Final Draft
of the original manuscript:

Brix, H.; Menemenlis, D.; Hill, C.; Dutkiewicz, S.; Jahn, O.; Wang, D.;
Bowman, K.; Zhang, H.:

**Using Green's Functions to initialize and adjust a global, eddying
ocean biogeochemistry general circulation model**

In: Ocean Modelling (2015) Elsevier

DOI: 10.1016/j.ocemod.2015.07.008

Using Green's Functions to Initialize and Adjust a Global, Eddying Ocean Biogeochemistry General Circulation Model

H. Brix^{a,*}, D. Menemenlis^b, C. Hill^c, S. Dutkiewicz^c, O. Jahn^c, D. Wang^c,
K. Bowman^b, H. Zhang^d

^a*Department of Atmospheric and Oceanic Sciences, The University of California, Los Angeles, CA, USA. Now at Helmholtz-Zentrum Geesthacht, Institute of Coastal Research, Tel: +49-4152-87-1523; fax: +49-4152-87-1525*

^b*California Institute of Technology, Jet Propulsion Laboratory, Pasadena, CA, USA*

^c*Massachusetts Institute of Technology, Cambridge, MA, USA*

^d*Joint Institute for Regional Earth System Science and Engineering, The University of California, Los Angeles, CA, USA*

Abstract

The NASA Carbon Monitoring System (CMS) Flux Pilot Project aims to attribute changes in the atmospheric accumulation of carbon dioxide to spatially resolved fluxes by utilizing the full suite of NASA data, models, and assimilation capabilities. For the oceanic part of this project, we introduce ECCO2-Darwin, a new ocean biogeochemistry general circulation model based on combining the following pre-existing components: (i) a full-depth, eddying, global-ocean configuration of the Massachusetts Institute of Technology general circulation model (MITgcm), (ii) an adjoint-method-based estimate of ocean circulation from the Estimating the Circulation and Climate of the Ocean, Phase II (ECCO2) project, (iii) the MIT ecosystem model “Darwin”, and (iv) a marine carbon chemistry model. Air-sea gas exchange

*Corresponding author

Email address: hbrix@ucla.edu (H. Brix)

coefficients and initial conditions of dissolved inorganic carbon, alkalinity, and oxygen are adjusted using a Green's Functions approach in order to optimize modeled air-sea CO₂ fluxes. Data constraints include observations of carbon dioxide partial pressure ($p\text{CO}_2$) for 2009–2010, global air-sea CO₂ flux estimates, and the seasonal cycle of the Takahashi et al. (2009) Atlas. The model sensitivity experiments (or Green's Functions) include simulations that start from different initial conditions as well as experiments that perturb air-sea gas exchange parameters and the ratio of particulate inorganic to organic carbon. The Green's Functions approach yields a linear combination of these sensitivity experiments that minimizes model-data differences. The resulting initial conditions and gas exchange coefficients are then used to integrate the ECCO2-Darwin model forward. Despite the small number (six) of control parameters, the adjusted simulation is significantly closer to the data constraints and to independent observations than is the baseline simulation. For example, the root-mean square difference with observed alkalinity decreased from 48.8 $\mu\text{mol/kg}$ for the baseline simulation to 28.9 $\mu\text{mol/kg}$ for the adjusted simulation, even though alkalinity observations were not used as data constraints. In addition to reducing biases relative to observations, the adjusted simulation exhibits smaller model drift than the baseline. For example, the volume-weighted drift reduction in the top 300 m is 12.5% for nitrate and 30% for oxygen. Although there remain several regions with large model-data discrepancies, for example, overly strong carbon uptake in the Southern Ocean and weak outgassing in the Equatorial Pacific, the adjusted simulation is a first step towards a more accurate representation of the ocean carbon cycle at high spatial and temporal resolution.

Keywords:

Carbon Monitoring System, ocean biogeochemical circulation model,
Green's function, data assimilation

1. Introduction

The components of the global carbon cycle interact through fluxes of natural and anthropogenic carbon between the atmosphere, ocean, and land reservoirs. Towards improved quantitative and mechanistic understanding of the exchange processes, the NASA Carbon Monitoring System (CMS) Flux Project¹ aims to attribute changes in the atmospheric accumulation of carbon dioxide (CO₂) to spatially-resolved fluxes by utilizing the full suite of NASA data, models, and assimilation capabilities. The oceanic component of these fluxes is of critical importance as it is estimated that the oceans have absorbed $48 \pm 9\%$ of the anthropogenic CO₂ emitted since the beginning of the industrial period (Sabine et al., 2004). The current oceanic CO₂ uptake is estimated to be about a quarter of the anthropogenic emissions (Le Quéré et al., 2010). The ECCO2-Darwin model, described herein, and the NASA Ocean Biogeochemical Model (NOBM; Gregg and Casey, 2007) provide data-constrained, time-varying, spatially-resolved estimates of air-sea CO₂ fluxes to the CMS Flux Project. These fluxes are calculated for the period 2009–2010, the time frame for which GOSAT (Kuze et al., 2009) atmospheric x CO₂ (dry mole air fraction of CO₂) measurements are available. The NOBM and ECCO2-Darwin air-sea CO₂ flux estimates serve as a priori surface forcing

¹see <http://carbon.nasa.gov> and <http://cmsflux.jpl.nasa.gov/>

20 for “top-down” atmospheric flux inversions carried out by the CMS Flux
21 Project (Liu et al., 2014).

22 Although the flux of CO_2 across the air-sea interface cannot be measured
23 directly, its magnitude is known to depend on the difference ($\Delta p\text{CO}_2$) be-
24 tween the partial pressures of CO_2 of the ocean ($p\text{CO}_2$) and of the atmosphere
25 ($p\text{CO}_2^{\text{atm}}$). The air-sea CO_2 flux for a given $\Delta p\text{CO}_2$ is usually estimated us-
26 ing a parameterization where the gas exchange coefficients depend on wind
27 speed at the surface of the ocean (e.g., Wanninkhof, 1992; Takahashi et al.,
28 2002; Sarmiento and Gruber, 2006; Krakauer et al., 2006).

29 Oceanic $p\text{CO}_2$ depends on temperature and on other components of the
30 oceanic carbon system, which in turn depend on ocean circulation and bio-
31 logical activity. Observations of $p\text{CO}_2$ are based on in-situ measurements and
32 have been compiled into large data bases (Takahashi et al., 2002, 2009; Pfeil
33 et al., 2012). Takahashi et al. (2002) used 940,000 measurements of surface-
34 water $p\text{CO}_2$ obtained since the International Geophysical Year of 1956–1959
35 to compile a gridded monthly atlas of surface $p\text{CO}_2$ distribution for the ref-
36 erence year 1995, later updated for the year 2000 (Takahashi et al., 2009).

37 On the basis of the available measurements numerous efforts have been
38 undertaken to quantify the oceanic uptake of CO_2 and its spatial structure
39 (e.g., Keeling et al., 1996; Battle et al., 2000; Takahashi et al., 2002, 2009),
40 its temporal variability, and the magnitude of its anthropogenic component
41 (e.g., Gloor et al., 2003; McNeil et al., 2003; Sabine et al., 2004).

42 Other approaches to quantify air-sea CO_2 fluxes are based on atmospheric
43 CO_2 mixing ratios (e.g., Conway et al., 1994) and transport inversions (e.g.,
44 Bousquet et al., 2000; Gurney et al., 2002; Rödenbeck et al., 2003; Baker

et al., 2006), which “estimate the regional distribution of air-surface CO₂ fluxes using the spatiotemporal variability in atmospheric CO₂ concentration measurements” (Le Quéré et al., 2009). With regard to the oceans, the results of these studies are difficult to interpret as the oceanic signal variability is small compared to terrestrial fluxes.

Another air-sea-CO₂-flux estimation method that does not require knowledge of a gas exchange coefficient is based on inverse calculations of oceanic transports using ocean-interior data of dissolved inorganic carbon (*DIC*) and on ocean circulation models (Gloor et al., 2003; Mikaloff Fletcher et al., 2006, 2007; Gruber et al., 2009). While this method allows a certain (coarse) regionalization of model results, it only yields long-term CO₂ fluxes without temporal variability.

Numerous biogeochemical ocean forward models have been developed (Aumont and Bopp, 2006; Le Quéré et al., 2007; Thomas et al., 2008; Doney et al., 2009; Buitenhuis et al., 2010; Assmann et al., 2010; Galbraith et al., 2010, to name a few recent ones). These models differ substantially in resolution as well as in the formulation of model physics, biology, and the components of the carbon cycle, depending on the questions they intend to answer. In light of societal interest in anthropogenic changes of carbon sources and sinks, substantial effort went into improving these models to address recent changes of oceanic carbon uptake (e.g., Le Quéré et al., 2007, 2009, 2010; Lenton et al., 2012) and possible future development and feedback under climate change scenarios (e.g., Roy et al., 2011; Séférian et al., 2012).

For the specific task of developing a Carbon Monitoring System, the requirement of an ocean component that represents the current state of the

oceans as realistically as possible leads to the following three criteria. First, to capture the fine scale structure of sources and sinks and their temporal changes in certain regions it is desirable to have a resolution that is at least eddy-permitting. The North Atlantic is an example of a region exhibiting such high variability on both temporal and spatial scales (see, for instance, Fig. 3 in Le Quéré et al., 2009). Second, biological and biogeochemical processes should be represented as realistically as computational constraints allow, and third, observational data should be incorporated into the model simulations to the highest degree possible. In order to address these three requirements, we developed the ECCO2-Darwin model (as described later in this paper), which combines eddying horizontal resolution with state-of-the-art biogeochemistry and data assimilation capability.

When setting up an Ocean Biogeochemistry General Circulation Model (OBGCM) with this type of specifications, one challenge is high computational cost. In the ocean, where circulation is slow compared to the atmosphere, it can take thousands of (model) years to reach an equilibrium state. As increased spatial resolution requires more computing time, increasing model resolution diminishes the ability to “spin-up” the model for a sufficiently long time. Adding complexity to models by, for instance, including biogeochemical cycles with many tracers, exacerbates the problem, as the model equations have to be solved for each additional tracer.

Another problem associated with long spin-ups is model drift, that is, the increasing biases between the model simulation and nature. Although initial conditions are traditionally chosen to be as close to nature as possible, systematic model or boundary condition errors will cause the simulation to

95 drift from a state near the attractor of nature to a state close to the attractor
96 of the model (Toth and Peña, 2007). If this model equilibrium state is too
97 far from nature, estimates of air-sea fluxes and the underlying physics and
98 biogeochemistry that drive these fluxes will be unrealistic, which means they
99 will not adequately represent the natural phenomena they are supposed to
100 model.

101 For short model integrations that circumvent these drift issues and allow
102 high resolution and model complexity, choosing initial conditions therefore
103 becomes a critical issue. The availability of oceanic data at any given point
104 in time chosen to be the model’s starting point is extremely limited. While
105 remote sensing has improved the data situation drastically, it only supplies
106 data for the top layer of the ocean and gives very little information that can
107 be used to initialize ocean biogeochemistry. To solve this initialization prob-
108 lem for all biogeochemical quantities simultaneously and to avoid producing
109 unrealistic air-sea carbon flux estimates, we use a simple-to-implement es-
110 timation approach based on the computation of model Green’s Functions,
111 that is, based on forward-model sensitivity experiments (Menemenlis et al.,
112 2005a).

