Dynamical interpolation of surface pCO_2 between lines of observation in the North Atlantic Ocean

Dissertation
zur Erlangung des Doktorgrades
der Mathematisch-Naturwissenschaftlichen Fakultät
der Christian-Albrechts-Universität zu Kiel

vorgelegt von

Tobias Friedrich

Kiel
2008

Zusammenfassung

Ziel der vorliegenden Dissertation ist es, einerseits die Steuerungsmechanismen des ozeanischen pCO_2 zu untersuchen und Methoden zu entwickeln und zu analysieren, um pCO_2 im Nordatlantik auf beckenweiter Skala zu bestimmen.
Zunächst werden Sensitivitätsuntersuchungen mit einem grob auflösenden biogeochemischen Modell des Nordatlantiks durchgeführt, wobei hier die Zeitspanne 1948–2002 simuliert wird. Die individuellen Anteile von Windstress und Wärmeflüssen an der gesamten zwischenjährlich-dekadischen Variabilität der CO_2- und O_2-Flüsse werden untersucht, indem mehrere Modellläufe mit unterschiedlichem atmosphärischen Antrieb integriert werden. Die Ergebnisse zeigen, dass die Variabilität der CO_2-Flüsse auf diesen Zeitskalen im gesamten Modellgebiet Windstress-getrieben ist. Die Hauptaufnahme von Kohlenstoff durch den Ozean findet in der Simulation im subpolaren Nordatlantik statt. Die zwischenjährlichen Schwankungen in der Aufnahme haben in der subpolaren, subtropischen und äquatorialen Region jedoch dieselbe Größenordnung. Für O_2 ist bezüglich der interannualen Variabilität eine klare Aufteilung in ein Windstress-dominiertes äquatoriales Becken und eine Wärmefluss-getriebene Region nördlich davon zu erkennen. Darüber hinaus wird die Möglichkeit untersucht, pCO_2-Messungen sogenannter Voluntary Observing Ships (VOS) mit Satellitenbeobachtungen und ARGO Float Daten zu kombinieren. Hierbei werden mit Hilfe eines Wirbelauflösenden biogeochemischen Modells die jeweiligen Beobachtungsdaten gemäß ihrer tatsächlichen Abdeckung im Referenzjahr 2005 simuliert. Die synthetischen Messungen der VOS-Linien bilden den Trainingsdatensatz für ein künstliches Neuronales Netz, das anschließend entweder auf die simulierten Satellitenbeobachtungen von SST und Chlorophyll oder ARGO Float SST- und Salzghaltsdaten angewandt wird. Im ersteren Fall ergeben sich direkt beckenweite pCO_2-Schätzungen, im letzteren punktuelle Approximationen, die durch Gaußsches Gewichten extrapoliert werden. Für die Satellitengestützten pCO_2-Karten ergibt sich ein beckenweiter RMS-Fehler von 19.0 ppm, wenn das Fehlen von Daten durch Wolken und zu geringe winterliche Einstrahlung in hohen Breiten vernachlässigt wird. Werden diese Datenlücken durch klimatologische Werte von SST und Chlorophyll ersetzt, steigt der RMS-Fehler auf 21.5 ppm an. Die extrapolierten Float-gestützten pCO_2 Schätzungen umfassen 70% der betrachteten Region (15°N bis 65°N) und besitzen einen RMS-Fehler von 15.9 ppm. Letztere ermöglichen eine genauere Wiedergabe der Jahresgänge von pCO_2 und den CO_2-Flüssen, da sie unabhängig von der Satellitenabdeckung sind.

