Simulation and assimilation of global ocean $pCO_2$ and air–sea CO$_2$ fluxes using ship observations of surface ocean $pCO_2$ in a simplified biogeochemical offline model

By VINU VALSALA* and SHAMIL MAKSYUTOV, National Institute for Environmental Studies (NIES), 16-2 Onogawa, Tsukuba 305-8506, Ibaraki, Japan

(Manuscript received 31 December 2009; in final form 6 July 2010)

ABSTRACT

We used an offline tracer transport model, driven by reanalysis ocean currents and coupled to a simple biogeochemical model, to synthesize the surface ocean $pCO_2$ and air–sea CO$_2$ flux of the global ocean from 1996 to 2004, using a variational assimilation method. This oceanic CO$_2$ flux analysis system was developed at the National Institute for Environmental Studies (NIES), Japan, as part of a project that provides prior fluxes for atmospheric inversions using CO$_2$ measurements made from an on-board instrument attached to the Greenhouse gas Observing SATellite (GOSAT). Nearly 250 000 $pCO_2$ observations from the database of Takahashi et al. (2007) have been assimilated into the model with a strong constraint provide by ship-track observations while maintaining a weak constraint of 20% on global averages of monthly mean $pCO_2$ in regions where observations are limited. The synthesized global air–sea CO$_2$ flux shows a net sink of 1.48 PgC yr$^{-1}$. The Southern Ocean air–sea CO$_2$ flux is a sink of 0.41 PgC yr$^{-1}$. The interannual variability of synthesized CO$_2$ flux from the El Niño region suggests a weaker source (by an amplitude of 0.4 PgC yr$^{-1}$) during the El Niño events in 1997/1998 and 2003/2004. The assimilated air–sea CO$_2$ flux shows remarkable correlations with the CO$_2$ fluxes obtained from atmospheric inversions on interannual time-scales.

1. Introduction

Ongoing anthropogenic emissions have caused an increase in the atmospheric concentrations of carbon dioxide (CO$_2$) from a pre-industrial value of 270 ppm to the present-day value of 380 ppm (Keeling et al., 1982). However, the present-day concentration of anthropogenic CO$_2$ in the atmosphere represents only half of the net emissions caused by human activities: the other half is absorbed by the terrestrial biosphere and the ocean, thereby moderating the net increase in atmospheric CO$_2$ concentrations. Uncertainties exist in any quantitative measure of the uptake of atmospheric CO$_2$ by the land biosphere and the ocean ecosystem, regardless of the calculation method. These uncertainties arise mainly because of the practical difficulties involved in direct estimates of fluxes between the atmosphere and the ocean or land surface. Moreover, there are few observational data with which to quantify the net exchange of CO$_2$. Here, we present a strategy of estimating the exchange of CO$_2$ between the ocean and the atmosphere. We adapt the usual strategy of biogeochemical cycle modelling in an offline ocean circulation model, with the important addition of constraining the $pCO_2$ of the modelled surface ocean using observations in a variational assimilation system.

Previous air–sea CO$_2$ flux estimates, as calculated from a large number of surface ocean $pCO_2$ observations, suggest that the ocean absorbs nearly 1.5 petagrams of carbon (PgC) each year (Takahashi et al., 2009). This annual mean uptake varies interannually depending on climate mode anomalies such as El Niño (during El Niño years, the uptake is generally 0.5 PgC more than the annual mean), the Pacific Decadal Oscillation, and the North Atlantic Oscillation (Feely et al., 1999; Le Quéré et al., 2000; Obata and Kitamura, 2003; Keeling et al., 2004; McKinley et al., 2004; Patra et al., 2005; Wetzel et al., 2005; Lenton and Matear, 2007; Le Quéré et al., 2007; Thomas et al., 2008).

The most common strategies employed in estimating air–sea CO$_2$ fluxes are as follows: (a) In the ‘top-down’ approach, sources and sinks are estimated from observations of atmospheric CO$_2$ concentrations and Greens functions (i.e. the transport of CO$_2$ concentrations) derived from transport models. This method is known as inverse modelling of sources and sinks (Gurney et al., 2004; Patra et al., 2005; Jacobson et al., 2007; Gruber et al., 2009). However, the efficiency of this method relies on the total number of observations from land-based tower...
stations and the regional distribution of the stations (Patra and Maksyutov, 2002). Because of the limited number of observation data, as well as the biases inherent in transport models, the inverse method involves large uncertainties. The direct assimilation of $CO_2$ observations into transport models also offers optimized estimates of $CO_2$ fluxes (Baker et al., 2006; Yumimoto and Uno, 2006). (b) The second method is a ‘bottom-up’ approach, in which process models of oceanic biogeochemical cycles are incorporated into a general circulation model, yielding estimates of $CO_2$ gas exchange. This approach involves uncertainty related to our limited knowledge of the key processes that govern the carbon cycle of the terrestrial biosphere and the ocean ecosystem. Previous studies have examined the assimilation of various types of data into biogeochemical models (Tjiputra et al., 2007). (c) It is also possible to measure the isotopic ratios of $CO_2$ in the atmosphere and thereby perform indirect measurements of $CO_2$ exchange among the terrestrial biosphere, ocean, and atmosphere (Keeling et al., 2001). (d) Mean global air–sea $CO_2$ fluxes can also be estimated from direct observations of oceanic partial pressure of $CO_2$ ($pCO_2$) and gas exchange coefficients (Takahashi et al., 2009). However, the sparse coverage of $pCO_2$ data requires large-scale interpolation to perform analyses in the global domain. Gruber et al. (2009) compared global estimates of air–sea $CO_2$ fluxes obtained using these various methods, revealing a general consistency, although with regional-scale inconsistencies.

Here, we present a bottom-up approach to estimating air–sea $CO_2$ fluxes, with an important modification compared with existing approaches: that is simulated surface ocean $pCO_2$ is constrained with available ship observations between 1996 and 2004. This approach provides model-predicted, interannually varying, error-minimized estimates of global air–sea $CO_2$ flux. This oceanic $CO_2$ flux analysis system was developed at the National Institute for Environmental Studies (NIES), Japan, as part of a project designed to provide prior fluxes for atmospheric inversions using $CO_2$ observations performed by an on-board instrument attached to the Greenhouse gas Observing SATellite (GOSAT).

Global oceanic $pCO_2$ observations have been carried out by several research programs. The World Ocean Circulation Experiments (WOCE), Joint Global Ocean Flux Study (JGOFS) and Voluntary Ship of Opportunity (VOP) observation programs, among others, are some of the major programs to collect surface ocean carbon data. Takahashi et al. (2009) combined all these $pCO_2$ observations into monthly climatological maps on a global scale. However, the aggregation of multi-year $pCO_2$ observations into monthly climatology results in the loss of valuable information regarding the interannual variability of $pCO_2$ and air–sea $CO_2$ fluxes. In contrast, by using multi-year individual observations of $pCO_2$ as a constraint on a process model, it would be possible to resolve valuable information regarding the interannual variability of air–sea $CO_2$ fluxes, at least on the regional scale, for which frequent observations of $pCO_2$ are available.

In this study, we used multi-year $pCO_2$ data to constrain the model $pCO_2$. The results demonstrate the potential benefits of using $pCO_2$ observations as constraints when minimizing model biases and errors.