113 This study is meant to (i) introduce the ECCO2-Darwin OBGCM and
114 (ii) demonstrate the application of the Green’s Functions approach to the
115 initialization and adjustment of high resolution biogeochemical ocean mod-
116 els. We start by introducing the various components of the ECCO2-Darwin
117 OBGCM (Section 2). We then describe the initial model simulation (Sec-
118 tion 3) and results from the various sensitivity experiments (Section 4). Next
119 we describe the observations used to evaluate and adjust the model solution

120 (Section 5). In Sections 6 and 7, we outline the data assimilation method,
121 our choice of control parameters, and the optimized model realization. Fi-
122 nally we evaluate this optimized solution relative to other studies and discuss
123 residual errors in Sections 8 and 9.

124 **2. The ECCO2-Darwin Model**

125 The ECCO2-Darwin Ocean Biogeochemistry General Circulation Model
126 is based on a global, eddying, ocean and sea ice configuration of the Mas-
127 sachusetts Institute of Technology general circulation model (MITgcm; Mar-
128 shall et al., 1997a,b) and on results from two separately funded projects: the
129 Estimating the Circulation and Climate of the Ocean, Phase II (ECCO2)
130 Project, which provides a data-constrained estimate of the time-evolving
131 physical ocean state, and the Darwin Project, which provides time-evolving
132 ocean ecosystem variables. Together, ECCO2 and Darwin provide time-
133 evolving physical and biological variables, which are used to drive a carbon
134 chemistry component model. We developed this new OBGCM in order to
135 compute surface fluxes of carbon at high spatial and temporal resolution for
136 the NASA CMS Flux Pilot Project.

137 *2.1. ECCO2*

138 The ECCO2 global-ocean model configuration uses a cube-sphere grid
139 (Adcroft et al., 2004) with 18-km horizontal grid spacing and 50 vertical lev-
140 els (Menemenlis et al., 2005b, 2008). It includes a dynamic/thermodynamic
141 sea ice model (Losch et al., 2010; Heimbach et al., 2010). In a first step, the
142 ECCO2 model configuration was adjusted using a low-dimensional (Green’s
143 Functions) estimation approach (Menemenlis et al., 2005a). In a second step,

144 the method of Lagrange multipliers, also known as the adjoint method (Wun-
145 sch and Heimbach, 2007), was used to adjust initial and time-evolving surface
146 boundary conditions. Data constraints for the adjoint-method optimization
147 include sea level anomaly from altimeters on Jason-1, Jason-2/Ocean Surface
148 Topography Mission (OSTM), and Environmental Satellite (Envisat); sea
149 surface temperature from the Advanced Microwave Scanning Radiometer-
150 EOS (AMSR-E); and temperature and salinity profiles from the Argo profil-
151 ing floats. The adjoint-method-based ECCO2 estimates of the global-ocean
152 circulation available at the beginning of this study were for year 2004 and
153 for 16 months starting in January 2009. These physical ocean circulation
154 estimates were used to drive the Darwin ecosystem model.

155 2.2. *Darwin and Ocean Chemistry*

156 The Darwin Project is an initiative to advance the development and ap-
157 plication of novel models of marine microbial communities, to identify the
158 relationships of individuals and communities to their environment, and to
159 connect cellular-scale processes to global microbial community structure (Fol-
160 lows et al., 2007; Follows and Dutkiewicz, 2011; Dutkiewicz et al., 2009). The
161 configuration used for the CMS Flux Pilot Project includes five phytoplank-
162 ton functional types (choices based on results from previous versions of the
163 model, see Section 3.1) and two zooplankton types. The particular setup we
164 use in this study prescribes properties for these phytoplankton types and thus
165 limits open competition between phytoplankton and zooplankton classes as
166 we did not, in this study, aim to investigate the impact of phytoplankton
167 diversity on carbon fluxes. Therefore the Darwin model is here used in a
168 configuration that is similar to a more conventional ocean biogeochemistry

169 model with multiple functional groups. The carbon cycle is explicitly rep-
170 resented, along with those of nitrogen, phosphorus, iron, silica, oxygen, and
171 alkalinity. The carbonate chemistry follows the simplified model proposed by
172 Follows et al. (2006) and air-sea CO₂ exchange is parameterized according
173 to Wanninkhof (1992).

174 **3. VERSION 1: The First ECCO2-Darwin Simulation**

175 To obtain a first estimate of air-sea carbon fluxes for 2009 and 2010,
176 ECCO2-Darwin was integrated using a “best guess” set of initial and bound-
177 ary conditions. We will refer to this simulation as “VERSION 1” or “V1”.
178 V1 results are available on the CMS Flux web site².

179 *3.1. Initial and Boundary Conditions*

180 The ECCO2-Darwin model was initialized in January 2004, the first
181 year for which optimized physical initial conditions provided by the adjoint-
182 method-based solution described in Section 2.1 were available. Biogeochem-
183 ical initial conditions for the spin-up were provided by an eight-year inte-
184 gration of Darwin driven by an earlier ECCO2 solution. The ecosystem
185 model of that simulation was that of Dutkiewicz et al. (2009). This earlier
186 ECCO2 simulation did not include carbon chemistry and had been initial-
187 ized with phosphate, nitrate, and silicate fields from the World Ocean Atlas
188 (Garcia et al., 2006b), and all other fields from an ECCO-GODAE solution
189 (Dutkiewicz et al., 2009), with the exception of iron fields that were pro-
190 vided from MIT’s Integrated Global System Modeling framework (IGSM;

²<http://cmsflux.jpl.nasa.gov/>

191 Dutkiewicz et al., 2012) in the Arctic, and ECCO-GODAE for the rest of the
 192 world. Phyto- and zooplankton biomass were from ECCO-GODAE, equally
 193 distributed over all types. The initial conditions for *DIC*, alkalinity, and
 194 oxygen for V1 were taken from a realization of the Community Climate Sys-
 195 tem Model (CCSM-3) ocean Biogeochemical Elemental Cycle (BEC) model
 196 (Blackmon et al., 2001; Lovenduski et al., 2007; Doney et al., 2009) that
 197 had been spun up to pre-industrial conditions (600 years after which it had
 198 reached a cyclostationary state (Lovenduski et al., 2007) and then continued
 199 with historical atmospheric CO₂ for 1765 through 2004.

200 *3.2. Model Spin-Up and Experimental Set-Up*

201 The ECCO2-Darwin model spin-up started January 2004 with two years
 202 of identical physical atmospheric boundary conditions provided by the 2004
 203 adjoint-method ECCO2 solution described in Section 2.1. The two year spin-
 204 up was followed by a “ramp-up” for the years 2004 through 2008. For this
 205 period the atmospheric boundary conditions were obtained from the Japan
 206 Meteorological Agency and Central Research Institute of Electric Power In-
 207 dustry 25-year reanalysis (JRA-25; Onogi et al., 2007).

208 The “V1” ECCO2-Darwin simulation, which is described and evaluated
 209 below, used physical boundary conditions from the ECCO2 adjoint solution
 210 for the year 2009 and for the first four months of 2010, followed by JRA-25
 211 forcing for the remainder of 2010. Surface boundary conditions for dust/iron
 212 deposition were provided by the climatology of Mahowald et al. (2006). At-
 213 mospheric $p\text{CO}_2$ forcing fields were constructed from spatially variable, daily
 214 atmospheric $p\text{CO}_2$ fields from 2009 GEOS-Chem results (Nassar et al., 2010)
 215 produced as part of the NASA CMS Flux Pilot Project. The spatial and

216 temporal anomalies of these fields were adjusted globally using the global
217 atmospheric monthly $p\text{CO}_2$ means from NOAA’s Earth System Research
218 Laboratory³.

219 3.3. Results and Evaluation

220 One of the most striking features of the simulations with ECCO2-Darwin
221 is the high temporal and spatial variability of air-sea CO_2 fluxes. The day-to-
222 day variability in the mid- and high latitudes is dominated by synoptic scale
223 atmospheric systems. An example from this simulation for six consecutive
224 days in January 2009 is shown in Fig. 1. The air-sea exchange processes
225 are dominated by wind speed variability, which is parameterized through a
226 quadratic dependence of the air-sea flux on the wind speed (Wanninkhof,
227 1992). While the reflection of atmospheric synoptic activity in air-sea gas
228 exchange fluxes is not surprising, it is worth pointing out that this high
229 variability is lost when using flux atlases (e.g., Takahashi et al., 2009) to
230 drive atmospheric models.

231 To evaluate the quality of our model calculations we compare our globally
232 integrated mean CO_2 exchange to published values derived from other ocean
233 models, for example, the Global Carbon Project (2011). There, the global
234 ocean uptake for 2010 is estimated to be $2.4 \pm 0.5 \text{ PgC yr}^{-1}$. Our V1 solution
235 yields 3.6 PgC yr^{-1} (see Tab. 1). The temporal evolution of this globally
236 integrated flux is shown in Fig. 2 for 2009 and 2010. The blue circles represent
237 V1 and the thick black lines monthly mean values from Takahashi et al.

³NOAA’s ESRL data provided through their website at <http://www.esrl.noaa.gov/gmd/ccgg/trends/>

(2009), which are referenced to 2000. The annual cycle of global uptake and outgassing can be explained by looking at processes in some key regions. As the processes shaping this annual cycle and the associated spatial patterns are the same for the other model realizations and only differ in magnitudes, we will discuss these processes later in Section 7.

With the overall CO_2 uptake of the ocean in V1 being clearly too strong, compared, for instance, to the Global Carbon Project (2011) values, we proceeded to adjust our solution, as is described next.

4. Sensitivity Experiments

To improve the air-sea CO_2 fluxes from our initial solution we performed a suite of sensitivity experiments spanning the years 2009 and 2010. The purpose of carrying out these experiments was twofold: (i) to investigate to what extent the magnitude and variability of air-sea carbon fluxes depend on initial conditions and empirical model parameters and (ii) to construct a suite of solutions that could be used to adjust these fluxes.

The physical initial conditions were identical for all sensitivity experiments (and the same as those described in Section 3.1). Five experiments (CCSM, GLODAP, KS, BLEND, and NOBM) varied in their biogeochemical initial conditions and were spun-up as described in Section 3.1, starting in 2004. Additional sensitivity experiments (PISVEL and PICPOC) varied in their model parameterizations while using the January 2009 biogeochemical fields of the GLODAP integration as initial conditions. A more detailed description of the sensitivity experiments is given in the Appendix.