Abstract

The present PhD thesis aims to elucidate driving mechanisms of oceanic surface pCO_2 variability and to develop and analyze techniques for mapping pCO_2 on a basinscale in the North Atlantic.
First of all, a number of sensitivity tests are carried out in a coarse resolution coupled ecosystem-circulation model simulating the period 1948–2002. The individual contributions by wind stress and surface heat fluxes to naturally driven interannual-to-decadal variability of air-sea fluxes of CO_2 and O_2 are examined using different atmospheric forcing fields. The model results reveal a pronounced dominance of wind stress in driving interannual-to-decadal variability of CO_2 fluxes in the entire model domain. Although the simulated mean carbon uptake takes place in the subpolar basin, interannual fluctuations are of the same magnitude in the subpolar region, the subtropics and the equatorial Atlantic. For O_2, mechanisms causing temporal variations can be separated into a wind-stress driven equatorial and a heat-flux driven subtropical and subpolar basin.
Subsequently, the potential of monitoring North Atlantic ocean-surface pCO_2 on a basin scale by combining Voluntary Observing Ship (VOS) observations with ARGO float and remote sensing data respectively is explored in the context of an eddy-resolving model. Here, model output is sampled according to realistic VOS-line, ARGO float and satellite coverage of the reference year 2005. The synthetic VOS-line observations form a training data set for a self-organizing neural network which is, in the first case, applied to simulated satellite data of SST and surface chlorophyll in order to derive basinwide monthly maps of surface pCO_2. In the second case the trained neural network is used to derive punctual pCO_2 estimates from ARGO float SST and salinity data which are extrapolated by objective mapping. For a remote-sensing based mapping the basinwide mean RMS-error amounts to 19.0 ppm when missing data in the satellite coverage due to clouds and low solar irradiation at high latitudes in winter is neglected and 21.1 ppm if this missing data is replaced by climatological SST and chlorophyll values. Extrapolated float-based estimates cover 70% of the considered area (15°N to 65°N) with a basinwide mean RMS-error of 15.9 ppm and provide a better accuracy in the reproduction of annual cycles of pCO_2 and CO_2 fluxes due to their independence of satellite coverage.

Hiermit erkläre ich, dass ich vorliegende Doktorarbeit selbständig, ohne unerlaubte Hilfe und unter Einhaltung der Regeln guter wissenschaftlicher Praxis angefertigt habe. Ferner habe ich weder diese noch eine ähnliche Arbeit an einer anderen Abteilung oder Hochschule im Rahmen eines Prüfungsverfahrens vorgelegt, veröffentlicht oder zur Veröffentlichung vorgelegt.

_____

(Tobias Friedrich)

Contents

Introduction

The beginning of the 21st century is marked by a raising awareness among all members of society on the consequences of a global climate change. In August 2005 hurricane Katrina hit the Southern Coast of the United States with devastating effect. Even though a direct connection between a single hurricane and a changing climate can not be established, the death of about 1,800 people and the costs of approximately 80 billion US dollars in damages have revealed the vulnerability of one of the wealthiest nations in the world towards a natural disaster whose increased occurrence is assumed in a global warming scenario. The *Stern Review on the Economics of Climate Change* released in October 2006 by economist Nicolas Stern became the most widely known and debated report on the effect of climate change and global warming on the world economy. It discussed the actions required in order to avoid the worst effects of climate change on the global gross domestic product as well as the economic risks associated with a failure to do so. The Stern review stated: "Our actions over the coming few decades could create risks of major disruption to economic and social activity, later in this century and in the next, on a scale similar to those associated with the great wars and the economic depression of the first half of the 20th century." Eventually, in October 2007 the Nobel Peace Prize was awarded to Al Gore and the Intergovernmental Panel on Climate Change "for their efforts to build up and disseminate greater knowledge about man-made climate change, and to lay the foundations for the measures that are needed to counteract such change". Climate change and global warming is inseparably linked to carbon dioxide (CO_2). CO_2 is the most important anthropogenically affected greenhouse gas with respect to the Earth's radiation balance whose content in the atmosphere has increased by 30% since the onset of the industrial revolution, causing an associated increase in average global near-surface air temperature of approximately 0.8°C. Although CO_2 concentrations have always been subject to variability in the planet's past, present atmospheric CO_2 concentration of about 383 ppm is unprecedented and has not been exceeded in the last 420,000 years [*Petit et al.*, 1999; *Fischer et al.*, 1999]. But still, the atmospheric increase constitutes only about half of the accumulated anthropogenic emissions which points to the second leading actor along with mankind in this process of global change: the ocean. The ocean covers more than 70% of the Earth's surface and its chemical properties allow to exceed by about 50-fold atmospheric carbon stocks. Deep water formation and ocean circulation can create time periods up to centuries between the uptake of CO_2 from the atmosphere and its release back to it, called the "physical pump". In addition, the "biological pump" forms a vertical gradient in CO_2 concentrations between surface and deep waters by transforming CO_2 into carbon-containing organic particles that sink through the water and allow for burial of carbon in sea floor sediments. According to studies by *Sarmiento and Toggweiler* [1984] atmospheric CO_2 would be approximately twice as