To minimize the model error via the assimilation of observed $pCO_2$, we derived a simple biogeochemical model from McKinley et al. (2004). In this model, the net export production in the euphotic zone is estimated from the relation between phosphorus, light availability and primary production, and is then scaled by a regional mask to represent other processes that limit the primary production, such as Fe availability and grazing efficiency. The net export production is then converted to equivalent dissolved inorganic carbon (DIC) consumption within the euphotic zone, using a Redfield ratio. Below the euphotic zone, re-mineralization returns DIC that is exported to the deep ocean. We adopt this simplified ecosystem model for two reasons: (1) our aim is to utilize observations of surface ocean $pCO_2$ as a main constraint on modelled $pCO_2$ values, and (2) the construction of an adjoint is simpler if the ecosystem is limited to one component and the adjoint can be limited to one control variable (i.e. DIC).

The remainder of the paper is organized as follows. Section 2 describes the physical and biogeochemical models, and the method of constraining modelled $pCO_2$. The results are presented in Section 3 and a comparison of our assimilated model output with data from other sources are given in Section 4. The results are discussed in Section 5.

2. Physical, biogeochemical models and experiments

2.1. Physical model

We used a simple biogeochemical model coupled to an ocean tracer transport model driven by reanalysis ocean data. The transport model is the Offline ocean Tracer Transport Model (OTTM), as documented in Valsala et al. (2008). In this model, two-dimensional circulation, temperature, salinity and other diagnostically physical parameters are taken from ocean reanalysis data and used to evolve a tracer. The tracer of interest is DIC, which is a collection of carbonate ($CO_3^{2-}$), bicarbonate ($HCO_3^-$) and dissolved $CO_2$ gas. We used the reanalysis products provided by the Geophysical Fluid Dynamics Laboratory (GFDL), NOAA, Princeton, USA, as offline physical circulation. This study uses the data available between 1975 and 2004.

OTTM has self-operating diagnostic mode vertical mixing and subgrid-scale process parameterization schemes, because the coefficients for these processes are not readily available from the reanalysis data; instead, they were estimated within the offline model. Vertical mixing is represented as a combination of the K-Profile Parameterization (KPP; Large et al., 1994) and a background vertical diffusion suggested by Bryan and Lewis (1979). Horizontal diffusion is a combination of
flow-dependent diffusion, in which the coefficients are proportional to the stress and strain experienced in the local fluid volume (Griffies and Hallberg, 2000), and advection fluxes due to eddy-induced transport (Gent and McWilliams, 1990).

The data for horizontal advection were taken from the offline data archive, whereas vertical velocities were calculated based on the principles of mass conservation. The free-surface kinematic boundary condition was determined from the sea surface height. OTTM transport was extensively tested in a previous study by simulating chlorofluorocarbon (see Valsala et al., 2008); this earlier study also contains details of the model design and validation. The exaggerated vertical mixing found in OTTM within the tropics, as reported by Valsala et al. (2008), was resolved by further tuning of the model in this study.

2.2. Chemical model

The chemical compartment of the biogeochemical model consists of a single tracer (DIC). pCO$_2$ was treated as being in near-equilibrium with the atmosphere at the sea surface via air–sea gas exchange. Within the ocean, DIC consists of CO$_3^{-2}$, HCO$_3^{-1}$, and dissolved gaseous CO$_2$, although they are transported as a single tracer (DIC). This approach corresponds to the solubility pump model described in Ocean Carbon Cycle Intercomparison Project-II (OCMIP-II; Orr et al., 1999, available at http://www.ipsl.jussieu.fr/OCMIP/).

Air–sea gas exchange depends on piston velocity $K_w$ and the difference in pCO$_2$ between the surface ocean and the ambient atmosphere above. Because the focus of this study is the contemporary air–sea CO$_2$ flux, a constant atmospheric pCO$_2$ of 368 µatm (as observed in 2000) was used. Alkalinity [taken as total alkalinity (TA)] was not transported but was estimated from the reanalysis salinity inputs with a constant conversion factor of 2310 µeq kg$^{-1}$, as $TA = 2310S/S_0$, in which $S$ is the local surface salinity and $S_0$ is the annual mean surface salinity, integrated globally. This relation is taken from the OCMIP-II protocol (Orr et al., 1999).

The air–sea flux in the chemical model was formulated as $\Phi_{GASEX} = K_w(p\text{CO}_2\text{OCEAN} - p\text{CO}_2\text{AIR})$. Model pCO$_2$ was calculated using the routines provided in OCMIP-II (Orr et al., 1999). The term $K_w$ is the piston velocity at which gaseous CO$_2$ enters or leaves the surface according to wind speed and CO$_2$ solubility, as formulated in Wanninkhof (1992). We used a gas exchange proportionality constant of $a = 0.337$ taken from OCMIP-II (Orr et al., 1999) when calculating $K_w$. Surface wind speed for the calculation of $K_w$ was the sum of 10-day averages of 6-hourly squared wind speed ($u_{10}^2$) and 6-hourly wind speed variance ($\sigma_{u_{10}}^2$). Here, $u_{10}$ indicates wind speed at 10 m height. An index for polar ice-caps was used to partially mask the air–sea gas exchange during sea-ice conditions. This was achieved by compiling a spatial map of the ice index, with values vary between 0.2 and 1, as taken from the OCMIP-II data set (Orr et al., 1999).

2.3. Biological model

The ecosystem model employed here is derived from McKinley et al. (2004). In the present model, net export production in the euphotic zone was calculated from climatological maps of phosphate ($P$) and light availability ($I$). An important difference between the present biological model and that of McKinley et al. (2004) is that we used monthly climatological phosphate to determine export production, whereas McKinley et al. transported phosphate and oxygen as two additional prognostic tracers. Given that McKinley et al. (2004) restored the biogeochemical tracers below 1200 m depth to observed climatology, our use of climatological phosphate means that the two studies show similar features below this depth.

A scaling factor was given as a regional mask ($\alpha$), which implicitly represented the Fe limitation, grazing efficiency and recycling. Export production in the euphotic zone (0–140 m) is formulated as

$$B(z) = -\alpha(x, y) \left( \frac{I_{(x, y, z, 12\text{–month})}}{I_{(x, y, z, 12\text{–month})} + I_0} \right) \times \left( \frac{P_{(x, y, z, 12\text{–month})}}{P_{(x, y, z, 12\text{–month})} + P_0} \right).$$

where $\alpha$ represents other controlling factors of export rates.

The value of the maximum export rate, $\alpha$, encapsulates all the processes leading to the export of DIC which were not explicitly represented by phosphate and light limitation (McKinley et al., 2004). In this case, the value of $\alpha$ should be consistent with the models circulation and climatological nutrient fields. The global ocean was divided into 14 regions and the value of $\alpha$ in each region was defined based on the assumption that a given model flow field will produce an annual mean phosphorus distribution consistent with climatological observations. The oceanic regions for the $\alpha$ mask were taken from McKinley et al. (2004), whereas the individual regional values of $\alpha$ were tuned for the circulation used in this study. Because we do not transport phosphate in the present model, we achieved an optimal $\alpha$ mask by tuning the individual regional values of $\alpha$ to obtain model pCO$_2$ values similar to the climatology of Takahashi et al. (2009).

The half-saturation values for phosphate ($P_0$) and light ($I_0$) were set to 0.01 µmol kg$^{-1}$ and 30 W m$^{-2}$, respectively, which are the same values as those used in McKinley et al. (2004). Fluxes of sinking particles [$F(z)$] were parameterized as in Dutkiewicz et al. (2001). The net export production was converted to an equivalent DIC consumption in the euphotic zone, based on the Redfield ratio, $R_{CP} = 117:1$ (McKinley et al., 2004). Within the euphotic zone, $Sb(z) = B(z) + F(z)$; below the euphotic zone, $Sb = F(z)$.