261 4.1. Results and Evaluation

262 Overall, most of our sensitivity experiments reproduced the spatial and
263 temporal patterns observed in biogeochemical quantities, although ampli-
264 tudes, either globally or with regard to specific regions, were not always well
265 matched. As an illustration, the two top rows of Fig. 3 show annual means of
266 surface *DIC* from four of our sensitivity experiments for 2009. The bottom
267 left shows the values from the GLODAP Atlas that had been used to initial-
268 ize our GLODAP sensitivity experiment (top left of Fig. 3). The differences
269 between the atlas and the model experiments are attributed to following
270 causes. First, the initial *DIC* fields of the model experiments differed from
271 each other quite substantially and thus led to a variety of patterns and am-
272 plitudes. Second, the GLODAP Atlas is based on measurements taken in
273 the 1990's and does not necessarily reflect contemporary *DIC* fields. It can
274 therefore be expected that the observational panel in Fig. 3 will not match
275 the model results for 2009. Furthermore, the GLODAP *DIC* Atlas does
276 not cover the Arctic Ocean, nor several marginal seas represented in our
277 model. For our GLODAP experiment, we created values for these areas by
278 extrapolating from nearest neighbors. The errors that can be introduced by
279 this extrapolation are reflected in the substantial differences between model
280 solutions in the Arctic Ocean and marginal seas.

281 Monthly mean air-sea CO₂ fluxes from our model experiments are com-
282 pared to each other and to the Takahashi et al. (2009) Atlas in Fig. 4 for July
283 2009 and in Fig. 5 for January 2010. The strongest discrepancies between
284 model and atlas are in the Southern Ocean. Measurements are notoriously
285 sparse in this region of the world ocean (especially in winter) and this lack

286 of information can bias an observationally-based atlas like that of Takahashi
287 et al. (2009) (lower left in Figs. 4 and 5). In these figures the coarse spatial
288 bins (5° by 4°) and absence of fine scale structure of the atlas contrast with
289 the finer horizontal grid spacing of the model (approximately 18 km). The
290 monthly means of the Takahashi et al. (2009) Atlas are referenced to year
291 2000 while model results are shown for July 2009 and January 2010. The
292 global oceanic uptake of CO_2 is modulated by the increase of atmospheric
293 CO_2 concentrations and by natural and anthropogenic climate variability
294 (e.g., Le Quéré et al., 2009). Due to all these factors the model-data dis-
295 crepancies are not surprising. Nevertheless, the comparison between model
296 simulations and the atlas highlights areas that warrant further investigation.

297 In July 2009 the main differences between the model runs in Fig. 4 lie
298 in the strength of the outgassing in the Equatorial Pacific, the spatial ex-
299 tension of the subtropical outgassing in the Northern Hemisphere, and the
300 structure and size of air-sea gas exchange fluxes south of 30°S . For January
301 2010 (Fig. 5) the model runs all share the strong oceanic carbon uptake in
302 the Southern Ocean (contrasted by the weak uptake in the climatology). The
303 differences in equatorial upwelling are similar to the July ones. The model
304 runs disagree on the structure and magnitude of air-sea carbon fluxes north
305 of 40°N , especially in the subpolar North Pacific (Fig. 5). While we cannot
306 rule out model deficiencies due to, for instance, incorrect circulation patterns
307 or the representation of alkalinity, discrepancies might also be due to differ-
308 ing reference years. Takahashi et al. (2009) show declining $p\text{CO}_2$ values for
309 this region between 2002 and 2006 which would lead to weaker outgassing or
310 even an uptake of CO_2 as shown in most of our model realizations.

311 The globally integrated mean ocean uptake for 2010 of our model sensi-
 312 tivity experiments ranges from 0.62 to 4.18 PgC yr⁻¹ (see Tab. 1) compared
 313 to the 2.4±0.5 PgC yr⁻¹ estimated by the Global Carbon Project (2011).
 314 The red and green circles in Fig. 2 represent two of our sensitivity exper-
 315 iments (GLODAP and NOBM), the blue dots represent V1, and the thick
 316 black lines show monthly mean values from Takahashi et al. (2009), which
 317 are referenced to 2000. A further discussion and evaluation of the spatial
 318 patterns and the time-series of air-sea CO₂ fluxes and their causes will be
 319 presented in Section 7.

320 **5. *p*CO₂ Data Sets**

321 In addition to evaluating our results against atlases (Key et al., 2004;
 322 Takahashi et al., 2009) and global mean values (Global Carbon Project, 2011)
 323 we also used surface *p*CO₂ data for 2009 and 2010 from the “Global Surface
 324 *p*CO₂ (LDEO) Database” of the Carbon Dioxide Information Analysis Center
 325 (Takahashi et al., 2011) to adjust our model results.

326 The available data for the time period of our model runs are very sparse
 327 and far from being regularly distributed (Fig. 6). We sampled our model
 328 sensitivity runs at the times and locations of the measurements. The dis-
 329 tribution and shape of the resulting scatter plots (not shown) indicates that
 330 there are substantial differences between the data and our model, which can
 331 be attributed to model deficiencies and observational errors.

332 The wide spread of our solutions, mainly due to the choice of initial
 333 conditions, illustrates the difficulty in estimating air-sea gas exchange at the
 334 ocean surface. To quantify the data-model mismatch and determine our

335 “best” solution we calculated a cost function (details in Section 6). The cost
 336 is defined as a weighted sum of the quadratic differences between observations
 337 and model. Values for the costs per observation (J/N_{OBS}) of the GLODAP
 338 experiment for this initial scenario were the lowest of the runs with variable
 339 initial conditions,. We therefore choose GLODAP to be the reference or
 340 “baseline” experiment for the model linearization and optimization, discussed
 341 next.

342 **6. Optimal Linear Combination**

343 Recently, two studies have assimilated oceanic $p\text{CO}_2$ data into numerical
 344 models: Valsala and Maksyutov (2010) used an offline tracer transport model
 345 driven by reanalysis ocean currents, and While et al. (2012) assimilated $p\text{CO}_2$
 346 into a biogeochemical ocean general circulation model. Both studies report
 347 improvements in representing $p\text{CO}_2$ fields. Valsala and Maksyutov (2010)
 348 also report “remarkable correlations with the CO_2 fluxes obtained from at-
 349 mospheric inversion”, that is, they find improvements also when comparing
 350 their results to independent data sets. Here we use $p\text{CO}_2$ measurements
 351 to constrain initial conditions and exchange coefficients of a global, eddying
 352 biogeochemical ocean model in order to improve air-sea CO_2 flux estimates.

353 *6.1. Green’s Function Approach*

354 The data assimilation approach used in this study is a least squares
 355 method based on the computation of model Green’s Functions. Green’s Func-
 356 tions have previously been used for atmospheric tracer inversions (Enting and
 357 Mansbridge, 1989; Tans et al., 1990; Bousquet et al., 2000), ocean circula-
 358 tion estimates (Stammer and Wunsch, 1996; Menemenlis and Wunsch, 1997),

359 ocean carbon inversions (Gloor et al., 2003; Mikaloff Fletcher et al., 2006,
 360 2007), joint ocean-atmosphere carbon dioxide inversions (Jacobson et al.,
 361 2007a,b), and anthropogenic CO₂ sequestration estimates (Khatiwala et al.,
 362 2009). Menemenlis et al. (2005a) showed that Green’s Functions are a very
 363 powerful methodology for adjusting general circulation model parameters.
 364 The traditional way of adjusting general circulation models is to run them
 365 repeatedly with perturbed model parameters until the solution appears ad-
 366 equate for the problem at hand or until one runs out of time, patience, or
 367 compute cycles. The problem with such a trial-and-error approach is that
 368 the final choice of model parameters is not based on objective criteria and,
 369 because the perturbed parameters are usually linearly dependent, it is diffi-
 370 cult to subjectively choose optimal parameters. A trial-and-error parameter
 371 search also usually requires several perturbation experiments per parameter,
 372 while a single perturbation experiment per parameter often suffices for the
 373 Green’s Functions approach.

374 As applied herein, the Green’s function approach involves the computa-
 375 tion of OBGCM sensitivity experiments followed by a recipe for constructing
 376 a solution that is the best linear combination of these sensitivity experiments.
 377 In other words, Green’s Functions are used to linearize the ocean biogeochem-
 378 istry model and discrete inverse theory is used to estimate uncertain model
 379 parameters. Algebraically, the OBGCM can be represented by a set of rules
 380 for time-stepping a state vector:

$$\mathbf{x}^f(t_{i+1}) = M_i [\mathbf{x}^f(t_i)] . \quad (1)$$

381 The OBGCM state vector $\mathbf{x}^f(t_i)$ includes the prognostic variables of the bio-
 382 geochemical model as described in Section 2.2, for instance, *DIC*, alkalinity,

383 and oceanic $p\text{CO}_2$, on a predefined grid at discrete time t_i . Function M_i rep-
 384 resents the known OBGCM time-stepping rules. The discretized dynamics
 385 of the true ocean $\mathbf{x}^t$ are assumed to differ from that of the numerical model
 386 (1) by a vector of stochastic perturbations:

$$\mathbf{x}^t(t_{i+1}) = M_i [\mathbf{x}^t(t_i), \boldsymbol{\eta}] , \quad (2)$$

387 where $\boldsymbol{\eta}$ represents uncertain model parameters, for example, gas exchange
 388 coefficients and initial and surface boundary conditions. It is assumed that $\boldsymbol{\eta}$
 389 is a noise process with zero mean and covariance matrix $\mathbf{Q}$. The optimization
 390 problem aims to minimize an objective quadratic function,

$$J = \boldsymbol{\eta}^T \mathbf{Q}^{-1} \boldsymbol{\eta} + \boldsymbol{\varepsilon}^T \mathbf{R}^{-1} \boldsymbol{\varepsilon}, \quad (3)$$

391 that penalizes differences between model and observations

$$\mathbf{y}^o = H \begin{bmatrix} \mathbf{x}^t(t_0) \\ \vdots \\ \mathbf{x}^t(t_N) \end{bmatrix} + \boldsymbol{\varepsilon}, \quad (4)$$

392 where vector $\mathbf{y}^o$ represents all available observations during the estimation
 393 period, $t_0 \leq t_i \leq t_N$, H is the measurement function, and residual $\boldsymbol{\varepsilon}$ is a
 394 noise process, which is assumed to have zero mean and covariance matrix
 395 $\mathbf{R}$. Vector $\boldsymbol{\varepsilon}$ represents measurement errors and all model errors that are not
 396 represented by $\boldsymbol{\eta}$ in (2) and superscript T is the transpose operator.

