JOURNAL OF GEOPHYSICAL RESEARCH, VOL. 114, C10012, doi:10.1029/2009JC005322, 2009

Basin-scale pCO_2 maps estimated from ARGO float data: A model study

T. Friedrich[1,2] and A. Oschlies[1]

Received 11 February 2009; revised 31 July 2009; accepted 6 August 2009; published 13 October 2009.

[1] A novel method for mapping surface pCO_2 on a basin scale using ARGO floats is presented and tested in the framework of an eddy-resolving biogeochemical model of the North Atlantic. Voluntary observing ship (VOS) and ARGO float coverage of the year 2005 is applied to the model to generate synthetic ''observations.'' The model-generated VOS line ''observations'' of pCO_2, SST, and SSS form a training data set for a self-organizing neural network. The trained neural network is subsequently applied locally to estimate pCO_2 from the model-generated ARGO float SST and SSS data. The local pCO_2 estimates at the simulated float positions are extrapolated using objective mapping. The accuracy of the nearly basinwide pCO_2 estimates is assessed by comparing against the pCO_2 output of the model that serves as synthetic ''ground truth.'' For an ARGO float coverage of the year 2005, the resulting monthly mean pCO_2 maps cover 70% of the considered area (15°N to 65°N) with an RMS error of 15.9 μatm. Compared to remote sensing-based estimates that suffer from large regional gaps in optical satellite data coverage, the RMS error in reproducing the annual cycle of pCO_2 can be reduced by 42% when the more evenly distributed ARGO float-based data are used.

Citation: Friedrich, T., and A. Oschlies (2009), Basin-scale pCO_2 maps estimated from ARGO float data: A model study, *J. Geophys. Res.*, *114*, C10012, doi:10.1029/2009JC005322.

1. Introduction

[2] The ocean is estimated to have taken up about 40% of the carbon dioxide (CO_2) emissions released to the atmosphere through anthropogenic activities since the beginning of the industrial revolution [*Sabine et al.*, 2004]. In the last decades large efforts have been undertaken to achieve a reliable and precise quantification of the marine carbon uptake, its temporal fluctuations, and its spatial patterns. To this extent, automated instruments on board Voluntary Observing Ships (VOS) measure partial pressure of CO_2 (pCO_2) at the ocean's surface. They have provided substantial quantitative information since detailed knowledge about the atmosphere-ocean pCO_2 difference is, together with the gas transfer velocity, needed for a direct calculation of air-to-sea CO_2 fluxes. However, seawater pCO_2 is highly variable in time and space due to changes in solubility, mainly driven by temperature, and marine biological processes altering the total dissolved inorganic carbon (DIC) content of the water. VOS line coverage depends on commercial ship tracks (and funding) rather than on observational requirements, and therefore results in an uneven coverage of the ocean surface that is not well suited for interpolation and extrapolation to basin-scale pCO_2 maps.

[3] A first attempt to overcome the problem of large data gaps between the individual VOS lines was to make use of satellite data of sea surface temperature (SST) and surface chlorophyll. In a pilot study, *Friedrich and Oschlies* [2009] showed that for an assumed perfect satellite coverage, a neural network trained with VOS line data could yield monthly pCO_2 estimates with an RMS error of about 19 μatm. When gaps in the satellite optical coverage due to clouds and low irradiation at high latitudes in winter were taken into account and interpolated using climatological values for surface temperature and chlorophyll, the RMS error increased to about 21 μatm.

[4] The above study motivated the search for a more evenly distributed data set that contains some information that can be correlated with surface water pCO_2, an obvious choice being the set of ARGO floats. There are currently more than 3300 ARGO floats operating in the ocean (www.argo.ucsd.edu). Current ARGO floats do not measure pCO_2 but they deliver relatively evenly distributed measurements of sea surface temperature (SST) and sea surface salinity (SSS) with a temporal resolution of about 10 days. Additionally, depth profiles of temperature and salinity can be used to calculate mixed layer depth (MLD). The ARGO floats thereby provide access to some parameters controlling pCO_2 variability at the ocean's surface without being subject to previously mentioned problems typically associated with remote sensing of SST and chlorophyll (Chl). In the present study we investigate to what extent the purely physical information provided by current ARGO floats can be used to estimate surface water pCO_2. This is investigated in the framework of an eddy-resolving biogeochemical ocean

[1] IFM-GEOMAR, Leibniz-Institut für Meereswissenschaften an der Universität Kiel, Kiel, Germany.

[2] Now at International Pacific Research Center, School of Ocean and Earth Science and Technology, University of Hawai'i at Manoa, Honolulu, Hawaii, USA.

0148-0227/09/2009JC005322$09.00

model combined with realistic sampling statistics of VOS lines and ARGO floats. Simulated VOS line observations of pCO_2, SST and SSS are utilized to train a self-organizing neural network that is subsequently applied to simulated ARGO float data in order to derive nearly basinwide monthly mean maps of surface pCO_2 for the North Atlantic.

[5] The paper is organized as follows: In the subsequent section we present the high-resolution biogeochemical circulation model and validate the model data against observations. Section 3 deals with the neural network and the objective mapping. Our results are presented and discussed in section 4. Section 5 summarizes our results.

2. Model Configuration and Validation

[6] A pelagic nitrogen-based nutrient-phytoplankton-zooplankton-detritus ecosystem model [*Oschlies and Garçon*, 1999] is coupled to an eddy-resolving regional ocean general circulation model of the North Atlantic. Dissolved inorganic carbon and dissolved oxygen are coupled to nitrogen via the Redfield ratios, alkalinity is diagnosed from a regional fit to salinity. For details, see *Eden and Oschlies* [2006]. The underlying regional ocean circulation model is based on MOM2 [*Pacanowski*, 1995]. The model domain spans the Atlantic Ocean from 18°S to 70°N at a horizontal resolution of $1/12^\circ \times 1/12^\circ$ cos(latitude) and 45 vertical geopotential levels ranging from 10 m thickness near the surface to 250 m near the maximum depth of 5500 m. Surface boundary forcing consists of monthly mean wind stress and a Haney-type heat flux condition as given by *Barnier et al.* [1995]. For the fresh water forcing a relaxation of the model surface salinity to monthly climatology provided by *Levitus* [1982] is used with a timescale of 30 days. In order to pragmatically account for the presence of sea ice, air-sea fluxes of heat, CO_2, and O_2 (but not of momentum) are set to zero whenever surface temperature falls below -1.8°C. Subgrid-scale parameterizations are biharmonic friction and diffusion (with biharmonic coefficients of 0.8×10^{10} m^4/s for diffusion and 2×10^{10} m^4/s for viscosity) and a level-1.5 closure scheme for vertical turbulent mixing following *Gaspar et al.* [1990].