high without this biological drawdown.
The progress and change of the Earth's climate is therefore closely connected to both the success in the efforts to reduce anthropogenic carbon emissions and the role of the ocean in buffering and alleviating man-made interferences. Thus, an extensive comprehension and monitoring as well as an effective modeling of the marine carbon cycle is a crucial point limiting our ability to forecast climate conditions in the future.

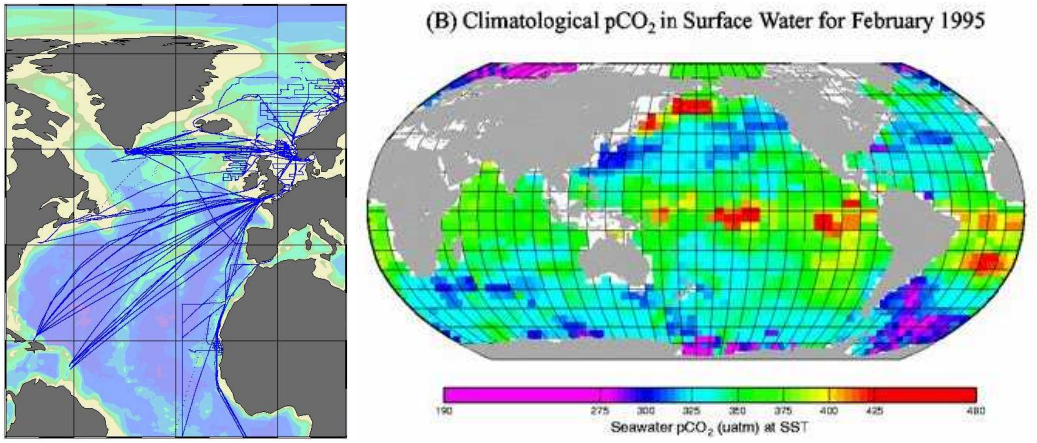

Figure 1: Left: Overview of pCO_2 data gathered within 2005 and 2006 in the North Atlantic by Voluntary Observing Ships, buoys, moorings and research vessels. Right: Climatological mean distribution of surface-water pCO_2 in February 1995 estimated by *Takahashi et al.* [2002].

In contrast to the well mixed atmosphere, seawater concentrations of CO_2 normally expressed as its partial pressure (pCO_2) are highly variable in time and space which impedes a calculation of air-to-sea CO_2 fluxes since a detailed knowledge of the air-sea pCO_2 difference is required here. In the Atlantic Ocean a pCO_2 monitoring network was initialized by the European project CAVASSOO (Carbon Variability Studies by Ships of Opportunity) in 2001 and is currently extended with better coverage and higher accuracy in the measurements by the European CARBOOCEAN project in cooperation with US-American partners. Figure 1 illustrates our state-of-the-art pCO_2 monitoring in the North Atlantic and knowledge about the global mean distribution of pCO_2.
Unfortunately, this knowledge only allows for estimating a mean climatological cycle of marine CO_2 uptake. Also the observational coverage of the years 2005 to 2006 in the North Atlantic, though appearing reasonable at first sight, should not mislead about the fact that for a calculation of interannual varying CO_2 fluxes basinwide estimates of pCO_2 are required on a monthly or at least seasonal scale. Hence, gaining an insight to the processes controlling pCO_2 variability and the potential to map basinwide pCO_2 on short ($\sim$ monthly) time scales are of paramount importance for reliable estimates of