397 For the Green's Functions approach, we seek to adjust a small, judiciously
 398 chosen subset of parameters $\boldsymbol{\eta}$. Here this subset includes piston velocity
 399 (PISVEL), the particulate inorganic to organic carbon ration (PICPOC),
 400 and initial conditions, to be estimated as a linear combination of the CCSM,

401 GLODAP, KS, BLEND, and NOBM sensitivity experiments described in
 402 Section 4. The objective is to adjust these parameters using the observations
 403 described in Section 5. To do this, Eqs. (2) and (4) are combined so that
 404 observations can be directly related to uncertain model parameters:

$$\mathbf{y}^o = G[\boldsymbol{\eta}] + \boldsymbol{\varepsilon}, \quad (5)$$

405 where G represents the convolution of observation operator H with OBGCM
 406 function M_i . We assume that (5) can be linearized about the GLODAP/BASELINE
 407 model trajectory. If the linearization assumption holds, (5) simplifies to

$$\mathbf{y}^d = \mathbf{y}^o - G[\mathbf{0}] = \mathbf{G}\boldsymbol{\eta} + \boldsymbol{\varepsilon}, \quad (6)$$

408 where $\mathbf{0}$ is the null vector, $G[\mathbf{0}]$ is the BASELINE integration sampled at
 409 the data locations, vector $\mathbf{y}^d$ is the model-data difference, and $\mathbf{G}$ is a matrix
 410 whose columns are the Green functions of G . Specifically, the j -th column
 411 of matrix $\mathbf{G}$ is

$$\mathbf{g}_{(j)} = \frac{G[\mathbf{e}_j] - G[\mathbf{0}]}{e_j}, \quad (7)$$

412 where $\mathbf{e}_j$ is a perturbation vector that is everywhere zero except for element
 413 j , which is set to e_j . That is, each column of $\mathbf{G}$ can be computed using an
 414 OBGCM sensitivity experiment. Matrix $\mathbf{G}$ is called the data kernel because
 415 it relates the data $\mathbf{y}^d$ with model parameters $\boldsymbol{\eta}$. The minimization of (3)
 416 given (6) is a discrete linear inverse problem with solution

$$\boldsymbol{\eta}^a = \mathbf{P}\mathbf{G}^T\mathbf{R}^{-1}\mathbf{y}^d \quad (8)$$

417 and uncertainty covariance matrix

$$\mathbf{P} = (\mathbf{Q}^{-1} + \mathbf{G}^T\mathbf{R}^{-1}\mathbf{G})^{-1}. \quad (9)$$

418 Although many of the simplifying assumptions listed above are not satisfied
 419 exactly, the Green’s Functions approach has proved to be relatively robust to
 420 these approximations and was successfully used to adjust highly non-linear
 421 systems, for example, an ocean circulation model (Menemenlis et al., 2005a),
 422 a coupled ocean and sea ice model (Nguyen et al., 2011), and coupled ocean-
 423 atmosphere model (Sugiura et al., 2008). We next discuss the application of
 424 this methodology to the ECCO2-Darwin ocean biogeochemistry model.

425 *6.2. Adjustment of ECCO2-Darwin Model*

426 One of the strengths of the Green’s Functions approach is that once the
 427 model sensitivity experiments have been computed, it is very easy and cheap
 428 to experiment with a large number of choices for the cost function, a very
 429 useful step as is discussed in detail in Menemenlis et al. (2005a). After some
 430 experimentation, we used the following data constraints ($\mathbf{y}^o$ in Eqs. 4, 5,
 431 and 6): (1) surface $p\text{CO}_2$ data for 2009 and 2010 from the “Global Surface
 432 $p\text{CO}_2$ (LDEO) Database” at the Carbon Dioxide Information Analysis Center
 433 (Takahashi et al., 2011) as described in Section 5, (2) a global mean air-sea
 434 CO_2 flux of 2.4 PgC yr^{-1} for 2010 from the Global Carbon Project (2011),
 435 and (3) the Takahashi et al. (2009) Atlas for the seasonal cycle of $p\text{CO}_2$
 436 after removing area-weighted monthly means from the original values. The
 437 standard errors (the square root of the diagonal elements of $\mathbf{R}$ in Eqs. 3,
 438 8, and 9) for the $p\text{CO}_2$ data and seasonal cycle atlas were set equal to the
 439 root-mean-square (RMS) difference between data and the BASELINE run.
 440 The error for the global mean air-sea CO_2 flux was set to 0.01 PgC yr^{-1} .
 441 It was assumed that nothing is known a-priori about the model errors, that
 442 is $\mathbf{Q}^{-1} = \mathbf{0}$ in Eqs. 3 and 9. One baseline (GLODAP) and six sensitivity

443 experiments (CCSM, KS, BLEND, NOBM, PISVEL, and PICPOC) were
 444 used to construct the data kernel matrix ($\mathbf{G}$ in Eqs. 6, 8, and 9).

445 The cost (J in Eq. 3) per observation of each of these runs, for the data
 446 sets and weights just discussed, is shown in Tab. 2. The optimized set of
 447 model parameters ($\boldsymbol{\eta}^a$ in Eq. 8) indicate the linear combination of the baseline
 448 and sensitivity experiments that minimizes cost function J . The cost per
 449 observation of this optimized linear combination of experiments is indicated
 450 in the second-to-last row of Tab. 2; it is reduced by approximately 20%
 451 relative to the KS simulation, which had the lowest overall cost prior to
 452 assimilation, but with too strong global carbon uptake ($-3.21 \text{ PgC yr}^{-1}$)
 453 relative to the Global Carbon Project (2011) estimate. The last step is to
 454 re-run the OBGCM model using the optimized initial conditions (a linear
 455 combination of GLODAP, CCSM, KS, BLEND, and NOBM) and optimized
 456 values for the PISVEL and PICPOC parameters. We call this optimized
 457 simulation “VERSION 2”, and its characteristics are described in Tables 1
 458 and 2. Notice that its cost per observation differs by only 2% relative to
 459 the cost of the optimized linear combination, 0.824 vs 0.808 in Tab. 2, hence
 460 verifying that the linearization assumption is largely satisfied.

461 7. VERSION 2: An Optimized Solution

462 The VERSION 2 model simulation yielded substantial cost (that is, data-
 463 model mismatch) reductions. While the cost per observation of the sensitivity
 464 experiments varied between 1.04 and 2.19, the VERSION 2 simulation has a
 465 cost per observation of 0.824 (see Tab. 2). This cost improvement is reflected
 466 in the $p\text{CO}_2$ scatter plot of Fig. 7a. While there is still considerable scatter in

the optimized solution (black dots) compared to V1 (red dots), the VERSION 2 solution is closer to the diagonal, as captured by the scatter plot density contour lines. The RMS difference between model and $p\text{CO}_2$ observations is 63.4 ppm for the BASELINE and 58.5 ppm for the VERSION 2 simulation.

To verify this improvement with independent data we choose to use *DIC* and alkalinity measurements for 2009 and 2010 from several cruises obtained from the CLIVAR & Carbon Hydrographic Data Office (CCHDO)⁴, as well as time-series data from Station ALOHA/Hawaii Ocean Time series program (HOT) in the North Pacific subtropical gyre (Karl and Lukas, 1996)⁵, Station "S"/Bermuda Atlantic Time series Site (BATS) in the North Atlantic subtropical gyre (Michaels and Knap, 1996)⁶, and data from Station Munda and surroundings in Sub-Antarctic Surface Water off the coast of New Zealand (Currie et al., 2009)⁷. These data sets cover the entire depth of the water column as compared to the surface $p\text{CO}_2$ data we used to constrain our solutions. Both *DIC* (not shown) and alkalinity (Fig. 7b) show a reduced model-data mismatch. For *DIC* the model-data RMS difference is 322.2 $\mu\text{mol/kg}$ for the BASELINE and 322.1 $\mu\text{mol/kg}$ for the VERSION 2 simulations. For alkalinity the RMS differences are 48.8 $\mu\text{mol/kg}$ and 28.9 $\mu\text{mol/kg}$ for BASELINE and VERSION 2, respectively.

The global mean oceanic CO_2 uptake for 2010 that was between 0.62 and 4.18 PgC yr^{-1} in the sensitivity runs, has a value of 2.54 PgC yr^{-1}

⁴through their website at <http://www.cchdo.ucsd.edu>

⁵<http://hahana.soest.hawaii.edu/hot/hot-dogs/interface.html>

⁶<http://bats.bios.edu/>

⁷Kim Currie, New Zealand Institute of Water and Atmospheric Research, personal communication

488 in the optimized run (Tab. 1). Note that we use a global mean oceanic
489 CO₂ constraint of 2.4 ± 0.01 PgC yr⁻¹, which forces this global mean bias
490 adjustment.

491 Another objective of the Green's Function optimization was to reduce
492 model drift. The Ocean Model Intercomparison Project (OCMIP, Orr et al.,
493 1999, and updates) has set standards for model drift. For air-sea carbon
494 fluxes, for instance, they require models in equilibrium runs to drift not
495 more than 0.01 PgC yr⁻¹. As our runs are not in a pre-industrial but a re-
496 cent environment our carbon values have to drift by definition. We therefore
497 choose to determine model drift from oxygen and nitrate as these can be
498 expected not to shift substantially over time. For this comparison we per-
499 formed additional model runs repeating the forcing for the year 2009 seven
500 times starting from the BASELINE and the VERSION 2 (optimized) 2009
501 initial conditions. The resulting, volume-weighted drift for the top 300 m of
502 the global ocean is shown in Figure 8. For nitrate, the drift has been reduced
503 by up to 12.5%, with the strongest drift reduction in the upper part of the
504 water column, while we found the strongest drift reduction of more than 30%
505 for oxygen in depths between 100 and 250 m. The drift reduction in depths
506 below 300 m were marginal for both quantities.

507 In typical model adjustment scenarios, it is relatively easy to reduce model
508 bias at the expense of increased model drift or, alternatively, to reduce model
509 drift at the expense of increased model bias. Despite the small number of
510 control variables that have been used in this Green's Function optimization,
511 the combined bias and drift reductions as well as the large reductions of
512 model-data differences relative to observations not used in the optimization

513 demonstrate that the approach is working.

514 The corresponding annual cycles of globally integrated air-sea CO₂ fluxes
515 for run VERSION 2 are shown as the large black circles in Fig. 2. The
516 optimization has led to values that are located between the extremes of the
517 sensitivity runs, but show the same overall structure in the seasonal cycle,
518 that is, a strong seasonal amplitude and high day-to-day variability. The
519 regional contributions to the seasonal cycle of globally integrated fluxes in
520 Figure 9 help interpret the seasonal cycle of the overall CO₂ uptake and
521 release.

522 Strong uptake in the Southern Ocean during austral spring and summer
523 dominates the global budget. The annual cycle there is mainly driven by
524 variations in vertical and horizontal transport processes as well as biological
525 production. Temperature differences between summer and winter are not
526 very pronounced (on average about 4°C), hence their influence on solubil-
527 ity and air-sea fluxes is limited. This globally integrated uptake maximum
528 during December and January is intensified by winter-time uptake in the
529 mid-latitudes of the northern hemisphere that is only slightly counteracted
530 by outgassing in the southern mid-latitudes. The equal phasing between the
531 mid-latitudes of the northern hemisphere and the Southern Ocean can also be
532 seen in boreal summer/austral winter when the combined (weak) outgassing
533 of these regions more than compensates for the slight uptake of the southern
534 hemispheric mid-latitudes. This interplay hence explains the seasonal cycles
535 of Fig. 2.

536 The spatial maps showing the VERSION 2 optimization for surface *DIC*
537 and the monthly mean *p*CO₂ are in the lower right plots of Figs. 3, 4, and

538 5, respectively. In general, the overall features of the biogeochemical fields
539 look realistic. Unrealistic features of single realizations (for instance, the
540 entire North Atlantic being a carbon sink in the CCSM run in January 2010)
541 have been ameliorated in the optimized run. The July 2009 CO₂ fluxes are
542 in closer agreement with the Takahashi et al. (2009) Atlas than our January
543 2010 results (that is, the root mean square difference of the monthly means for
544 July is 2.85 $gC/m^2/month$, compared to 4.69 $gC/m^2/month$ for January).
545 Features like the strength of the Southern Ocean uptake and the almost
546 complete lack of outgassing in the northernmost Pacific are characteristics of
547 our solutions that need to be explored in more detail.

