for CaCO₃ export flux proposed. In addition, the mean fraction of exported POC reaching the deep ocean as well as the shallower remineralisation length-scale diagnosed from sediment trap analysis (Lutz et al., 2002) are consistent with our inverse-modelled parameter values (Sect. 3.2).

We confirm our assumptions regarding the century-scale (and beyond) applicability of an 8-level configuration by running an OCMIP-2 test, in which an abiotic ocean is brought to steady state with an atmospheric CO₂ concentration of 278 ppm, atmospheric CO₂ instantaneously doubled, and the invasion of CO₂ into the ocean then tracked for 1000 years (Orr, 2002). We find that atmospheric CO₂ concentrations diverge between our 8-level model and the OCMIP-2 models over the first few decades of uptake, but the trajectories re-converge thereafter (Fig. 8), making the century-scale (and longer) CO₂ behavior in GENIE-1 indistinguishable from the differing OCMIP-2 model behaviors. This is consistent with simulations we have made of radiocarbon distributions in the ocean (not shown) in which predicted (natural) Δ¹⁴C properties of deep ocean water masses also fall within the range of OCMIP-2 model behavior (Matsumoto et al., 2004). That we reproduce the year 2100 prediction for the strength of “CO₂-calcification” feedback (the enhancement of CO₂ uptake from the atmosphere due to reduced calcification) made by a much higher resolution ocean model (Ridgwell et al., 2006) also supports our assertion regarding the time-scales of applicability of the fast 8-level version of the GENIE-1 model.

5 Conclusions and perspectives

We have constructed a 3-D ocean based Earth System Model that includes key feedbacks between marine carbon cycling, atmospheric CO₂, and climate, yet even when simulating 12 biogeochemical tracers in the ocean together with the exchange with the atmosphere of 6 gaseous tracers, achieves better than 1000 yr simulated per (2.4 GHz) CPU hour on a desktop computer. The computational speed of the GENIE-1 model has enabled us to calibrate via an Ensemble Kalman Filter (EnKF) method the critical parameters controlling marine carbon cycling. The EnKF has also allowed us to judge the consistency of available global alkalinity data-sets with our mechanistic representation of global biogeochemical cycling, and has determined the a priori unknown relationship between η and \( r_0^{CaCO_3:POC} \) in our parameterization of carbonate production. Overall, global particulate organic carbon and inorganic (carbonate) carbon export predicted by the calibrated model is consistent with recent data- and model-based estimates and despite the very coarse resolution of the ocean model, the distribution of dissolved O₂ in the ocean generally compares favorably with the data and with the results of much more computationally expensive 3-D ocean circulation models.

However, important caveats are raised by our analysis regarding very low resolution configurations of the GENIE-1 model. As highlighted by the results of the OCMIP-2 tests presented in Sect. 4.3, the transient tracers (CFCs, anthropogenic CO₂, and bomb Δ¹⁴C) delineate apparent deficiencies in the ventilation of intermediate waters in the Southern Ocean and North Pacific. This means that estimates made by this version of the model of the anthropogenic CO₂ uptake by the ocean on time-scales much shorter than ~100 years should be treated with caution. We find that an increased number of vertical levels produces a significan improvement in simulated transient tracer behaviour, suggesting that there is a limit to how low the vertical resolution of a GCM may be and yet still maintain adequate water column stability everywhere. Because our results highlight the importance of global biogeochemical cycling in providing a sensitive diagnostic of ocean circulation in models, we propose that more physically realistic models might be achieved by “co-tuning” key physical and biogeochemical parameters.

A second caveat concerns our highly-parameterized single nutrient (PO₄) scheme for marine biological production. While this simple scheme gave us important advantages in developing the EnKF calibration of the model against global biogeochemical observations, it has the obvious disadvantage that there is no representation of iron limitation, which is critical in the modern ocean in restricting nutrient depletion in the HNLC regions of the ocean such as the North and Eastern Equatorial Pacific, and Southern Ocean. The consequence of this is that the overall quality of fitting of global PO₄ distributions in the EnKF is somewhat degraded because the appropriate mechanism for achieving relatively high PO₄ in the HNLC regions is absent. This will be addressed in the first instance by incorporating the biogeochemical scheme and Fe cycle of Parekh et al. (2005) into the GENIE-1 model. Further planned developments will include an optional functional type ecosystem model as well as the addition of the limiting nutrients nitrate and silicic acid.

The current PO₄ scheme is also somewhat non-ideal when considering long (>10 000 years) time-scales because the marine organic carbon cycle is configured as a closed system, which can have a very different dynamical response from that of an open system in which there is a continual throughput (Hotinski et al., 2000). In-progress developments to this model will address both the sink of PO₄, via a mechanistic prediction of the burial of organic matter in marine sediments as well as its source, via a representation of terrestrial rock weathering. Further extensions to the representation of marine sediments and their interaction with the ocean will include the incorporation of a representation of suboxic
diagenesis following Archer et al. (2002) and the regeneration of phosphate from anoxic sediments.

Finally, although the particular configuration of the GENIE-1 model presented here is extremely fast compared to other coupled and even off-line GCMs models of ocean carbon cycling, it is still too numerically expensive to enable application to >1 Myr time-scales. Box models will this remain essential tools for addressing the very longest time-scale questions. However, the speed and multi-tracer capabilities make the GENIE-1 model a viable alternative to marine biogeochemical box models for questions regarding the controls on atmospheric $\text{CO}_2$ on $\sim$100 to multi 100 000 year time-scales.

Acknowledgements. A. Ridgwell acknowledges a University Research Fellowship from the Royal Society as well as support from Canada Research Chairs and the Canadian Foundation for Climate and Atmospheric Sciences. Development of the model was supported by the NERC e-Science programme (NER/T/S/2002/00217) through the Grid ENAbleq Integrated Earth system modelling (GENIE) project (http://www genie.ac.uk) and by the Tyndall Centre for Climate Change Research (Project TC IT 1.31). Computer facilities for the EnKF calculations were provided by JAMSTEC. We are indebted to the generosity of OCMP2 participants who made their unpublished model $\text{CO}_2$ results available to us: A. Ishida and Y. Yamanaka at the Frontier Research Center for Global Change, Japan (formerly “IGCR”), F. Joos at Climate and Environmental Physics, Bern, and R. Schlitzer of AWI-Bremenhaven.

Edited by: T. W. Lyons

References