The use of climatological $P_{(x, y, z, 12\text{–month})}$ may have a detrimental effect on the predicted pCO$_2$ because of the absence of interannual variability of export production in our model. One may interpret that those areas in the ocean where our simplified parameterization would influence the predicted pCO$_2$ are those
areas with $P$ values similar to or less than the half-saturation coefficient $P_0$. In these regions, export production depends on $P/P_0$. Climatological phosphate is already in balance between the supply of $P$ from the bottom layer and deposition of $P$ by removal from organic matter. We compared the air–sea CO$_2$ fluxes from our model (non-assimilated run) with those reported by McKinley et al. (2004), revealing that they are well correlated over a seasonal cycle as well as regarding interannual variability [see also Fig. 3.12 in Maksyutov et al. (2010) for a comparison of the seasonal cycle]. Except for equatorial regions and to some extent the North Atlantic, we obtained a statistically significant correlation between the air–sea CO$_2$ fluxes of the present study and of McKinley et al. (2004).

In addition to air–sea gas exchange, the surface dilution of DIC by rainfall ($P$) and its concentration by evaporation ($E$) were included in the model ($\phi_{FWEX}$). Although the models surface boundary condition is kinematic, we have given an additional estimate of surface dilution to compensate for salinity errors (and corresponding alkalinity errors in the model) in the reanalysis ocean data (Whit Anderson, 2008, personal communication). In this case, to eliminate the net addition of DIC due to virtual fluxes, we subtracted the global mean of $P-E$ before computing the virtual fluxes. This protocol is taken from OCMIP-II. The data for virtual dilution were derived from the same reanalysis data. Therefore, the total (i.e. physical, chemical and biological) components of the model have the following form:

$$\frac{D[DIC]}{dt} = R_{CP} \times Sb(z) + \phi_{GASEX} + \phi_{FWEX},$$

where $\frac{D[DIC]}{dt}$ represents the total changes due to advection, mixing, diffusion and eddy-induced transport.

### 2.4. Data and model setup

We used the ocean reanalysis data prepared by GFDL as the physical driver of OTTM. These data were produced by Modular Ocean Model-4 (MOM4) forced with Coordinated Ocean Research Experiments (CORE) data sets (Large and Yeager, 2008) and with an assimilation of in situ temperature profiles from the National Oceanographic Data Center (NODC) archives using a 3D variational scheme (see http://www.gfdl.noaa.gov/ocean-data-assimilation). Details of the physical part of the ocean model can be found in Delworth et al. (2006) and Gnanadesikan et al. (2006).

The physical parameters borrowed from the reanalysis data are velocity ($u, v$), temperature, salinity, mixed layer depth, evaporation and precipitation rates, surface heat flux, surface wind stress (to derive vertical mixing) and sea surface height. The reanalysis data have a zonal resolution of 1° with a total of 360 grid points. Meridional resolution is 1° at higher latitudes and 0.8° in the tropics, with a total of 200 grid points. The data have 50 vertical levels with a 10 m increment in the upper 225 m and stretched vertical intervals below this layer. The study domain extends from 81.5°S to 85.5°N. The reanalysis data are available for each month between 1960 and 2004. Physical variables at monthly time scales were interpolated into the models time-step of 2 h.

The physical and biogeochemical model was run for the mean state of DIC (pre-run) using a 12-month mean circulation and other physical parameters derived for the 10 yr between 1975 and 1984. We chose this period for deriving the mean circulation and other physical parameters for the pre-run because our interannual DIC simulation started from 1980. Thus, 12-month mean circulations derived between 1975 and 1984 are expected to yield a reasonable mean state of DIC.

The pre-run of the model was initialized with annual mean values of DIC derived from the GLODAP data set (Key et al., 2004). For the pre-run, surface wind speed and its variance for the air–sea gas exchange calculation were kept as climatological values. We took these climatological data from the OCMIP-II forcing database (available at http://www.ipsl.jussieu.fr/OCMIP/).

The data were derived from a 5-year average of SSMI monthly squared wind speed and 30-day variance of the instantaneous wind speed from the corresponding monthly period (see also README.sdat provided by OCMIP-II, available at http://www.ipsl.jussieu.fr/OCMIP/).

Surface wind-speed data for the air–sea gas exchange calculations in the interannual run (Section 2.2) were taken from ERA-40 (Uppala et al., 2004). We note that the physical circulation used in this study is produced by forcing with optimal surface fluxes derived from CORE, which takes most of its atmospheric data from NCEP/NCAR. However, in CORE data, the NCEP winds are modified with QSCAT satellite winds to eliminate biases (Large and Yeager, 2008). Similarly, other components of the surface fluxes in the CORE data are corrected with available satellite measurements. Therefore, CORE is an optimal surface flux combined from various sources, and the offline circulation used in our study does not restrict us from using ERA-40 winds for the air–sea CO$_2$ flux calculation.

The pre-run was carried out for 20 years using monthly mean circulation. The monthly mean wind for the surface gas exchange calculations was repeated at each year of the pre-run. The surface DIC concentration and air–sea CO$_2$ flux of the model reached a quasi-equilibrium state at the end of the first 10 years of the pre-run and thereafter showed a minimum departure from the mean state. The global integral of CO$_2$ flux from the model was 1.3 PgC yr$^{-1}$ in the first year of the spin-up and 2.25 PgC yr$^{-1}$ at the end of 9 year. Therefore, the estimated drift in the first 9 years of spin-up is approximately 0.1 PgC yr$^{-1}$. After 16 years, the model CO$_2$ flux reached 2.35 PgC yr$^{-1}$. Therefore, the drift between 10 and 16 years is 0.014 PgC yr$^{-1}$, which represents 14% of the drift during the first 10 years of spin-up. From 16 years onward, the model CO$_2$ flux remained close to 2.35 PgC yr$^{-1}$. Although the surface CO$_2$ flux reached a quasi-equilibrium state after a 20-year spin-up, deeper adjustments (below 2000 m) continued because of the slow communication between surface and
The assimilation involves the following steps.

The average of the last 5 years of the pre-run was taken as the restart condition for the interannual simulation. The simulation was started with the restart condition and continued through the monthly fields for the year 1980. The simulation was then continued from 1980 to 1995 using real-time data. The model results deviated from the restart condition as soon as the data for 1980 were introduced. A near-steady-state was reached by the time the model had passed through the fields for the year 1980. The restart condition for January 1996 was used for the experimental runs.

We analysed the pre-run and interannual-run $pCO_2$ and air–sea $CO_2$ fluxes, and calculated the correlations with the observations reported by Takahashi et al. (2007) and Takahashi et al. (2009). Model performance was reasonable on seasonal to interannual time scales. The pre-run and interannual-run results are basically similar to those of McKinley et al. (2004).

### 2.5. Constraining the model $pCO_2$ using observations

The use of a simple biogeochemical model, in which biological and chemical cycles of $CO_2$ are related to a single tracer (DIC), simplified the construction of adjoint equations for the assimilation. These adjoints are used to constrain model $pCO_2$ to the corresponding observations. A variational assimilation method was used to minimize the cost function. We employed a descendant algorithm technique similar to that described by Ikeda and Sasai (2000). The assimilation involves the following steps.