year-to-year ocean carbon uptake.
My PhD studies were focused on exactly these issues. First of all, driving mechanisms of North Atlantic surface pCO_2 have been analyzed in a coarse-resolution coupled ecosystem-circulation model with regard to the individual contributions of wind stress and surface heat fluxes to naturally driven interannual-to-decadal variability. Interannual fluctuations in Oxygen (O_2) were also examined in this context. In contrast to seawater pCO_2, surface O_2 concentrations can react quickly to changing equilibrium conditions. This often simplifies an establishment of an understanding of the processes controlling O_2 fluctuations and since O_2 is physically and biologically coupled to CO_2 also those controlling pCO_2. Furthermore air-sea O_2 fluxes are relevant for methods that attempt to separate terrestrial and marine sinks of anthropogenic CO_2 by employing the atmospheric Oxygen-Nitrogen mixing ratio [*Keeling and Shertz*, 1992].
Subsequently two different methods for mapping pCO_2 on a basinwide scale have been developed and tested. Present day monitoring of pCO_2 is mainly carried out by Voluntary Observing Ships, whose coverage depends on commercial ship tracks rather than observational requirements. Therefore an interpolation between VOS-line observations remains difficult. As seawater pCO_2 is strongly related to physical, biological and chemical variables such as sea surface temperature (SST), chlorophyll (Chl) and total alkalinity (TA), this discloses the opportunity to estimate it from parameters with a better availability. However, a stringent mathematical relation connecting for example SST and Chl (which can be derived from satellite data) or SST and salinity (which can be derived from ARGO float data) to pCO_2 does not exist. In the two presented studies a neural network approach was accomplished. A self-organizing neural network (Kohonen Feature Map, [*Kohonen*, 1982]) was trained with synthetic VOS-line data to reproduce pCO_2 from parameters being available on a basinwide or nearly basinwide scale.
In the first case the focus was on a remote-sensing based mapping of pCO_2 from satellite data. VOS-line observations of the reference year 2005 were simulated in an eddy-resolving North Atlantic model and a self-organizing map (SOM) was trained to reproduce pCO_2 from synthetic samples of SST, Chl and corresponding positions in time and space. For the subsequent application remote sensing of SST and Chl with realistic satellite coverage for the reference year was also simulated in the model. The accuracy of the maps with respect to the model's "true" simulated pCO_2 was analyzed with regard to VOS-line coverage, remote sensing errors and the interpolation of missing data due to cloud cover and low solar irradiation at high latitudes in winter.
In the second case an alternative procedure was established which was highly motivated by the fact that availability of remote sensing data is strongly limited and the replacement of missing data increases mapping errors significantly. A SOM was trained with the year 2005 synthetic VOS-line observations to reproduce pCO_2 from SST and sea surface salinity (SSS) and applied to ARGO float data. The punctual pCO_2 estimates were extrapo-

lated using objective mapping. The resulting pCO_2 maps cover 70% of the considered area (15°N to 65°N) and allow for a calculation of mean annual cycles of pCO_2 and CO_2 fluxes. The advantage of the method is given by the independence of remote sensing and the associated handling of missing data.

The individual research papers are:

- Friedrich, T., A. Oschlies and C. Eden, 2006: Role of wind stress and heat fluxes in interannual-to-decadal variability of air-sea CO_2 and O_2 fluxes in the North Atlantic. *Geophys. Res. Lett.*, 33, (21), doi:10.1029/2006GL026538.

  - Examines the contributions by wind stress and surface heat fluxes to naturally driven interannual-to-decadal variability of air-sea fluxes of CO_2 and O_2 during 1948–2002 and shows a geographical separation into heat-flux and wind-stress driven regions for O_2 fluxes and a mainly wind-stress driven variability of CO_2 fluxes over the entire model domain.

- 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.*, in press.

  - Explores the potential of monitoring ocean-surface pCO_2 on a basin scale by combining VOS-line and remote sensing data of SST and Chl.

- Friedrich, T. and A. Oschlies, 2008: How to estimate North Atlantic surface pCO_2 from ARGO float data, *J. Geophys. Res.*, submitted.