548 Major differences between the regional time-series of our model run and
549 the Takahashi et al. (2009) Atlas in Fig. 9 exist in the Equatorial Pacific, the
550 North Atlantic and Pacific, and the Southern Ocean. These are reflected in
551 the comparison of the VERSION 2 run’s annual mean CO₂ flux map with
552 the annual mean Takahashi climatology in Fig. 10 and will be discussed in
553 the following section.

554 The air-sea fluxes from this VERSION 2 run have been published to the
555 research community (v2.0 and v2.1, differing only in the method of interpo-
556 lation to a lat-lon grid)⁸.

557 8. Comparison with Other Studies

558 To further assess the adjusted ECCO2-Darwin simulation, we compare
559 the VERSION 2 simulation to other studies. Since other studies have mostly

⁸<http://cmsflux.jpl.nasa.gov/>

560 reported annual mean results, we refer to our annual mean air-sea CO₂ flux
 561 in Fig. 10. The ocean inversion study by Gruber et al. (2009) shows only
 562 slight discrepancies between the ocean inversion estimate of uptake in the
 563 Southern Ocean (defined as south of 44°S) and the Takahashi et al. (2009)
 564 Atlas with regard to the overall uptake values, but a very different spatial
 565 structure. In the atlas, CO₂ is taken up near the continent (very weak to
 566 neutral in our VERSION 2 run), and released into the atmosphere north of
 567 the Antarctic Polar Front (strong outgassing in VERSION 2). The values in
 568 the ocean inversion for the uptake in that area are approximately between
 569 0.5 and 3 $\text{molCm}^{-2}\text{yr}^{-1}$ (nominal for 1995), our model produces values be-
 570 tween an outgassing of up to 3 $\text{molCm}^{-2}\text{yr}^{-1}$ in the Eastern Pacific Sector
 571 of the Southern Ocean to an uptake of up to 7 $\text{molCm}^{-2}\text{yr}^{-1}$ off the coast
 572 of Argentina. These overall higher values appear to be more in line with the
 573 majority of the OCMIP-2 ocean forward models (Watson and Orr, 2003) that
 574 are referred to in Gruber et al. (2009) with a decidedly higher uptake in the
 575 Southern Ocean. The inverse calculations of Mikaloff Fletcher et al. (2006)
 576 find a strongly elevated uptake of anthropogenic CO₂ south of 44°S compared
 577 to results from model forward calculations. Other recent model results from
 578 the Community Climate System Model (CCSM-3) ocean Biogeochemical El-
 579 emental Cycle (BEC) model (Doney et al., 2009) and an isopycnic carbon
 580 cycle model (Assmann et al., 2010) show higher overall carbon uptake in the
 581 Southern Ocean compared to the Takahashi et al. (2009) Atlas. Although
 582 spatial patterns and amplitudes vary from model to model, there is general
 583 agreement that the atlas underestimates the carbon sink in the Southern
 584 Ocean. The carbon uptake of the ECCO2-Darwin solutions in the Southern

585 Ocean is on the high end compared to most other models and to the clima-
586 tology, causing an overly pronounced amplitude of the seasonal cycle of the
587 globally integrated CO₂ flux.

588 In other areas of the world ocean the ECCO2-Darwin solutions were more
589 consistent with other studies with differences only in magnitudes. The equa-
590 torial outgassing in the Eastern Pacific of run VERSION 2, for instance, has
591 a shape similar to Doney et al. (2009), the values are closer to those shown
592 by Assmann et al. (2010). The Northern North Pacific in our model does
593 not show the outgassing in the annual mean that is found in the other two
594 models as well as in the climatology.

595 As we constrained our model to achieve a “target” global carbon uptake of
596 about 2.4 to 2.5 PgC yr⁻¹, it needs to compensate for the too strong carbon
597 uptake in the Southern Ocean by weaker uptake or intensified outgassing
598 in other regions. Additional Green’s functions sensitivity experiments will
599 be required to “repair” these regional problems in representing air-sea CO₂
600 fluxes. Preliminary conclusions hint to the model treatment of nutrient fields
601 in combination with the parameterization of biological production as the
602 most likely candidate for the overly strong Southern Ocean carbon uptake.

603 **9. Concluding Remarks**

604 This proof-of-concept study has shown that the Green’s function method
605 can help adjust OBGCM model parameters and reduce model-data mis-
606 match. The existing ECCO2-Darwin solution remains, however, preliminary
607 in many ways. Future investigations should be directed to address the fol-
608 lowing set of improvement possibilities.

609 First, we only used a very small set of observational data to constrain
610 the simulation. Adding data constraints from observed, full-depth profiles
611 of *DIC*, alkalinity, and oxygen and the utilization of remotely-sensed prod-
612 ucts like chlorophyll and other satellite-derived biological or biogeochemical
613 quantities will vastly expand the number of data constraints.

614 Second, we used a very small number of control parameters; only seven
615 model sensitivity experiments (or Green's Functions) were used in the opti-
616 mization. Forward-model sensitivity experiments varying additional model
617 parameters will lead to additional control parameters for the optimization.
618 For example, we are exploring the creation of new sets of biogeochemical
619 initial conditions that will expand the control parameter space. One specific
620 set of initial conditions that we are taking into consideration is the construc-
621 tion of self-consistent data sets of *DIC* alkalinity, and density. We plan
622 to enforce this self-consistency by assuming that the biogeochemical fields
623 correlate with density and by estimating their variation through regression.
624 This method has been used successfully in regional ocean modeling studies
625 (Gruber et al., 2012; Lachkar and Gruber, 2012).

626 Third, the model parameterizations, especially with regard to the choice
627 of phytoplankton species and air-sea exchange piston velocity, can be evalu-
628 ated in greater detail and adjusted to observations. In particular, there are
629 strong indications that the Wanninkhof (1992) quadratic dependence of air-
630 sea flux on the wind speed may be exaggerating spatio-temporal variability
631 of ECCO2-Darwin air-sea gas exchange. For this reason, we are exploring a
632 piston velocity formulation that is closer to linear, e.g., as in Krakauer et al.
633 (2006).

634 Lastly, a limitation of the Green’s function approach discussed herein
 635 is that computational cost increases linearly with the number of control pa-
 636 rameters. Using this approach for detailed regional adjustments in the global
 637 simulation would be impractical. For this reason, we are also exploring the
 638 application of the adjoint method to the ECCO2-Darwin model. This strat-
 639 egy is similar to that adopted by the ECCO2 project, whereby preliminary
 640 model adjustments were carried out using a Green’s function approach while
 641 the adjoint method was subsequently used to further fine tune the solution.
 642 The cost of the adjoint method, while substantial, is largely independent
 643 from the number of control parameters; hence it possible to simultaneously
 644 adjust a large number of control parameters. For example, in the ECCO2
 645 physical-ocean optimizations, some two billion model parameters have been
 646 adjusted using the adjoint method. For ECCO2-Darwin, we plan to adjust
 647 the initial and surface biogeochemical boundary conditions in order to fit the
 648 available observations.

649 Despite the caveats listed above and the preliminary nature of the ECCO2-
 650 Darwin simulation, the present study nevertheless remains an important first
 651 step towards more realistic air-sea CO₂ exchange estimates for the NASA
 652 CMS Flux project as well as for the underlying physical, biological, and
 653 chemical processes that drive these air-sea fluxes. Of particular note is the
 654 fact that Green’s Function optimization was able to reduce the drift of the
 655 optimization (Fig. 8) while also substantially reducing the model data differ-
 656 ences both vs observations used as constraints (Table 2 and Fig. 7a) as well
 657 as independent observations (e.g., Figure 7b).

658 Acknowledgments

659 We wish to thank Watson Gregg for providing data from his NOBM runs
660 and Corinne Le Quéré for making available her method to calculate *DIC*
661 initial conditions corrected for the anthropogenic component. This work was
662 funded by the NASA CMS Flux Project and made use of NASA High-End
663 Computing (HEC) resources.

664 Appendix A: Description of Sensitivity Experiments

665 The initial conditions that distinguish the sensitivity experiments were
666 created by interpolating the data sets described below to the ECCO2 cube
667 sphere grid. An overview of the parameterizations and initial conditions used
668 in the integrations is given in Tab. 1.

669 *CCSM experiment.* The CCSM experiment is identical to V1 except that
670 we use two complete years, 2009 and 2010, of ECCO2 physical ocean forc-
671 ing. For V1, only 16 months of ECCO2 forcing was available so we padded
672 the last 8 months of 2010 with JRA-25 forcing. The initial conditions for
673 *DIC*, alkalinity, and oxygen were taken from a realization of the Community
674 Climate System Model (CCSM-3) ocean Biogeochemical Elemental Cycle
675 (BEC) model (Lovenduski et al., 2007; Doney et al., 2009). More details can
676 be found in the description of V1 in Section 3.1.

677 *GLODAP (BASELINE) experiment.* For the GLODAP experiment the ini-
678 tial conditions for *DIC* and alkalinity were taken from the Global Ocean
679 Data Analysis Project (GLODAP) annual mean climatology provided by the
680 Carbon Dioxide Information Analysis Center (CDIAC; Key et al., 2004).

681 These GLODAP data represent measurements taken in the 1990's. We nev-
682 ertheless used them unaltered, that is, without anthropogenic correction, to
683 create a contrast to the KS experiment, for which such a correction was ap-
684 plied. The oxygen values were derived from the World Ocean Atlas 2005
685 (WOA05; Garcia et al., 2006a) by taking the January values where available,
686 that is, for the top 1500m, and the annual values for deeper depths. This
687 model realization will also be referred to as "BASELINE".

688 *KS experiment.* Alkalinity and oxygen initial conditions for the Key & Sabine
689 (KS) experiment were identical to the ones used for GLODAP. For *DIC* the
690 CDIAC data set was blended with an anthropogenic correction using the
691 data set of Sabine et al. (2004). The values for initial conditions for January
692 2009 were calculated according to the method described in Le Quéré et al.
693 (2010).

694 *BLEND experiment.* For the BLEND experiment, *DIC*, alkalinity, and oxy-
695 gen were blended from the initial conditions of the three previous experiments
696 based on subjective criteria to obtain realistic-looking fields of *DIC*, alkalin-
697 ity, and oxygen. The fields north of 40°N were identical to CCSM, between
698 35°N and 15°N to GLODAP, and south of 10°N to KS. The areas between
699 these regions (from 40°N to 35°N and 15°N to 10°N) were linearly interpo-
700 lated in between the data sets. In addition, *DIC* values in Antarctic Bottom
701 Water were increased by 10 $\mu\text{mol kg}^{-1}$.

702 *NOBM experiment.* Values for the initial conditions of *DIC* and dissolved
703 organic carbon (*DOC*) were taken from a realization of the NASA Ocean
704 Biogeochemical Model (NOBM; Gregg et al., 2003; Gregg and Casey, 2007,

2009) provided by W. Gregg, NASA Goddard Space Flight Center. As this model does not carry alkalinity and oxygen, these quantities were initialized with the data sets used in the BLEND run.