(a) The forward model is run through an assimilation window of 2 months, starting from the initial condition. A 2-month window was chosen because the surface $pCO_2$ shows rapid equilibration with the atmosphere. Therefore, the adjoint has relatively negligible magnitude to correct the initial condition in a longer backward integration. The surface ocean $pCO_2$, DIC and other variables are saved along the ship observation tracks. The corresponding global-scale variables are saved at monthly intervals.

(b) The observation-model misfit values are run backward in a time–space domain in the adjoint simulation, which starts from the end point of day-60 and ends at day-1. We used a simple adjoint of physical circulation by reversing the direction of the offline currents and integrated the model backward in time while the direction of mixing and diffusion remained unchanged. This adjoint method is similar to that discussed previously by Fukumori et al. (2004), Hourdin and Talagrand (2002) and Hourdin et al. (2002). The adjoint model contains only one control variable: DIC. The surface ocean $pCO_2$ observations are converted into equivalent $CO_2$ concentrations using model sea surface temperature (SST), sea surface salinity (SSS) and alkalinity, all of which are reanalysis offline input data. The cost function was chosen to be the squared difference between the data and modelled $pCO_2$, as follows:

$$J = w_1 \Sigma (pCO_2 - pCO_2_{ST})^2 + w_2 \Sigma (pCO_2 - pCO_2_{CL})^2,$$

where $pCO_2_{ST}$ denotes the ship-track $pCO_2$ data and $pCO_2_{CL}$ denotes the climatology of Takahashi et al. (2009). The weighting $w_1$ was kept equal to 1.

In addition to the along-track $pCO_2$ data, we used the monthly mean surface ocean $pCO_2$ derived from Takahashi et al. (2009) as an additional constraint. However, the weighting for this mean constraint ($w_2$) was retained only up to 20% of the weighting of the along-ship-track data assimilation. Moreover, the weighting for this mean constraint was spatially varied according to the inverse variance ($\sigma^2$) of the model interannual $CO_2$ flux variability, which we inferred from our forward model simulation for 1980–1995. This spatially varying weighting helped to constrain the model to the mean $pCO_2$ only in regions where the interannual variance was minimized. Thus, regions with higher interannual variability in $pCO_2$ (e.g., the eastern equatorial Pacific) were constrained only when ship-track data were available. Therefore, $w_2$ in the above equation is assumed the form min(0.2, $\sigma_{min}/\sigma$) in which $\sigma_{min} = 0.02$ (mole m$^{-2}$ yr$^{-1}$)$^2$. A map of $w_2$ is shown in Fig. 5 (bottom panel).

The adjoint equations are deduced from

$$- \frac{\lambda_n - \lambda_{n-1}}{\Delta t} + w_1 \frac{\delta_{st}}{\Delta t} f_s (pCO_2 - pCO_2_{ST}) + w_2 \frac{\delta_{cl}}{\Delta t} f_s (pCO_2 - pCO_2_{CL}) + \frac{\partial F_n}{\partial C} = 0,$$

where the adjoint operator $\lambda$ at nth time-step is given as

$$\lambda_n = W_f \left( \frac{C_{n+1} - C_n}{\Delta t} - F_n \right),$$

where $C_n$ is the control variable at the nth time step, $W_f$ is a weighting coefficient, and $F_n$ is forcing, including $CO_2$ gas exchange, virtual fluxes and biological export of DIC. $f_s$ is a conversion factor for converting $pCO_2$ to DIC using SST, SSS and alkalinity (Weiss, 1974). The last term in eq. (4) represents the derivative of forcing with respect to the control variable, as follows:

$$\frac{\partial F_n}{\partial C} = \lambda_n K_a K \left[ \frac{(2C - A)(3A - 2C)}{(A - C)^2} \right] + \lambda_n B(z) + \lambda_a (E - P),$$

where $A$ represents alkalinity, $K$ is a $CO_2$ solubility coefficient (Weiss, 1974) and $B(z)$ is derived from eq. (1). The first term on the right side of eq. (6) represents the derivative of the $pCO_2$ formulation with respect to DIC.

(c) The cost function is minimized using a descendant algorithm in which at each iteration, the initial condition is
corrected by the gradient of the cost function (Bennett, 2002). The weighted correction to the forcing in the assimilation run incorporates corrections to DIC induced as a combination of corrections to the biology and other surface fluxes such as fresh water and salt errors. However, these corrections are not separable because we use only one control variable (eqs 3–6). The iteration is repeated until the along-track model-to-observation misfit of $p_{CO_2}$ is minimized below a threshold, defined as the iteration at which the cost function falls below 10% of the initial value.

The model was restarted from January 1996 and $p_{CO_2}$ was synthesised for a monthly period with an iteration window of 2 months. The whole procedure was repeated every 2 months within the assimilation period.

3. Simulation and synthesis of $p_{CO_2}$

3.1. $p_{CO_2}$ data

The number of surface $p_{CO_2}$ ship observations has expanded considerably since 1995. In Takahashi et al. (2007), compilation of global ocean $p_{CO_2}$ database, approximately $3 \times 10^5$ quality-controlled observations were used. This study made use of Takahashi et al. (2007) data (Version 1.0) as ship-track $p_{CO_2}$ data (i.e. $p_{CO_2,ST}$ in eq. 3). In addition to the research vessels participating in $p_{CO_2}$ measurements, Ship-Of-Opportunity measurements made by cargo liners have provided a wealth of information about the surface ocean $p_{CO_2}$, especially in regions such as the North Pacific (Zeng et al., 2002) and North Atlantic. Sampling density in the Southern Ocean has been improved in Version 1.0 $p_{CO_2}$ data (Takahashi et al., 2007) compared to the older versions. Figure 1 summarizes the $p_{CO_2}$ observation points used in this study, comprising 236 777 observations between 1996 and 2004. Although not all data in Takahashi et al. (2007) database are sampled along ship tracks (some are from station-data sampled on research cruises), we use the term ‘ship-track’ to represent individual $p_{CO_2}$ observations taken from Takahashi et al. (2007) database.

The $p_{CO_2,ST}$ in eq. (3) is taken from Takahashi et al. (2009). Here we note that, there are following inconsistencies exist (inherited from the Takahashi et al., 2007) in the $p_{CO_2}$ data that used in this study. (1) Different corrections were applied to 13 981 observations from the southern Indian Ocean lines collected between 1998 and 2000 (file name “OISO” in the LDEO database), which were subjected to a repeated (twice) SST correction in the Version 1.0 of Takahashi et al. (2007). The average error caused by this discrepancy is $-9.27 \pm 3.43 \mu atm$. (2) In all OISO data files, only CO$_2$ fugacity ($f_{CO_2}$) values are reported without $p_{CO_2}$ values or CO$_2$ concentrations in dry equilibrated gas. The approximate formula for the conversion of $f_{CO_2}$ to $p_{CO_2}$ yields $p_{CO_2}$ values greater than $f_{CO_2}$ by about 1.3 $\mu atm$. These two uncertainties were common for both ship-track data and the climatology mean used here. (3) There is a slight change in $p_{CO_2}$ interpolation near the meridian, which generates an error of less than 10 $\mu atm$ in several grid points around 0°E in Takahashi et al. (2009). For this study, we derived the climatological mean constraint from Takahashi et al. (2009) and obtained ship-track data from Takahashi et al. (2007). The magnitude of data errors is minor compared with the residual error in the assimilation.

