

# Skill assessment of three earth system models with common marine biogeochemistry

Roland Séférian · Laurent Bopp · Marion Gehlen · James C. Orr ·  
Christian Ethé · Patricia Cadule · Olivier Aumont · David Salas y Mélia ·  
Aurore Voldoire · Gurvan Madec

Received: 8 November 2011 / Accepted: 7 April 2012 / Published online: 8 June 2012  
© The Author(s) 2012. This article is published with open access at Springerlink.com

**Abstract** We have assessed the ability of a common ocean biogeochemical model, PISCES, to match relevant modern data fields across a range of ocean circulation fields from three distinct Earth system models: IPSL-CM4-LOOP, IPSL-CM5A-LR and CNRM-CM5.1. The first of these Earth system models has contributed to the IPCC 4th assessment report, while the latter two are contributing to the ongoing IPCC 5th assessment report. These models differ with respect to their atmospheric component, ocean subgrid-scale physics and resolution. The simulated vertical distribution of biogeochemical tracers suffer from biases in ocean circulation and a poor representation of the sinking fluxes of matter. Nevertheless, differences between

upper and deep ocean model skills significantly point to changes in the underlying model representations of ocean circulation. IPSL-CM5A-LR and CNRM-CM5.1 poorly represent deep-ocean circulation compared to IPSL-CM4-LOOP degrading the vertical distribution of biogeochemical tracers. However, their representations of surface wind, wind stress, mixed-layer depth and geostrophic circulations (e.g., Antarctic Circumpolar Current) have been improved compared to IPSL-CM4-LOOP. These improvements result in a better representation of large-scale structure of biogeochemical fields in the upper ocean. In particular, a deepening of 20–40 m of the summer mixed-layer depth allows to capture the 0–0.5  $\mu\text{gChl L}^{-1}$  concentrations class of surface chlorophyll in the Southern Ocean. Further improvements in the representation of the ocean mixed-layer and deep-ocean ventilation are needed for the next generations of models development to better simulate marine biogeochemistry. In order to better constrain ocean dynamics, we suggest that biogeochemical or passive tracer modules should be used routinely for both model development and model intercomparisons.

This paper is a contribution to the special issue on the IPSL and CNRM global climate and Earth System Models, both developed in France and contributing to the 5th coupled model intercomparison project.

**Electronic supplementary material** The online version of this article (doi:10.1007/s00382-012-1362-8) contains supplementary material, which is available to authorized users.

R. Séférian (✉) · L. Bopp · M. Gehlen ·  
J. C. Orr · C. Ethé · P. Cadule  
Laboratoire des Sciences du Climat et de l'environnement  
(LSCE/IPSL), Batiment 712, 91191 Gif sur Yvette, Paris, France  
e-mail: roland.seferian@lsce.ipsl.fr

R. Séférian · D. Salas y Mélia · A. Voldoire  
CNRM-GAME (Météo-France, CNRS), Toulouse, France

C. Ethé · G. Madec  
IPSL/LOCEAN, Paris, France

O. Aumont  
LPO/IRD/CNRS, Plouzané, France

G. Madec  
NOC, Southampton, UK

**Keywords** Marine biogeochemistry · Skill assessment · Earth system model · Climate · Ocean dynamics

## 1 Introduction

Earth system models are key to the IPCC assessment process but present multiple differences in term of resolution, subgrid-scale physics, and biogeochemical components. These differences affect their simulated modern climate and future projections. Thus it is essential to assess model skill, i.e., agreement between simulated and observed fields (Meehl et al. 2007; Randall et al. 2007).

Systematic assessment of numerical ocean models with data-based (e.g., Saba et al. 2010) and satellite-derived products (e.g., SeaWiFS, Carr et al. 2006) reveals large differences between ocean carbon models in terms of their ability to match relevant data fields (e.g., Sarmiento et al. 2000; Doney 2004; Schneider et al. 2008; Steinacher et al. 2010). Yet greater model complexity does not necessarily imply greater model skill (e.g., Friedrichs et al. 2007, 2009; Kriest et al. 2010). Unfortunately, when comparing ocean carbon models it is difficult to disentangle the main causes for disagreement because of their many differences, including resolution, parametrizations for subgrid-scale physics and biogeochemical components. This problem is exacerbated in Earth system models, where there are more model components, which differ as well.

To shed light on how simulated ocean circulation affects simulated ocean biogeochemistry, the Ocean Carbon-Cycle Model Intercomparison Project (OCMIP) compared simulations from different circulation models but with the same embedded diagnostic biogeochemical model (e.g., Najjar et al. 2007). This study illustrated how differences in surface forcing, subgrid-scale parametrizations, and resolution can drive discrepancies in simulated biogeochemical fields, rivaling those due to unrepresented biogeochemical processes. Since the OCMIP effort, marine biogeochemistry models have evolved (e.g. Hood et al. 2006) from models with one nutrient, one phytoplankton, one zooplankton, and one detritus (e.g., NPZD, Fasham et al. 1993) to multiple phytoplankton functional types and several limiting micronutrients (e.g., Le Qu  r   et al. 2005; Aumont and Bopp 2006). However, the OCMIP approach of comparing models with different circulation fields and the same biogeochemical model remains useful. Here, we use the same approach but in the context of the new generation of Earth system models that are being used in the ongoing IPCC assessment process in the framework of the Coupled Model Intercomparison Project (CMIP5, Taylor et al. 2009, 2011).

Although the full range of Earth system models taking part in CMIP5 use different ocean biogeochemical components, we take the OCMIP approach with three of them that use a common ocean biogeochemical model, PISCES (Aumont and Bopp 2006). In comparison to the diagnostic ocean biogeochemical model used in OCMIP, the nutrient cycling of PISCES is prognostic (i.e., it is not restored to observed nutrient concentrations). This biogeochemical model further includes a representation of the iron cycle and limitation by multiple nutrients (Aumont 2003; Schneider et al. 2008). Additionally, in the three selected Earth System Models, each of the ocean circulation models is based on the same code, although versions differ as to resolution and subgrid-scale physics (Madec 2008a, b; Marti et al. 2010; Voldoire et al. 2012; Dufresne et al. 2012). The largest differences in simulated ocean

circulation fields probably stem from each ocean model being coupled to a different atmospheric model. The three atmospheric models differ in horizontal and vertical resolution, physics parametrizations, and architecture.

This effort has two objectives: (1) to evaluate models in terms of how well they simulate present-day ocean biogeochemical fields, and (2) to shed light on the relative roles of physical or biogeochemical controls on marine biogeochemistry. Within this framework, we examine differences between modeled and observed physical fields (sea surface temperature, sea surface salinity, wind stress, short wave radiation, net air-sea exchange of heat, mixed-layer depth and water mass transport) and marine biogeochemical fields (air-sea exchange of CO<sub>2</sub> and O<sub>2</sub>, export of organic matter, and the 3-D distribution of biogeochemical tracers). These physical fields are known to have first-order effects on marine biogeochemistry (Sarmiento and Gruber 2006); these fields and their effects are sensitive to resolution, sub-scale physics, and subgrid-scale parametrizations. Model-data and model-model differences between the three models offer hints as for how to improve simulated marine biogeochemistry in the next generation of Earth system models.

## 2 Methods

### 2.1 Climate models

This study exploits three fully coupled 3-D atmosphere-ocean models, each using OASIS coupler (Valcke 2006). OASIS ensures space and time interpolation between models grids and time-steps. IPSL-CM4-LOOP was used during the IPCC Fourth Assessment Report (AR4; Meehl et al. 2007). The other two models, IPSL-CM5A-LR and CNRM-CM5.1, are the latest generation of climate models that will contribute to the IPCC Fifth Assessment Report (AR5). Main models characteristics are listed in Table 1 and detailed below.

#### 2.1.1 IPSL-CM4-LOOP

The IPSL-CM4-LOOP (IPSLCM4) model combines four major dynamical components of the climate system: atmosphere, land, ocean, and sea-ice. The atmospheric component is LMDZ-4 developed at LMD (Laboratoire de M  t  orologie Dynamique) with a horizontal resolution of 3.75° × 2.5° and 19 vertical levels (Hourdin et al. 2006). The land component is the global land-vegetation model ORCHIDEE (Organizing Carbon and Hydrology in Dynamic Ecosystems, Krinner et al. 2005), which simulates interactions between land and atmosphere. The ocean model, OPA8, has a horizontal resolution of 2°–0.5° at the Equator and 31 vertical levels (Madec et al. 1998). Vertical eddy and viscosity coefficients using a 1.5-order turbulent

**Table 1** Name, resolution and reference for the atmospheric, oceanic, sea-ice and land components of the three Earth system models

Model	Atmosphere	Ocean	Sea-Ice	Land
IPSLCM4 (Marti et al. 2010)	LMDZ-4 3.75° × 2.5° × 19L (Hourdin et al. 2006)	OPA8 2° × 2°cos( $\phi$ ) × 31L (Madec et al. 1998)	LIM2 2° × 2°cos( $\phi$ ) (Fichefet et al. 1997)	ORCHIDEE 3.75° × 2.5° (Krinner et al. 2005)
IPSLCM5 (Dufresne et al. 2012)	LMDZ-4 3.6° × 1.8° × 39L (Hourdin et al. 2006)	NEMOv3.2 2° × 2°cos( $\phi$ ) × 31L (Madec 2008a, b)	LIM2 2° × 2°cos( $\phi$ ) (Fichefet et al. 1997)	ORCHIDEE 3.6° × 1.8° (Krinner et al. 2005)
CNRMCM5 (Voldoire et al. 2012)	ARPEGE-CLIMATv5.2 T127 ~ 1.4° × 31L (Voldoire et al. 2012)	NEMOv3.2 1° × 1°cos( $\phi$ ) × 42L (Madec 2008a, b; Hewitt et al. 2011)	GELATO 1° × 1°cos( $\phi$ ) (Salas y Méliá 2002)	SURFEX ~ 1.4° (Noilhan and Mahfouf 1996)

closure scheme are prognostically computed (Blanke and Delecluse 1993). The sea-ice component is LIM2 (Louvain Ice Model, Fichefet and Morales-Maqueda 1997), which resolves three layers (one for snow and two others for sea-ice) and accounts for sea-ice dynamics. The model is described in more detail in Marti et al. (2010), but this version includes the marine carbon cycle model PISCES (detailed below). The IPSLCM4 is a fully coupled climate and carbon cycle model (Cadule et al. 2010). Thus carbon sources and sinks over land and ocean are included.

### 2.1.2 IPSL-CM5A-LR

The current version of the coupled model developed at IPSL is reference as to IPSL-CM5A-LR (IPSLCM5, Dufresne et al. 2012). It represents the most recent development of the previous version IPSLCM4. As before, IPSLCM5 combines the four major components of the climate system, but it now also includes a module for tropospheric chemistry, aerosols, and online interactions with land vegetation. The atmosphere and land models of IPSLCM5 are updated versions of those used in IPSLCM4, namely, the atmospheric general circulation model LMDZ (Hourdin et al. 2006) and the ORCHIDEE land-surface model (Krinner et al. 2005). The atmospheric and land components use the same regular horizontal grid with 96 × 96 points, representing a resolution of 3.6° × 1.8°, while the atmosphere has 39 vertical levels. The INCA (INteraction between Chemistry and Aerosol, e.g., Schulz 2007 and Szopa et al. 2012) model is used to simulate tropospheric greenhouse gases and aerosol concentrations, while stratospheric ozone is modeled by REPROBUS (Reactive Processes Ruling the Ozone Budget in the Stratosphere, Lefevre et al. 1994, 1998).

The oceanic component is based on the ocean part of the “Nucleus for European Modelling of the Ocean” (NEMOv3.2, Madec 2008a, b). It has a horizontal resolution of 2°–0.5° at the Equator and 31 vertical levels. NEMOv3.2 also includes the sea ice model LIM2 (Fichefet and

Maqueda 1997), and the marine biogeochemistry model PISCES. NEMOv3.2 uses a partial-step formulation (Barnier et al. 2006), which improves the representation of bottom bathymetry and thus water flow and effects of friction at the bottom of the ocean. Compared to the parametrizations of the mixed-layer dynamics used in OPA8, those used in NEMOv3.2 accounts for a double diffusion process (Merryfield et al. 1999), Langmuir cell (Axell 2002), and the contribution of surface wave breaking (Mellor and Blumberg 2004; Burchard and Rennau 2008). It also includes a parameterization of bottom intensified tidal-driven mixing, similar to Simmons et al. (2004), in combination with a specific tidal mixing parameterization in the Indonesian area (Koch-Larrouy et al. 2007, 2010). In addition, NEMOv3.2 includes prognostic interaction between the penetration of incoming shortwave radiation into the ocean and phytoplankton (Lengaine et al. 2009).

### 2.1.3 CNRMCM5.1

The CNRM-CM5.1 model (CNRMCM5, Voldoire et al. 2012 this issue) couples atmosphere, land, ocean, and sea-ice components. The atmospheric component is ARPEGE-CLIMATv5.2, which derives from ARPEGE (action de recherche petite echelle grande echelle, which means research project on small and large scales, Déqué et al. 1994). ARPEGE-CLIMATv5.2 is a spectral model using a resolution of about 1.4° in longitude and latitude, while vertically there are 31 levels. The land-surface component is SURFEX (SURFace EXternalisée, Manzi and Planton 1994; Noilhan and Mahfouf 1996), which prognostically simulates three types of surfaces: land surfaces, inland water bodies (e.g., lakes), ocean or seas. Atmospheric chemistry in CNRMCM5 includes only stratospheric ozone chemistry (Cariolle and Teysseire 2007).

The ocean component also uses NEMOv3.2 but in the ORCA1 configuration (Hewitt et al. 2011). In ORCA1, a nominal resolution of 1° at the equator is chosen to which a latitudinal grid refinement of 1/3° is added in the tropics. The vertical discretization accounts for 42 levels and also

includes partial-steps. Compared to IPSLCM5, this configuration includes an improvement of the Turbulent Kinetic Energy closure scheme (Madec 2008a, b) from Blanke and Delecluse (1993). That is, it allows a fraction of surface wind energy to penetrate below the base of the mixed-layer, improving the coupling between surface wind and mixed-layer depth. The sea-ice model is GELATO (Global Experimental Leads and ice for ATmosphere and Ocean, Salas M  lia 2002). GELATO is a multi-category model. Sea-ice dynamics is represented by the EVP rheology (Hunke and Dukowicz 2012). In CNRMCM5, the interaction between incoming shortwave radiation penetration into the ocean and phytoplankton (Lengaine et al. 2009) is assumed constant (chlorophyll concentration is equal to  $0.05 \text{ gChl L}^{-1}$ ).

## 2.2 Marine biogeochemistry model: PISCES

PISCES (Pelagic Interaction Scheme for Carbon and Ecosystem Studies) simulates the biogeochemical cycles of oxygen, carbon and nutrients using 24 state variables. Macronutrients (nitrate and ammonium, phosphate, and silicate) and the micronutrient iron limit phytoplankton growth and thus improve the representation of their dynamics (Aumont et al. 2003; Aumont and Bopp 2006).

PISCES has two sizes for phytoplankton (nanophytoplankton and diatoms), for which growth is parametrized using the Eppley et al. (1969) formulation but limited by external availability of nutrients. Diatoms differ from nanophytoplankton because they need silicon and more iron (Sunda and Huntsman 1997) and because they have higher half-saturation constants due to their larger mean size. In PISCES, zooplankton is distinguished by two size-classes: microzooplankton and mesozooplankton. The ratios between carbon, nitrate, and phosphate are considered constant and held to the values proposed by Takahashi et al. (1985) for all plankton classes. However, internal concentrations of iron are simulated prognostically in both phytoplankton classes as is silica in diatoms. Phytoplankton growth depends on external concentrations of nutrients and light availability, while chlorophyll concentration is simulated prognostically following Geider et al. (1998). PISCES also simulates semi-labile dissolved organic matter as well as small and big sinking particles, which differ in their sinking speeds (i.e., 3 and 50–200  $\text{m d}^{-1}$ , respectively). For these three compartments, PISCES imposes fixed stoichiometric ratios of O:C:N:P after Takahashi et al. (1985). However, concentrations of iron, silicon, and calcite are simulated prognostically inside the respective pools of sinking particles.

The fate of mortality and aggregation of nanophytoplankton depends on the proportion of the calcifying organisms. In PISCES, we assume that half of the organic

matter of the calcifiers is associated with the shell. Since calcite is denser than organic matter, 50 % of the dying calcifiers biomass is routed to the fast sinking particles. The same is assumed for the mortality of diatoms as a consequence of the higher density of biogenic silica. The sinking speed of the particles is not altered by their content in calcite and biogenic silicate, while their degradation depends on the local temperature. Remineralization of dissolved organic matter can occur within both oxic and anoxic waters. It depends on the local oxygen concentration and thus assumes that the degradation rates for oxic respiration and denitrification are identical. Differential sedimentations, turbulence coagulation mechanisms, and aggregation processes are considered for dissolved and particulate organic matter. Only the differential sedimentation is omitted for dissolved organic matter, because its contribution is almost negligible compared to turbulence.