  - Introduces a nearly basinwide ARGO-float based pCO_2 mapping which is independent of remote sensing and associated handling of missing data.

References

Fischer, H., M. Wahlen, J. Smith, D. Mastroianni and B. Deck (1999), Ice core records of atmospheric CO_2 around the last three glacial terminations, *Science*, *283*, 1712–1714.

Keeling, R. F. and S. R. Shertz (1992), Seasonal and interannual variations in atmospheric oxygen and implications for the global carbon cycle, *Nature*, *358*, 723–727.

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

Petit, J.R., J. Jouzel, D. Raynaud, N.I. Barkov, J.-M. Barnola, I. Basile, M. Benders, J. Chappellaz, M. Davis, G. Delayque, M. Delmotte, V.M. Kotlyakov, M. Legrand, V.Y. Lipenkov, C. Lorius, L. Ppin, C. Ritz, E. Saltzman, and M. Stievenard (1999), Climate and atmospheric history of the past 420000 years from the Vostok ice core, Antarctica, *Nature*, *399*, 429–436.

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

Sarmiento, J. L. and J. R. Toggweiler (1984), A new model for the role of oceans in determining atmospheric pCO_2, *Nature*, *308*, 621–624.

Takahashi, T., S. C. Sutherland, C. Sweeney, A. Poisson, N. Metzl, B. Tilbrook, N. Bates, R. Wanninkhof, R. A. Feely, C. Sabine, J. Olafsson and Y. Nojiri (2002), Global seaair CO_2 flux based on climatological surface ocean pCO_2, and seasonal biological and temperature effects, *49*, *1601–1622*.

Role of wind stress and heat fluxes in interannual-to-decadal variability of air-sea CO_2 and O_2 fluxes in the North Atlantic

–*Geophysical Research Letters*, published[1]–

Tobias Friedrich[1]
(corresponding author)

Andreas Oschlies[1]

Carsten Eden[1]

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

Abstract

A coupled ecosystem-circulation model of the North Atlantic is used to examine the individual contributions by wind stress and surface heat fluxes to naturally driven interannual-to-decadal variability of air-sea fluxes of CO_2 and O_2 during 1948–2002. The model results indicate that variations in O_2 fluxes are mainly driven by variations in surface heat fluxes in the extratropics (15°N to 70°N), and by wind stress in the tropics (10°S to 15°N). Conversely, variations in simulated CO_2 fluxes are predominantly wind-stress driven over the entire model domain (18°S to 70°N); while variability in piston velocity and surface heat fluxes is less important. The simulated uptake of O_2 by the North Atlantic amounts to 70±11 Tmol yr^{-1} to which the subpolar region (45°N to 70°N) contributes by 62±10 Tmol yr^{-1}. Whereas the subpolar North Atlantic takes up more than 2/3 of the total carbon absorbed by the North Atlantic in our model (about 0.3 Pg C yr^{-1}), interannual variability of air-sea CO_2 fluxes reaches similar values (about 0.01 Pg C yr^{-1} each) in the subpolar (45°N to 70°N), the subtropical (15°N to 45°N) and the equatorial (10°S to 15°N) Atlantic.

[1] *Geophys. Res. Lett.*, 33, L21S04, doi:10.1029/2006GL026538

Introduction

Reliable estimates of air-sea fluxes of CO_2 are needed not only for a better understanding of the global carbon cycle and the long-term fate of anthropogenic CO_2, but also to better constrain inverse models that help to monitor sources and sinks of anthropogenic CO_2 on politically and economically more relevant shorter ($\sim$ annual) time scales. Time-series data [*Bates et al.*, 1999; *Gruber et al.*, 2002] and model studies [*LeQuéré et al.*, 2000; *Bopp et al.*, 2002; *McKinley et al.*, 2003, 2004a; *Wetzel et al.*, 2005], have revealed substantial interannual to decadal fluctuations in air-sea fluxes of CO_2 and also of O_2 . Air-sea O_2 fluxes are relevant for methods that attempt to separate terrestrial and marine sinks of anthropogenic CO_2 by employing the atmospheric mixing ratio of O_2/N_2 [*Keeling and Shertz*, 1992]. To better understand the mechanisms that drive these fluctuations in air-sea fluxes of CO_2 and O_2, we made a number of sensitivity studies with different combinations of climatological and interannually varying forcing fields applied to a prognostic, 4-component ecosystem-circulation model of the North Atlantic. This complements earlier model studies [*Wetzel et al.*, 2005] including applications of interannually varying atmospheric forcing fields to one-component [*McKinley et al.*, 2000, 2004a], or multi-component off-line biogeochemical models [*Bopp et al.*, 2002] and to models restored to climatological temperature and salinity values beneath the surface mixed layer [*LeQuéré et al.*, 2000]. While the simulated variability in air-sea fluxes of CO_2 and O_2 is of similar magnitude in all these models, our sensitivity studies aim to elucidate the processes that cause the temporal fluctuations.