To enable post-assimilation comparisons, we combined Takahashi et al. (2007) $p_{CO_2}$ data into $1^\circ \times 1^\circ$ bins. This approach leaves gaps in regions for which no observations are available, especially in the southern sub-tropical regions of all the major oceans and in the Southern Ocean. The North Atlantic and North Pacific have the largest number of observations. We

![Fig. 1. Total number of ship observations and tracks used in the assimilation of surface ocean $p_{CO_2}$ in the model. A total of 236 777 observations are used from Takahashi et al. (2007) database.](image-url)
Fig. 2. (Top panel) 1980–1995 mean model simulated surface ocean \( \text{pCO}_2 \) in the free run and (bottom) observations from Takahashi et al. (2007) regridded into \( 1^\circ \times 1^\circ \) bins. The free run refers to the non-assimilated interannual simulation from 1980 to 1995. Units are in \( \mu \text{atm} \).

3.2. Model \( \text{pCO}_2 \) error

The free-run of the model produced patterns of global mean surface ocean \( \text{pCO}_2 \) that are consistent with observations (Fig. 2). Here, the free-run refers to interannual model simulations without any constraint on surface ocean \( \text{pCO}_2 \). Figure 2 shows the annual mean \( \text{pCO}_2 \) of the model for the period 1980–1995. The free-run performs reasonably well in representing the elevated \( \text{pCO}_2 \) in the tropical Pacific and depressed \( \text{pCO}_2 \) in the Southern Ocean and North Atlantic. However, the modelled \( \text{pCO}_2 \) contains regional-scale errors in magnitude and in phase. For example, \( \text{pCO}_2 \) in the tropical Atlantic is overestimated in the free-run compared with the observations of Takahashi et al. (2007) (Fig. 2, bottom panel). The North Pacific has a limited springtime bloom related reduced \( \text{pCO}_2 \) in the free-run compared with observations. In the Indian Ocean, the Arabian Sea is poorly represented and biased toward a reduced \( \text{pCO}_2 \).

Figure 3 shows the misfit between the annual mean \( \text{pCO}_2 \) of the free-run and the observations of Takahashi et al. (2007). Here, errors are calculated as the difference between the annual mean \( \text{pCO}_2 \) of the free-run and the corresponding mean from the observations of Takahashi et al. (2007). Large errors are found in the tropical Atlantic, North Atlantic, Northwest Pacific and Indian Ocean (Fig. 3, top panel). In the southern subtropics, however, model bias cannot be assessed because of a lack of observed data.

The annual mean error does not capture the seasonal drift in model \( \text{pCO}_2 \) from observations. To show the total error arising from seasonal biases in the model, we calculated the sum of the absolute misfit of free-run \( \text{pCO}_2 \) during each month from January to December (Fig. 3, bottom panel), compared with the observations of Takahashi et al. (2007).
The cumulated seasonal error in the model $p_{CO_2}$ is large in the subarctic North Pacific and North Atlantic, which is not apparent in the annual mean error. The seasonal bias is also strong in the tropical Atlantic and eastern tropical Pacific, contributing to the annual mean error. The largest error in the seasonal cycle is observed in the North Pacific.

The lack of a one-to-one correspondence between regional-scale observed and simulated surface ocean $p_{CO_2}$ may reflect the limited and simplified representation of the ecosystem model, which is inefficient in resolving processes such as spring bloom and variability in the euphotic zone. This limitation is reflected in errors in the northeastern Pacific and Arabian Sea, and may explain the errors and biases in the model. Given that our ecosystem model has only one component, the seasonal cycle of biology is not well represented. Moreover, the $\alpha$ values used to obtain export production are too coarse to represent the complete features of the regional ecosystem. For example, Fig. 4 compares the $p_{CO_2}$ seasonal cycle of our free-run and that of Takahashi et al. (2009) (black and green lines, respectively), averaged over each of the 14 $\alpha$ regions used in the export production calculation (see also Section 2.3). The subdivisions are the same as those used in table 1 of McKinley et al. (2004). In the present free-run, the tropical to mid-latitude oceans have a reasonable seasonal cycle of $p_{CO_2}$, although biases are apparent. For example, the seasonal biases in the equatorial Atlantic and Indian Ocean are about 25 $\mu$atm. Good agreement between the free-run and the observed $p_{CO_2}$ seasonal cycle is found for the subtropical South Pacific, subtropical South Atlantic, eastern equatorial Pacific, western equatorial Pacific, subtropical North Pacific and subtropical North Atlantic. However, the model performance is poor in high-latitude oceans (e.g. subpolar North Pacific and subpolar North Atlantic). The free-run seasonal cycle of $p_{CO_2}$ of the Southern Ocean is in phase with observations, although with relatively weak amplitudes.

In addition to the limitations of the ecosystem model, which explain some of the discrepancies between the model and
SYNTHESIS OF GLOBAL OCEAN PCO₂

Fig. 4. Seasonal cycle of pCO₂ from 14 α-regions of the biological model are shown. Three lines are shown and they are from the model (black line), Takahashi et al. (2009) (green-line) and the assimilation (red-line). Units are in μatm.

3.3. Constraining the surface pCO₂

The upper panel in Fig. 5 shows the interannual variance of air–sea CO₂ flux as resolved in the 16-year period of the free-run. The inverse of this map is normalized to vary between zero and 0.2, and is used for the weighting coefficients ($w_2$) as a constraint on monthly mean pCO₂ (Fig. 5, bottom panel; see also Section 2.5). This approach guarantees the following conditions: (1) the model biases in the mean seasonal cycle are reduced because the pCO₂ is constrained to a mean cycle, and (2) interannual variability in modelled pCO₂ is relatively weakly influenced by...

observations, other sources of errors are likely to be associated with biases in surface salt and fresh water fluxes. A more important source of error, especially at high latitudes, arises from alkalinity. The simple linear relation between alkalinity and salinity does not incorporate regional differences in the alkalinity–salinity relation, as noted by Lee et al. (2006). A comparison of the present alkalinity maps with the map provided by Lee et al. (2006), for the 14 regions described earlier, reveals discrepancies in the seasonal cycle of alkalinity at high latitudes. Note that the present model used the alkalinity–salinity relation from OCMIP-II.
Fig. 5. (Top panel) 1980–1995 interannual variance of air–sea CO2 flux. Variance is calculated after subtracting the monthly climatology. Units are in (mole m$^{-2}$ yr$^{-1}$)$^2$.

(Bottom panel) The weight $w_2$ used for the climatological constraint (non-dimensional).

We verified the influence of the climatological constraint on the fit of modelled data to observations along ship tracks. The assimilation was run by switching off the climatological data constraint. The results are analysed for regions with the greatest coverage of data. In the case of no climatological constraint, the cost function along the ship track is reduced by 10% compared with the case of a climatological constraint. The remaining error in the assimilated $p$CO$_2$ values due to the climatological constraint is less than 10 μatm.

The assimilation is run from 1996 to 2004, yielding monthly $p$CO$_2$ and air–sea CO$_2$ flux. The only assimilation in the system is performed for $p$CO$_2$. Air–sea fluxes are calculated from the assimilated $p$CO$_2$ and gas exchange coefficients. We refer to this calculated air–sea gas exchange as ‘assimilated air–sea CO$_2$ flux’.