In PISCES, inorganic carbon is represented by dissolved inorganic carbon, alkalinity and calcite. Aragonite is not considered for the chemical dissolution of calcium carbonate in the water column. The proportion of calcifying phytoplankton is parametrized following the formulation of Moore et al. (2002), because PISCES does not prognostically represent this phytoplankton functional type. On the other hand, the dissolution rate of calcite follows the parametrization of Gehlen et al. (2007): there is no dissolution is allowed when waters are supersaturated, but dissolution increases with the level of under-saturation. Total alkalinity in PISCES includes contributions from carbonate, bicarbonate, borate, hydrogen, and hydroxide ions (Practical alkalinity). Oxygen is simulated prognostically using two different oxygen-to-carbon ratios, one accounting when ammonium is converted to organic matter, the other when oxygen is consumed while nitrification.

For carbon and oxygen pools, air-sea exchange follows the Wanninkhof (1992) quadratic wind-speed formulation. To ensure conservation of nitrate in the ocean, annual total nitrogen fixation balances denitrification following Lipschultz et al. (1990), Middelburg et al. (1996), and Soetaert et al. (2000). Thus, carbon and nitrogen cycles are partly decoupled.

The geochemical boundary conditions also account for nutrient supply from three different sources: atmospheric dust deposition of iron and silicon (Tegen and Fung 1995; Jickells and Spokes 2001; Moore et al. 2004), rivers for macronutrients, dissolved carbon, and alkalinity. Ludwig et al. (1996) and sediment mobilization for sedimentary iron (Johnson et al. 1999; de Baar and de Jong 2001).

For all macronutrients except iron and nitrogen, the external sources are balanced yearly by an equivalent loss from the bottom of the ocean. This loss, which represents geological burial within the sediments, is proportional to the sinking rate. For nitrogen, this loss balances yearly the

amount of nitrogen provided by atmospheric deposition and riverine input since nitrogen fixation is hypothesized to compensate for denitrification at each time-step. Due to its short residence time, iron budget is not balanced in PISCES. A fraction of dissolved iron is scavenged by sinking organic matter. Then, all the particulate iron that reaches the bottom of the ocean is definitely removed from the ocean.

### 2.3 Simulations

Our results derive from historical simulations that have been obtained in three sequential steps:

- First, there is the 3000-year spinup, an off-line integration forced by monthly climatological fields repeated annually and atmospheric CO<sub>2</sub> held at 278 ppm, to obtain a quasi-steady state for the preindustrial ocean of each climate model. For each model, monthly climatological fields have been generated from a first preindustrial simulation, in which only dynamical components have been integrated starting from observed climatologies (e.g., Locarnini et al. 2009; Antonov et al. 2009). The methodology used for integrating IPSLCM4, IPSLCM5, and CNRMCM5 are further detailed in Marti et al. (2010), Dufresne et al. (2012), and Voldoire et al. (2012, this issue), respectively.
- Second each model is then integrated online to allow for interannual variations in biogeochemical tracer fields. These on-line integrations are shorter (100 years IPSLCM4 and 300 years for IPSLCM5 and CNRMCM5). At the end of this on-line spin-up, the linear drift of the global ocean carbon flux is about 0.013 PgC y<sup>-1</sup> per year for IPSLCM4, 0.001 PgC y<sup>-1</sup> per year for IPSLCM5 and 0.002 PgC y<sup>-1</sup> per year for CNRMCM5. These drifts are comparable or even smaller than the OCMIP-2 threshold, i.e. 0.01 PgC y<sup>-1</sup> per year (Orr 2002). Compared to the estimation of preindustrial global air-sea fluxes from forward model (e.g., Aumont et al. 2001) or inverse model (e.g., Mikalhof-Fletcher et al. (2007), IPSLCM4 overestimates the net ocean outgassing of about 0.8 PgC y<sup>-1</sup>, while the two CMIP5 models are within the range of models estimates: 0.4 PgC y<sup>-1</sup> for CNRMCM5 and 0.3 PgC y<sup>-1</sup> for IPSLCM5.
- Third, there is the transient simulation over the industrial era (beginning in 1860 for IPSLCM4, and 1850 for IPSLCM5 and CNRMCM5).

External greenhouse gas forcing for the historical scenario differs between the CMIP3 model and the two CMIP5 models. As IPSLCM4 is a fully coupled climate-carbon cycle Earth system model, it is forced only by

anthropogenic CO<sub>2</sub> emissions from fossil fuel burning and land-use change from 1860 to 2005 (Houghton and Marland 2007). IPSLCM5 and CNRMCM5 are forced by the historical anthropogenic CO<sub>2</sub> concentrations from 1850 to 2005 (Taylor et al. 2009, 2011). During 1985–2005, historical anthropogenic CO<sub>2</sub> concentrations is about 379 ppm for IPSLCM5 and CNRMCM5. Despite being forced by emission, the prognostically computed historical anthropogenic CO<sub>2</sub> concentrations is close to the latter, facilitating comparison among models and with observations. For IPSLCM5 and CNRMCM5, other greenhouse gases (e.g., nitrous oxide, chlorofluorocarbon) and aerosols are also taken into account. In both, five tropospheric aerosol types are considered: sea salt, sulfate, organic, black carbon, and desert aerosols. Volcanic aerosols are specified as a stratospheric aerosol type. These fields were estimated from observed emissions by the LMDZ-INCA chemical model (Szopa et al. 2012).

In parallel, we completed this set of model experiments by making offline simulations of CFC11 and CFC12 using a common passive tracer module (Dutay et al. 2002). This passive tracer module has been forced by monthly circulation fields from the three Earth system models and by observed atmospheric concentrations of CFC11 and CFC12 from 1950 to 2005.

### 2.4 Assessment methods

To evaluate simulated ocean dynamics, we first relied on data for physical variables clearly tied to the circulation, which controls ocean biogeochemistry. For observed sea surface temperature and sea surface salinity, we used the World Ocean Atlas 2009 (WOA 2009, Locarnini et al. 2009; Antonov et al. 2009), a database that combines observations from 1965 to 2007. For mixed-layer depth we used the observed climatology of de Boyer Montégut et al. (2004, updated in September 2008), in particular their vertical density criterion of 0.03 kg m<sup>-3</sup> computed from in situ temperature and salinity collected since 1961. For wind stress, our reference is the NCEP 1957–1996 reanalysis of Kalnay et al. (1996). Simulated short wave radiation is compared to the global climatology of Gupta et al. (1999), i.e., a monthly climatology from the Earth Radiation Budget Experiment. For assessing modeled ocean-atmosphere net heat fluxes (i.e., sum of solar radiation, long-wave radiation, sensible heat transfer by conduction and convection, and latent heat transfer by evaporation of sea surface water), we use the 1981–2005 reanalysis of Yu et al. (2007), which provides a monthly climatology combining observations from surface meteorological stations and satellite data. Finally, for data-based estimates of watermass transport of North Atlantic Deep Water, Antarctic Bottom Water, and the Antarctic Circumpolar

Current, we relied on estimations from Talley et al. (2003), Ganachaud (2003), Cunningham (2003).

As references for simulated biogeochemistry, we also based our analysis on global data products. We use observed climatologies of silicate ( $\text{SiO}_2$ ), phosphate ( $\text{PO}_4^{3-}$ ), nitrate ( $\text{NO}_3^-$ ), and oxygen  $\text{O}_2$  from the World Ocean Atlas 2009 (Garcia et al. 2009a, b). For the concentrations of dissolved inorganic carbon (DIC) and alkalinity (Alk), we used the GLODAP climatology (Key et al. 2004), which is based on discrete measurements collected during 1972–1999 but centered in 1994. For the difference in the partial pressure of  $\text{CO}_2$  between the ocean and the atmosphere ( $\Delta p\text{CO}_2$ ) we relied on the Takahashi (2009) climatology referenced to year 2000. Simulated surface chlorophyll ( $\text{CHL}_a$ ) was compared to the SeaWiFS climatology based on satellite observations from 1997 to 2003 (Carr et al. 2006). Additionally, we used multiple references for data-based estimates of marine primary productivity (Behrenfeld et al. 2006) as well as export of organic carbon, calcite, and silicate (e.g., Laws et al. 2000; Schlitzer 2002; Heinze et al. 2003; Moore et al. 2004; Jin et al. 2006; Dunne et al. 2007).

To provide quantitative assessment of model skill, we used several common statistical metrics (e.g., Stow et al. 2009): (1) correlation coefficient ( $R$ ) for assessing dominant large-scale patterns between modeled and observed fields; (2) normalized standard deviation for quantifying differences between modeled and observed spatial variations; (3) bias (average error) as a measure of average mismatch between modeled and observed fields; and (4) root mean square error (RMSE) for assessing the total mismatch between modeled and observed fields. All of these computations are performed after regridding model output to the observed grid ( $2^\circ \times 2^\circ$  for partial pressure of  $\text{CO}_2$  and  $1^\circ \times 1^\circ$  for other variables) and weighting by grid-cell area.

Statistical metrics (i.e., RMSE, bias, correlation coefficients and standard deviation) are computed from modeled and observed fields at surface, 500, 1,000, 1,500, 2,000, 2,500 and 3,000 m. Correlations and standard deviations are summarized in Taylor diagrams for distributions at the surface, 500, 1,000 and 3,000 m.

### 3 Results

#### 3.1 Ocean physics

##### 3.1.1 Ocean-atmosphere interactions and ocean surface properties

For an atmospheric perspective, we present at first the ability of the Earth system models results to match the

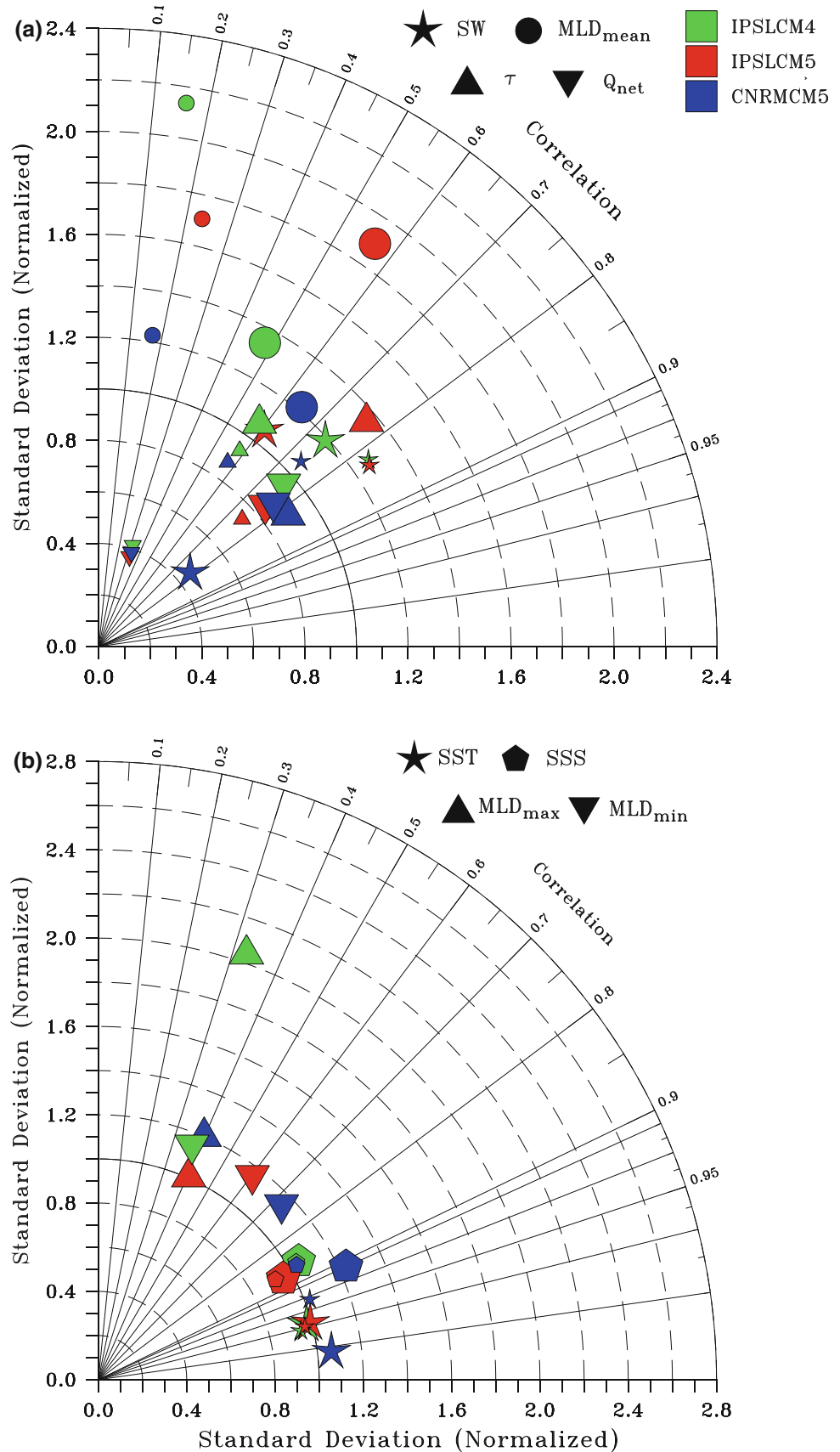
observed net heat flux ( $Q_{\text{net}}$ ), incoming shortwave radiation (SW), and wind stress ( $\tau$ ). It allows assessing surface forcings that the atmosphere exerts on the surface ocean controlling the dynamics of several ocean fields and those of air-sea exchange of  $\text{O}_2$  and  $\text{CO}_2$  (e.g., Sarmiento and Gruber 2006). For an oceanic perspective, we show the maximum annual of mixed-layer depth (MLD, during winter) as well as its minimum (during summer). The maximum and the minimum of MLD are commonly used to diagnose how a physical model is suited to simulate proper ocean biogeochemistry (e.g., Doney 2004; Schneider et al. 2008). The maximum MLD controls nutrient supply within photic zone while minimum MLD controls light limitation (e.g., Sarmiento and Gruber 2006; Behrenfeld 2010).

Figure 1a shows that for both  $\tau$  and MLD, correlations between simulated and data-based annual-mean fields range between 0.5 and 0.8 (Fig. 1a). Generally, correlations for both IPSLCM5 and CNRMCM5 are higher than those for IPSLCM4; normalized standard deviations ( $\sigma$ ) range between 0.4 and 1.9. Models broadly replicate  $Q_{\text{net}}$  ( $R > 0.7$  and  $\sigma \sim 0.9$ ). Including the annual cycle in the analysis decreases the correlation between modeled and observed fields. This appears clearly for correlations and spatiotemporal variations of  $Q_{\text{net}}$ , which are poorly represented by the models ( $\sigma \sim 0.3$  and  $R \sim 0.3$ ). In contrast, spatiotemporal variations of modeled SW are better reproduced when taking into account the annual cycle.

All three models exhibit reasonable large-scale structure of annual-mean sea surface temperature (SST) and sea surface salinity (SSS) (Fig. 1b). Models generally reproduce observed SST (Loccardini et al. 2009) with correlation coefficients above 0.95 and normalized standard deviations close to 1. The same generally holds for SST when we include the annual cycle. The largest differences between modeled and observed SST are found in the Equatorial Pacific Ocean and The Southern Ocean (supplementary materials, Figure S1). In the Equatorial Pacific, IPSLCM4 and IPSLCM5 exhibit higher than observed SST, while SST is slightly colder in CNRMCM5. The opposite happens in the Southern Ocean. As SST, annual-mean fields and annual cycle for SSS presents a good agreement with the observations ( $R > 0.95$  and  $\sigma \sim 1$ ). The largest biases between modeled and observed SSS are found in high latitude oceans, especially in the Southern Ocean.

Figure 2 reveals that models broadly capture the general zonal average structure of maximum MLD ( $\text{MLD}_{\text{max}}$ ), but still exhibit substantial differences compared to the observed climatology (de Boyer Mont  gut et al. 2004): (1) In the northern oceans, IPSLCM4 and IPSLCM5 fail to simulate the deep  $\text{MLD}_{\text{max}}$  in Labrador Sea, an important deep-convection site; (2) In the tropical ocean, the  $\text{MLD}_{\text{max}}$

**Fig. 1** Taylor diagram showing the correspondence between model results and observations for **a** shortwave radiation ( $SW$ , Gupta et al. 1999), net heat flux ( $Q_{net}$ , Yu et al. 2007), wind stress ( $\tau$ , Kalnay et al. 1996) and mixed-layer depth ( $MLD_{mean}$ , de Boyer Montégut et al. 2004) and **b** sea surface temperature (SST, Locarnini et al. 2009), minimum (summer) mixed-layer depth ( $MLD_{min}$ ), maximum (winter) mixed-layer depth ( $MLD_{max}$ ) and sea surface salinity (SSS, Antonov et al. 2009). *Large symbols* indicate statistics from annual mean distributions while *small symbols* indicate that computations include annual cycle



in IPSLCM4 and IPSLCM5 roughly agrees with observed values, but CNRMCM5 overestimates it by 20–40 m; (3) Southern Ocean (<30°S)  $MLD_{max}$  is poorly simulated by the models, all of which fail to simulate the deep  $MLD_{max}$  that is observed in the Pacific sector; and (4) In western boundary currents (e.g., Gulf Stream, Kuroshyo, Malvinas-Brazil confluence),  $MLD_{max}$  is overestimated in all models, particularly CNRMCM5. Figure 2 also shows that there are also large discrepancies between modeled and observed minimum MLD ( $MLD_{min}$ ). The  $MLD_{min}$  is better simulated in CNRMCM5 than in IPSLCM4 and IPSLCM5, because they capture the deep  $MLD_{min}$  in the Southern Ocean. Although it remains too shallow, simulated  $MLD_{min}$  in the Southern Ocean improves with the evolution from IPSLCM4 to IPSLCM5 (Fig. 2). In the Equatorial Pacific,  $MLD_{min}$  is overestimated in all models, particularly CNRMCM5.