Model and Forcing

The numerical model consists of an NPZD-type (nitrate, phytoplankton, zooplankton, detritus) ecosystem model [*Oschlies and Garçon*, 1999] embedded into a $\frac{4}{3} \times \frac{4}{3}$ cos(latitude) resolution circulation model with 45 vertical levels based on the Modular Ocean Model version 2.1 [*Pacanowski*, 1995][2]. The model domain spans the Atlantic Ocean from 70°N to 18°S and is the same as that used by *Eden and Willebrand* [2001], except that northern and southern boundaries as well as the Straits of Gibraltar are closed with restoring zones for temperature, salinity and biogeochemical tracers. Advection is modelled using the QUICKER scheme [*Leonard*, 1992] with an additional flux correction [*Lafore et al.*, 1998] to ensure positive biogeochemical tracer concentrations. The nitrogen-based ecosystem model is coupled to the oxygen and carbon cycles via the respective Redfield ratios.Particulate inorganic carbon is not modelled explicitly, and surface alkalinity is computed from an observational fit to surface salinities [*Eden and Oschlies*, 2006].

[2]The numerical code together with all configurations used in this study can be accessed at http://www.ifm-geomar.de/index.php?id=spflame

name of model run	wind-stress	surface heat flux	wind for piston velocity
full forcing	ECMWF clim. + NCEP anom.	ECMWF clim. + NCEP anom.	ECMWF clim. + NCEP anom.
WIND	ECMWF clim. + NCEP anom.	ECMWF clim.	ECMWF clim.
HFLX	ECMWF clim.	ECMWF clim. + NCEP anom.	ECMWF clim.
PISTON	ECMWF clim. + NCEP anom.	ECMWF clim. + NCEP anom.	ECMWF clim.

Table 1: Model runs and forcing used for wind-stress, surface heat flux and piston velocity. Note that full "ECMWF clim. + NCEP anom." forcing for wind was used in the heat flux formulation in full forcing, PISTON and HFLX run.

The formulation of air-sea gas exchange follows *Wanninkhof* [1992]. For computational and conceptional simplicity, the atmospheric CO_2 concentrations were set to a preindustrial value of 278 ppm. Given the apparent absence of, up to now, strong non-linear responses of the ocean to the steady rise of atmospheric CO_2 our conclusions about mechanisms of interannual variability should also apply for a run that includes anthropogenic CO_2. Initial conditions for nitrate and oxygen are taken from *Boyer and Levitus* [1997], for DIC from the observational estimate of preindustrial DIC from GLODAP [*Sabine et al.*, 2004]. Initial conditions for phytoplankton, zooplankton and detritus are set to small values [*Oschlies and Garçon*, 1999]. During a 30-year physical spin-up followed by a 20-year biogeochemical spin-up, the model was forced with climatological monthly wind-stress and heat-flux fields computed from ECMWF analysis data [*Barnier et al.*, 1995].

Subsequently, a number of sensitivity experiments were run over a 54-year period (simulating the period 1948 to 2002) using different combinations of climatological and interannually varying monthly wind-stress and heat-flux anomaly fields (Table 1). To be consistent with the model's spin up we have added anomaly fields taken from the NCEP/NCAR reanalysis [*Kalnay et al.*, 1996] to the climatological ECMWF data as used during the spinup. Interannual fluctuations in the model results are calculated by subtracting a control simulation with climatological forcing to differentiate between forced signal and model drift, although the latter is small, i.e., an annual increase of CO_2 uptake by 0.0015 Pg C yr^{-2}.