[7] Initial conditions for nitrate are taken from *Conkright et al.* [2002] and for dissolved inorganic carbon from the preindustrial estimate of the GLODAP data set [*Key et al.*, 2004]. The atmospheric pCO_2 remains on a preindustrial level, but varies seasonally and latitudinally according to a nonlinear fit to estimates by *Conway et al.* [1994].

[8] The eddy-resolving coupled biogeochemical-physical model is integrated over a 10 year spin-up period, and model data used here for simulating VOS line and ARGO float-based observations are taken from the model year 11.

[9] At Bermuda the annual cycle of pCO_2 for the reference year is in good agreement with the mean annual cycle observed at BATS (Figure 1a), although the model overestimates the amplitude of the annual cycle by about 10 μatm. In a basinwide comparison of the simulated annual variance to the Takahashi climatology [*Takahashi et al.*, 2008] the general patterns are well reproduced for the subpolar North Atlantic (Figures 1c and 1d) showing a small annual variance for the region 45°N to 55°N and a significantly larger one further north. Overall the model overestimates the annual variance almost everywhere. For example, for the region 45°N to 65°N this overestimate amounts to about 36% (6 μatm). This discrepancy remains essentially unchanged if we map the model monthly means onto the coarser grid of the Takahashi climatology before computing the annual variance.

[10] In the subtropics and in the North Atlantic Current region a significantly larger amplitude in the annual cycle is found for simulated pCO_2 (Figures 1b–1d) although the annual cycles for observed and modeled SST are almost identical. One possible explanation for this overestimated amplitude in the simulated annual cycle of surface pCO_2 is the use of a fixed Redfield stoichiometry by the biogeochemical model. Models have failed to reproduce the observed summer drawdown of dissolved inorganic carbon that occurs at Bermuda in the absence of significant nutrient supply, unless a variable carbon-to-nitrogen stoichiometry specifically for the dissolved organic matter is taken into account [*Anderson and Pondaven*, 2003]. A related decoupling of carbon and nitrogen dynamics has also been observed at higher latitudes during summer, when the drawdown of dissolved inorganic carbon can considerably exceed the Redfield equivalent of the drawdown of dissolved inorganic nitrogen [*Körtzinger et al.*, 2001]. We therefore speculate that the absence of a parametrization of non-Redfield processes in our model is at least partly responsible for the overestimated pCO_2 variance in large parts of the subtropical and mid-latitude North Atlantic. The lack of a DOC-generated late-summer pCO_2 drawdown leads to an overestimation of the pCO_2 summer-maximum as well as to an associated underestimation of entrainment of DIC-rich waters in winter.

[11] For a reliable methodological study the variability of simulated pCO_2 and simulated related parameters must be representative of the variability of the real data. In our previous study [*Friedrich and Oschlies*, 2009] it was shown that the temporal variability along the VOS cruise tracks of carcarrier M/V Falstaff in the years 2002/2003 [*Lüger et al.*, 2004] was similar for observed and model-simulated pCO_2. Autocorrelation scales for SST were found to be larger than those of pCO_2 both in the observations and in the model. However, along-track decorrelation scales for simulated pCO_2 (decorrelation scale of 59 km) were found to be slightly larger compared to observations (43 km) supposably due to the absence of a diurnal cycle and the use of climatological forcing in the model which leads to less small-scale variability in the simulated fields.

[12] For the physical model fields, eddy energy spectra were calculated by *Eden* [2007]. In the subpolar North Atlantic, simulated eddy kinetic energy was found to have a maximum at a marginally smaller scale than in the observations. Overall, the simulated energy spectra and corresponding length scales, as well as their lateral variations, were found to be in remarkably good agreement with observations.

3. Data and Methods

3.1. Self-Organizing Neural Network

[13] In the year 2005 approximately 740,000 VOS line measurements of pCO_2, SST and SSS were gathered in the area 10°S to 70°N (∼our model domain). As shown by *Lefèvre et al.* [2005] and *Friedrich and Oschlies* [2009] these observations can be used for the training of a self-organizing map (SOM) which, in turn, can be used to estimate pCO_2