### 3.1.2 Water mass transport and general circulation

Table 2 shows that the magnitude of simulated NADW transport is low relative to observations in IPSLCM4, while those of IPSLCM5 and CNRMCM5 are within the low end of the observational uncertainty (e.g., Talley et al. 2003; Ganachaud 2003). The northward transport of AABW is underestimated in IPSLCM5 and CNRMCM5 relative to observational estimates (e.g., Talley et al. 2003). All models underestimate the observed strength of the ACC (e.g., Talley et al. 2003; Cunningham 2003), although simulated transport in IPSLCM5 and CNRMCM5 is more similar to the observed transport relative to the much lower transport of IPSLCM4 (i.e., 60–70 Sv lower).

### 3.2 CFC11 and CFC12 distribution

CFC11 is an inert, abiotic, and entirely man-made tracer in seawater, which is set solely by ocean dynamics and uptake at air-sea interface, not by biological activity. Modeled CFC11 in surface waters is generally quite similar, being essentially set by atmospheric CFC11 and SST patterns. But, modeled CFC11 in subsurface waters and inventories can differ greatly providing information on ocean dynamics (e.g., Matsumoto et al. 2004). Global inventories of CFC11 are about  $4.3 \times 10^8$ ,  $6.3 \times 10^8$ , and  $6.4 \times 10^8$  mol for CNRMCM5, IPSLCM4 and IPSLCM5, respectively. This shows that models bracket the observed estimate of  $5.2 \times 10^8$  mol. The simulated and observed CFC11 inventories (GLODAP, Key et al. 2004) shown in Fig. 3 illustrate how the largest discrepancies occur in the Southern Ocean. In this region, models manage to capture the location of subduction zone of mode waters (~40–50°S), especially in IPSLCM5 and CNRMCM5.

However, south of 50°S, models exhibit large differences in their CFC11 inventories. Compared to the observations, the simulated CFC11 inventories are low in CNRMCM5 and high in IPSLCM4. These biases occurring in the zone of deep-convection reveal strong differences in their ability to ventilate subsurface waters. In particular, Fig. 3 indicates that models have trouble to replicate the convection zones in the Weddell Sea for IPSLCM4 and in the North Atlantic for IPSLCM5 and CNRMCM5.

### 3.3 Marine biogeochemistry

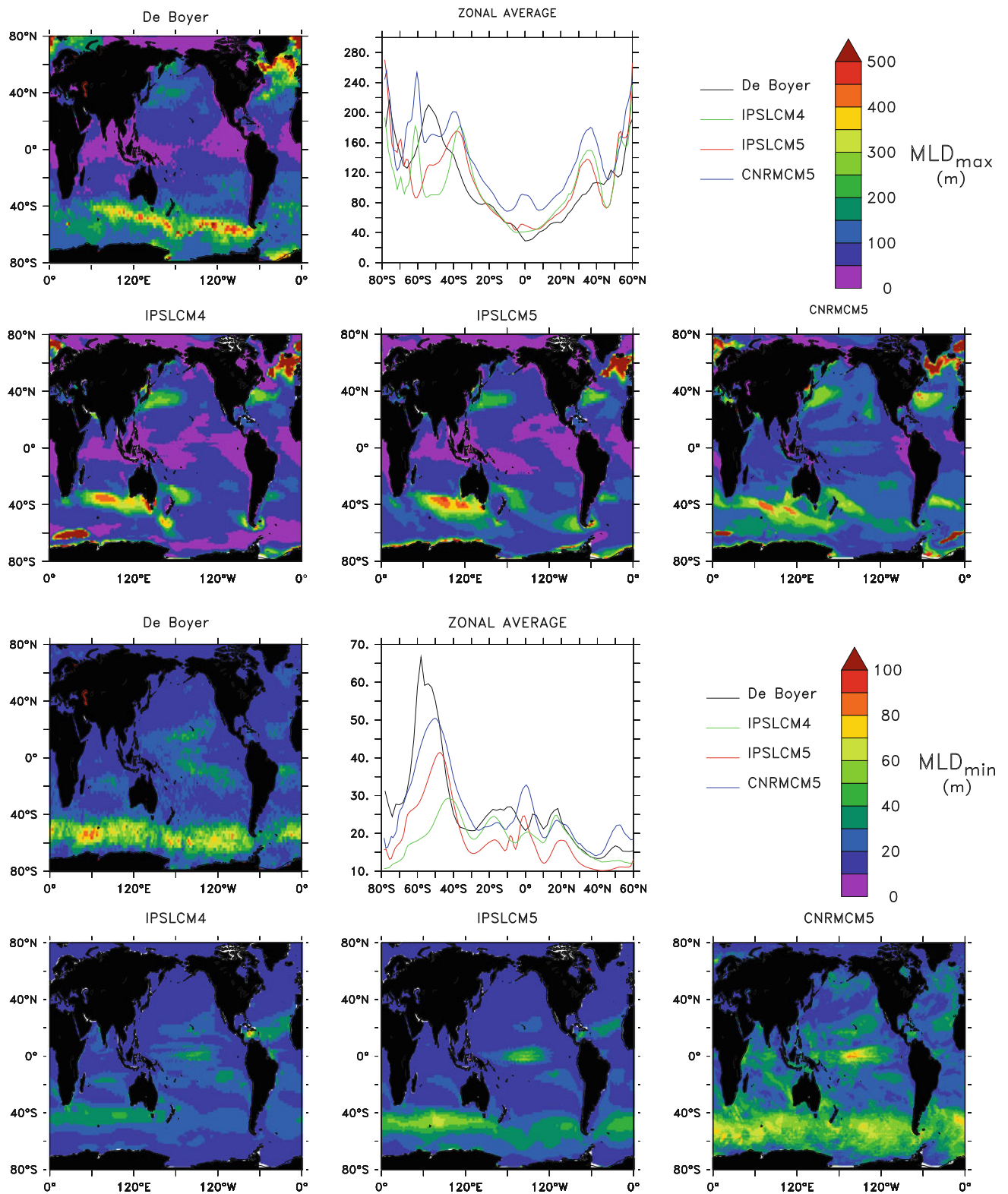
#### 3.3.1 Surface properties

**3.3.1.1 Annual-mean macronutrient distributions** The three models generally capture the large-scale structure of annual-mean surface concentrations of macronutrients (Table 3). Correlation coefficients between modeled and observed surface concentrations of macronutrients are high ( $R \sim 0.9$ ), except for  $SiO_2$  in IPSLCM4 ( $R \sim 0.6$ ). Global average bias and RMSE reveal that models are skillful to simulate surface macronutrients concentrations (Table 3). Only  $NO_3^-$  surface concentrations in CNRMCM5 are strongly overestimated by about  $5 \mu mol L^{-1}$  (~16 % of the global mean).

All models generally capture the observed dominant surface patterns for  $PO_4^{3-}$  and  $SiO_2$ ; for  $NO_3^-$ , IPSLCM4 and IPSLCM5 similarly agree with overall observed patterns but CNRMCM5 does not (Fig. 4). In the tropics, models tend to underestimate the high surface concentrations that are associated with the strong Equatorial Pacific upwelling (Fig. 4). In the high latitudes, all models underestimate observed zonal-mean concentrations between 40°N and 60°N, mostly because surface concentrations in North Pacific are too low. In the Southern Ocean, models diverge in their ability to properly simulate surface macronutrients. Only IPSLCM5 yields sufficiently high surface  $SiO_2$  concentrations observed close to the Antarctic shelf. Zonal averages illustrate that simulated gradients in macronutrients in CNRMCM5 are shifted southward relative to those for IPSLCM4 and IPSLCM5 (Fig. 4), a feature that is related to the position of maximum wind stress, stemming from different atmospheric components of the coupled models (see Sect. 4.2).

**3.3.1.2 Annual-mean surface  $CHL_a$**   $CHL_a$  provide evidence of nutrient and light limitation controlled by the ocean mixed-layer dynamics (e.g., Behrenfeld 2005, 2010; Behrenfeld et al. 2006). Due to its global scale monitoring over several decades (e.g., Siegel et al. 2002; Carr et al. 2006),  $CHL_a$  consists in a useful indicator for assessing skill of ocean biogeochemical models.





**Fig. 2** Comparison between modeled and observed depth (in m) of the maximum of mixed-layer ( $MLD_{max}$ ) and the minimum mixed-layer ( $MLD_{min}$ ). Observed climatology is based on an update of the

climatology of de Boyer Montégut et al. (2004), which incorporate Argo's data until september 2009

**Table 2** Estimation from observations (e.g., Talley et al. 2003; Cunningham 2003) and models of the Atlantic Meridional overturning circulation (AMOC, maximum of meridional streamfunction between 300 and 2,000 m and 0–75°N), Antarctic bottom waters export (AABW, minimum of meridional streamfunction between 2,000 m and the bottom of the ocean), Antarctic circumpolar current (ACC, maximum of barotropic streamfunction at Drake) in Sverdrup ( $Sv = 10^6 \text{ m}^3 \text{ s}^{-1}$ )

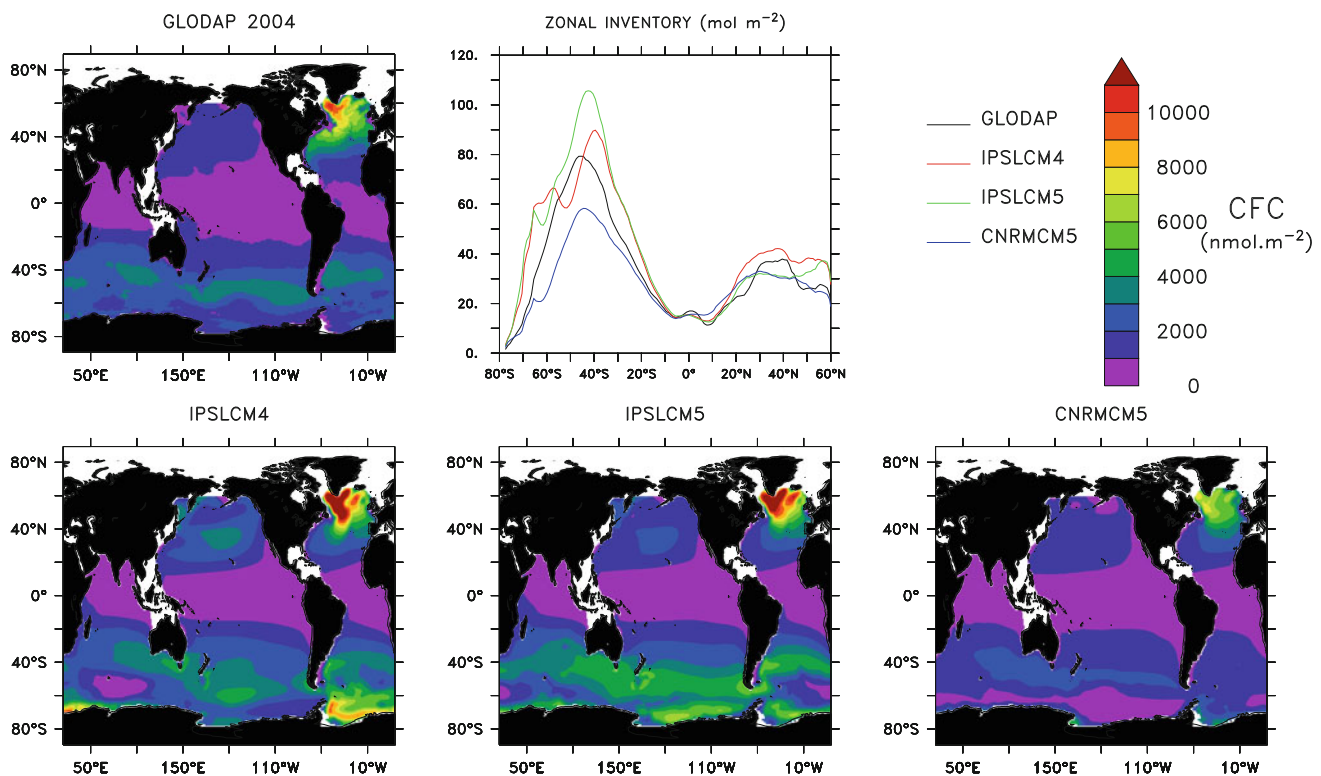
Ocean transport	Observations	IPSLCM4	IPSLCM5	CNRMCM5
AMOC	$18 \pm 5$	$11.4 \pm 1.3$	$13.2 \pm 2$	$14.8 \pm 2$
AABW	$10 \pm 2$	$12 \pm 9$	$9.1 \pm 7.1$	$4.8 \pm 1.4$
ACC	134 – 148	$60 \pm 10$	$110 \pm 10$	$110 \pm 20$

Generally, models roughly agree with large-scale observed patterns ( $R \sim 0.4$ , Table 3). This is especially the case for high  $\text{CHL}_a$  surface concentrations associated with western boundary currents (Fig. 5). However, models systematically underestimate  $\text{CHL}_a$  concentration in the oligotrophic gyres of the subtropical oceans and the upwelling regions such as the Eastern Equatorial Pacific. Yet, in these regions, SeaWiFS may detect the deep  $\text{CHL}_a$  maximum (e.g., Garcia et al. 2005). Models differ most in the Southern Ocean, where IPSLCM4 and IPSLCM5 simulate values between 0.3 and  $0.8 \mu\text{gChl L}^{-1}$ , much higher than those estimated by satellite ( $0.2 \mu\text{gChl L}^{-1}$ ).

However, models fall within the observation uncertainties knowing that SeaWiFS underestimates  $\text{CHL}_a$  surface concentration in the Southern Ocean (e.g., Garcia et al. 2005).

**3.3.1.3 Annual-mean  $\Delta p\text{CO}_2$**  A global data product for the ocean minus atmospheric difference in the partial pressure of carbon dioxide ( $p\text{CO}_2$ ), known as  $\Delta p\text{CO}_2$ , has been computed from millions of discrete measurements of oceanic  $p\text{CO}_2$  (Takahashi et al. 2002, 2006; Takahashi 2009). It is known that  $\Delta p\text{CO}_2$  drives the air-sea  $\text{CO}_2$  flux and that it is affected by solubility (function of SST, SSS, DIC and Alk), ocean transport, and biological activity (photosynthesis and respiration).

Compared to the climatology of Takahashi (2009), models broadly capture large-scale structures of annual mean  $\Delta p\text{CO}_2$  with correlation coefficients of about  $0.6 \pm 0.1$  (Table 3). However, global averaged RMSE values range between 16 and 24 ppm and contrast with global averaged bias values, which range between  $-2$  and  $-16$  ppm. The latter reflects regional biases that tend to cancel each other. Regional model-data discrepancies shown in Fig. 6, illustrate that the largest differences occur in three main areas: (1) in the northern oceans,  $\Delta p\text{CO}_2$  in IPSLCM4 and IPSLCM5 is much stronger than observed, whereas CNRMCM5's  $\Delta p\text{CO}_2$  is roughly on target; (2) in the Equatorial Pacific, the high observed  $\Delta p\text{CO}_2$  is



**Fig. 3** Annual mean of CFC11 inventory (in  $\text{nmol m}^{-2}$ ) and zonal integral inventory (in  $\text{mol m}^{-2}$ ) for IPSLCM4, IPSLCM5 and CNRMCM5. Observed Climatology is based on GLODAP climatology. Model outputs and observations use 1994 as reference year (Key et al. 2004)

**Table 3** Weighted-area correlations coefficients (R), root mean square error (RMSE) and bias/average error (AE) of surface concentrations of annual-mean phosphate, nitrate, silicate, chlorophyll and  $\Delta p\text{CO}_2$  compute between models results on observational data (Carr et al. 2006; Takahashi 2009; Garcia et al. 2009a, b, for chlorophyll,  $\Delta p\text{CO}_2$ , and nutrients, respectively)

Surface fields	Skill score	IPSLCM4	IPSLCM5	CNRMCM5
CHL <sub>a</sub>	R	0.38	0.42	0.43
	RMSE	0.18	0.18	0.16
	AE	-0.08	-0.07	-0.12
$\Delta p\text{CO}_2$	R	0.64	0.67	0.59
	RMSE	24.29	18.10	16.05
	AE	-16.02	-2.05	-1.87
$\text{NO}_3^-$	R	0.87	0.94	0.89
	RMSE	2.39	1.64	5.5
	AE	-1.31	-0.61	4.9
$\text{PO}_4^{3-}$	R	0.85	0.91	0.86
	RMSE	0.19	0.21	0.33
	AE	-0.03	-0.17	-0.32
$\text{SiO}_2$	R	0.58	0.8	0.78
	RMSE	6.89	5.58	3.92
	AE	2.13	3.07	-0.65

underestimated by IPSLCM4, whose Equatorial tongue is too confined toward the Eastern part of the basin (conversely IPSLCM5 and CNRMCM5 exhibit excessive outgassing); and (3) in the Southern Ocean, models poorly represent the observed structure of slight outgassing over large areas. Nevertheless, the Southern Ocean data may be biased due to seasonal sampling: most measurements have been taken during summer.

### 3.3.2 Vertical distributions

Models broadly capture vertical observed gradients in macronutrients between the surface and 1000 m, but they have trouble to reproduce subsurface maxima in  $\text{NO}_3^-$  and  $\text{PO}_4^{3-}$  as observed (Fig. 7, see also supplementary materials Figure S2 and S3).