PISVEL experiment. The PISVEL experiment was designed to identify the sensitivity of air-sea CO₂ fluxes to air-sea gas exchange piston velocity (PISVEL) parameterization. To this end we modified the air-sea gas exchange piston velocity from a value of 0.337 used in the other simulations to the original value of 0.31 used by Wanninkhof (1992). This experiment was started in 2009 using initial conditions from the BASELINE experiment.

PICPOC experiment. The PICPOC experiment takes into account the dependency of air-sea CO₂ fluxes on alkalinity. One big driver of alkalinity changes is the cycling of particulate inorganic carbon (PIC), which is formed as calcium carbonate (CaCO₃) in shells in some phytoplankton species and subsequently dissolved at depth. The amount of PIC produced by these species relative to particulate organic carbon (POC) is the parameter we choose to manipulate from 0.04 in the BASELINE experiment to 0.1 for the PICPOC sensitivity experiment. For an overview of the variability of this ratio for a variety of ocean regions see Sarmiento et al. (2002). The PICPOC experiment was also started in 2009 using initial conditions from the BASELINE experiment.

725 Adcroft, A., Campin, J., Hill, C., Marshall, J., 2004. Implementation of an
726 Atmosphere-Ocean general circulation model on the expanded spherical
727 cube. *Monthly Weather Review* 132, 2845–2863.

728 Assmann, K.M., Bentsen, M., Segschneider, J., Heinze, C., 2010. An isopy-
729 cnic ocean carbon cycle model. *Geoscientific Model Development* 3, 143–
730 167.

731 Aumont, O., Bopp, L., 2006. Globalizing results from ocean in situ iron
732 fertilization studies. *Global Biogeochemical Cycles* 20, GB2017, 15 pp.

733 Baker, D.F., Law, R.M., Gurney, K.R., Rayner, P., Peylin, P., Denning, A.S.,
734 Bousquet, P., Bruhwiler, L., Chen, Y., Ciais, P., Fung, I.Y., Heimann,
735 M., John, J., Maki, T., Maksyutov, S., Masarie, K., Prather, M., Pak,
736 B., Taguchi, S., Zhu, Z., 2006. TransCom 3 inversion intercomparison:
737 Impact of transport model errors on the interannual variability of regional
738 CO₂ fluxes, 1988-2003. *Global Biogeochemical Cycles* 20, GB1002, 17 pp.

739 Battle, M., Bender, M.L., Tans, P.P., White, J.W.C., Ellis, J.T., Conway,
740 T., Francey, R.J., 2000. Global carbon sinks and their variability inferred
741 from atmospheric O₂ and $\delta^{13}\text{C}$. *Science* 287, 2467–2470.

742 Blackmon, M., Boville, B., Bryan, F., Dickinson, R., Gent, P., Kiehl, J.,
743 Moritz, R., Randall, D., Shukla, J., Solomon, S., Bonan, G., Doney, S.,
744 Fung, I., Hack, J., Hunke, E., Hurrell, J., Kutzbach, J., Meehl, J., Otto-
745 Bliesner, B., Saravanan, R., Schneider, E.K., Sloan, L., Spall, M., Taylor,
746 K., Tribbia, J., Washington, W., 2001. The Community Climate System
747 Model. *Bull. Am. Meteorol. Soc.* 82, 2357–2376.

- 748 Bousquet, P., Peylin, P., Ciais, P., Le Quéré, C., Friedlingstein, P., Tans,
749 P.P., 2000. Regional changes in carbon dioxide fluxes of land and oceans
750 since 1980. *Science* 290, 1342–1346.
- 751 Buitenhuis, E.T., Rivkin, R.B., Sailley, S., Le Quéré, C., 2010. Biogeochem-
752 ical fluxes through microzooplankton. *Global Biogeochemical Cycles* 24,
753 GB4015, 16 pp.
- 754 Conway, T.J., Tans, P.P., Waterman, L.S., Thoning, K.W., Kitzis, D.R.,
755 Masarie, K.A., Zhang, N., 1994. Evidence for interannual variability of
756 the carbon cycle from the National Oceanic and Atmospheric Administra-
757 tion/Climate Monitoring and Diagnostics Laboratory global air sampling
758 network. *J. Geophys. Res.* 99, 22831–22855.
- 759 Currie, K.I., Reid, M.R., Hunter, K.A., 2009. Interannual variability of
760 carbon dioxide drawdown by subantarctic surface water near New Zealand.
761 *Biogeochemistry* 104, 23–34.
- 762 Doney, S.C., Lima, I., Feely, R.A., Glover, D.M., Lindsay, K., Mahowald, N.,
763 Moore, J.K., Wanninkhof, R., 2009. Mechanisms governing interannual
764 variability in upper-ocean inorganic carbon system and air-sea CO₂ fluxes:
765 Physical climate and atmospheric dust. *Deep Sea Research Part II: Topical*
766 *Studies in Oceanography* 56, 640–655.
- 767 Dutkiewicz, S., Follows, M.J., Bragg, J.G., 2009. Modeling the coupling
768 of ocean ecology and biogeochemistry. *Global Biogeochemical Cycles* 23,
769 GB4017.

- 770 Dutkiewicz, S., Scott, J., Follows, M., 2012. Winners and losers: Ecological
771 and biogeochemical changes in a warming ocean. *Global Biogeochemical*
772 *Cycles* , submitted.
- 773 Enting, I.G., Mansbridge, J.V., 1989. Seasonal sources and sinks of atmo-
774 spheric CO₂ direct inversion of filtered data. *Tellus B* 41B, 111–126.
- 775 Follows, M.J., Dutkiewicz, S., 2011. Modeling diverse communities of marine
776 microbes. *Annual Review of Marine Science* 3, 427–451.
- 777 Follows, M.J., Dutkiewicz, S., Grant, S., Chisholm, S.W., 2007. Emergent
778 biogeography of microbial communities in a model ocean. *Science* 315,
779 1843–1846.
- 780 Follows, M.J., Ito, T., Dutkiewicz, S., 2006. On the solution of the carbonate
781 chemistry system in ocean biogeochemistry models. *Ocean Modelling* 12,
782 290–301.
- 783 Galbraith, E.D., Gnanadesikan, A., Dunne, J.P., Hiscock, M.R., 2010. Re-
784 gional impacts of iron-light colimitation in a global biogeochemical model.
785 *Biogeosciences* 7, 1043–1064.
- 786 Garcia, H.E., Locarnini, R., Boyer, T., Antonov, J., 2006a. *World Ocean*
787 *Atlas 2005, Volume 3: Dissolved Oxygen, Apparent Oxygen Utilization,*
788 *and Oxygen Saturation. Technical Report NOAA Atlas NESDIS 64. U.S.*
789 *Government Printing Office. Washington, D.C. 342 pp.*
- 790 Garcia, H.E., Locarnini, R., Boyer, T., Antonov, J., 2006b. *World Ocean*
791 *Atlas 2005, Volume 4: Nutrients (phosphate, nitrate, silicate). Techni-*

792 cal Report NOAA Atlas NESDIS 64. U.S. Government Printing Office.
 793 Washington, D.C. 396 pp.

794 Global Carbon Project, 2011. Carbon budget and trends 2010. [http://www.](http://www.globalcarbonproject.org/carbonbudget)
 795 [globalcarbonproject.org/carbonbudget](http://www.globalcarbonproject.org/carbonbudget).

796 Gloor, M., Gruber, N., Sarmiento, J.L., Sabine, C.L., Feely, R.A.,
 797 Rödenbeck, C., 2003. A first estimate of present and pre-industrial air-
 798 sea CO₂ fluxes patterns based on ocean interior carbon measurements and
 799 models. *Geophys. Res. Lett.* 30, 1010.

800 Gregg, W.W., Casey, N., 2007. Modeling coccolithophores in the global
 801 oceans. *Deep-Sea Res. II* 54, 447–477.

802 Gregg, W.W., Casey, N., 2009. Skill assessment of a spectral ocean-
 803 atmosphere radiative model. *Journal of Marine Systems* 76, 49–63.

804 Gregg, W.W., Ginoux, P., Schopf, P.S., Casey, N., 2003. Phytoplankton
 805 and iron: Validation of a global three-dimensional ocean biogeochemical
 806 model. *Deep-Sea Res. II* 50, 3143–3169.

807 Gruber, N., Gloor, M., Mikaloff Fletcher, S.E., Doney, S.C., Dutkiewicz, S.,
 808 Follows, M.J., Gerber, M., Jacobson, A.R., Joos, F., Lindsay, K., Men-
 809 emenlis, D., Mouchet, A., Mller, S.A., Sarmiento, J.L., Takahashi, T.,
 810 2009. Oceanic sources, sinks, and transport of atmospheric CO₂. *Global*
 811 *Biogeochemical Cycles* 23, GB1005.

812 Gruber, N., Hauri, C., Lachkar, Z., Loher, D., Frölicher, T.L., Plattner, G.,
 813 2012. Rapid progression of ocean acidification in the california current
 814 system. *Science* .

815 Gurney, K.R., Law, R.M., Denning, A.S., Rayner, P.J., Baker, D., Bousquet,
816 P., Bruhwiler, L., Chen, Y.H., Ciais, P., Fan, S., Fung, I.Y., Gloor, M.,
817 Heimann, M., Higuchi, K., John, J., Maki, T., Maksyutov, S., Masarie, K.,
818 Peylin, P., Prather, M., Pak, B.C., Randerson, J., Sarmiento, J., Taguchi,
819 S., Takahashi, T., Yuen, C.W., 2002. Towards robust regional estimates of
820 CO₂ sources and sinks using atmospheric transport models. *Nature* 415,
821 626–630.

822 Heimbach, P., Menemenlis, D., Losch, M., Campin, J., Hill, C., 2010. On the
823 formulation of sea-ice models. part 2: Lessons from multi-year adjoint sea-
824 ice export sensitivities through the Canadian Arctic Archipelago. *Ocean*
825 *Modelling* 33, 145–158.

826 Jacobson, A.R., Mikaloff Fletcher, S.E., Gruber, N., Sarmiento, J.L., Gloor,
827 M., 2007a. A joint atmosphere-ocean inversion for surface fluxes of carbon
828 dioxide: 1. Methods and global-scale fluxes. *Global Biogeochem. Cycles*
829 21, GB1019.

830 Jacobson, A.R., Mikaloff Fletcher, S.E., Gruber, N., Sarmiento, J.L., Gloor,
831 M., 2007b. A joint atmosphere-ocean inversion for surface fluxes of carbon
832 dioxide: 2. Regional results. *Global Biogeochem. Cycles* 21, GB1020.

833 Karl, D.M., Lukas, R., 1996. The Hawaii Ocean Time-series (HOT) program:
834 Background, rationale and field implementation. *Deep-Sea Res. II* 43, 129–
835 156.

836 Keeling, R.F., Piper, S.C., Heimann, M., 1996. Global and hemispheric CO₂

837 sinks deduced from changes in atmospheric oxygen concentration. *Nature*
838 381, 218–221.

839 Key, R.M., Kozyr, A., Sabine, C.L., Lee, K., Wanninkhof, R., Bullister,
840 J.L., Feely, R.A., Millero, F.J., Mordy, C., Peng, T., 2004. A global ocean
841 carbon climatology: Results from global data analysis project (GLODAP).
842 *Global Biogeochemical Cycles* 18, GB4031, 23 pp.