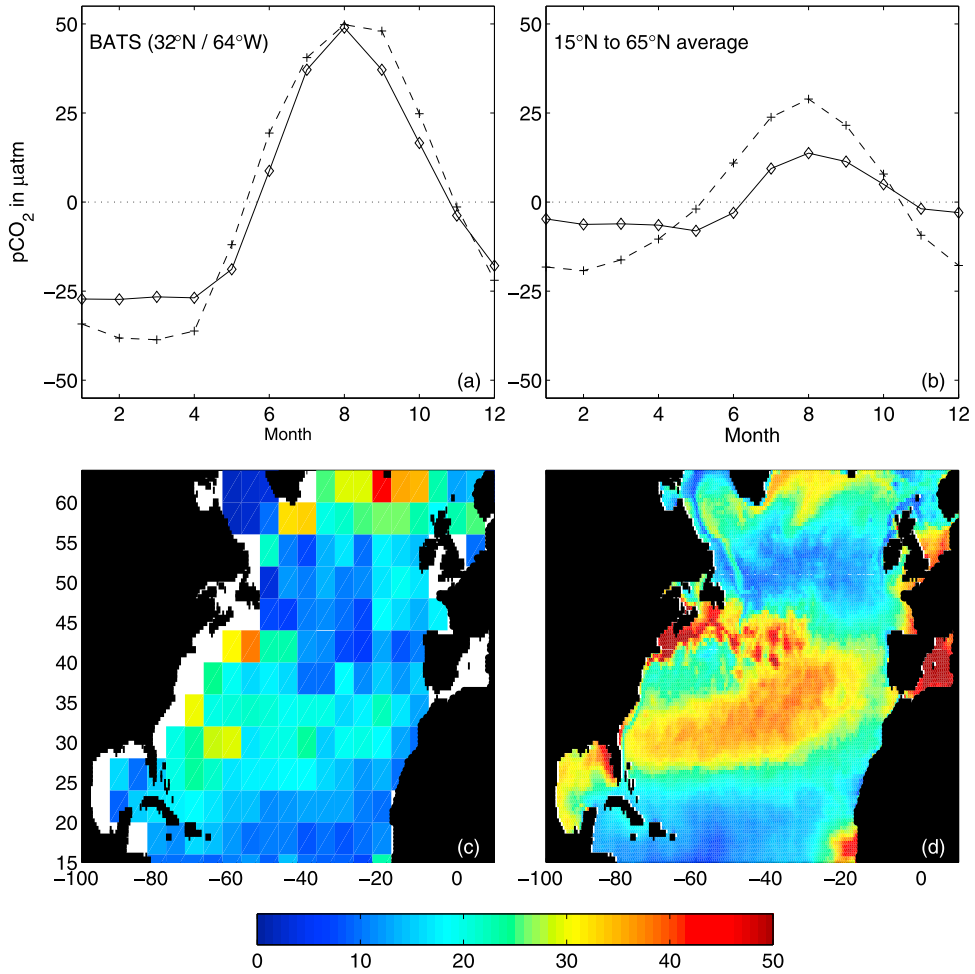

Figure 1. (a) Seasonal cycle (annual mean removed) of observed (solid) and model-generated (dotted) oceanic pCO_2 at Bermuda. (b) Seasonal cycle (annual mean removed) of observed [*Takahashi et al.*, 2008] (solid) and model-generated (dotted) oceanic pCO_2 averaged over 15°N to 65°N. Annual variance of (c) observed [*Takahashi et al.*, 2008] and (d) model-generated pCO_2 in μatm. Annual variance of model-generated data was calculated from monthly mean model data. Bermuda pCO_2 data were kindly provided by Niki Gruber. Note the preindustrial setup of the model.

from related parameters such as SST and SSS over the area represented by the training data. SOMs are a subtype of neural networks. The SOM used here is the same as the neural network used in the work of *Friedrich and Oschlies* [2009]. SOMs can recognize relationships in the data during the training process and are able to associate an input vector (here SST, SSS, position, time) with a target value (here pCO_2). In our case, simulated VOS line samples of SST and SSS and corresponding latitude, longitude, and day of the year are used to form a five-dimensional parameter space representation of simulated VOS line samples of pCO_2 in the training process (Figures 2a–2d). After the second step of the training process, pCO_2 values from the training data set are assigned to the SOM following a minimum distance criterion between SOM and training data vectors. Thus the pCO_2 values in the SOM form a discrete net over the training data in which the grade of discretization (and hence the resolution) depends on the training data density (Figure 2e). In the application of the SOM, an input vector (here consisting of month, latitude, longitude, SST, SSS) is associated with the pCO_2 value of the best matching vector (minimum distance criterion) of the SOM's parameter space representation (month, latitude, longitude, SST, SSS). A more detailed description of the training process and the functioning of the SOM can be found in the works of *Kohonen* [1982], *Lefèvre et al.* [2005] and *Friedrich and Oschlies* [2009].

[14] Surface water pCO_2 is, besides its dependence on sea level pressure, a function of DIC, total alkalinity, SST and SSS. Because for any individual ocean basin total alkalinity can, to good accuracy, be estimated from SSS using a nonlinear empirical fit [e.g., *Eden and Oschlies*, 2006], ARGO SST and SSS data already provide substantial (though local) information about parameters that determine pCO_2. In addition to SST and SSS, we here use the position and the month of the respective data points. Even though their use appears unfounded since month, latitude and longitude have no direct impact on pCO_2, the SOM can exploit this as information about different pCO_2 dynamics in different regions or seasons respectively and about an underlying background pCO_2 field and which is then altered by SST and SSS. Thereby the training of the SOM with physical input data implicitly accounts for biological processes that

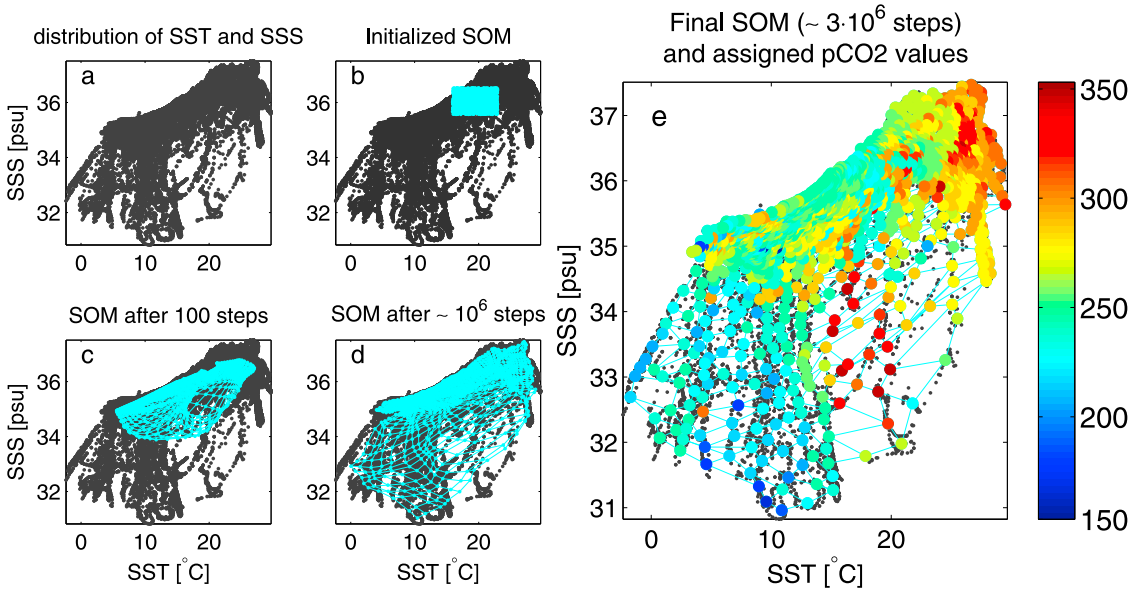

Figure 2. Formation of a SOM. (a) Distribution of SST and SSS in model-generated VOS line data of the year 2005. (b) Randomly initialized SOM network. (c and d) Inflation during training process. (e) Final SOM network with assigned pCO_2 values in μatm. Dimension of the SOM was reduced for illustration to SST and SSS only. Note the preindustrial setup of the model.

affect pCO_2 and that are, in turn, ultimately driven by the physical environment. Although there is no direct functional form available to relate biological carbon uptake to physical variables, the training process allows the SOM to extract a substantial part of the net effect of physics and biology on surface pCO_2 for given SST, SSS and space-time coordinates. A limitation of the training process to SST and SSS would cause substantially larger mapping errors. Obviously, the information about the location and time of the SST and SSS measurements contains useful information about surface pCO_2 even for a data set taken from an eddy-resolving model with intense small-scale variability.

3.2. Objective Mapping

[15] VOS line pCO_2 observations and pCO_2 estimates at ARGO float positions provide a patchy view, with, however, approximately uniform spatial coverage of North Atlantic surface pCO_2. As surface water pCO_2 is subject to high spatial variability, any interpolation or extrapolation to surrounding areas has to account for the typical scales of this variability.