For both these macronutrients, the fact that models underestimate concentrations between 500 and 1,000 m, while they overestimate the concentration between 1,500 and 2,000 m indicates that models poorly represent the remineralization depth. This appears clearly for all of the models in the RMSE profiles: with the largest values between 500 and 1,000 m, except for  $\text{SiO}_2$  where RMSE is larger between for 1500 and 2000 m, but to a lesser degree. In CNRMCM5, deep  $\text{NO}_3^-$  concentrations are too high, worsening with the proximity of the Southern Ocean, indicating a problem with the ventilation of deep waters in that model. Simulated concentrations of  $\text{PO}_4^{3-}$  and  $\text{SiO}_2$  in all 3 models are similar to those observed, especially for

$\text{PO}_4^{3-}$  with correlation coefficients always above 0.8 and normalized standard deviations close to 1 (Figs. 7 and S2). But in CNRMCM5, for  $\text{NO}_3^-$  both of these metrics decline sharply with depth.

Fields of  $\text{O}_2$ , dissolved inorganic carbon (DIC), and total alkalinity (Alk) are of particular interest for assessing the skill of ocean biogeochemistry models to simulate air-sea fluxes. Unlike other tracers, surface  $\text{O}_2$  and DIC are affected by air-sea gas exchange; conversely, below the surface they act like other tracers being affected by ocean circulation and interior sources minus sinks.

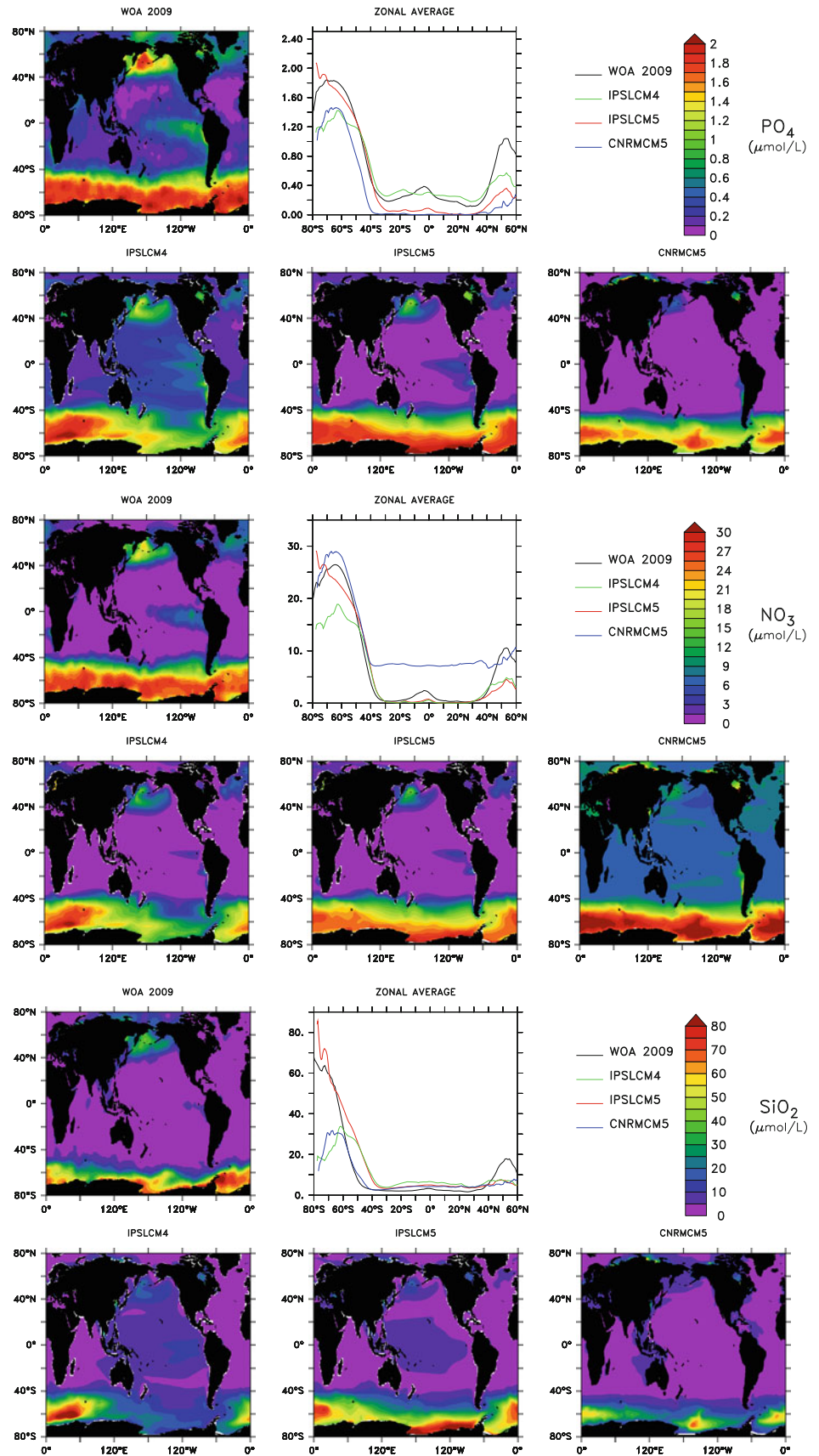
Large model-model and model-data differences are found for vertical profiles of DIC, Alk and  $\text{O}_2$  as illustrated by profiles of global average bias and RMSE (Fig. 8). IPSLCM4 simulates the most realistic profiles for  $\text{O}_2$  and DIC (supplementary materials Figure S4 and S5). IPSLCM5 strays further from observed DIC and  $\text{O}_2$  below 1,500 m, but its deep Alk profile shows substantial improvement. CNRMCM5 shows the largest disagreement for profiles of DIC and  $\text{O}_2$ , whereas for Alk, its skill is as good or better than the other models. Despite these strong biases in DIC and  $\text{O}_2$  profiles for IPSLCM5 and CNRMCM5, their correlation coefficients are larger than for IPSLCM4.

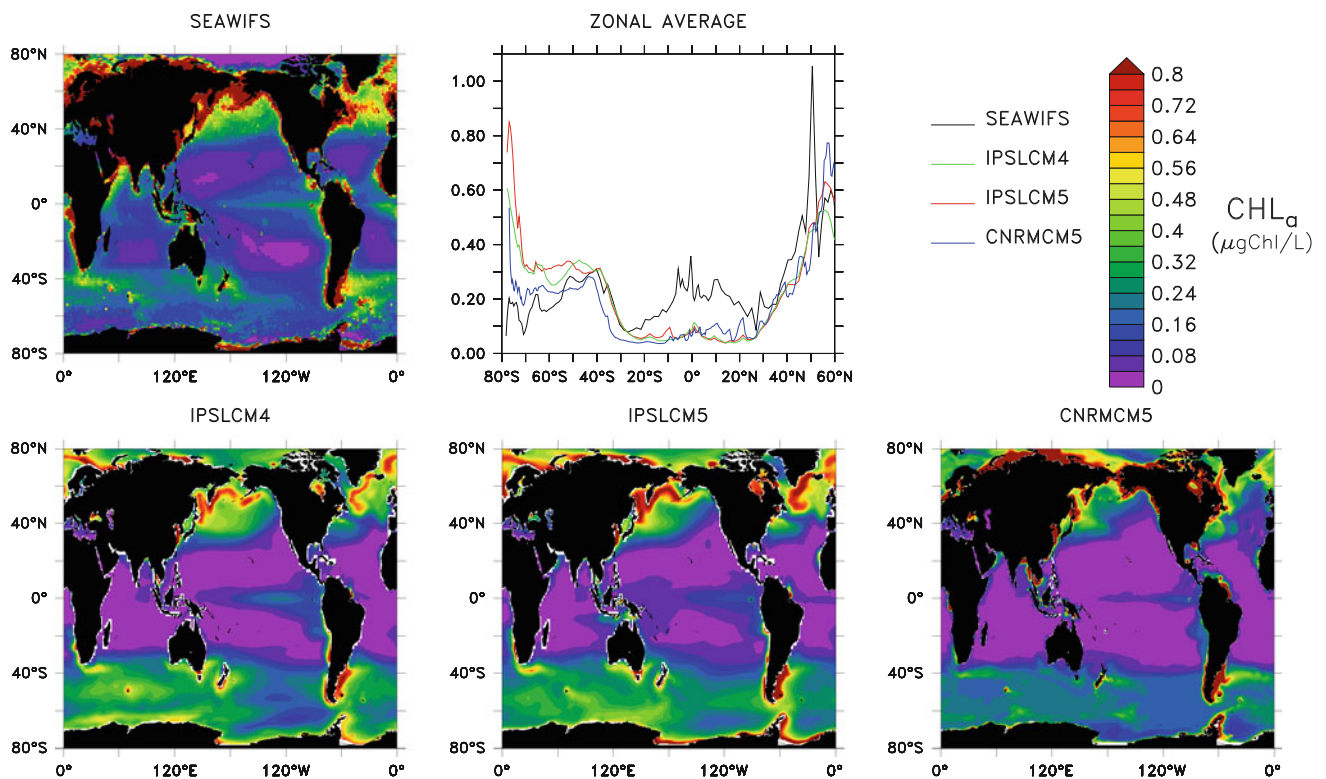
### 3.4 Air-sea fluxes and biological production

Models differ substantially in their simulated air-sea  $\text{CO}_2$  fluxes, primary productivity, export production at 100 m and attenuation length (Table 4). For primary productivity, models global estimates mirror differences in Equatorial Pacific primary productivity. Otherwise, model estimates span globally between the low values from CNRMCM5 and high values from IPSL models. Yet our simulated ranges are within the range of published estimates based on observations (e.g., Laws et al. 2000), observations with inverse models (e.g., Schlitzer 2002), and forward models (e.g., Heinze et al. 2003; Moore et al. 2004; Jin et al. 2006; Dunne et al. 2007).

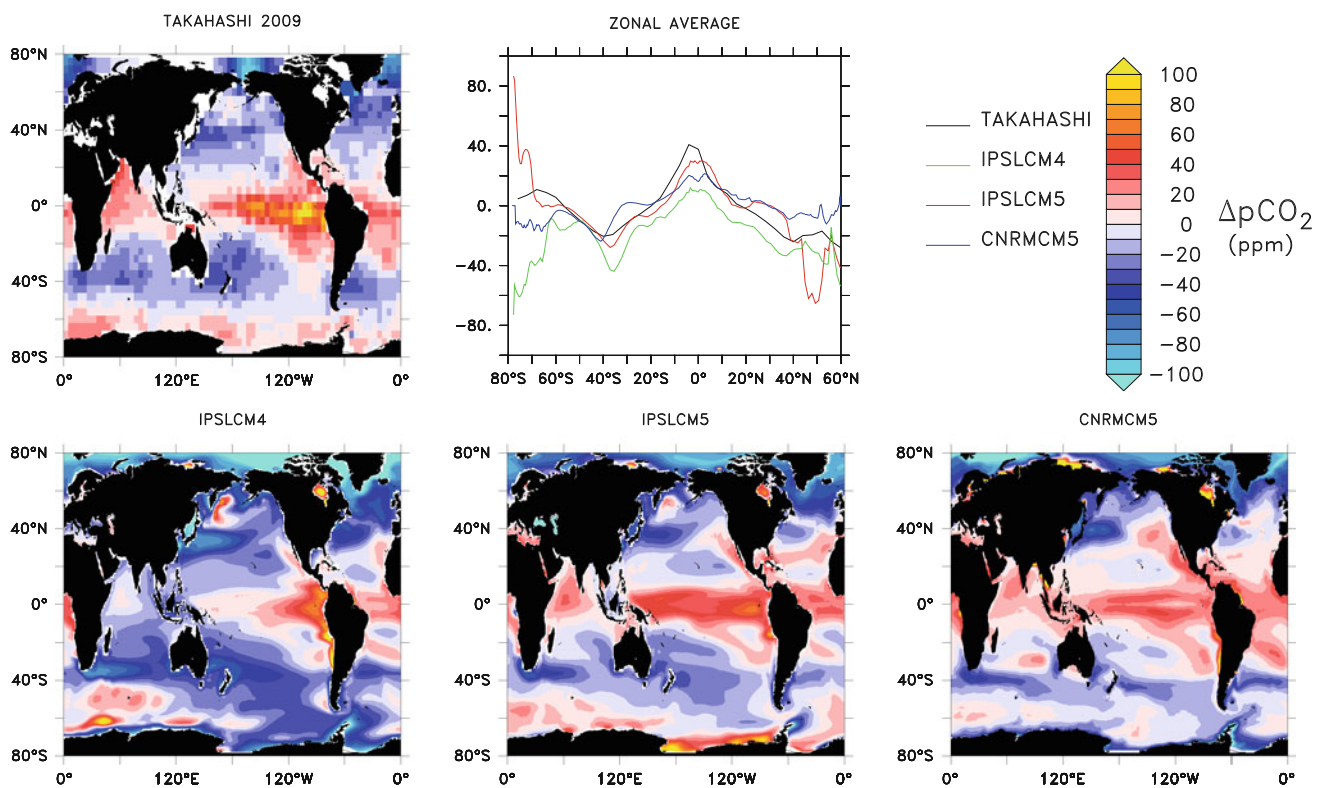
However, models exhibit large differences at depth compared to the previous estimates (Table 4). Globally, models underestimate the depth at which waters are saturated with the respect to the calcium carbonate mineral calcite, the calcite saturation horizon depth (CSH). This underestimation of the depth of CSH is strong in IPSLCM4 and IPSLCM5, but to a lesser degree in CNRMCM5. On the contrary, Table 4 shows that the remineralization length ( $Z_{remin}$ ) is overestimated in all of the three models compared to the estimation based on the Martin's exponent law (Martin et al. 1987). Note that we define  $Z_{remin}$  as the depth at which 10% of the export of organic matter at 100 m remain following the exponent law of Martin et al. (1987).

**Fig. 4** Annual-mean of seasonal climatology of surface concentrations (in  $\mu\text{mol L}^{-1}$ ) of phosphate ( $\text{PO}_4^{3-}$ ), nitrate ( $\text{NO}_3^-$ ), and silicate ( $\text{SiO}_2$ ) over 1985–2005 period for IPSLCM4, IPSLCM5 and CNRMCM5. Observed climatology is based on the World Ocean Atlas 2009 (Garcia et al. 2009a, b)



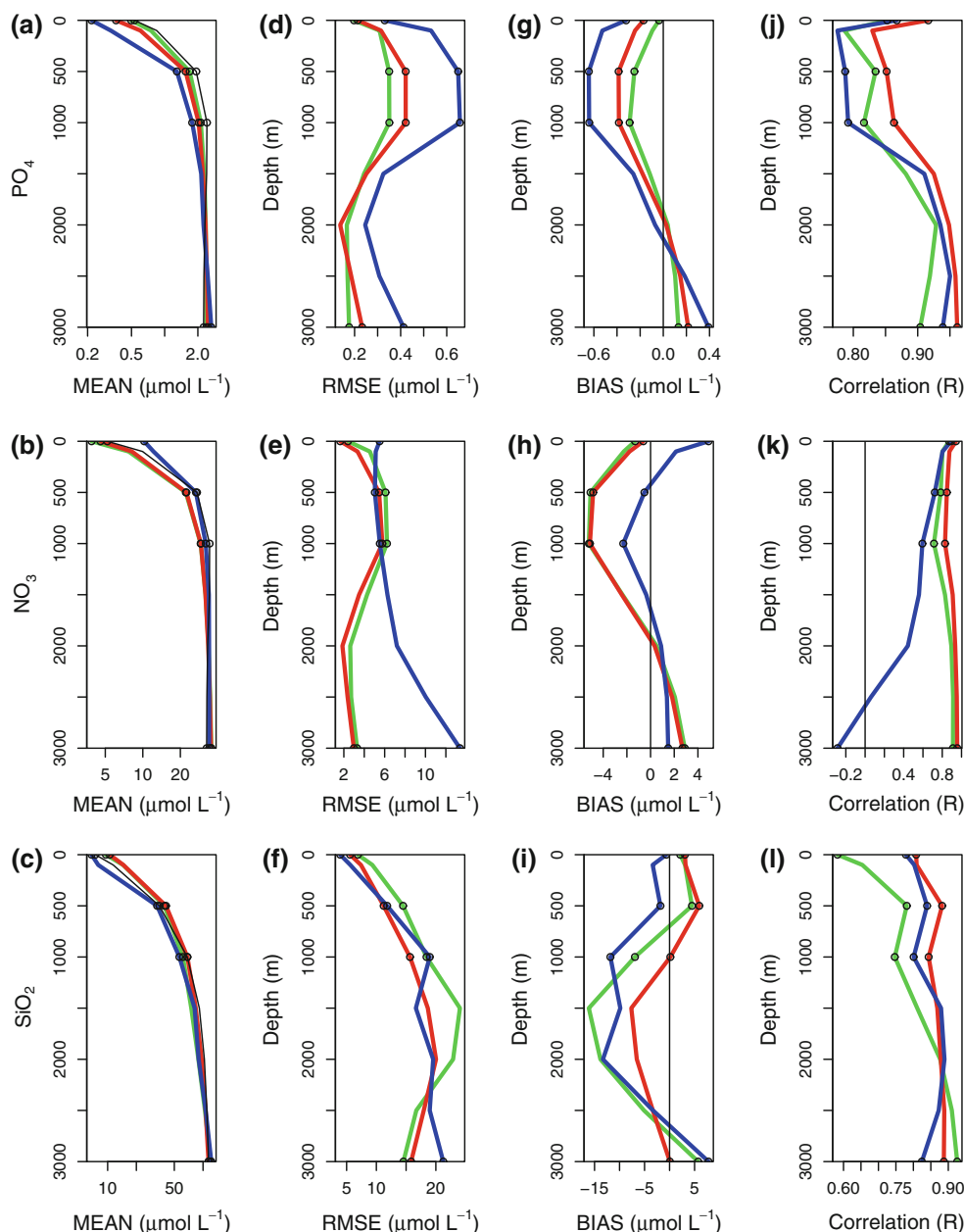


**Fig. 5** Annual mean of seasonal climatology of surface chlorophyll ( $\text{CHL}_a$ , in  $\mu\text{gChl L}^{-1}$ ) concentrations over 1985–2005 period for IPSLCM4, IPSLCM5 and CNRMCM5. Observed climatology is based on SeaWiFS satellite product (Carr et al. 2006)