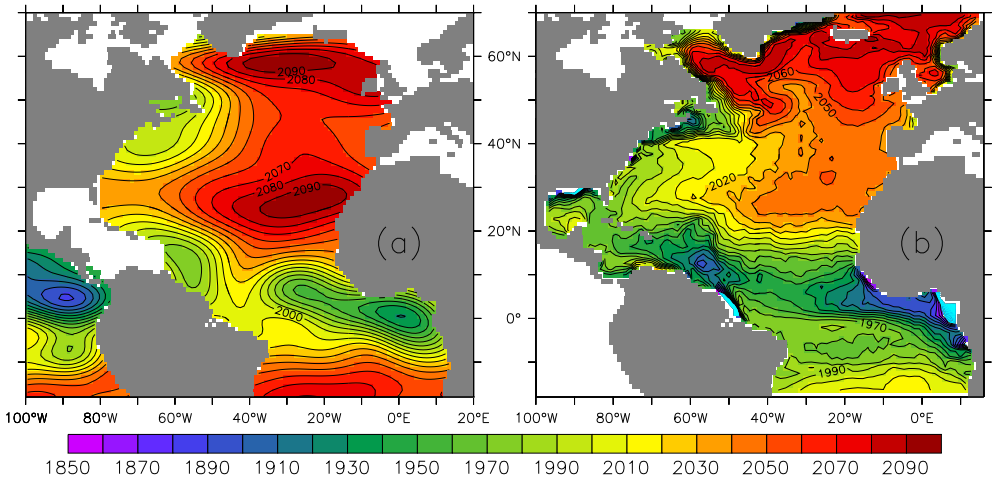

Figure 2: (a): Estimated [*Sabine et al.*, 2004] preindustrial and (b) mean surface DIC simulated by the full forcing run, in mmol/m^3.

Results and Discussion

The general patterns of observational estimates of pre-industrial surface concentrations of dissolved inorganic carbon (DIC) [*Sabine et al.*, 2004] are generally reproduced by the model (Figure 2). However there are systematically lower values in simulated surface DIC concentrations (< 1800 mmol m^{-3}) in the eastern equatorial Atlantic and in the Amazon river plume. At Bermuda the mean annual cycle of pCO_2 agrees closely with Takahashi's compilation of pCO_2 measurements [*Takahashi et al.*, 1997] as does simulated O_2 with observed O_2 from the World Ocean Atlas (WOA) [*Conkright et al.*, 2002], (Figure a,b). Similar to previous ocean model simulations compiled by *Raynaud et al.* [2005], the amplitude of interannual variations of surface pCO_2 at Bermuda is underestimated by approximately a factor of two (Figure c).We speculate that this might be caused by the lack of meso- and sub-mesoscale variability in the model simulations.

Oxygen

Interannual to decadal fluctuations in simulated air-sea O_2 fluxes are largest in the subpolar North Atlantic (45°N to 70°N, Figure 4). In the zonal integral, anomalies reach $\pm 25\,\mathrm{Mmol\,O_2\,m^{-1}\,yr^{-1}}$ which results in basin-scale fluctuations of $\pm 20\,\mathrm{Tmol\,O_2\,yr^{-1}}$ about the basinwide uptake of $70\,\mathrm{Tmol\,O_2\,yr^{-1}}$ (0°N to 70°N). A comparison of the full forcing run with the HFLX run and the WIND run reveals that variations in wind stress explain only 35% of the variability in air-sea O_2 fluxes generated by the full forcing run, while variations in surface heat fluxes are the driving force of variability of O_2 fluxes everywhere north of 15°N. Oxygen uptake between 15°N and 70°N in the