[16] In the present study, a simple spatial objective analysis routine was used to obtain (almost) basin-scale estimates of surface pCO_2. For each month, all simulated VOS line observations and ARGO-based estimates of pCO_2 were interpolated monthly to a $1/2^\circ \times 1/2^\circ$ grid using a simple Gaussian weighting based on autocorrelation scales derived from observations by *Li et al.* [2005]. A cutoff radius of 2° was chosen, following *Li et al.* [2005] who found minimum spatial autocorrelations of ~0.2 for lags of 140 km to 200 km for pCO_2 measurements in the North Atlantic. For the influence radius (R_{inf}) we use a value of 1/2° which corresponds to a mean autocorrelation value of 0.5 [*Li et al.*, 2005]. We are aware of the problem that using a constant cutoff and influence radius neglects any regional differences in spatial variability, but so far the sparseness of pCO_2 observations does not allow reliable estimates of regional autocorrelation lengths on a monthly timescale. The Gaussian weights (κ_i) for all data within the cutoff radius are then calculated by:

$$\kappa_i = \frac{e^{-\left(\frac{d_i}{R_{inf}}\right)^2}}{\sqrt{2\pi}}. \quad (1)$$

with (d_i) being the distance between the grid point and the ith data point within the cutoff radius.

[17] Calculating decorrelation lags in time on a monthly scale would require several months long uninterrupted time series of in situ pCO_2 data. As reported by *Li et al.* [2005], not even the Bermuda time series (Hydrostation "S" and BATS) is sufficient for this analysis. Although temporal decorrelation scales (and hence temporal cutoff radii) could have been derived from our eddy-resolving model, in the absence of validation data no attempt has been made to interpolate samples in time other than binning all data available in a calendar month into monthly means.

4. Results and Discussion

[18] The VOS line coverage available for the year 2005 could already ensure a monthly monitoring along the route UK-Caribbean, and also between the North Sea and the southern tip of Greenland. Furthermore, pCO_2 data were gathered along several transects between the northeastern United States and Iceland and between the UK and Cape Town. However, when simulating this VOS line coverage in the model and extrapolating the model-generated "observations" using an objective mapping method with a 2° cutoff scale derived from the work of *Li et al.* [2005] approximately 70% of the region 15°N to 65°N remains uncovered in the

Table 1. Seasonal Spatial Coverage of pCO_2 Estimates, Mean RMS Error for Extrapolated Model-Generated VOS Line "Observations" of the Year 2005 and Float-Based Mapping of pCO_2, RMS Error in Reproduction of Basin Mean (15°N to 65°N) pCO_2 Annual Cycle, and Deviations From Annual Mean pCO_2

	Extrapolated VOS Line Data		ARGO Float-Based Mapping	
Coverage	15°N:40°N	40°N:65°N	15°N:40°N	40°N:65°N
Spring (MAM)	25.3%	42.0%	67.0%	79.3%
Summer (JJA)	22.0%	38.2%	62.7%	78.8%
Fall (SON)	26.3%	42.3%	64.7%	80.0%
Winter (DJF)	19.5%	36.2%	64.5%	79.5%
Mean RMS error of pCO_2 over area covered by estimates	14.4 μatm		15.9 μatm	
RMS error of annual cycle of basin mean (15°N:65°N) pCO_2 (see Figure 3)	5.02 μatm		2.01 μatm	
Annual mean of simulated "true" pCO_2 − annual mean of estimated pCO_2 (15°N:65°N)	2.21 μatm		−0.03 μatm	

monthly maps. In our simulation this insufficient and unrepresentative data coverage leads to a large RMS error in reproduction of the annual cycle of basin mean pCO_2 as well as a large deviation from the annual mean of basin averaged pCO_2 (Table 1 and Figure 3). Thus relying exclusively on (extrapolated) VOS line pCO_2 data impedes reliable estimates of basinwide pCO_2.

[19] In a first attempt we assessed the potential of filling in areas not covered by extrapolated VOS line observations with climatological values. Climatological pCO_2 was simulated on a grid of 4° × 5° resolution (as originally used by *Takahashi et al.* [1997]) by a model run with slightly different forcing (daily forcing for the considered year instead of monthly climatological mean forcing as described in the work of *Eden and Jung* [2006]) to account for possible biases in our climatology. In order to simulate the impact of uncertainties in the climatological data sets, a second "climatology" was derived from the first one by simply doubling the difference between the model-generated climatological and the model-generated "true" pCO_2 value for every month at every grid point. Although this analysis cannot provide a reliable error estimate, it helps to illustrate the error sensitivity due to uncertainties in the pCO_2 climatology. The range the of basinwide RMS error associated with these two climatologies accounts for about 10 μatm indicating that the quality of the pCO_2 fields used as missing data replacement is crucial even for the, until then, unprecedented VOS line coverage of the year 2005.

[20] In addition to the regional high-density coverage of pCO_2 measurements achieved by VOS lines, ARGO floats provide a large number of approximately evenly distributed sampling points on a global scale. For the considered area (15°N to 65°N) the number of ARGO surface sampling points increased from 7195 in the year 2004 and 8277 in 2005, to 10445 for 2006. To evaluate the potential of float-based pCO_2 estimates to extend the restricted VOS line-based coverage, the self-organizing neural net was trained with the model-generated VOS line "observations" of SST, SSS and pCO_2 of the year 2005 and applied to the simulated ARGO float measurements of SST and SSS of the same year. The local pCO_2 estimates were extrapolated by objective mapping with a 2° cutoff and a 0.5° influence radius and subsequently used to fill in regions not covered by extrapolated VOS lines observations. This combination of extrapolated model-generated "observations" and float-based estimates of pCO_2 increases the monthly coverage of the region 15°N to 65°N to approximately 70%.