**Fig. 6** Annual-mean of seasonal climatology of surface  $\Delta p\text{CO}_2$  (in ppm) over 1985–2005 period for IPSLCM4, IPSLCM5 and CNRMCM5. Observed climatology is based on Takahashi (2009)

**Fig. 7** Vertical profiles of mean concentration, root mean square error (RMSE), bias and correlation (R) of annual mean phosphate ( $\text{PO}_4^{3-}$ ), nitrate ( $\text{NO}_3^-$ ), and silicate ( $\text{SiO}_2$ ) concentrations at surface, 500, 1,000, 1,500, 2,000, 2,500 and 3,000 m for IPSLCM4 (in green), IPSLCM5 (in red) and CNRMCM5 (in blue). Circles represent depths used in Figs. S3 and S5. Observed profile of mean concentration of phosphate, nitrate, and silicate (Garcia et al. 2009b) are represented in thin black line



## 4 Discussion

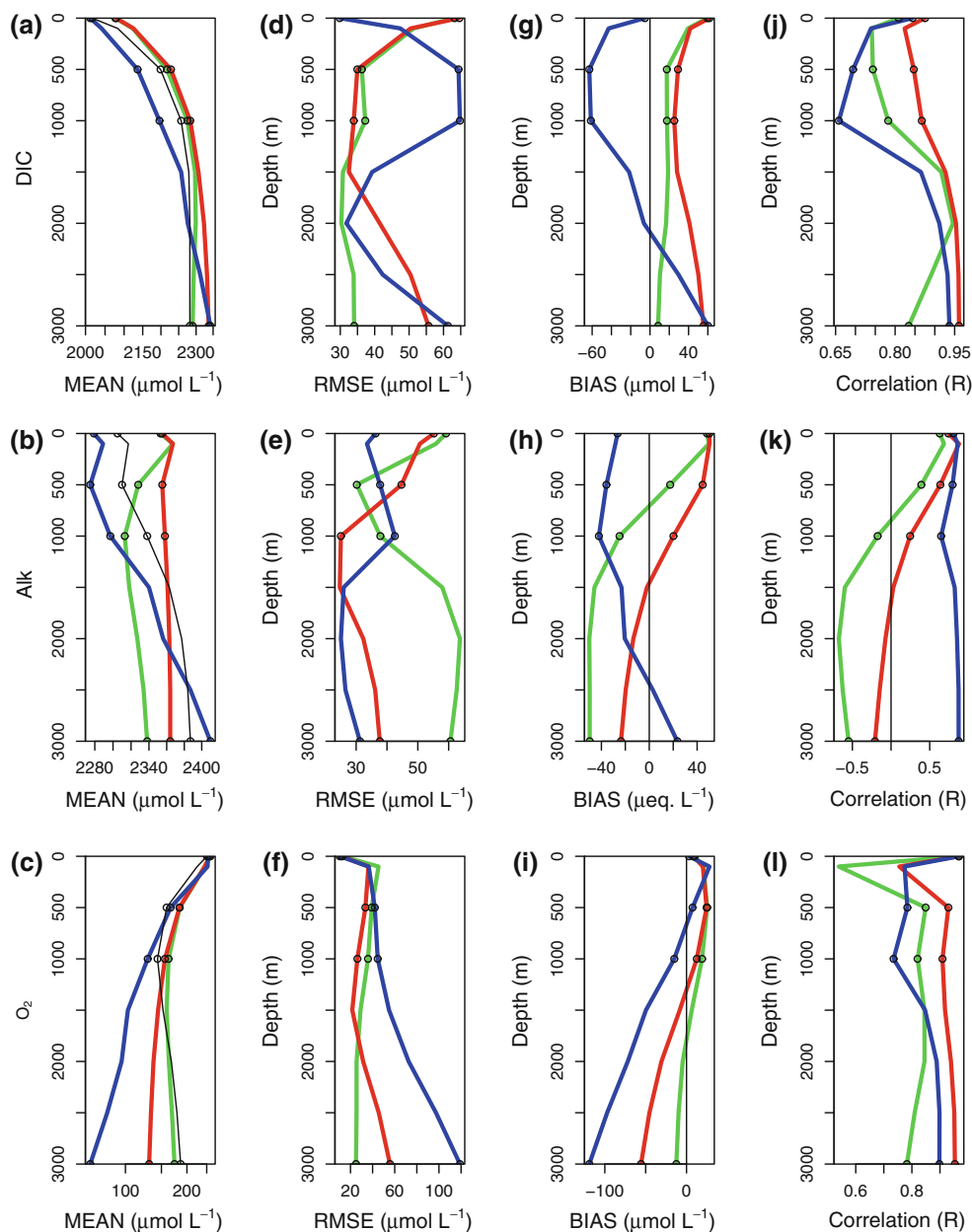
### 4.1 Model skill

The three Earth system models studied here exhibit reasonable skill in simulating SST and SSS, similar to other state-of-the-art Earth system models (e.g., Meehl et al. 2007; Randall et al. 2007). Although IPSLCM5 and CNRMCM5 are more skillful in simulating the dynamics of upper ocean than is IPSLCM4, all models have trouble to adequately simulate mixed layer depths, both the summer minimum and winter maximum, as well as wind stress and the dominant characteristics of the large-scale geostrophic circulation (e.g., intensity and position of the

ACC). These physical deficiencies degrade the ability of the models to simulate biogeochemical fields, both in the upper ocean and at depth. Deep distributions of biogeochemical tracers reveal signatures of unrealistic simulated ocean dynamics, and ventilation of the deep waters over decadal to centennial timescales (supplementary materials Figure S2 to S5).

Although the three models differ in multiple ways, this exercise helps summarize how developments in the IPSL model have altered simulated ocean biogeochemical fields between IPCC AR4 and AR5. The major differences between IPSLCM4 and IPSLCM5 are as follows: (1) Surface macronutrients have improved, illustrated by lower biases and higher correlation coefficients, while large-scale

**Fig. 8** Vertical profiles of mean concentration (MEAN), root mean square error (RMSE), bias and correlation (R) of annual mean dissolved inorganic carbon (DIC), Alkalinity (Alk) and oxygen ( $O_2$ ) concentrations at surface, 100, 500, 1,000, 1,500, 2,000, 2,500 and 3,000 m for IPSLCM4 (green), IPSLCM5 (red) and CNRMCM5 (blue). Circles represent depths used in Figs. S3 and S5. Observed profile of mean concentration of DIC and Alk (Key et al. 2004) and that of  $O_2$  (Garcia et al. 2009a) are represented in thin black line



structures of surface  $SiO_2$ ,  $CHL_a$ , and  $\Delta pCO_2$  are also closer to those observed; (2) Vertical profiles of alkalinity have improved; and (3) Deep-ocean ventilation has become weaker, which in turn has degraded deep distributions of biogeochemical tracers, especially  $O_2$ . Relative to the two IPSL models, CNRMCM5 has larger biases and RMSE in  $NO_3^-$  and  $O_2$ , attributable to sluggish deep-ocean ventilation. Yet for other biogeochemical fields, CNRMCM5 results are comparable.

Despite model differences in biogeochemical fields, a model's chemical potential to take up atmospheric  $CO_2$  is ultimately determined by alkalinity minus DIC (Alk-DIC). That difference generally serves as a good approximation for the carbonate ion concentration [ $CO_3^{2-}$ ] (Sarmiento and

Gruber 2006). And [ $CO_3^{2-}$ ] is inversely related to the Revelle factor, making it a proxy for the buffer capacity of seawater and hence the chemical capacity for the ocean to take up atmospheric  $CO_2$ . This aspect is interesting because Earth system models are used in the IPCC AR5 (CMIP5) framework to estimate compatible emissions and climate-carbon cycle feedbacks (e.g., Dufresne et al. 2012) from simulations with future Representative Concentration Pathways (RCP, Taylor et al. 2009, 2011).

All models simulate large-scale structures of Alk-DIC in the upper ocean that are similar to those observed (Fig. 9). Model-data differences reveal that simulated Alk-DIC is generally too weak in the high latitudes, particularly the Southern Ocean ( $<30^\circ S$ ). The anomalies of Alk-DIC

**Table 4** Modeled and observation-based (i.e., Behrenfeld et al. 2006; Laws et al. 2000; Schlitzer 2002; Heinze et al. 2003; Moore et al. 2004; Jin et al. 2006; Dunne et al. 2007) annual-mean fluxes of

matter at the ocean-atmosphere interface and fluxes of sinking particles toward the deep ocean

Biogeochemical fluxes		Observations	IPSLCM4	IPSLCM5	CNRMCM5
At surface	Carbon	2	1.95	2.15	1.9
	Primary Prod.	47.50	30.05	33.95	25.31
At 100 m	Carbon	5.8–12.9	7.89	7.09	4.9
	Calcite	0.38–1.64	0.23	0.35	0.13
	Silicate	0.07–0.17	0.14	0.12	0.07
Attenuation length	$Z_{remin}$	1,460	1,840	2,150	3,450
	CSH	3,300	1,320	1,820	2,660

All of the carbon fluxes are in  $\text{PgC y}^{-1}$ , silicate fluxes are respectively in  $\text{PmolSi y}^{-1}$ . The attenuation length of the biological fluxes is mentioned in m. The calcite saturation horizon (CSH) is the depth at which water masses are saturated in carbonate ions. The remineralization length ( $Z_{remin}$ ), is the depth at which 10 % of the organic matter fluxes at 100 m remain

indicate that all of the three models fail in replicating the chemical properties of NADW and AABW. The simulated NADW, characterized by positive Alk-DIC anomalies, contains much more  $\text{CO}_3^{2-}$  than expected from observation-based estimates. The opposite happens for the simulated AABW. The fact that deep waters overall contains much more DIC than observed seems to be related to the coarse resolution of models (Orr 2002). Especially in the North Atlantic, the vertical structure of Alk-DIC anomalies shows that the NADW export is shallower than estimated in the three models, especially in IPSLCM4 and IPSLCM5.

Although all of the three models present large discrepancies in ocean circulation, we cannot attribute bias in tracer distribution to solely the ocean dynamics. Our analysis suggested that the remineralization length and calcite saturation horizon are poorly represented systematically in all of the three models (Table 4). The Martin's exponent estimated from model output (Martin et al. 1987) indicates that the mean Martin exponent is about 0.79 for IPSLCM4, 0.75 for IPSLCM5 and 0.65 for CNRMCM5, while the expected Martin exponent is close to 0.858. This means that a large fraction of the sinking organic matter is remineralized at depth inducing bias in the simulated nutricline (Table 4, Figures S2 and S3). Such points highlight that the representation of the biological pump shall be improved in order to better simulated nutricline, carbochlorine and oxycline.

#### 4.2 Mixed-layer depth

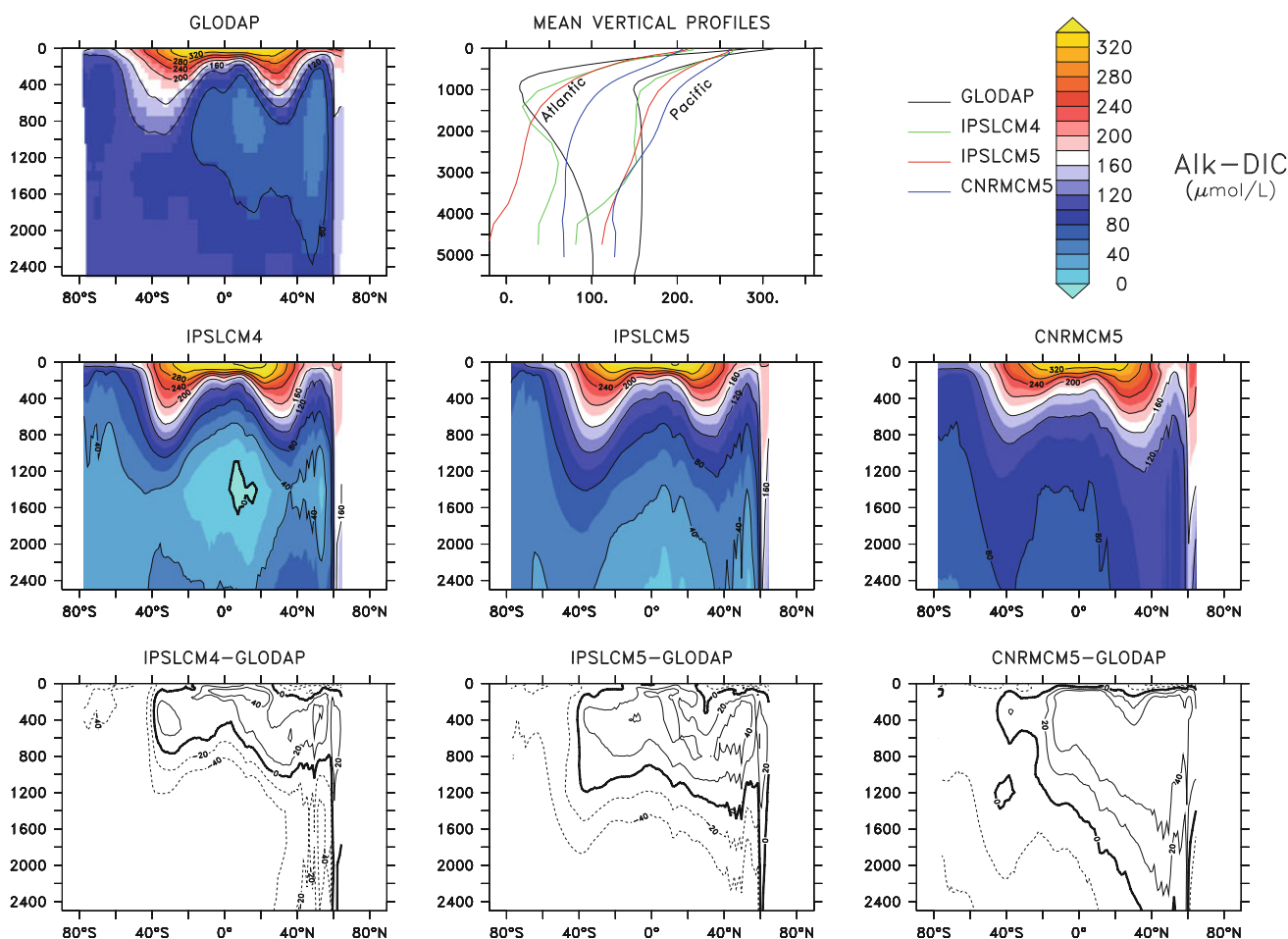
The mixed-layer depth is more realistic in IPSLCM5 and CNRMCM5 relative to the older IPSLCM4, particularly in summer. Hence the two AR5-era models provide better dynamics for phytoplankton growth in the light-limited high latitudes. Light limitation occurs when the depth of the mixed-layer is deeper than the critical depth. A summer

mixed-layer that is shallower than observed can induce excess simulated productivity, particularly in the regions where nutrients are replenished by the ocean dynamics (e.g., Sarmiento et al. 2004, Behrenfeld 2005, 2010; Behrenfeld et al. 2006). Thus biases in surface  $\text{CHL}_a$  often indicate problems with a model's local representation of light-to-nutrient limitation.

Even if SST is known to control the growth of phytoplankton impacting  $\text{CHL}_a$  surface concentrations, the correlation between simulated SST and  $\text{CHL}_a$  ( $R \sim 0.4$ ) is lower than the correlation between simulated mixed-layer depth and  $\text{CHL}_a$  ( $R > 0.7$ ) in high latitudes. In particular, mismatches between observed and simulated  $\text{CHL}_a$  concentrations in IPSLCM4 and IPSLCM5 in the Southern Ocean are linked to the mixed-layer (Fig. 10). The summer mixed-layer depth is shallower in both IPSLCM4 and IPSLCM5 than observed. These same models overestimate the prevalence of low surface  $\text{CHL}_a$  concentrations (i.e.,  $0.12\text{--}0.5 \mu\text{gChl L}^{-1}$ ) at surface due to the light-to-nutrient limitation. In IPSLCM5, the summer mixed-layer depth has improved relative to IPSLCM4 (Figs. 3, 10), but surface  $\text{CHL}_a$  remains higher than observed (Figs. 8, 10) due to the improved nutrient supply (see Sect. 4.3).