Figure 6 summarizes the assimilated $p$CO$_2$ annual mean calculated between 1996 and 2004, as well as the observation data reported by Takahashi et al. (2007). The assimilation resulted in reduced amplitude errors in annual mean $p$CO$_2$. The regional variability of $p$CO$_2$ is better represented in the assimilation compared with the free-run (cf. Fig. 1). The elevated amplitudes of $p$CO$_2$ in the eastern tropical Pacific are regulated in the assimilation compared with the free-run. The errors in the northern Pacific and Atlantic Ocean are improved in the assimilation, as is the regional $p$CO$_2$ distribution in the Southern Ocean.

Figure 7 summarizes the total error remaining in the assimilation compared with the observations reported by Takahashi et al. (2007). Compared with the free-run, an improvement is apparent in the seasonal $p$CO$_2$ of the subarctic North Pacific (cf. Fig. 3). This part of the ocean has a dense coverage of $p$CO$_2$ sampling contributed by Ship-Of-Opportunity measurements in which the instruments are attached to commercial vessels operating between Japan and North America, and among other Asia-Pacific countries. This wealth of information is readily assimilated into the model, resulting in a marked reduction in errors. Thus, the assimilation performed well in constraining the
model, especially in regions where large amounts of observation data are available.

A comparison of Figs 3 and 7 shows that error reduction in the assimilation is concentrated in the tropical Atlantic, northern Pacific, northern Atlantic and equatorial Pacific. In the eastern equatorial Pacific, however, observations are limited and do not cover a long enough period to resolve interannual variability related to ENSO. Moreover, in this region our model employed only a minimum constraint on the mean \( pCO_2 \), depending on the inverse variance map used for weighting. Thus, the amplitude mismatch remains in this region, compared with the data in Takahashi et al. (2007), reflecting interannual signals rather than the residual error of the model.

Figure 4 also compares assimilated \( pCO_2 \) as average values derived from the 14 regions used in the biological export calculation (Section 3.2). The assimilation contributed to correcting the phase of the seasonal cycle and resulted in reduced bias; remarkable improvements are seen at high latitudes. For example, the assimilation corrected the seasonal cycle of the sub-polar North Pacific, because of the large amount of data incorporated from this region. However, the seasonal cycle of the sub-polar North Atlantic was only corrected to a small degree. The LDEO ship-track database (Version 1.0; Takahashi et al., 2007) has not incorporated all the available \( pCO_2 \) observations from the North Atlantic.

Figure 8 summarizes the percentage reduction of error in annual mean and seasonal \( pCO_2 \) in the assimilation, with respect to the total error in the free-run. The error reductions are calculated as the difference between the free-run error (i.e. Fig. 3) and residual error (i.e. Fig. 7), expressed as a ratio relative to the free-run error. A positive value indicates that the assimilation has improved the model \( pCO_2 \) values, with respect to the observed data provided by Takahashi et al. (2007). Sixty percent of the annual mean model biases are eliminated in the assimilation. Forty to sixty percent of the cumulative seasonal errors are also minimized at a regional scale. Marked reductions in error are seen in the tropical Atlantic and North Pacific. In the Indian Ocean, the errors are reduced in both the seasonal and annual
mean $p$CO$_2$. However, the density of observation tracks within the Indian Ocean is less than that in other oceans.

There are few locations for which the assimilation results in an increase in model $p$CO$_2$ error (e.g., regions of the eastern equatorial Pacific). However, large interannual variations in $p$CO$_2$ remain in such regions (Fig. 7).

### 3.4. Conservation of DIC in the assimilation

Here, we consider the following question: To what degree does the adjoint disturb the total balance of globally integrated DIC in the model? The cost function is designed based on the surface ocean $p$CO$_2$ mismatch, whereas DIC is adjusted to minimize the cost function. In this case, it is instructive to assess whether the adjoint introduced any DIC imbalance in the model. The only permissible change in DIC within the model is at the surface, via air–sea gas exchange. Although we disturbed the total DIC in the ocean via adjoints, we optimized the near-surface $p$CO$_2$ according to observations, which probably resulted in air–sea fluxes of the desired directions and magnitude. This approach serves the purpose of our assimilation system, which is designed to optimally estimate the net air–sea CO$_2$ fluxes. The imbalance of DIC produced in the ocean by the adjoints can be considered as corrections to biases in model circulation. In fact, the initial condition taken from GLODAP contains biases, and the model seeks to attain its own equilibrium via the given biogeochemical cycle and transport. In this case, the adjoints compensate for the mismatch in surface ocean DIC (at least at distances for which the adjoints can reach in a 2-month iteration) that originates from the mismatch between modelled and observed $p$CO$_2$.

To quantify the model DIC budget and its variations, we calculated the global average DIC concentration (the weighted average in the $x$, $y$ and $z$ directions) from our assimilation and compared it with area-integrated global air–sea CO$_2$ fluxes (Fig. 9). The global DIC concentration varies in proportion to the air–sea CO$_2$ flux. When the air–sea flux is toward the ocean, the DIC concentration increases, and vice versa. This relation is equally apparent in the seasonal cycle (top panel in Fig. 9) and in

--

**Fig. 7.** Same as Fig. 3, but for assimilated run.
3.5. Assimilated air–sea CO$_2$ flux

The reduction in biases and errors in the synthesized pCO$_2$ would enhance the quality of air–sea CO$_2$ fluxes retrieved from the model. The assimilated CO$_2$ flux shows an annual mean sink of 1.48 Pg C yr$^{-1}$ (1996–2004). The corresponding value from Takahashi et al. (2009) is an annual mean sink of 1.41 Pg C yr$^{-1}$. Thus, the assimilated estimate is closer to the estimate based on observations of Takahashi et al. (2009).

Except for regions of marked interannual variability, the seasonal cycle of the assimilated air–sea CO$_2$ flux shows a strong correlation with observed data (Fig. 10). The majority of areas in tropical to high-latitude oceans shows a seasonal correlation above 0.8 (above the 99% level of significance). Biases remain in the assimilation for the equatorial Pacific and Arabian Sea, compared with observations. The Indian Ocean has the poorest coverage of pCO$_2$ measurements; consequently, the assimilation did not yield any considerable improvements in this region. There are limited observations in the equatorial Pacific with which to calculate the interannual signal; thus, the poor correlation between the seasonal cycle in this region and the observations of Takahashi et al. (2009) does not reflect the models inability; in fact, it may have originated from interannual variability.

The global integral of seasonal CO$_2$ flux in the assimilation has a minimum sink during June, compared with August and September in the observations presented by Takahashi et al. (2009). The Southern Ocean shows a major phase mismatch in the assimilation, especially during the austral summer. The model performs poorly in representing the sink observed during the austral summer, which may reflect the limitations of the
ecosystem model regarding the biological cycle. However, the annual mean of the assimilation shows a net sink, as found in the observation data. The mean sink in the Southern Ocean (south of 40°S) in the assimilation is $-0.41 \text{ PgC yr}^{-1}$; the equivalent figure in Takahashi et al. (2009) is $-0.45 \text{ PgC yr}^{-1}$.

Figure 11 (top panel) shows the annual mean CO$_2$ flux from the assimilation for 1996–2004. The regional patterns of this flux are consistent with the results shown in Takahashi et al. (2009), as a result of two factors: (1) the $p_{\text{CO}_2}$ track observations of Takahashi et al. (2007) are fully utilized in the model to synthesis the surface ocean $p_{\text{CO}_2}$, and (2) the monthly mean map provided by Takahashi et al. (2009) $p_{\text{CO}_2}$ is used as a weak constraint for regions with limited $p_{\text{CO}_2}$ observations. However, the mean constraint was given a weighting of only 20% of the track assimilation, and it is further masked by the inverse of the interannual variability of modelled air–sea CO$_2$ flux.