[21] Figure 3 and Table 1 respectively compare the observational coverage and accuracy in the reproduction of the pCO_2 annual cycle and annual mean pCO_2 for extrapolated

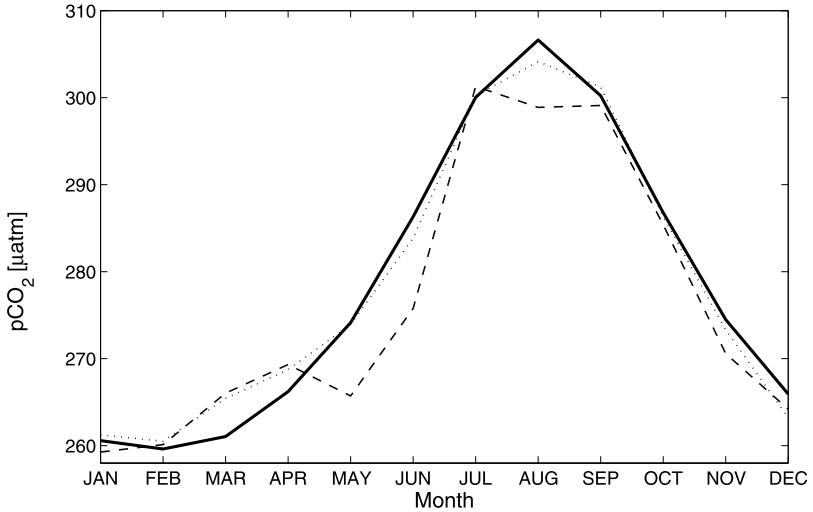

Figure 3. Annual cycle of basin mean (15°N to 65°N) pCO_2 in μatm. (solid) Model-generated "true" annual cycle of pCO_2. (dashed) Extrapolated model-generated VOS line data of the year 2005. (dotted) Mapping based on ARGO float data of SST and SSS and extrapolated VOS line data. Note the preindustrial setup of the model.

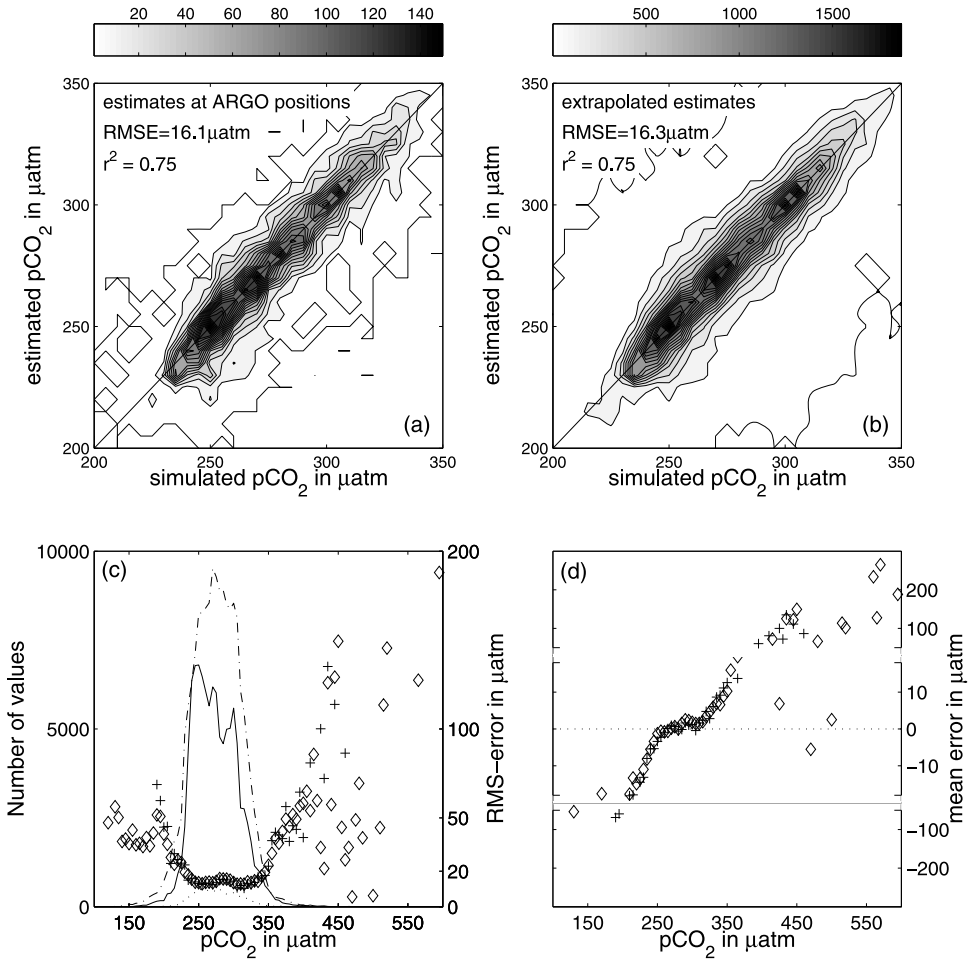

Figure 4. Comparison of simulated and estimated pCO_2 in numbers of data for (a) estimates at ARGO float positions and (b) extrapolated estimates (not area-weighted). The area-weighted basin mean RMS error associated with Figure 4b amounts to 15.9 μatm. (c) Number of values in modeled pCO_2 (dash-dot), model-generated VOS lines of the year 2005 (solid), and model-generated ARGO floats (dotted). RMS error in μatm versus modeled pCO_2 in μatm (5 μatm intervals) for estimates at ARGO float positions (crosses) and extrapolated estimates (diamonds). (d) Mean error in μatm versus modeled pCO_2 in μatm (5 μatm intervals) for estimates at ARGO float positions (crosses) and extrapolated estimates (diamonds). Note the split y axis in Figure 4d and the preindustrial setup of the model.

VOS line data and extrapolated float-based estimates. Even though extrapolated VOS line observations yield good pCO_2 estimates for the area covered by the objective analysis, their sparse spatial and temporal coverage leads to a lack of representativeness of the entire basin. ARGO float-based pCO_2 estimates ensure a more accurate reproduction of both the annual pCO_2 cycle and the annual mean pCO_2 map. Outside the ice-covered areas, their observational coverage is independent of season, relatively evenly distributed and, for the year 2005, turns out to be representative of the basin considered.

[22] The distributions of estimated and model-generated "true" pCO_2 values as well as the RMS errors for both the SOM-based mapping at the floats' positions and the extrapolated estimates are shown in Figure 4. The SOM-based pCO_2 estimate results in a RMS error of 16.1 μatm with respect to the model's "true" pCO_2 at the individual float positions. This is not markedly altered by the subsequent interpolation (Figures 4a and 4b) onto the $1/2° \times 1/2°$ basin-scale grid. For simulated VOS lines (Figure 4c, solid line), basin wide modeled pCO_2 (dash-dot) and simulated ARGO floats (dotted) most pCO_2 values are located in the interval of 230–330 μatm (note the preindustrial setup of the model). In this range, resolution of the SOM is best and the RMS errors resulting from the application of the SOM are found to be minimum. Largest RMS errors occur for $pCO_2 < 200$ μatm and for >400 μatm, which are not well represented by the training data set of simulated VOS data. Outside the range of the training data, lower pCO_2 values tend to be estimated too high and higher pCO_2 values tend to be estimated too low (Figure 4d), a phenomenon already reported in the application of a SOM to simulated satellite data [*Friedrich and Oschlies*, 2009].