843 Khatiwala, S., Primeau, F., Hall, T., 2009. Reconstruction of the history of
844 anthropogenic CO₂ concentrations in the ocean. *Nature* 462, 346–349.

845 Krakauer, N.Y., Randerson, J.T., Primeau, F., Gruber, N., Menemenlis, D.,
846 2006. Carbon isotope evidence for the latitudinal distribution and wind
847 speed dependence of the air-sea gas transfer velocity. *Tellus B* 58B, 390–
848 417.

849 Kuze, A., Suto, H., Nakajima, M., Hamazaki, T., 2009. Thermal and near
850 infrared sensor for carbon observation fourier-transform spectrometer on
851 the greenhouse gases observing satellite for greenhouse gases monitoring.
852 *Applied Optics* 48, 6716–6733.

853 Lachkar, Z., Gruber, N., 2012. Response of biological production and air-
854 sea CO₂ fluxes to upwelling intensification in the California and Canary
855 Current Systems. *Journal of Marine Systems* in press.

856 Le Quéré, C., Raupach, M.R., Canadell, J.G., Marland, G., Bopp, L., Ciais,
857 P., Conway, T.J., Doney, S.C., Feely, R.A., Foster, P., Friedlingstein, P.,
858 Gurney, K., Houghton, R.A., House, J.I., Huntingford, C., Levy, P.E.,
859 Lomas, M.R., Majkut, J., Metzl, N., Ometto, J.P., Peters, G.P., Prentice,

860 I.C., Randerson, J.T., Running, S.W., Sarmiento, J.L., Schuster, U., Sitch,
 861 S., Takahashi, T., Viovy, N., van der Werf, G.R., Woodward, F.I., 2009.
 862 Trends in the sources and sinks of carbon dioxide. *Nature Geoscience* 2,
 863 831–836.

864 Le Quéré, C., Rödenbeck, C., Buitenhuis, E.T., Conway, T.J., Langenfelds,
 865 R., Gomez, A., Labuschagne, C., Ramonet, M., Nakazawa, T., Metzl, N.,
 866 Gillett, N., Heimann, M., 2007. Saturation of the Southern Ocean CO₂
 867 sink due to recent climate change. *Science* 316, 1735–1738.

868 Le Quéré, C., Takahashi, T., Buitenhuis, E.T., Rödenbeck, C., Sutherland,
 869 S.C., 2010. Impact of climate change and variability on the global oceanic
 870 sink of CO₂. *Global Biogeochemical Cycles* 24, GB4007, 10 pp.

871 Lenton, A., Metzl, N., Takahashi, T., Kuchinke, M., Matear, R.J., Roy, T.,
 872 Sutherland, S.C., Sweeney, C., Tilbrook, B., 2012. The observed evolu-
 873 tion of oceanic pCO₂ and its drivers over the last two decades. *Global*
 874 *Biogeochemical Cycles* 26, GB2021, 14 pp.

875 Liu, J., Bowman, K.W., Lee, M., Henze, D.K., Bousserez, N., Brix, H.,
 876 Collatz, G.J., Menemenlis, D., Ott, L., Pawson, S., Jones, D., Nassar, R.,
 877 2014. Carbon monitoring system flux estimation and attribution: impact
 878 of ACOS-GOSAT XCO₂ sampling on the inference of terrestrial biospheric
 879 sources and sinks. *Tellus B* 66, 22486.

880 Losch, M., Menemenlis, D., Campin, J., Heimbach, P., Hill, C., 2010. On
 881 the formulation of sea-ice models. part 1: Effects of different solver imple-
 882 mentations and parameterizations. *Ocean Modelling* 33, 129–144.

- 883 Lovenduski, N.S., Gruber, N., Doney, S.C., Lima, I.D., 2007. Enhanced CO₂
884 outgassing in the Southern Ocean from a positive phase of the Southern
885 Annular Mode. *Global Biogeochem. Cycles* 21, GB2026.
- 886 Mahowald, N.M., Yoshioka, M., Collins, W.D., Conley, A.J., Fillmore, D.W.,
887 Coleman, D.B., 2006. Climate response and radiative forcing from min-
888 eral aerosols during the last glacial maximum, pre-industrial, current and
889 doubled-carbon dioxide climates. *Geophysical Research Letters* 33, L20705,
890 4 pp.
- 891 Marshall, J., Adcroft, A., Hill, C., Perelman, L., Heisey, C., 1997a. A finite-
892 volume, incompressible navier stokes model for studies of the ocean on
893 parallel computers. *Journal of Geophysical Research* 102, 5753–5766.
- 894 Marshall, J., Hill, C., Perelman, L., Adcroft, A., 1997b. Hydrostatic, quasi-
895 hydrostatic, and nonhydrostatic ocean modeling. *Journal of Geophysical*
896 *Research* 102, 5733–5752.
- 897 McNeil, B.I., Matear, R.J., Key, R.M., Bullister, J.L., Sarmiento, J.L., 2003.
898 Anthropogenic CO₂ uptake by the ocean based on the global chlorofluoro-
899 carbon data set. *Science* 299, 235–239.
- 900 Menemenlis, D., Campin, J., Heimbach, P., Hill, C., Lee, T., Nguyen, A.,
901 Schodlock, M., Zhang, H., 2008. ECCO2: High resolution global ocean
902 and sea ice data synthesis. *Mercator Ocean Quarterly Newsletter* 31, 13–
903 21.
- 904 Menemenlis, D., Fukumori, I., Lee, T., 2005a. Using green’s functions to

905 calibrate an ocean general circulation model. *Monthly Weather Review*
906 133, 1224–1240.

907 Menemenlis, D., Hill, C., Adcroft, A., Campin, J., Cheng, B., Ciotti, B.,
908 Fukumori, I., Heimbach, P., Henze, C., Khl, A., Lee, T., Stammer, D.,
909 Taft, J., Zhang, J., 2005b. NASA supercomputer improves prospects for
910 ocean climate research. *EOS* 86, 89, 96.

911 Menemenlis, D., Wunsch, C., 1997. Linearization of an oceanic general cir-
912 culation model for data assimilation and climate studies. *Journal of At-*
913 *mospheric and Oceanic Technology* 14, 1420–1443.

914 Michaels, A.F., Knap, A.H., 1996. Overview of the U.S. JGOFS Bermuda
915 Atlantic Time-series Study and the Hydrostation S program. *Deep-Sea*
916 *Res. II* 43, 157–198.

917 Mikaloff Fletcher, S.E., Gruber, N., Jacobson, A.R., Doney, S.C., Dutkiewicz,
918 S., Gerber, M., Follows, M., Joos, F., Lindsay, K., Menemenlis, D.,
919 Mouchet, A., Müller, S.A., Sarmiento, J.L., 2006. Inverse estimates of
920 anthropogenic CO₂ uptake, transport, and storage by the ocean. *Global*
921 *Biogeochem. Cycles* 20, GB2002.

922 Mikaloff Fletcher, S.E., Gruber, N., Jacobson, A.R., Gloor, M., Doney, S.C.,
923 Dutkiewicz, S., Gerber, M., Follows, M., Joos, F., Lindsay, K., Menemen-
924 lis, D., Mouchet, A., Miller, S.A., Sarmiento, J.L., 2007. Inverse estimates
925 of the oceanic sources and sinks of natural CO₂ and the implied oceanic
926 carbon transport. *Global Biogeochem. Cycles* 21, GB1010.

927 Nassar, R., Jones, D.B.A., Suntharalingam, P., Chen, J.M., Andres, R.J.,
 928 Wecht, K.J., Yantosca, R.M., Kulawik, S.S., Bowman, K.W., Worden,
 929 J.R., Machida, T., Matsueda, H., 2010. Modeling global atmospheric CO₂
 930 with improved emission inventories and CO₂ production from the oxida-
 931 tion of other carbon species. *Geoscientific Model Development* 3, 689–716.

932 Nguyen, A., Menemenlis, D., Kwok, R., 2011. Arctic ice-ocean simulation
 933 with optimized model parameters: approach and assessment. *J. Geophys.*
 934 *Res.* 116, C04025.

935 Onogi, K., Tsuitsui, J., Koide, H., Sakamoto, M., Kobayashi, S., Hat-
 936 sushika, H., Matsumoto, T., Yamazaki, N., Kamahori, H., Takahashi, K.,
 937 Kadokura, S., Wada, K., Kato, K., Oyama, R., Ose, T., Mannoji, N.,
 938 Taira, R., 2007. The JRA-25 reanalysis. *Journal of the Meteorological*
 939 *Society of Japan* 85, 369–432.

940 Orr, J., Najjar, R., Sabine, C., Joos, F., 1999. Abiotic-HOWTO. Internal
 941 OCMIP Report. LSCE/CEA Saclay, Gif-sur-Yvette, France.

942 Pfeil, B., Olsen, A., Bakker, D., et al., 2012. A uniform, quality controlled,
 943 Surface Ocean CO₂ Atlas (SOCAT). submitted to *Earth System Science*
 944 *Data* .

945 Rödenbeck, C., Houweling, S., Gloor, M., Heimann, M., 2003. CO₂ flux
 946 history 1982-2001 inferred from atmospheric data using a global inversion
 947 of atmospheric transport. *Atmos. Chem. Phys.* 3, 1919–1964.

948 Roy, T., Bopp, L., Gehlen, M., Schneider, B., Cadule, P., Frölicher, T.L.,
 949 Segschneider, J., Tjiputra, J., Heinze, C., Joos, F., 2011. Regional impacts

950 of climate change and atmospheric CO₂ on future ocean carbon uptake: A
 951 multimodel linear feedback analysis. *Journal of Climate* 24, 2300–2318.

952 Sabine, C.L., Feely, R.A., Gruber, N., Key, R.M., Lee, K., Bullister, J.L.,
 953 Wanninkhof, R., Wong, C.S., Wallace, D.W.R., Tilbrook, B., Millero, F.J.,
 954 Peng, T.H., Kozyr, A., Ono, T., Rios, A.F., 2004. The oceanic sink for
 955 anthropogenic CO₂. *Science* 305, 367–371.

956 Sarmiento, J.L., Dunne, J., Gnanadesikan, A., Key, R.M., Matsumoto, K.,
 957 Slater, R., 2002. A new estimate of the CaCO₃ to organic carbon export
 958 ratio. *Global Biogeochemical Cycles* 16, 1107, 12 pp.

959 Sarmiento, J.L., Gruber, N., 2006. *Ocean Biogeochemical Dynamics*. Prince-
 960 ton University Press.

961 Séférian, R., Iudicone, D., Bopp, L., Roy, T., Madec, G., 2012. Water mass
 962 analysis of effect of climate change on air-sea CO₂ fluxes: The Southern
 963 Ocean. *Journal of Climate* 25, 3894–3908.

964 Stammer, D., Wunsch, C., 1996. The determination of the large-scale cir-
 965 culation of the Pacific Ocean from satellite altimetry using model Green’s
 966 functions. *Journal of Geophysical Research* 101, 18,409–18,432.