The modified mixed-layer parametrization used in CNRMCM5 leads to better represent phytoplankton dynamics in the high-latitudes. In the Southern Ocean, the simulated area of light-limited zone is  $150 \times 10^6 \text{ km}^2$  in CNRMCM5, while it is  $74 \times 10^6$  and  $91 \times 10^6 \text{ km}^2$  in IPSLCM4 and IPSLCM5, respectively. The deeper summer mixed-layer depth due to the wind energy means that phytoplankton become more light limited. Yet this coupling between surface winds and mixed-layer depth amplifies effects from atmosphere biases. Thus, in the Equatorial Pacific, CNRMCM5 overestimates the depth of the summer mixed-layer, which results in less primary productivity, thereby reducing its the estimation of global





**Fig. 9** Zonal-average annual-mean difference between alkalinity and dissolved inorganic carbon (Alk-DIC, in  $\mu\text{mol L}^{-1}$ ) for the GLODAP data-based estimate (Key et al. 2004), IPSLCM4, IPSLCM5, and CNRMCM5. *Bottom panels* represent zonal-average anomalies of

Alk-DIC (Models-GLODAP). *Thin solid lines* are positive anomalies, while *thin dashed lines* are negative. *Vertical profiles* represent basin-averages for the Atlantic and Pacific Ocean

marine production (Table 4); given that the Equatorial Pacific is one of the most productive region of this ocean (e.g., Behrenfeld et al. 2006).

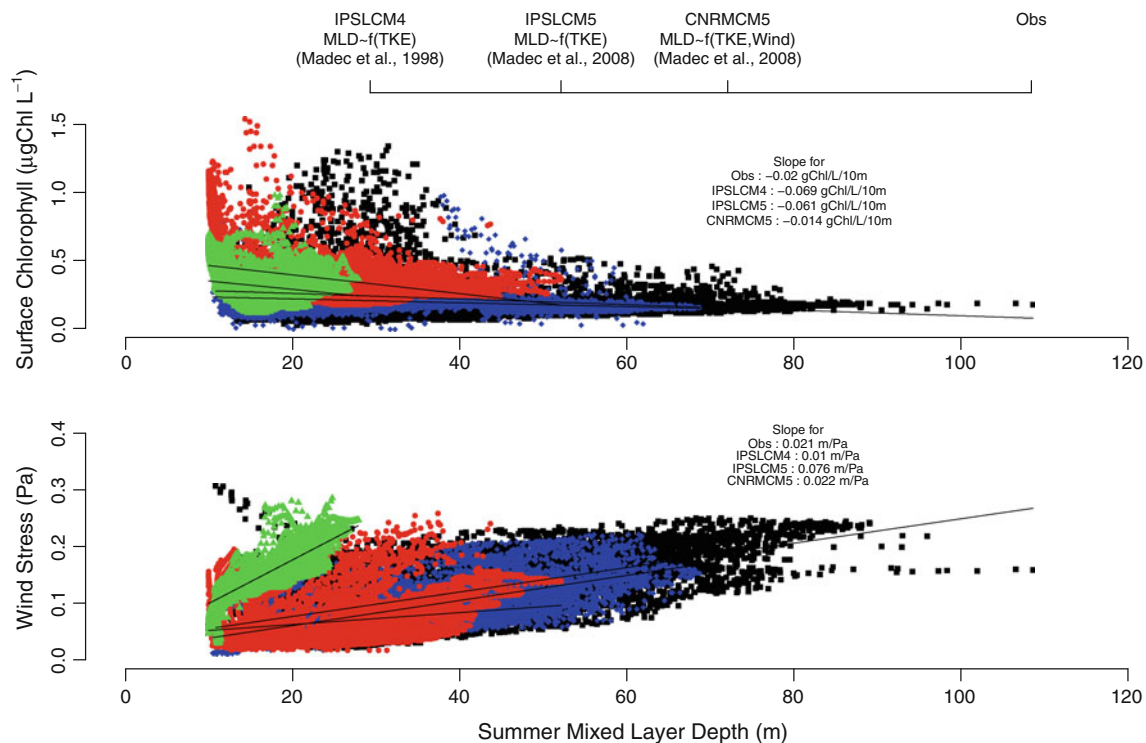
### 4.3 Surface winds

Our results suggest that surface wind patterns and large-scale geostrophic circulation are better in IPSLCM5 and CNRMCM5 compared to IPSLCM4. Surface winds affect on the ocean dynamics through Ekman transport, which in turn affects the distributions of tracers in the upper ocean. But the strength of nutrient supply from Ekman-driven upwelling is difficult to diagnose. As soon as nutrient-rich waters upwell, nutrients brought into the euphotic zone are rapidly consumed by the biota and advected by upper ocean circulation (Sarmiento et al. 2004).

Hourdin et al. (2012) that in climate models, horizontal resolution of the atmospheric component affects the location of westerly winds and hence the strength of the Antarctic

Circumpolar Current at Drake Passage. Thus atmospheric model resolution affects simulated ocean Ekman transport and thus nutrient supply. Furthermore, our results suggest a link between the location of maximum (winter) zonally-averaged wind stress (a proxy of westerly winds) and the maximum of zonal inventory of annual-mean macronutrients of the upper 500 m (Fig. 11). Although CNRMCM5 exhibits different resolution in its ocean component, we found that the location of maximum zonal inventory of annual-mean macronutrients and the location of maximum zonally-averaged wind stress are also shifted southward. Generally then, improving surface winds should *à priori* lead to more realistic distributions of biogeochemical tracers.

Nonetheless, we must be careful not to over-generalize. First, coarse-resolution ocean models tend to overestimate the response to surface winds (e.g., Farneti et al. 2010; Ito et al. 2010). Second, the three Earth system models studied here use ocean component, that does differ in terms of horizontal resolution, vertical resolution and parametrizations. A model's



**Fig. 10** Scatter-plot of surface chlorophyll ( $\mu\text{gChl L}^{-1}$ , *top panel*) and wind stress (in Pa, *bottom panel*) versus summer mixed-layer depth (m) for observations (*black*), IPSLCM4 (*green*), IPSLCM5 (*red*), and CNRMCM5 (*blue*) over the Southern Ocean ( $<30^\circ\text{S}$ ). For the observations, we use mixed-layer depth climatology (de Boyer Mont  gut et al. 2004), NCEP wind stress reanalysis (Kalnay et al.

1996) and the SeaWiFS chlorophyll climatology (Carr et al. 2006). *Black thin lines* represent the linear relationship between surface chlorophyll or wind stress and summer mixed-layer depth. Summer mixed-layer depths do not exceed 100 m typically; deep observed summer mixed-layer depths are due to the relatively low density difference used to define mixed-layer depth

response to surface wind forcing and resulting Ekman transport is highly dependent on the coefficients used for the Gent and McWilliams (1990) parametrization (Farneti et al. 2010; Sall  e et al. 2010). The choice of these coefficients in the three Earth system models, varies from  $600 \text{ m}^2 \text{ s}^{-1}$  in CNRMCM5 to  $2,000 \text{ m}^2 \text{ s}^{-1}$  in IPSLCM4 and IPSLCM5. Thus we cannot now separate effects purely due to resolution of the atmospheric component. A priority for future research should be to compare different versions of an Earth system model differing only by its atmospheric component resolution, as it was done in Russell et al. (2006), i.e., with the same ocean component and Gent and McWilliams (1990) parametrization.

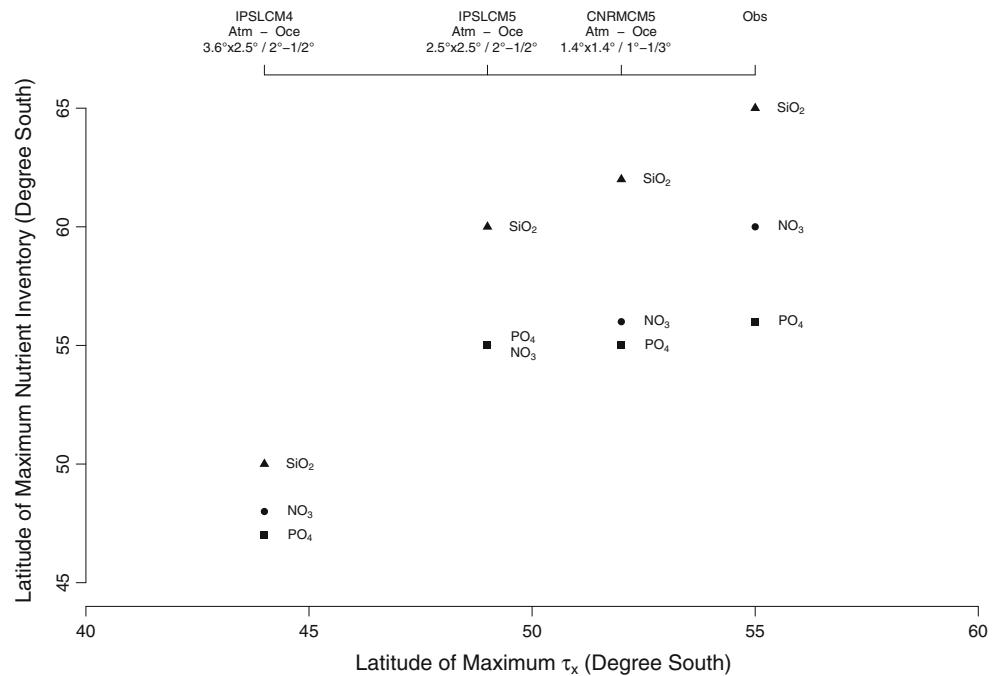
#### 4.4 Deep-ocean ventilation

The three models have difficulties to properly ventilate deep water masses. Thus they fail to adequately represent some ocean-atmosphere or ocean-sea ice interactions that drive light-to-dense water transformation, especially in the Southern Ocean. Our analysis indicates that most of these misrepresentations are fueled from biases in thermodynamic properties of the intermediate (i.e., between 800 and 1,200 m) and deep (i.e., between 1,200 and 2,500 m) water masses (Fig. 12). Particularly, all models fail to reproduce

the salinity of the intermediate and deep water masses. This points toward the poor representation of freshwater flux in models, although exchange of freshwater is one of the most important interactions between the ocean, the atmosphere, and the sea-ice in high latitudes. Biases in freshwater fluxes stem from a combination of misrepresentation or biases found in the different components of Earth system models. For example, it is likely the warm biases over the Southern Ocean, diagnosed from sea surface temperature anomalies (Fig. S1) inhibit subsurface waters from ventilating of the deep ocean. Yet model differences in high latitude salinity can stem from the different sea-ice models; knowing they exhibit large differences in term of sea-ice dynamics, thermodynamics, and rheology. For instance, sea-ice dynamics contributes to the formation of intermediate water mass that is characterized by a minimum of salinity (Sloyan and Rintoul 2001). Concerning CNRMCM5, it is likely that the stronger than observed freshwater flux could lighten surface density hence inducing a slow-down of the Southern Ocean deep ventilation.

Such differences in water mass properties induce biases in CFCs inventory (Figs. 12, S7). As IPSL models underestimate the light-to-dense formation of intermediate waters, the CFCs are stored in the mode waters (i.e.,

**Fig. 11** Scatter-plot of latitudinal locations of maximum (winter) wind stress and maximum (winter) on zonal macronutrients inventory between 0 and 500 m for observations, IPSLCM4, IPSLCM5, and CNRMCM5. Observations are from the wind stress reanalysis of NCEP (Kalnay et al. 1996) and the nitrate climatology of WOA 2009 (Locarnini et al. 2009)



between 0 and 1,000 m) and the deep waters (i.e., below 2,000 m). On the contrary, CFCs are stored in the upper layer of the intermediate waters in CNRMCM5, because the volume of this water mass is overestimated in CNRMCM5. The mean CFC11/CFC12 age indicates that mode and upper layer of the Southern Ocean intermediate waters (i.e., between 500 and 1,000 m) are older than observed by 5 years for IPSLCM5 and 12 years for CNRMCM5. This bias is typically found in z-coordinate ocean models, which struggle to produce adequate flow of dense waters along continental slopes (Doney 2004).

To evaluate models in terms of their export of young subsurface waters to the deep ocean, we use scatter plot of surface density anomalies versus anomalies of  $\text{O}_2$  inventory between 1,000 and 2,000 m (Fig. 13). This shows (1) how models mean-state compare to the observations at 95 % of significance and (2) how models compare in regards their skill in ventilating the deep ocean. The IPSL models broadly capture the distribution of surface density. Yet they slightly underestimate the inventory of  $\text{O}_2$  contained between 1,000 and 2,000 m. Conversely, CNRMCM5 completely fails to reproduce the distribution of  $\text{O}_2$  inventory with the right surface density. This result is consistent with Fig. 12 showing that CNRMCM5's surface density is too low; thus does not ventilate deep waters with an adequate supply of  $\text{O}_2$ .

Thus improved ocean-sea ice coupling should lead to a better representation of the sea surface salinity in the high latitudes, a major weakness of these models, helping to improve the representation of the winter mixed-layer

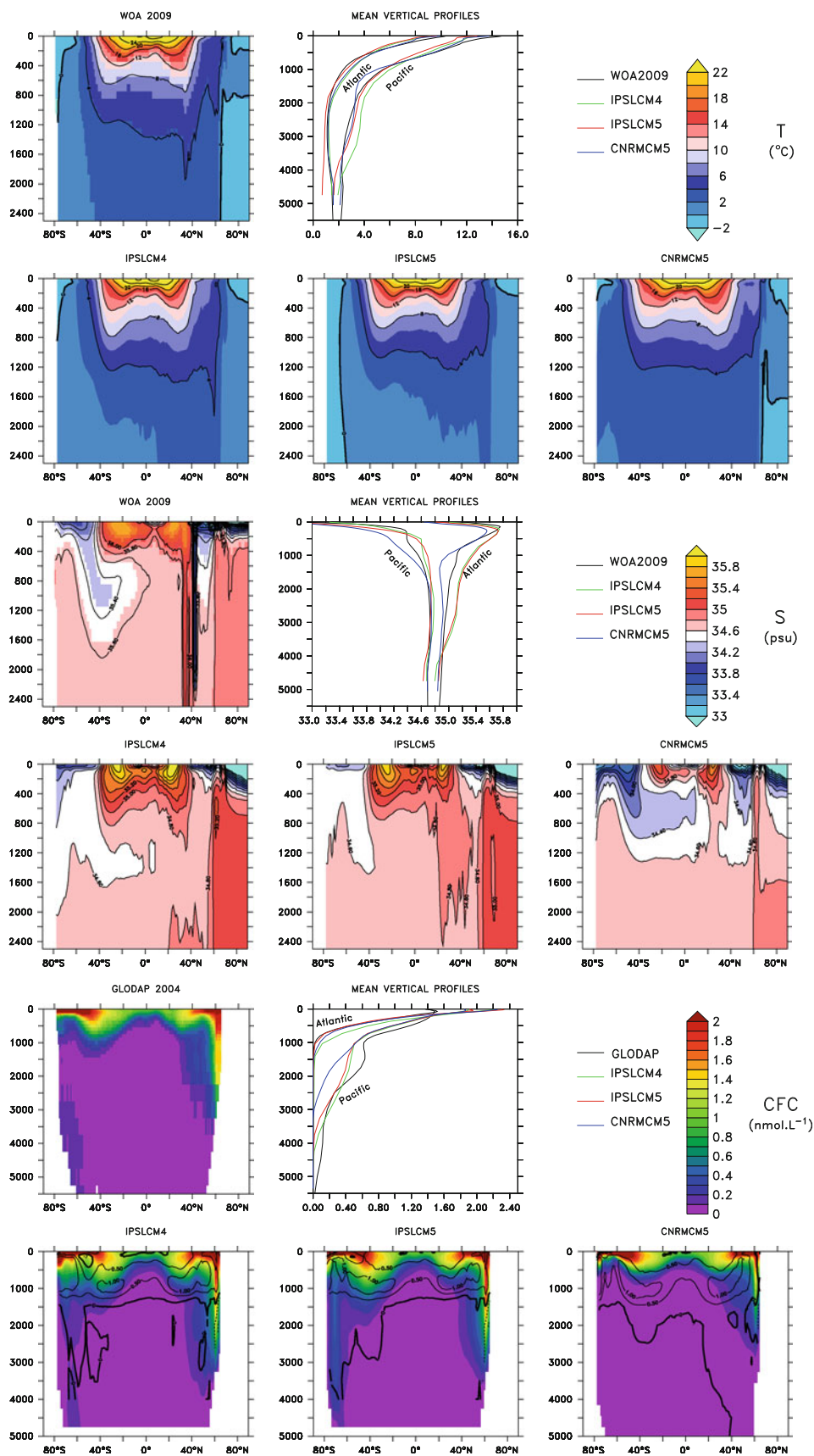
depths and deep-ocean ventilation, especially in the Southern Ocean.

Biases in deep-ocean ventilation affect subsurface distribution of tracers (e.g., CFC11 concentrations, Figs. 4, 12). In CNRMCM5, sluggish deep ventilation results in suboxic deep waters ( $\text{O}_2$  concentration are lower than  $60 \mu\text{mol L}^{-1}$ ), particularly in the North Pacific. Suboxic waters stimulate denitrification, a sink for  $\text{NO}_3^-$  at depth. In the PISCES model, the nitrogen that is removed by denitrification is compensated by nitrogen fixation at surface. Thus nitrogen is conserved but errors in deep-ocean  $\text{O}_2$  result in errors in denitrification and those propagate to the surface ocean through compensating nitrogen fixation.