[23] Figure 5 illustrates the seasonally averaged RMS error for the float-based pCO_2 estimates. It turns out that areas being most crucial for the North Atlantic carbon uptake, i.e., the Labrador Sea, the North Atlantic Current, and the West African upwelling region are covered by the present method

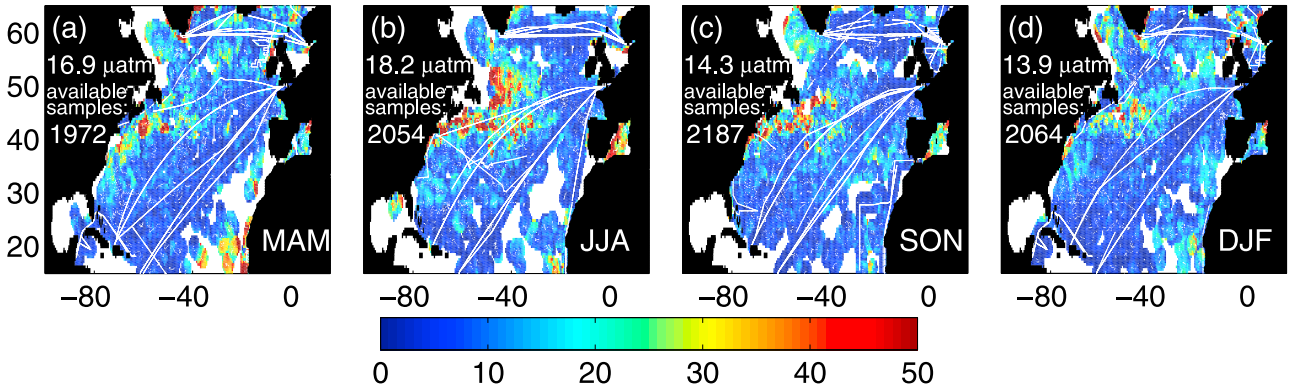

Figure 5. Seasonal mean RMS error in μatm and number of available ARGO float samples for mapping based on VOS lines and ARGO float coverage of the year 2005. (white lines and dots) Corresponding seasonal VOS lines and ARGO float positions of the year 2005.

throughout the entire year. Only for the Gulf of Mexico and off the South American coast are no estimates available. The RMS error patterns are similar to those presented by *Friedrich and Oschlies* [2009] for pCO_2 estimates based on remote sensing data of SST and Chl. Largest errors are found near the Grand Banks in summer and in the West African upwelling region caused by a high spatial variability of pCO_2 which cannot be reproduced by the SOM.

4.1. Comparison to Remote Sensing-Based Mapping

[24] The methodological study by *Friedrich and Oschlies* [2009] showed that a basinwide view on pCO_2 can be obtained by combining VOS line observations and remote sensing data of SST and chlorophyll using neural network-based mapping. Besides the still insufficient VOS line coverage, one crucial limiting factor of this method turned out to be the lack of optical coverage in the remote sensing of Chl. The availability of remotely sensed Chl north of 50°N drops to zero between November and February and reaches availabilities of more than 15 days per month in only a few parts of the basin for every month of the year. This uneven distribution of data gaps necessitates a replacement of missing data, e.g., with climatological SST and Chl values. Unfortunately, this induces significant biases in the monthly mean maps.

[25] Figure 6 compares the annual cycles of CO_2 fluxes derived from the different pCO_2 estimates using the piston velocity formulation of *Wanninkhof* [1992]. Besides the model's "true" pCO_2 field, these include remote sensing-based pCO_2 estimates in which gaps in optical satellite coverage were replaced by climatological values of SST and Chl [*Friedrich and Oschlies*, 2009] and the ARGO float-based mapping of this study. Results are shown for different latitudinal bands of the basin. For all latitudinal bands the RMS errors in the estimated annual cycle of CO_2 fluxes are smaller for the float-based mapping compared to the remote sensing-based one. Particularly noticeable is the improvement in the northernmost region in late winter and spring time. The error is also reduced substantially in the subtropical North Atlantic between 25°N and 45°N. The RMS error in the basinwide annual cycle of pCO_2 can be reduced by 42% and the deviation from the simulated "true" mean pCO_2 and mean CO_2 flux is decreased by 38%, although this effect is hardly visible in the basinwide RMS error of CO_2 fluxes. However, the float-based mapping can achieve a good reproduction of the annual cycle of basinwide CO_2 fluxes through a good performance in all parts of the basin, whereas the remote sensing-based mapping partly benefits from a fortuitous canceling of errors over the basin.

[26] The remote sensing-based mapping as presented in the work of *Friedrich and Oschlies* [2009] and the ARGO-based method presented here in principle have a similar potential for estimating pCO_2 and CO_2 fluxes in all areas in which both ARGO and remote sensing data are available. Because of the good spatial coverage of the ARGO floats, the ARGO-based pCO_2 maps reach essentially the same accuracy as one could reach by remote sensing if there were not data gaps due to clouds or low solar irradiation at high latitudes in winter. The advantage over remote sensing-based mapping is thus offered by the relatively uniform and global data coverage of the ARGO floats. The impact of uncertainties on pCO_2 estimates can be large when gaps in the satellite coverage must be filled with climatological values of SST and Chl. For the currently available ARGO float coverage, float-based estimates allow for more profound estimates since they further reduce the area over which no information about pCO_2-related properties is available.

[27] The results presented here are not attained by a fortunate distribution of sampling points but are nearly identical for simulating the coverage of the years 2004, 2005 and 2006. It is hoped that at least a similar coverage will be obtained by the ongoing ARGO as well as follow-on projects in the future.