3.6. Interannual variability of assimilated CO$_2$ flux

Area-integrated CO$_2$ fluxes from the global ocean and the Niño region show that the assimilation has weakened the strength of the interannual variability compared with the atmospheric inversions of Patra et al. (2005) and other forward model simulations (Le Quere et al., 2000; Obata and Kitamura, 2003; McKinley et al., 2004). This result may reflect the weak constraint of the model on the monthly mean maps of $p_{\text{CO}_2}$, which could suppress the interannual signals. In addition, the amplitudes of interannual signals in the assimilation are smaller than those in the present free-runs. The observation data reported by Feely et al. (2002) and the inverse modelling of Patra et al. (2005) and several OGCMs (Le Quere et al., 2000; Obata and Kitamura, 2003; McKinley et al., 2004) show stronger interannual variability than that obtained in this study. However, the magnitude of the interannual signal in the OGCM results is only half that in the atmospheric inversions.

The Niño region shows an interannual anomaly with a magnitude of 0.4 PgC yr$^{-1}$ during the 1997–1998 and 2003–2004 El Niño events. This is smaller in magnitude than the inversions reported by Patra et al. (2005). The years 2002–2004 were marked by weak El Niño conditions in the tropical Pacific, and the CO$_2$ flux anomalies for these years are negative. Thus, the assimilated CO$_2$ flux anomalies from the eastern tropical Pacific are consistent with observations, as well as with our general consensus regarding the evolution of CO$_2$ flux during El Niño years.

4. Comparison with other data sources

Here, we compare our assimilated air–sea CO$_2$ flux with other data sources. The following sources are used for the intercomparison: the atmospheric inversions reported by Rödenbeck et al. (2003) (Version s96_v3.1) and Patra et al. (2005), and the biogeochemical general circulation model fluxes of Le Quere et al. (2005). The seasonal cycle of global air–sea CO$_2$ fluxes is compared among these four products, as well as the seasonal cycle...
Fig. 10. (Top panel) Seasonal correlation coefficients calculated between assimilated sea-air CO2 fluxes and Takahashi et al. (2009). The line plots compare the seasonal cycle of area integrated sea-air CO2 fluxes from assimilation and Takahashi et al. (2009).

The seasonal cycle of globally integrated air-sea CO2 fluxes differs considerably among the above four products (Fig. 12). The assimilated fluxes show a minimum sink during the boreal summer. The seasonal cycle of Le Quere et al. (2005) shows a minimum sink during September and October, and a maximum sink during April and May, consistent with the findings of Takahashi et al. (2009). Among the atmospheric inversions, Rödenbeck et al. (2003) reported a seasonal cycle that is largely opposite to that of the present study and to observations. The oceanic seasonal cycle in the inversion performed by Rödenbeck et al. (2003) is probably affected by land signals, as such signals are much larger than oceanic signals and it is generally difficult to separate the two. Moreover, the Rödenbeck et al. (2003) inversion does not employ any prior flux for seasonality. The inversion fluxes reported by Patra et al. (2005) show a seasonal cycle consistent with that reported by Takahashi et al. (2009).

The error bars in Fig. 12 show the interannual standard deviations for each month of the year. The interannual amplitudes are weakest in our assimilated air-sea CO2 fluxes, possibly because of the weak mean constraint applied to the model. The standard deviations are comparable between our product and Le Quere et al. (2005). Interannual variability is greater in the atmospheric inversions performed by Patra et al. (2005) compared to other products compared here. The inversion fluxes
Fig. 11. (Top panel) 1996–2004 mean assimilated sea–air CO$_2$ fluxes. Units are in mol m$^{-2}$ yr$^{-1}$. (Middle panel) Interannual CO$_2$ fluxes from the global ocean as given by the assimilation. (Bottom panel) Area integrated air–sea CO$_2$ flux from the Nino region (area is shown as box in the top panel). Dashed line represents a 12-month running mean.

Fig. 12. Seasonal cycle of global air–sea CO$_2$ fluxes. Units are in PgC yr$^{-1}$. The error bars show the interannual standard deviations of each month calculated from a span of 1996–2004. The y-axis varies for each data product.
calculated by Rödenbeck et al. (2003) show the largest interannual amplitudes among the four products.

The seasonal cycle of the North Pacific alone is worth comparing because our model benefited from a large number of observations in this region. Figure 13 shows the integrated air–sea CO₂ fluxes from the four data sources, for the North Pacific area. The ocean models show comparable seasonal cycles. The error bars in the figure indicate the interannual standard deviations for each month. The amplitude of the interannual signal in the North Pacific CO₂ flux is weaker in the OGCM of Le Quere et al. (2005), and is weaker during the boreal summer in our data product. This finding is consistent among the two ocean models. Among the atmospheric inversions, Patra et al. (2005) reported large interannual variability. The seasonal cycle reported by Patra et al. (2005) is comparable to that in the ocean models. However, the seasonal cycle in atmospheric inversion fluxes reported by Rödenbeck et al. (2003) is inconsistent with that of the other products.

4.1. Intercomparison of interannual variability

This section compares the interannual anomalies of area-integrated air–sea CO₂ fluxes obtained from four data sources. First, we compare the interannual variability in the North Pacific. Figure 14 shows a remarkable agreement between the assimilated air–sea CO₂ fluxes of the present study and atmospheric inversions; the cycles of interannual anomalies are in phase in all cases. The magnitude of interannual CO₂ flux anomalies in the atmospheric inversion reported by Patra et al. (2005) is larger than those from other products, and is shown in the figure using a separate y-axis. Except for the relatively high amplitudes in Patra et al. (2005), the atmospheric inversions generally agree with the interannual variability of assimilated air–sea CO₂ fluxes. The inversion performed by Rödenbeck et al. (2003) yields a similar interannual cycle to that of the present assimilation, although the seasonal cycles are opposite. Given that the inversion method employed by Rödenbeck et al. (2003) is linear, seasonal and interannual variability are possibly independent of each other, meaning that the comparison of interannual variability may be reasonable even if the seasonality is different (C. Rödenbeck, 2009, personal communications). The comparison demonstrates that the global atmospheric inversions agree with the assimilated air–sea CO₂ fluxes of the present study. The interannual variability derived from an OGCM by Le Quere et al. (2005), in contrast, is weak and out of phase compared with the other data sources.

Figure 15 compares the globally integrated air–sea CO₂ flux interannual anomalies between three products: the assimilated air–sea CO₂ fluxes of the present study, the atmospheric inversions of Rödenbeck et al. (2003), and OGCM results presented by Le Quere et al. (2005). We excluded Patra et al. (2005) from this comparison because of the relatively large amplitude of the
anomalies in their study. The three data products show comparable interannual variability (both in magnitude and phase). The assimilation system of the present study agrees with the atmospheric inversions on interannual scales. The assimilated fluxes offer a better comparison with atmospheric inversions than do the OGCM results of Le Quere et al. (2005).