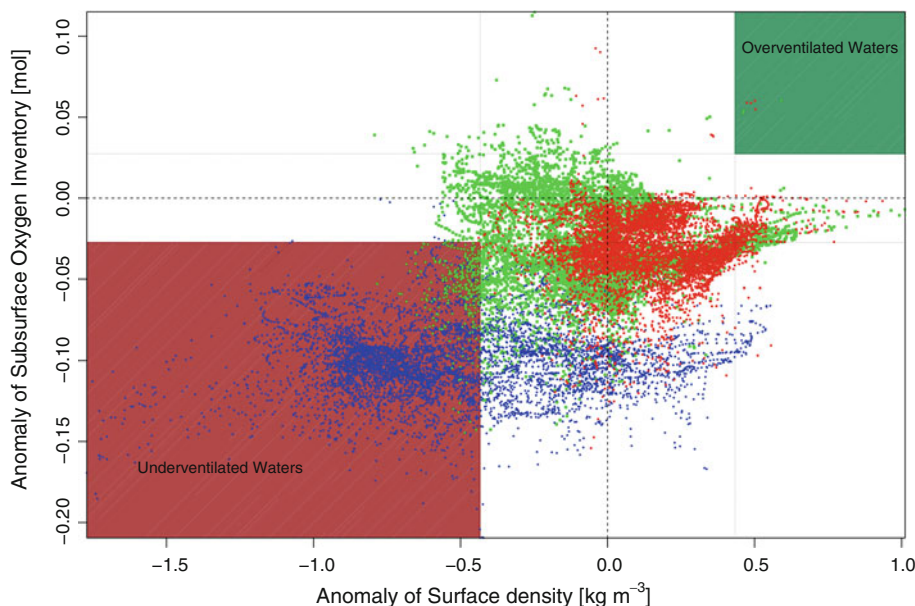
#### 4.5 Spinup strategy versus skill assessment

Using physical fields to diagnose problems in deep-ocean ventilation is not straightforward. It requires for example a complex water mass framework (Iudicone et al. 2008; Downes et al. 2011). In comparison, biogeochemical tracers that are integrated for thousand years or more, in order to reach near-steady state conditions, contain strong signatures from the circulation. In that context, our results demonstrate that passive tracers like CFCs, which are integrated during  $\sim 50$  years, provide a good basis for assessing models. It allows to enrich the observation mix hence better constrain ocean dynamics. Thus there are several advantages to using biogeochemical tracers in addition to diagnose only the ocean dynamics; especially because some biogeochemical data like surface chlorophyll

**Fig. 12** Zonal average of seasonal climatology of temperature ( $T$ , in  $^{\circ}\text{C}$ ), salinity ( $S$ ), and CFC12 ( $\text{nmol L}^{-1}$ ) over 1995-2005. Observed climatology is based on GLODAP (Key et al. 2004) for CFC12 and World Ocean Atlas 2009 for temperature (Locarnini et al. 2009) and salinity (Antonov et al. 2009). Anomalies between modeled and observed CFC concentrations are superimposed in *black thin lines* over the IPSLCM4, IPSLCM5 and CNRMCM5 panels of zonal averaged CFC concentrations



**Fig. 13** Scatter plot of the winter averaged surface density anomalies (in  $\text{kg m}^{-3}$ ) versus anomalies of annual-mean oxygen inventory between 1,000 and 2,000 m. Dashed black lines represent the 95 % confidence interval for surface density anomalies and oxygen inventory anomalies. Green, red, blue point represents IPSLCM4, IPSLCM5 and CNRMCM5 respectively



or primary productivity are continuously monitored several decades.

However, it does matter how an ocean biogeochemical model is initialized. By first integrating an ocean biogeochemical model offline for thousands of years, until it reaches near steady state “natural” conditions, resulting biogeochemical fields are consistent with its ocean circulation fields. At this state, the intrinsic variability of the biogeochemical fields is stronger than their drift. Alternatively, the more economical approach of initializing the ocean biogeochemical model with observed climatologies and integrating for much shorter periods before it is run as part of the Earth system model guarantees that simulated ocean biogeochemical fields are close to those observed but at the cost of substantial model drift. The “steady-state” spin-up is preferable for analyzing the seasonal cycle, climate variability, and regional sensitivities as well as the model’s true natural state.

However, our results illustrate that the same approach may be problematic if a model poorly represents deep-ocean circulation. Thus not only model skill but also projections depend on the approach selected for initializing the ocean biogeochemical model. Yet in CMIP5, there is no recommended protocol for initializing biogeochemical models. In subsequent international efforts to compare Earth system models, we recommend that model groups discuss this issue to determine if a common approach is possible and to evaluate the consequences if one is not taken. Meanwhile, sensitivity tests using both approaches to initialize the ocean biogeochemical component of one or more Earth system models would be of great interest for the community.

## 5 Conclusions

We have assessed three Earth system models with a common marine biogeochemistry: IPSL-CM4-LOOP, IPSL-CM5A-LR and CNRM-CM5.1. The former contributed to IPCC AR4 (CMIP3) while the latter two are contributing to IPCC AR5 (CMIP5). Evaluation of physical fields (wind stress, short-wave radiation, net heat flux, SST, SSS, MLD, and water mass transport) and biogeochemical fields (macronutrients,  $\text{O}_2$ , DIC, Alk,  $\text{CHL}_a$ ,  $\Delta p\text{CO}_2$ , air-sea exchange of  $\text{CO}_2$ , export of organic matter, silicate and calcite production, and primary productivity) demonstrated that models generally differ most in the Southern Ocean. In that region, IPSL-CM5A-LR and CNRM-CM5.1 offer much poorer representations of deep-ocean circulation than does IPSL-CM4-LOOP. Model biases in deep-ocean ventilation affect deep tracer concentrations, which are lower than observed. Nonetheless, improvements in surface winds, mixed-layer depth, and large-scale geostrophic circulation for IPSL-CM5A-LR and CNRM-CM5.1 contribute to produce more accurate simulated surface-ocean biogeochemical fields (e.g.,  $\text{CHL}_a$ , meridional gradients of macronutrients, location of ingassing and outgassing regions). Our results emphasize the need of Earth system models to better represent the summer mixed-layer depth, which plays a fundamental role when simulating marine biogeochemistry, particularly in the high latitudes.

Our study demonstrates that a systematic methodology based on basic skill assessment metrics (e.g., bias, root mean square error) can be valuable in understanding ocean dynamics and biogeochemistry of Earth system models. Indeed, much could be learned from basic statistical metrics and ocean tracers in a routinely skill assessment process; especially because the use of biogeochemical or

passive tracers can help to constrain the ocean dynamics through additional data. Concerning the skill assessment of biogeochemical models, the use of derived metrics like Alk-DIC allows focusing on models properties, not their biases.

For the next IPCC assessment, much could be learned if Earth system models would include one simple passive tracer, like CFC11, in their ocean component (e.g., run optionally for 50 years as one additional passive tracer in the online ocean component). With this one anthropogenic tracer, it would be straightforward to assess ocean model skill in ventilating subsurface waters (e.g., England 1995; Smethie et al. 2001; Dutay et al. 2002; Matsumoto et al. 2004, 2010) which is crucial for ocean biogeochemistry. Such simulations were performed routinely a decade ago for evaluating ocean circulation component of ocean carbon models during OCMIP. It would be highly advantageous to institute the same practice for subsequent phases of CMIP and even in the ongoing ocean models development.

However, skill assessment of ocean dynamics is not the whole story. Earth system models contributing to CMIP5 differ enormously, both in their physical components (atmosphere, ocean, sea-ice) and in their biogeochemical components. As atmosphere, ocean, land and sea-ice models, biogeochemical models also need to be improved to better represent subtle biogeochemical processes (Doney 1999). Although the PISCES ocean biogeochemical model that we have used here may be considered state of the art, there is still need for improvements (e.g., biological pump) and greater diversity (e.g., non-Redfield models, phytoplankton type models). In that context, sensitivity experiments and parametrization comparison (e.g., Gehlen et al. 2006; Krist et al. 2010) would be conducted in parallel to the ongoing coupled models skill assessment.

**Acknowledgments** The authors thank the four anonymous reviewers for helpful comments on the manuscript as well as the Climate team at M  t  o-France and IPSL. This work was supported by GENCI (Grand Equipement National de Calcul Intensif) at CCRT (Centre de Calcul Recherche et Technologie), allocation 016178, and M  t  o-France/DSI supercomputing center. This work was supported through EU FP7 project CARBOCHANGE Changes in carbon uptake and emissions by oceans in a changing climate which received funding from the European Community's Seventh Framework Programme under grant agreement no. 264879 and also EU FP7 project EPOCA (European Project on Ocean Acidification) under grant agreement no. 211384.

**Open Access** This article is distributed under the terms of the Creative Commons Attribution License which permits any use, distribution, and reproduction in any medium, provided the original author(s) and the source are credited.

## References

- Antonov JI et al (2009) World ocean database 2009, vol 2: salinity. US Government Printing Office, Washington, D.C
- Aumont O (2003) An ecosystem model of the global ocean including Fe, Si, P colimitations. *Global Biogeochem Cycles* 17(2). doi: [10.1029/2001GB001745](https://doi.org/10.1029/2001GB001745)
- Aumont O, Bopp L (2006) Globalizing results from ocean in situ iron fertilization studies. *Global Biogeochem Cycles* 20(2). doi: [10.1029/2005GB002591](https://doi.org/10.1029/2005GB002591)
- Aumont O, Orr JC, Monfray P, Ludwig W, Amiotte-Suchet P, Probst J-L (2001) Riverine-driven interhemispheric transport of carbon. *Global Biogeochem Cycles* 15(2):393–405. doi: [10.1029/1999GB001238](https://doi.org/10.1029/1999GB001238)
- Axell L (2002) Wind-driven internal waves and Langmuir circulations in a numerical ocean model of the southern Baltic Sea. *J Geophys Res Oceans* 107. doi: [10.1029/2001JC000922](https://doi.org/10.1029/2001JC000922)
- Barnier B et al (2006) Impact of partial steps and momentum advection schemes in a global ocean circulation model at eddy-permitting resolution. *Ocean Dyn* 56(5–6):543–567. doi: [10.1007/s10236-006-0082-1](https://doi.org/10.1007/s10236-006-0082-1)
- Behrenfeld MJ (2005) Carbon-based ocean productivity and phytoplankton physiology from space. *Global Biogeochem Cycles* 19(1). doi: [10.1029/2004GB002299](https://doi.org/10.1029/2004GB002299)
- Behrenfeld MJ (2010) Abandoning Sverdrup's critical depth hypothesis on phytoplankton blooms. *Ecology* 91(4):977–989. doi: [10.1890/09-1207.1](https://doi.org/10.1890/09-1207.1)
- Behrenfeld MJ et al (2006) Climate-driven trends in contemporary ocean productivity. *Nature* 444(7120):752–755. doi: [10.1038/nature05317](https://doi.org/10.1038/nature05317)
- Blanke B, Delecluse P (1993) Variability of the tropical Atlantic ocean simulated by a general circulation model with two different mixed-layer physics. *J Phys Oceanogr* 23(7):1363–1388
- Burchard H, Rennau H (2008) Comparative quantification of physically and numerically induced mixing in ocean models. *Ocean Model* 20(3):293–311. doi: [10.1016/j.ocemod.2007.10.003](https://doi.org/10.1016/j.ocemod.2007.10.003)
- Cadule P, Friedlingstein P, Bopp L, Sitch S, Jones CD, Ciais P, Piao SL, Peylin P (2010) Benchmarking coupled climate-carbon models against long-term atmospheric CO<sub>2</sub> measurements. *Global Biogeochem Cycles* 24(2):GB2016. doi: [10.1029/2009GB003556](https://doi.org/10.1029/2009GB003556)
- Cariolle D, Teyssedre H (2007) A revised linear ozone photochemistry parameterization for use in transport and general circulation models: multi-annual simulations. *Atmos Chem Phys* 7(9): 2183–2196
- Carr M-E et al (2006) A comparison of global estimates of marine primary production from ocean color. *Deep-Sea Res Part II* 53(5–7):741–770. doi: [10.1016/j.dsr2.2006.01.028](https://doi.org/10.1016/j.dsr2.2006.01.028)
- Cunningham SA (2003) Transport and variability of the Antarctic circumpolar current in drake passage. *J Geophys Res* 108(C5). doi: [10.1029/2001JC001147](https://doi.org/10.1029/2001JC001147)
- de Baar HJW, de Jong JTM (2001) The biogeochemistry of iron in seawater. Distributions, sources and sinks of iron in seawater. Wiley, Hoboken, NJ
- de Boyer Mont  gut C (2004) Mixed layer depth over the global ocean: an examination of profile data and a profile-based climatology. *J Geophys Res* 109(C12). doi: [10.1029/2004JC002378](https://doi.org/10.1029/2004JC002378)
- D  qu   M, Dreveton C, Braun A, Cariolle D (1994) The ARPEGE/IFS atmosphere model: a contribution to the French community climate modelling. *Clim Dyn* 10:249–266
- Doney S (1999) Major challenges confronting marine biogeochemical modeling. *Global Biogeochem Cycles* 13(3):705–714
- Doney SC (2004) Evaluating global ocean carbon models: the importance of realistic physics. *Global Biogeochem Cycles* 18(3). doi: [10.1029/2003GB002150](https://doi.org/10.1029/2003GB002150)
- Downes SM, Bindoff NL, Rintoul SR (2011) Impacts of climate change on the subduction of mode and intermediate water masses in the southern ocean. *J Clim* 22(12):3289–3302. doi: [10.1175/2008JCLI2653.1](https://doi.org/10.1175/2008JCLI2653.1)