4.2. Potential Benefit of Mixed Layer Depth and Chlorophyll Observations

[28] In addition to the relatively evenly distributed horizontal sampling points, ARGO floats also provide information about the vertical dimension. Depth profiles of SST and SSS can be used to calculate mixed layer depth (MLD). The use of MLD as training parameters for the SOM has also been simulated in our methodological study. MLD can be derived from ARGO float profiles, but there is still no validated technique to estimate MLD in the open ocean or along VOS line tracks. An attempt to use MLD from ARGO floats being close (in time and space) to VOS lines did not succeed. The

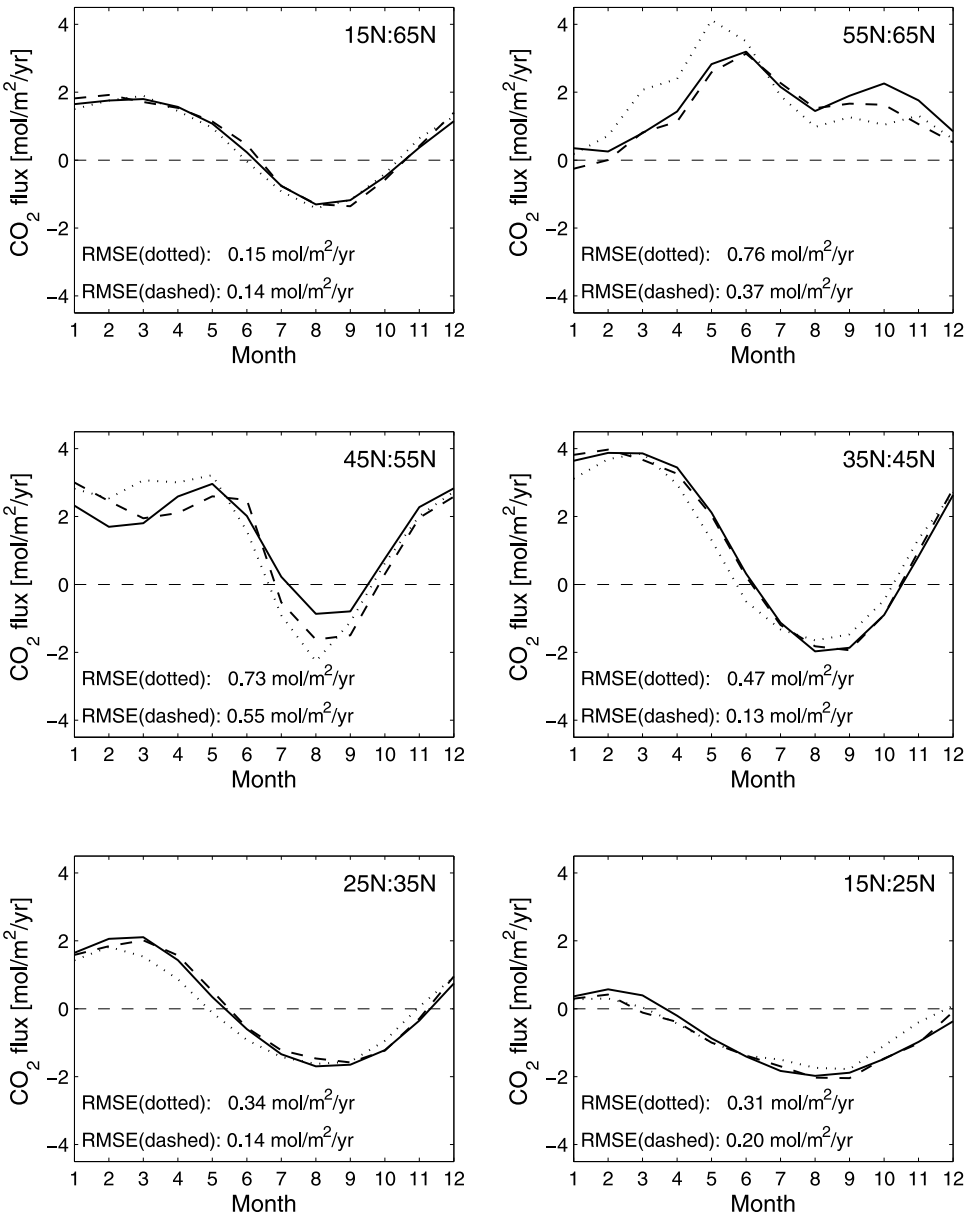

Figure 6. Annual cycle of model-generated "true" (solid) and estimated CO_2 fluxes averaged over given latitude ranges in mol/m^2/a. Underlying pCO_2 fields correspond to Figure 5 for dashed line or were mapped from remote sensing of SST and Chl replacing missing data with climatological values for SST and Chl (dotted line), respectively. Positive values denote fluxes into the ocean. Note the preindustrial setup of the model.

amount of training data is either reduced drastically by setting the conditions for "being close to VOS lines" to small limits, or the training data becomes contaminated with significant errors in MLD. For our eddy-resolving model simulation, a window of $1^\circ \times 1^\circ$ in space and two days in time already resulted in a mean MLD RMS error of 50 meters.

[29] The potential gain of using MLD information, however, turns out to be small. Assuming that perfect in situ MLD data were available not only at the ARGO sites, but also along the VOS lines (the latter are needed for training of the SOM), and using in situ MLD additionally to SST and SSS reduced the RMS error in reproducing the annual cycle of pCO_2 by about 10%.

[30] At first sight, it might appear incomplete and somehow unconvincing to estimate pCO_2 without explicitly taking the biology into account. Unfortunately, ARGO floats cannot yet measure biological parameters such as Chl. One currently would have to rely on Chl observations remotely sensed at the floats' positions in order to add the biology to the training parameters. As described in section 4.1, this would necessitate a replacement of missing data due to the limited availability of satellite Chl observations. The presented method would lose its major advantage provided by the relatively uniform data coverage. On the other hand, it should be kept in mind that in addition to SST and SSS also information about time and space is used to train the neural network. The SOM

can memorize that pCO_2 dynamics are different in different regions and seasons, respectively. Thus biological processes are to some extent implicitly accounted for via the training process of the neural network.

[31] In order to estimate the potential benefit of Chl observations we added a simulation assuming perfect Chl observations were available at ARGO float positions without data gaps. In this hypothetical scenario the RMS error in reproducing the annual cycle of pCO_2 was reduced by about 12%.

5. Conclusions

[32] We have developed and analyzed a novel method that combines VOS line and ARGO float SST and SSS data by a self-organizing neural network in order to generate nearly basinwide monthly maps of surface pCO_2. In contrast to remote sensing of SST and Chl, the data coverage is approximately uniform and does not exhibit large data gaps due to clouds and low solar irradiation. Our results show that about 70% of the considered domain (15°N to 65°N) can be covered by objectively mapped float-based pCO_2 estimates at monthly intervals with an RMS error of 15.9 μatm. The good coverage allows for reliable pCO_2 estimates representative of the basin mean. Analysis of the estimated pCO_2 fields reveals large errors in the North Atlantic Current and the West African upwelling region. Compared to the remote sensing-based mapping introduced by *Friedrich and Oschlies* [2009], the accuracy in reproducing the annual cycle of pCO_2 and basinwide mean values of pCO_2 and CO_2 fluxes can be improved by 42% and 38% when employing ARGO float SST and SSS data at a coverage obtained in year 2005. Furthermore, our study shows that the additional use of Chl or MLD respectively for estimating pCO_2 only results in a minor reduction of the mapping error. Besides the relatively uniform data coverage it is the nearly basinwide availability of SSS and SST data in combination with information about time and position of the data, that allows the neural net to estimate pCO_2 with little loss of accuracy compared to the case of additional, but as yet hypothetical, chlorophyll measurements by the ARGO floats.