5. Discussion and Conclusion

This study differs from previous attempts at modelling of the ocean carbon cycle in the following key aspects. (1) We used an offline tracer transport model for physical circulation, which uses observationally constrained reanalysis data products. This approach performs well in representing carbon transport in the ocean. (2) The ecosystem model, although a simple model derived from McKinley et al. (2004), is acceptable in terms of representing interannual variability in air–sea CO$_2$ fluxes and was implemented to constrain values of surface ocean $p$CO$_2$ with available observations. (3) The assimilation resulted in improved model performance, yielding a reduction in systematic biases and errors. (4) An intercomparison of the present data product with atmospheric inversions revealed reasonable agreement in the interannual variability of air–sea fluxes. (5) Ultimately, we obtained an observationally constrained air–sea CO$_2$ flux on an interannual time scale for the global ocean.

The use of reanalysis physical circulation enables us to extend the assimilated CO$_2$ flux data in a near-real-time basis. This approach provides a wealth of information in terms of ocean flux priors for upcoming atmospheric inversions using CO$_2$ measurements made by the on-board instrument attached to GOSAT. Thus, the proposed model has the potential to contribute interannually varying near-real-time ocean flux priors, constrained by available observations of $p$CO$_2$ and with reduced bias, for near-real-time atmospheric inversions.

The proposed model can be used to set up for the operational forecasts of CO$_2$ fluxes, depending on the availability of operational ocean circulation. The recent increase in the amount of $p$CO$_2$ observation data is helpful in setting up this model for operational forecasts of CO$_2$ fluxes. Moreover, autonomous instruments such as Argo may contain sensors for DIC and $p$CO$_2$ in the future. In such a case, the proposed model could be used to provide continuously updated, observationally constrained air–sea CO$_2$ flux data.

5.1. Limitations

The method used to constrain values of surface ocean $p$CO$_2$ contains a few uncertainties. Accordingly, we discuss the limitations of the present system, with a view to improving the proposed CO$_2$ assimilation system.

First, the model uses only one adjoint, meaning that the control variable (DIC) is corrected irrespective of the source of errors. This assimilation system considers that DIC determines the surface ocean $p$CO$_2$ in the model, given that SST, SSS and surface alkalinity are close to observed values. However, this assumption has caveat because errors may occur in these variables even though they are derived from the reanalysis offline inputs. The modelled DIC is corrected to match $p$CO$_2$ values with observations; however, there exists little guidance regarding whether the correction is in the desired direction. To verify that the corrections applied to DIC are valid, we compared the assimilated DIC with observations in several sectors of the Pacific. In most cases, the assimilation corrected the DIC in the desired direction. As an example, Fig. 16 shows the surface ocean DIC from three cruises carried out in the North Pacific. The figure compares the assimilated DIC of the surface ocean with observations, revealing a reasonable agreement.

Another limitation of the system is the short term (2 months) of the assimilation windows. We found that the adjoint loses its sensitivity in longer backward runs. Here, the data are considered only at the surface ocean, where vertical mixing is maximized. Therefore, the model-data misfit shows a relatively rapid downward “dilution”. During the 2-month period of adjoint runs, the sensitivities are not propagated to depths greater than the euphotic zone. Therefore, the correction applied to the surface ocean DIC is confined to within 100 m of the surface ocean.

This limits the assimilation system in terms of using surface ocean $p$CO$_2$ observations to eliminate errors in simulated DIC below 100 m depth. However, we cannot be sure that surface ocean $p$CO$_2$ is indeed controlled by deep DIC. Vertical mixing plays a crucial role in bringing the subsurface DIC signal to the surface and thereby influencing $p$CO$_2$. However, the present
assimilation system only considers corrections that are possible within the distance across which adjoints can propagate in a 2-month iteration. Moreover, DIC is a non-conservative tracer, which may contribute to the weak sensitivity of adjoints to the longer backward integration.

Although the present method has limitations, it represents an important step toward additional future improvements. The incorporation of satellite-derived chlorophyll-a data into the assimilation would be a promising future step, along with expanding the ecosystem model to include the full dynamics of alkalinity.

6. Acknowledgments

This work was carried out as part of the GOSAT carbon cycle research project at CGER, NIES. Computational resources were provided by supercomputer facility at NIES. The offline fields for the physical model were provided by GFDL, Princeton, NJ. The authors are thankful to the comments received from Andrew Jacobson, NOAA, U.S.A. and one anonymous reviewer which were helpful to improve the quality of the paper. The paper is proof read by Stallard Scientific Editing, Nelson, New Zealand. Global ocean air–sea CO2 flux data from 1996 to 2004 are available for research purposes; please contact vinu.valsala@nies.go.jp.

References

Baker, D. F., Doney, S. C. and Schimel, D. S. 2006. Variational data assimilation for atmospheric co2. Tellus 58B, 359–365.
Bennett, A. F. 2002. Inverse Modeling of the Ocean and Atmosphere. Cambridge University Press, New York, NY, 234 pp.
Bryan, K. and Lewis, L. J. 1979. A watermass model of the World ocean. J. Geophys. Res. 84, 2503–2517.
Delworth, T. L., Broccoli, A. J., Rosati, A., Stouffer, R. J., Balaji, V. and co-authors. 2006. GFDL’s CM2 Global coupled climate models. Part I: Formulation and simulation characteristics. J. Clim. 19, 643–674.
Dutkiewicz, S., Marshall, M. F. J. and Gregg, W. W. 2001. Interannual variability of phytoplankton abundances in the north Atlantic. Deep-Sea Res. 48, 2323–2344.
Feely, R. A., Boutin, J., Cosca, E. C., Dandonneau, Y., Etcheto, J. and co-authors. 2002. Seasonal and interannual variability of co2 in the equatorial pacific. Deep-Sea Res. 14, 2443–2469.
Feely, R. A., Wanninkhof, R., Takahashi, T. and Tans, P. 1999. Influence of el niño on the equatorial pacific contribution to atmospheric CO2 accumulation. Nature 398, 365–386.
Fukumori, I., Cheng, T. L. and B. and Menemenlis, D. 2004. Theorigine, pathway and destination of Niño-3 water estimated by a simulated passive tracer and its adjoint. J. Phys. Oceanogr. 34, 582–604.
Gent, P. R. and McWilliams, J. C. 1990. Isopycnal mixing in ocean circulation models. J. Phys. Oceanogr. 20, 150–155.
Gnanadesikan, A., Dixon, K. W., Griffies, S. M., Balaji, V., Barreiro, M. and co-authors. 2006. GFDL’s CM2 global coupled climate models. Part II: The baseline ocean simulation. J. Clim. 19, 675–697.
Griffies, S. M. and Hallberg, R. W. 2000. Biharmonic friction with a smagorinsky-like viscosity for use in large-scale eddy-permitting ocean models. J. Phys. Oceanogr. 128, 2935–2946.
Gruber, M. Gloor, M., Fletcher, S. E. M., Doney, S. C., Dutkiewicz, S. and co-authors. 2009. Oceanic sources, sinks and transport of atmospheric CO2. Global Biogeochem. Cycles 23, doi:10.1029/2008GB003349.
Gurney, K. R., Law, R. M., Denning, A. S., Rayner, P. J., Pak, B. and Transom-3-L2-modelers 2004. Transcom-3 inversion intercomparison: control results for the estimation of seasonal carbon sources and sinks. Global Biogeochem. Cycles 18, doi:10.1029/2003GB002111.
Hourdin, F. and Talagrand, O. 2002. Eulerian backtracking of atmospheric tracers: I adjoint derivation and parameterization of subgrid-scale transport. Quart. J. Roy. Meteor. Soc. 128, doi:10.1256/qj.01.53.