- Dufresne J-L et al (2012) The IPSL-CM5A earth system model: general description and climate change projections. *Clim Dyn* (submitted)
- Dunne JP, Sarmiento JL, Gnanadesikan A (2007) A synthesis of global particle export from the surface ocean and cycling through the ocean interior and on the seafloor. *Global Biogeochem Cycles* 21(4). doi:10.1029/2006GB002907
- Dutay J-C et al (2002) Evaluation of ocean model ventilation with CFC-11: comparison of 13 global ocean models. *Ocean Model* 4(2):89–120. doi:10.1016/S1463-5003(01)00013-0
- England MH (1995) The age of water and ventilation timescales in a global ocean model. *J Phys Oceanogr* 25(11):2756–2777
- Eppley RW, Rogers JN, McCarthy JJ (1969) Half-saturation constants for uptake of nitrate and ammonium by marine phytoplankton. *Limnol Oceanogr* 14:912–920
- Farneti R, Delworth TL, Rosati AJ, Griffies SM, Zeng F (2010) The role of Mesoscale Eddies in the rectification of the southern ocean response to climate change. *J Phys Oceanogr* 40(7):1539–1557. doi:10.1175/2010JPO4353.1
- Fasham M, Sarmiento J, Slater R, Ducklow H, Williams R (1993) Ecosystem behavior at Bermuda station-S and Ocean Weather Station India—a general-circulation model and observational analysis. *Global Biogeochem Cycles* 7(2):379–415
- Fichefet T, Maqueda MAM (1997) Sensitivity of a global sea ice model to the treatment of ice thermodynamics and dynamics. *J Geophys Res* 102(C6):12609–12646
- Friedrichs MAM, et al. (2007) Assessment of skill and portability in regional marine biogeochemical models: role of multiple planktonic groups. *J Geophys Res* 112(C8). doi:10.1029/2006JC003852
- Friedrichs MAM et al (2009) Assessing the uncertainties of model estimates of primary productivity in the tropical Pacific Ocean. *J Mar Syst* 76(1–2):113–133. doi:10.1016/j.jmarsys.2008.05.010
- Ganachaud A (2003) Large-scale mass transports, water mass formation, and diffusivities estimated from World ocean circulation experiment (WOCE) hydrographic data. *J Geophys Res Oceans* 108. doi:10.1029/2002JC001565
- Garcia C, Garcia V, McClain C (2005) Evaluation of SeaWiFS chlorophyll algorithms in the Southwestern Atlantic and Southern Oceans. *Remote Sens Environ* 95(1):125–137. doi:10.1016/j.rse.2004.12.006
- Garcia HE, Locarnini RA, Boyer TP, Antonov JI, Baranova OK, Zweng MM, Johnson DR (2009a) World ocean database 2009, vol 3: dissolved oxygen, apparent oxygen utilization, and oxygen saturation. US Government Printing Office, Washington, DC
- Garcia HE, Locarnini RA, Boyer TP, Antonov JI, Zweng MM, Baranova OK, Johnson DR (2009b) World ocean database 2009, vol 4: nutrients (phosphate, nitrate, silicate). US government printing office, Washington, DC
- Gehlen M, Bopp L, Emprin N, Aumont O, Heinze C, Ragueneau O (2006) Reconciling surface ocean productivity, export fluxes and sediment composition in a global biogeochemical ocean model. *Biogeosciences* 3(4):521–537
- Gehlen M, Gangstør R, Schneider B, Bopp L, Aumont O, Ethe C (2007) The fate of pelagic CaCO<sub>3</sub> production in a high CO<sub>2</sub> ocean: a model study. *Biogeosciences* 4(4):505–519. doi:10.5194/bg-4-505-2007
- Geider R, MacIntyre H, Kana T (1998) A dynamic regulatory model of phytoplankton acclimation to light, nutrients, and temperature. *Limnol Oceanogr* 43(4):679–694
- Gent PR, McWilliams JC (1990) Isopycnal mixing in ocean circulation models. *J Phys Oceanogr* 20(1):150–155
- Gupta S, Ritchey N, Wilber A, Whitlock C (1999) A climatology of surface radiation budget derived from satellite data. *J Clim* 12:2691–2710
- Heinze C, Hupe A, Maier-Reimer E, Dittert N, Ragueneau O (2003) Sensitivity of the marine biospheric Si cycle for biogeochemical parameter variations. *Global Biogeochem Cycles* 17(3). doi:10.1029/2002GB001943
- Hewitt HT, Copsey D, Culverwell ID, Harris CM, Hill RSR, Keen AB, McLaren AJ et al (2011) Design and implementation of the infrastructure of HadGEM3: the next-generation Met Office climate modelling system. *Geosci Model Dev* 4(2):223–253. doi:10.5194/gmd-4-223-2011
- Hood RR et al (2006) Pelagic functional group modeling: progress, challenges and prospects. *Deep Sea Res Part II Top Stud Oceanogr* 53:459–512
- Houghton R, Marland G (2007) Contributions to accelerating atmospheric CO<sub>2</sub> growth from economic activity, carbon intensity, and efficiency of natural sinks. *Proceedings of the national academy of sciences*
- Hourdin F et al (2006) The LMDZ4 general circulation model: climate performance and sensitivity to parametrized physics with emphasis on tropical convection. *Clim Dyn* 27:87–813. doi:10.1007/s00382-006-0158-0
- Hourdin F et al (2012) Climate and sensitivity of the IPSL-CM5A coupled model: impact of the LMDZ atmospheric grid configuration. *Clim Dyn* (submitted)
- Hunke EC, Dukowicz JK (2012) An elastic–viscous–plastic model for sea ice dynamics. *J Phys Oceanogr* 27(9):1849–1867
- Ito T, Woloszyn M, Mazloff M (2010) Anthropogenic carbon dioxide transport in the Southern Ocean driven by Ekman flow. *Nature* 463(7277):80–83. doi:10.1038/nature08687
- Iudicone D, Madec G, McDougall TJ (2008) Water-mass transformations in a neutral density framework and the key role of light penetration. *J Phys Oceanogr* 38(7):1357–1376. doi:10.1175/2007JPO3464.1
- Jickells T, Spokes L (2001) The biogeochemistry of iron in seawater. Atmospheric iron inputs to the oceans. Wiley, Hoboken, NJ
- Jin X, Gruber N, Dunne JP, Sarmiento J (2006) Diagnosing the contribution of phytoplankton functional groups to the production and export of particulate organic carbon, CaCO<sub>3</sub>, and opal from global nutrient and alkalinity distributions. *Global Biogeochem Cycles* 20. doi:10.1029/2005GB002532
- Johnson K, Chavez F, Friederich G (1999) Continental-shelf sediment as a primary source of iron for coastal phytoplankton. *Nature* 398(6729):697–700
- Kalnay E et al (1996) The NCEP/NCAR 40-year reanalysis project. *Bull Am Meteor Soc* 77(3):437–472
- Key R et al (2004) A global ocean carbon climatology: results from global data analysis project (GLODAP). *Global Biogeochem Cycles* 18(4):GB4031. doi:10.1029/2004GB002247
- Koch-Larrouy A, Lengaigne M, Terray P, Madec G, Masson S (2010) Tidal mixing in the Indonesian Seas and its effect on the tropical climate system. *Clim Dyn* 34:891–904. doi:10.1007/s00382-009-0642-4
- Koch-Larrouy A, Madec G, Bouruet-Aubertot P, Gerkema T, Bessières L, Molcard R (2007) On the transformation of Pacific water into Indonesian throughflow water by internal tidal mixing. *Geophys Res Lett* 34(4). doi:10.1029/2006GL028405
- Kriest I, Khattiwala S, Oschlies A (2010) Towards an assessment of simple global marine biogeochemical models of different complexity. *Progr Oceanogr* 86(3–4):337–360. doi:10.1016/j.pocyan.2010.05.002
- Krinner G et al (2005) A dynamic global vegetation model for studies of the coupled atmosphere-biosphere system. *Global Biogeochem Cycles* 19(1):1–33
- Laws E, Falkowski P, Smith W, Ducklow H, McCarthy J (2000) Temperature effects on export production in the open ocean. *Global Biogeochem Cycles* 14(4):1231–1246
- Le Quééré C et al (2005) Ecosystem dynamics based on plankton functional types for global ocean biogeochemistry models. *Global Change Biol* 11(11):2016–2040. doi:10.1111/j.1365-2468.2005.01004.x

- Lef  vre F, Brasseur GP, Folkins I, Smith AK, Simon P (1994) Chemistry of the 1991–1992 stratospheric winter: three-dimensional model simulations. *J Geophys Res* 99(D4):8183–8195. doi:[10.1029/93JD03476](https://doi.org/10.1029/93JD03476)
- Lef  vre F, Figarol F, Carslaw KS, Peter T (1998) The 1997 Arctic Ozone depletion quantified from three dimensional model simulations. *Geophys Res Lett* 25(13):2425–2428. doi:[10.1029/98GL51812](https://doi.org/10.1029/98GL51812)
- Lengaigne M, Madec G, Bopp L, Menkes C, Aumont O, Cadule P (2009) Bio-physical feedbacks in the Arctic Ocean using an earth system model. *Geophys Res Lett* 36(21):L21602. doi:[10.1029/2009GL040145](https://doi.org/10.1029/2009GL040145)
- Lipschultz F, Wofsy S, Ward B, Codispoti L, Friedrich G, Elkins J (1990) Bacterial transformations of inorganic nitrogen in the oxygen-deficient waters of the Eastern Tropical South Pacific Ocean. *Deep-Sea Res* 37(10):1513–1541. doi:[10.1016/0198-0149\(90\)90060-9](https://doi.org/10.1016/0198-0149(90)90060-9)
- Locarnini RA, Mishonov AV, Antonov JI, Boyer TP, Garcia OHE, Baranova OK, Zweng MM (2009) World ocean database 2009, vol 1: temperature. US government printing office, Washington, DC
- Ludwig W, Probst J, Kempe S (1996) Predicting the oceanic input of organic carbon by continental erosion. *Global Biogeochem Cycles* 10(1):23–41
- Madec G. (2008a) NEMO ocean engine. Institut Pierre-Simon Laplace (IPSL), France, institut pierre-simon laplace (IPSL), France ed
- Madec G (2008b) Global ocean storage of anthropogenic carbon. Institut Pierre-Simon Laplace (IPSL), France, institut pierre-simon laplace (IPSL), Gif-sur-Yvette, France
- Madec G, Delecluse P, Levy C (1998) Ocean general circulation model reference manual—note du p  le de mod  lisation. Institut Pierre Simon Laplace (IPSL)
- Manzi AO, Planton S (1994) Implementation of the ISBA parameterization scheme for land surface processes in a GCM—an annual cycle experiment. *J Hydrol* 155(3–4):353–387. doi:[10.1016/0022-1694\(94\)90178-3](https://doi.org/10.1016/0022-1694(94)90178-3)
- Marti O et al (2010) Key features of the IPSL ocean atmosphere model and its sensitivity to atmospheric resolution. *Clim Dyn* 34(1):1–26. doi:[10.1007/s00382-009-0640-6](https://doi.org/10.1007/s00382-009-0640-6)
- Martin J-H, Knauer G, Karl D, Broenkow W (1987) VERTEX: carbon cycling in the northeast Pacific. *Deep-Sea Res* 34(2):267–285. doi:[10.1016/0198-0149\(87\)90086-0](https://doi.org/10.1016/0198-0149(87)90086-0)
- Matsumoto K, Sarmiento JL, Key RM, Aumont O, Bullister JL, Caldeira K, Campin J-M et al (2004) Evaluation of ocean carbon cycle models with data-based metrics. *Geophys Res Lett* 31(7):L07303. doi:[10.1029/2003GL018970](https://doi.org/10.1029/2003GL018970)
- Matsumoto K, Tokos KS, Chikamoto MO, Ridgwell A (2010) Characterizing post-industrial changes in the ocean carbon cycle in an Earth system model. *Tellus B* 62(4):296–313. doi:[10.1111/j.1600-0889.2010.00461.x](https://doi.org/10.1111/j.1600-0889.2010.00461.x)
- Meehl G, Covey C, Delworth T, Latif M, McAvaney B, Mitchell J, Stouffer R, Taylor K (2007) The WCRP CMIP3 multimodel dataset: a new era in climate change research. *Bull Am Meteorol Soc* 88:1383–1394
- Mellor G, Blumberg A (2004) Wave breaking and ocean surface layer thermal response. *J Phys Oceanogr* 34(3):693–698. doi:[10.1175/2517.1](https://doi.org/10.1175/2517.1)
- Merryfield WJ, Holloway G, Gargett AE (1999) A global ocean model with double-diffusive mixing. *J Phys Oceanogr* 29(6):1124–1142
- Middelburg J, Soetaert K, Herman P, Heip C (1996) Denitrification in marine sediments: a model study. *Global Biogeochem Cycles* 10(4):661–673
- Mikalhof-Fletcher S, et al (2007) Inverse estimates of the oceanic sources and sinks of natural CO<sub>2</sub> and the implied oceanic carbon transport. *Global Biogeochemical Cycles* 21(1). doi:[10.1029/2006GB002751](https://doi.org/10.1029/2006GB002751)
- Moore J, Doney S, Kleypas J, Glover D, Fung I (2002) An intermediate complexity marine ecosystem model for the global domain. *Deep Sea Res Part II* 49:403–462
- Moore J, Doney S, Lindsay K (2004) Upper ocean ecosystem dynamics and iron cycling in a global three-dimensional model. *Global Biogeochem Cycles* 18(4). doi:[10.1029/2004GB002220](https://doi.org/10.1029/2004GB002220)
- Najjar R, et al (2007) 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 inter-comparison project (OCMIP-2). *Global Biogeochem Cycles* 21(3). doi:[10.1029/2006GB002857](https://doi.org/10.1029/2006GB002857)
- Noilhan J, Mahfouf JF (1996) The ISBA land surface parameterisation scheme. *Global Planet Change* 13(1–4):145–159. doi:[10.1016/0921-8181\(95\)00043-7](https://doi.org/10.1016/0921-8181(95)00043-7)
- Orr JC (2002) Global Ocean Storage of Anthropogenic Carbon (GOSAC), final report, EC Environment and Climate Programme, Institut Pierre Simon Laplace, Paris
- Randall DA, Wood RA, UK (2007) Climate models and their evaluation. Climate change 2007: the physical science basis. Contribution of working group I to the fourth assessment report of the intergovernmental panel on climate change, Cambridge University Press, Cambridge, UK, New York, NY, USA, pp 1–74
- Russell JL, Dixon KW, Gnanadesikan A, Stouffer RJ, Toggweiler JR (2006) The Southern Hemisphere westerlies in a warming world: propping open the door to the deep ocean. *J Clim* 19(24):6382–6390
- Saba VS et al (2010) Challenges of modeling depth-integrated marine primary productivity over multiple decades: a case study at BATS and HOT. *Global Biogeochem Cycles* 24(3). doi:[10.1029/2009GB003655](https://doi.org/10.1029/2009GB003655)
- Salas y M  lia D (2002) A global coupled sea ice-ocean model. *Ocean Model* 4:137–172
- Sall  e J-B, Speer K, Rintoul S, Wijffels S (2010) Southern ocean thermocline ventilation. *J Phys Oceanogr* 40(3):509–529. doi:[10.1175/2009JPO4291.1](https://doi.org/10.1175/2009JPO4291.1)
- Sarmiento J, Gruber N, Brzezinski M, Dunne JP (2004) High-latitude controls of thermocline nutrients and low latitude biological productivity. *Nature* 427(6969):56–60
- Sarmiento J, Monfray P, Maier-Reimer E, Aumont O, Murnane R, Orr J (2000) Sea-air CO<sub>2</sub> fluxes and carbon transport: a comparison of three ocean general circulation models. *Global Biogeochem Cycles* 14(4):1267–1281. doi:[10.1029/1999GB900062](https://doi.org/10.1029/1999GB900062)
- Sarmiento JL, Gruber N (2006) Ocean biogeochemical dynamics. Princeton University Press, Princeton
- Schlitzer R (2002) Carbon export fluxes in the Southern Ocean: results from inverse modeling and comparison with satellite-based estimates. *Deep-Sea Research Part II, The Southern Ocean I: climatic changes in the cycle of carbon in the Southern Ocean* 49(9–10):1623–1644. doi:[10.1016/S0967-0645\(02\)00004-8](https://doi.org/10.1016/S0967-0645(02)00004-8)
- Schneider B et al (2008) Climate-induced interannual variability of marine primary and export production in three global coupled climate carbon cycle models. *Biogeosciences* 5(2):597–614
- Schulz M (2007) Constraining model estimates of the aerosol radiative forcing. Ph. D. thesis, Universit   Pierre et Marie Curie, Paris VI
- Siegel DA, Doney SC, Yoder JA (2002) The North Atlantic spring phytoplankton bloom and the Sverdrup’s critical depth hypothesis. *Science* 296:730–733. doi:[10.1126/science.1069174](https://doi.org/10.1126/science.1069174)
- Simmons H, Jayne S, St Laurent L, Weaver A (2004) Tidally driven mixing in a numerical model of the ocean general circulation. *Ocean Model* 6:245–263. doi:[10.1016/S1463-5003\(03\)00011-8](https://doi.org/10.1016/S1463-5003(03)00011-8)
- Sloyan BM, Rintoul SR (2001) The southern ocean limb of the global deep overturning circulation. *J Phys Oceanogr* 31(1):143–173



- Smethie W, Fine R (2001) Rates of north Atlantic deep water formation calculated from chlorofluorocarbon inventories. *Deep-Sea Res Part I* 48(1):189–215
- Soetaert K, Middelburg J, Herman P, Buis K (2000) On the coupling of benthic and pelagic biogeochemical models. *Earth Sci Rev* 51:173–201
- Steinacher M et al (2010) Projected 21st century decrease in marine productivity: a multi-model analysis. *Biogeosciences* 7(3):979–1005. doi:[10.5194/bg-7-979-2010](https://doi.org/10.5194/bg-7-979-2010)
- Stow CA, Jolliff J, McGillicuddy DJJ, Doney SC, Allen JJ, Friedrichs MAM, Rose KA, Wallhead P (2009) Skill assessment for coupled biological/physical models of marine systems. *J Mar Syst* 76:4–15. doi:[10.1016/j.jmarsys.2008.03.011](https://doi.org/10.1016/j.jmarsys.2008.03.011)
- Sunda W, Huntsman S (1997) Interrelated influence of iron, light and cell size on marine phytoplankton growth. *Nature* 390(6658):389–392
- Szopa S, Cozic A, Shulz M, Balkanski Y, Hauglustaine D et al (2012) Aerosol and Ozone changes as forcing for Climate Evolution between 1850 and 2100. *Clim Dyn* (submitted)
- Takahashi T (2009) Climatological mean and decadal change in surface ocean pCO<sub>2</sub>, and net sea–air CO<sub>2</sub> flux over the global oceans. *Deep Sea Res Part II Top Stud Oceanogr* 56(8–10):554–577. doi:[10.1016/j.dsr2.2008.12.009](https://doi.org/10.1016/j.dsr2.2008.12.009)
- Takahashi T, Broecker W, Langer S (1985) Redfield ratio based on chemical-data from isopycnal surfaces. *J Geophys Res Oceans* 90:6907–6924
- Takahashi T, Sutherland SC, Feely RA, Wanninkhof R (2006) Decadal change of the surface water pCO<sub>2</sub> in the North Pacific: a synthesis of 35 years of observations. *J Geophys Res* 111(C7). doi:[10.1029/2005JC003074](https://doi.org/10.1029/2005JC003074)
- Takahashi T et al (2002) Global sea–air CO<sub>2</sub> flux based on climatological surface ocean pCO<sub>2</sub>, and seasonal biological and temperature effects. *Deep-Sea Res Part II* 49(9–10):1601–1622. doi:[10.1016/S0967-0645\(02\)00003-6](https://doi.org/10.1016/S0967-0645(02)00003-6)
- Talley LD, Reid JL, Robbins PE (2003) Data-based meridional overturning streamfunctions for the global ocean. *J Clim* 16(19):3213–3226. doi:[10.1175/1520-0442\(2003\)016](https://doi.org/10.1175/1520-0442(2003)016)
- Taylor KE, Stouffer RJ, Meehl GA (2009) A summary of the CMIP5 experiment design, vol 4. December 2009 edn, International CLIVAR Project Office
- Taylor KE, Stouffer RJ, Meehl GA (2011) An overview of CMIP5 and the experiment design. *Bull Am Meteor Soc.* doi:[10.1175/BAMS-D-11-00094.1](https://doi.org/10.1175/BAMS-D-11-00094.1)
- Tegen I, Fung I (1995) Contribution to the atmospheric mineral aerosol load from land-surface modification. *J Geophys Res Atmos* 100:18707–18726
- Valcke S (2006) OASIS3 user guide (prism\_2-5). PRISM Support Initiative 3:68
- Voltaire A, et al (2012) The CNRM-CM5.1 global climate model: description and basic evaluation. *Clim Dyn.* doi:[10.1007/s00382-011-1259-y](https://doi.org/10.1007/s00382-011-1259-y)
- Wanninkhof R (1992) A relationship between wind speed and gas exchange over the ocean. *J Geophys Res* 97(C5):7373–7382
- Yu L, Weller RA (2007) Objectively analyzed air–sea heat fluxes for the global ice-free oceans (1981–2005). *Bull Am Meteor Soc* 88(4):527–539. doi:[10.1175/BAMS-88-4-527](https://doi.org/10.1175/BAMS-88-4-527)