967 Sugiura, N., Awaji, T., Masuda, S., Mochizuki, T., Toyoda, T., Miyama,
 968 T., Igarashi, H., Ishikawa, Y., 2008. Development of a four-dimensional
 969 variational coupled data assimilation system for enhanced analysis and
 970 prediction of seasonal to interannual climate variations. *Journal of Geo-
 971 physical Research* 113, C10017.

- 972 Takahashi, T., Sutherland, S., Kozyr, A., 2011. Global Ocean Surface Wa-
 973 ter Partial Pressure of CO₂ Database: Measurements Performed During
 974 1957-2010 (Version 2010). Technical Report ORNL/CDIAC-159, NDP-
 975 088(V2010). Carbon Dioxide Information Analysis Center, Oak Ridge Na-
 976 tional Laboratory, U.S. Department of Energy. Oak Ridge, Tennessee.
- 977 Takahashi, T., Sutherland, S.C., Sweeney, C., Poisson, A., Metzl, N.,
 978 Tilbrook, B., Bates, N., Wanninkhof, R., Feely, R.A., Sabine, C., Olaf-
 979 sson, J., Nojiri, Y., 2002. Global sea-air CO₂ flux based on climatologi-
 980 cal surface ocean pCO₂, and seasonal biological and temperature effects.
 981 Deep-Sea Res. II 49, 1601–1622.
- 982 Takahashi, T., Sutherland, S.C., Wanninkhof, R., Sweeney, C., Feely, R.A.,
 983 Chipman, D.W., Hales, B., Friederich, G., Chavez, F., Sabine, C., Wat-
 984 son, A., Bakker, D.C., Schuster, U., Metzl, N., Yoshikawa-Inoue, H., Ishii,
 985 M., Midorikawa, T., Nojiri, Y., Krtzinger, A., Steinhoff, T., Hoppema,
 986 M., Olafsson, J., Arnarson, T.S., Tilbrook, B., Johannessen, T., Olsen,
 987 A., Bellerby, R., Wong, C., Delille, B., Bates, N., de Baar, H.J., 2009.
 988 Climatological mean and decadal change in surface ocean pCO₂, and net
 989 sea-air CO₂ flux over the global oceans. Deep Sea Research Part II: Topical
 990 Studies in Oceanography 56, 554–577.
- 991 Tans, P.P., Fung, I.Y., Takahashi, T., 1990. Observational constraints on the
 992 global atmospheric CO₂ budget. Science 247, 1431–1438.
- 993 Thomas, H., Prowe, A.E.F., Lima, I.D., Doney, S.C., Wanninkhof, R., Great-
 994 batch, R.J., Schuster, U., Corbière, A., 2008. Changes in the North At-

- 995 lantic Oscillation influence CO₂ uptake in the North Atlantic over the past
996 2 decades. *Global Biogeochem. Cycles* 22, GB4027.
- 997 Toth, Z., Peña, M., 2007. Data assimilation and numerical forecasting with
998 imperfect models: The mapping paradigm. *Physica D: Nonlinear Phenom-*
999 *ena* 230, 146–158.
- 1000 Valsala, V., Maksyutov, S., 2010. Simulation and assimilation of global ocean
1001 pCO₂ and air-sea CO₂ fluxes using ship observations of surface ocean pCO₂
1002 in a simplified biogeochemical offline model. *Tellus B* 62, 821–840.
- 1003 Wanninkhof, R., 1992. Relationship between wind speed and gas exchange
1004 over the ocean. *J. Geophys. Res.* 97, 7373–7382.
- 1005 Watson, A.J., Orr, J.C., 2003. Carbon dioxide fluxes in the global ocean,
1006 in: Fasham, M.J.R. (Ed.), *Ocean Biogeochemistry: A JGOFS Synthesis*.
1007 Springer-Verlag, Berlin. chapter 5, pp. 123–141.
- 1008 While, J., Totterdell, I., Martin, M., 2012. Assimilation of pCO₂ data into a
1009 global coupled physical-biogeochemical ocean model. *Journal of Geophys-*
1010 *ical Research* 117, C03037, 11 pp.
- 1011 Wunsch, C., Heimbach, P., 2007. Practical global oceanic state estimation.
1012 *Physica D: Nonlinear Phenomena* 230, 197–208.

Table 1: Initial conditions and parameterizations used in sensitivity and optimized experiments plus air-sea CO₂ fluxes for 2010. **Gas Exchange Parameter:** Air-sea gas exchange is calculated following Wanninkhof (1992): Gas transfer velocity $k = fact\ u^2 (Sc/660)^{-0.5}$, with Sc being the Schmidt number of CO₂ in seawater and $fact$ the numerical value in this table; **CCSM:** values are from a realization of the Community Climate System Model (Blackmon et al., 2001; Lovenduski et al., 2007; Doney et al., 2009) in T62 resolution with historic atmospheric CO₂ forcing; **GLODAP:** Global Ocean Data Analysis Project, Carbon Dioxide Information Analysis Center (Key et al., 2004); **WOA05:** World Ocean Atlas (Garcia et al., 2006a); **GLODAP, Sabine:** *DIC* data from Global Ocean Data Analysis Project, Carbon Dioxide Information Analysis Center (Key et al., 2004) with anthropogenic correction of Sabine et al. (2004); values for initial conditions calculated according to method described in Le Quéré et al. (2010); **Blended:** *DIC*, *Alk*, and *O₂* blended from other runs: North of 40°N from CCSM, between 35°N and 15°N from GLODAP, South of 10°N from KS, with linear interpolation in between data sets from 40°N to 35°N and 15°N to 10°N. In addition, *DIC* values in Antarctic Bottom Water were increased by 10 $\mu mol\ kg^{-1}$; **NOBM:** Values taken from a realization of the NASA Ocean Biogeochemical Model (NOBM; Gregg et al., 2003; Gregg and Casey, 2007, 2009) provided by Watson Gregg, NASA Goddard Space Flight Center; **VERSION 2:** Multiplication factors for all biogeochemical initial conditions: $VERSION\ 2 = 0.56404 \cdot GLODAP + 0.16033 \cdot CCSM - 0.077899 \cdot KS + 0.10418 \cdot BLEND + 0.24935 \cdot NOBM$.

Run name	Initial Conditions			Gas Exchange	PIC/POC	Flux 2010	Remarks
	<i>DIC</i>	<i>Alk</i>	<i>O₂</i>	Parameter		[PgC/yr]	
V1	CCSM	CCSM	CCSM	0.337	0.04	-3.61	Published as version 1.0
GLODAP	GLODAP	GLODAP	WOA05	0.337	0.04	-3.63	BASELINE run
CCSM	CCSM	CCSM	CCSM	0.337	0.04	-4.18	
KS	GLODAP, Sabine	GLODAP	WOA05	0.337	0.04	-3.21	
BLEND	Blended	Blended	Blended	0.337	0.04	-3.87	
NOBM	NOBM	Blended	Blended	0.337	0.04	-0.62	DOC IC also from NOBM
PISVEL	BASELINE	BASELINE	BASELINE	0.31	0.04	-3.55	
PICPOC	BASELINE	BASELINE	BASELINE	0.337	0.1	-3.79	
VERSION 2	Optimization			0.34822	0.133	-2.54	Published as versions 2.0 and 2.1

Table 2: Costs per observation (J/N_{OBS}) and multiplication factors for initial condition calculation for sensitivity and optimized model realizations. The “Linear Combination” refers to a calculation of cost and fluxes simply based on a linear combination of the factors from the sensitivity runs, calculated without actually performing a run.

Run	VERSION 2	
name	J/N_{OBS}	Factors
GLODAP	1.28	0.564
CCSM	1.97	0.160
KS	1.04	-0.0779
BLEND	1.66	0.104
NOBM	2.19	0.249
PISVEL	1.29	
Lin.Combi.	0.808	
VERSION 2	0.824

Figure 1: Sequence of 6 daily means of sea-air CO_2 gas fluxes in $\text{molCm}^{-2}\text{yr}^{-1}$ from the “VERSION 1” simulation for January 4–9, 2009. Positive values denote upward fluxes, i.e., outgassing, negative values oceanic uptake. Note the rotating propagation of high and low values from west to east following the paths of atmospheric synoptic systems.

Figure 2: Globally integrated sea-air CO₂ fluxes for 2009 and 2010 in $PgCyr^{-1}$ for four different model realizations: Initial conditions taken from VERSION 1 (blue), NOBM (green), GLODAP (baseline integration, red), VERSION 2 (open black circles), and the monthly means derived from Takahashi et al. (2009) (thick black lines).

Figure 3: Annual mean of surface dissolved inorganic carbon (DIC) in mmolCm^{-2} for 2009 for model runs GLODAP and CCSM (top), BLEND and NOBM (middle), from the Global Ocean Data Analysis Project data set (Key et al., 2004), and the optimized run VERSION 2 (bottom).

Figure 4: Monthly mean CO₂ gas fluxes in $\text{molCm}^{-2}\text{yr}^{-1}$ for July 2009 for model runs GLODAP and CCSM (top), BLEND and NOBM (middle), from the Takahashi et al. (2009) data set, and the optimized run VERSION 2 (bottom). Positive values (red) are upward.

Figure 5: Monthly mean CO₂ gas fluxes in $\text{molCm}^{-2}\text{yr}^{-1}$ for January 2010 for model runs GLODAP and CCSM (top), BLEND and NOBM (middle), from the Takahashi et al. (2009) data set, and the optimized run VERSION 2 (bottom). Positive values (red) are upward.

Figure 6: Positions of surface $p\text{CO}_2$ measurements during 2009 and 2010 from the Global Surface $p\text{CO}_2$ database at the Lamont-Doherty Earth Observatory (Takahashi et al., 2011) represented by the blue dots and lines. Graph was created using ODV software (<http://odv.awi.de>).

Figure 7: Scatter plots for (a) locations of $p\text{CO}_2$ measurements as depicted in Fig. 6, (b) locations of alkalinity measurements as described in the text. The graphs relate model output (vertical axis) to observations (horizontal axis). Red dots and contour line represent VERSION 1, black dots and contour line represent the optimized simulation VERSION 2. The contour lines denote regions where scatter plot density is greater than 0.5 points per ppm^2 or $(\mu\text{mol kg}^{-1})^2$, respectively.

Figure 8: Global volume-weighted trend vs. depth plots for (a) nitrate and (b) oxygen for seven-year model realizations repeating 2009 forcing for the BASELINE (red) and the OPTIMIZED/VERSION 2 (black) experiments.

Figure 9: Time-series of regionally integrated sea-air CO_2 fluxes for 2010 from the optimized model run (VERSION 2) in $\text{PgC}\cdot\text{yr}^{-1}$ (blue dots) and, for reference, the monthly mean values from the Takahashi et al. (2009) climatology (red dots). Positive values (red) are upward. The black lines in the time-series plots denote “0”, values above this line represent outgassing, below it ocean uptake; the scales are identical for all time-series plots.

Figure 10: Annual mean CO₂ gas fluxes in $\text{molCm}^{-2}\text{yr}^{-1}$ for 2010 for the optimized model run VERSION 2 (top) and the Takahashi et al. (2009) data set. Positive values (red) are upward.