[33] **Acknowledgments.** Our thanks go first to the numerous scientists and personnel responsible for the collection of VOS line and ARGO float data that form the basis of our model simulation. We are grateful to Carsten Eden for running the biogeochemical ocean model. This work was funded by the European Union via CarboOcean 511176 (GOCE). We thank the two anonymous reviewers for their constructive comments that have helped to improve the paper. This is International Pacific Research Center contribution 624 and School of Ocean and Earth Science and Technology contribution 7785.

References

Anderson, T. R., and P. Pondaven (2003), Non-redfield carbon and nitrogen cycling in the Sargasso Sea: Pelagic imbalances and export flux, *Deep Sea Res., Part I*, *50*, 573–591.

Barnier, B., L. Siefridt, and P. Marchesiello (1995), Surface thermal boundary condition for a global ocean circulation model from a three-year climatology of ECMWF analyses, *J. Mar. Syst.*, *6*, 363–380.

Conkright, M. E., H. E. Garcia, T. D. O'Brien, R. A. Locarnini, T. P. Boyer, C. Stephens, and J. I. Antonov (2002), *World Ocean Atlas 2001, vol. 4: Nutrients*, edited by S. Levitus, *NOAA Atlas NESDIS 54*, 392 pp., U.S. Gov. Print. Off., Washington, D. C.

Conway, T. J., P. P. Tans, L. S. Waterman, K. W. Thoning, D. R. Kitzis, K. A. Masarie, and N. Zhang (1994), Evidence for interannual variability of the carbon cycle from the National Oceanic and Atmospheric Administration/Climate Monitoring and Diagnostics Laboratory Global Air Sampling Network, *J. Geophys. Res.*, *99*(D11), 22,831–22,855.

Eden, C. (2007), Eddy length scales in the North Atlantic Ocean, *J. Geophys. Res.*, *112*, C06004, doi:10.1029/2006JC003901.

Eden, C., and T. Jung (2006), Wind driven eddies and plankton blooms in the North Atlantic Ocean, *Tech. Memo. 490*, ECMWF, Reading, U. K.

Eden, C., and A. Oschlies (2006), Adiabatic reduction of circulation-related CO_2 air-sea flux biases in a North Atlantic carbon-cycle model, *Global Biogeochem. Cycles*, *20*, GB2008, doi:10.1029/2005GB002521.

Friedrich, T., and A. Oschlies (2009), Neural network-based estimates of North Atlantic surface pCO_2 from satellite data: A methodological study, *J. Geophys. Res.*, *114*, C03020, doi:10.1029/2007JC004646.

Gaspar, P., Y. Gregoris, and J.-M. Lefevre (1990), A simple eddy kinetic energy model for simulation of the oceanic vertical mixing: Tests at Station Papa and Long-Term Ocean Study Site, *J. Geophys. Res.*, *95*(C9), 16,179–16,193.

Key, R. M., A. Kozyr, C. L. Sabine, K. Lee, R. Wanninkhof, J. L. Bullister, R. A. Feely, F. J. Millero, C. Mordy, and T.-H. Peng (2004), A global ocean carbon climatology: Results from Global Data Analysis Project (GLODAP), *Global Biogeochem. Cycles*, *18*, GB4031, doi:10.1029/2004GB002247.

Kohonen, T. (1982), Self-organized formation of topologically correct feature maps, *Biol. Cybern.*, *43*, 59–69.

Körtzinger, A., W. Koeve, P. Kähler, and L. Mintrop (2001), C:N ratios in the mixed layer during the productive season in the Northeast Atlantic Ocean, *Deep Sea Res., Part I*, *48*, 661–688.

Lefèvre, N., A. J. Watson, and A. R. Watson (2005), A comparison of multiple regression and neural network techniques for mapping in situ pCO_2 data, *Tellus*, *57*, 375–384.

Levitus, S. (1982), Climatological Atlas of the World Ocean, *Prof. Paper 13*, U.S. Dept. of Commerce, NOAA, U.S. Gov. Print. Off., Washington, D. C.

Li, Z., D. Adamec, T. Takahashi, and S. C. Sutherland (2005), Global autocorrelation scales of the partial pressure of oceanic CO_2, *J. Geophys. Res.*, *110*, C08002, doi:10.1029/2004JC002723.

Lüger, H., D. W. R. Wallace, A. Körtzinger, and Y. Nojiri (2004), The pCO_2 variability in the midlatitude North Atlantic Ocean during a full annual cycle (2004), *Global Biogeochem. Cycles*, *18*, GB3023, doi:10.1029/2003GB002200.

Oschlies, A., and V. Garçon (1999), An eddy-permitting coupled physical-biological model of the North Atlantic: 1. Sensitivity to advection numerics and mixed layer physics, *Global Biogeochem. Cycles*, *13*(1), 135–160.

Pacanowski, R. (1995), MOM 2 documentation user's guide and reference manual, *Tech. Rep. 3*, 232 pp., GFDL Ocean Group, GFDL, Princeton, N. J.

Sabine, C. L., et al. (2004), The ocean sink for anthropogenic CO_2, *Science*, *305*, 367–371.

Takahashi, T., R. A. Feely, R. F. Weiss, R. H. Wanninkhof, D. W. Chipman, and S. C. Sutherland (1997), Global air-sea flux of CO_2: An estimate based on measurements of sea-air pCO_2 difference, *Proc. Natl. Acad. Sci.*, *94*, 8292–8299.

Takahashi, T., et al. (2008), Climatological mean and decadal changes in surface ocean pCO_2, and net sea-air CO_2 flux over the global oceans, *Deep Sea Res., Part II*, *56*(8-10), 554–577.

Wanninkhof, R. (1992), Relationship between windspeed and gas exchange over the ocean, *J. Geophys. Res.*, *97*(C5), 7373–7382.

T. Friedrich, International Pacific Research Center, School of Ocean and Earth Science and Technology, University of Hawai'i at Manoa, East-West Road 1680, Post Building 412A, Honolulu, HI 96822, USA. (tobiasf@hawaii.edu)

A. Oschlies, Leibniz-Institut für Meereswissenschaften an der Universität Kiel, Düsternbrooker Weg 20, D-24105 Kiel, Germany. (aoschlies@ifm-geomar.de)