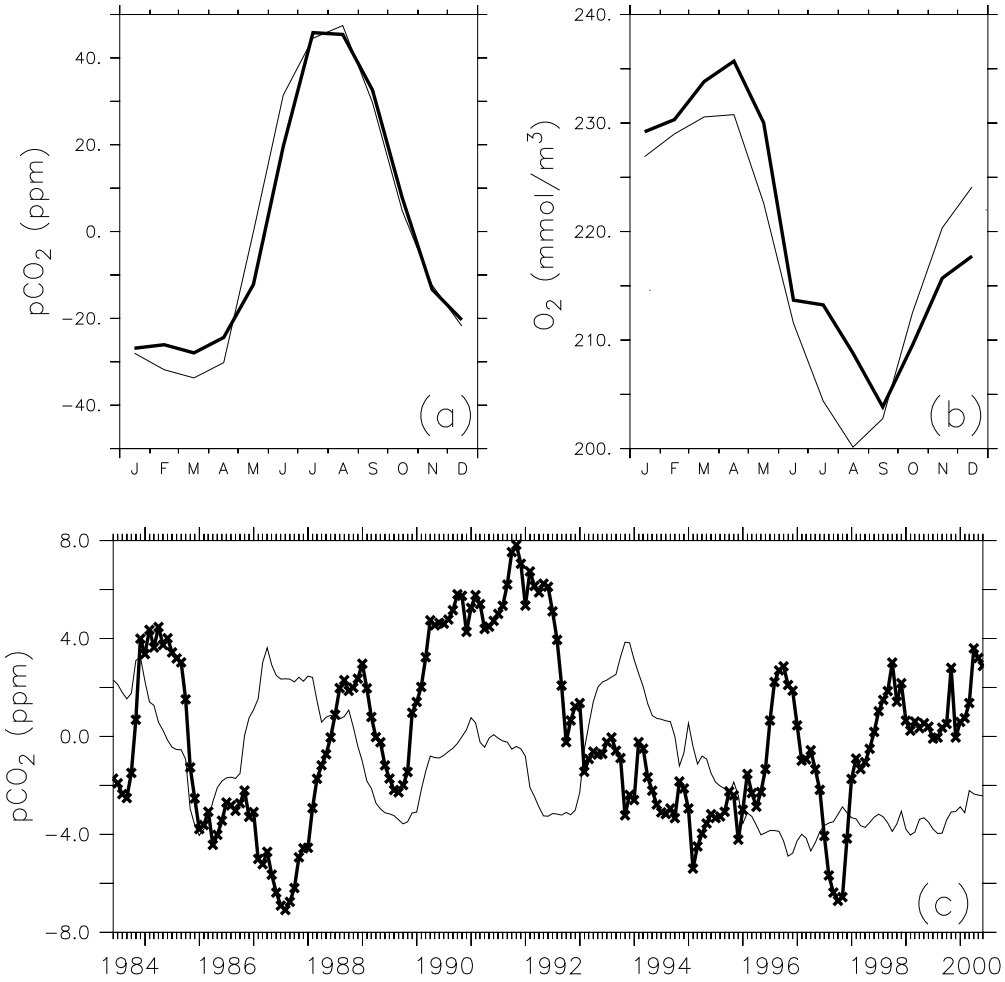

Figure 3: (a): Mean seasonal cycle of oceanic pCO_2 [ppm] (annual mean removed): observed [*Takahashi et al.*, 1997] (bold), simulated by the full forcing run (thin) , (b): mean seasonal cycle of surface O_2 [mmol/m^3]: WOA [*Conkright et al.*, 2002] (bold); simulated by the full forcing run (thin) at Bermuda (32°N, 64°W). (c): Comparison of simulated (thin) to observed (bold, crosses) oceanic pCO_2 at Bermuda Hydrostation S. Mean seasonal cycles (and anthropogenic trend for the observations) were removed. A 12-month running mean was applied. Bermuda pCO_2 data were kindly provided by Niki Gruber.

HFLX-run is highly correlated (r = 0.94) with the uptake simulated by the full forcing run.

In agreement with the results presented by *McKinley et al.* [2004] and *Körtzinger et al.* [2004] this variability is largely driven by interannual vari-

ations in wintertime convection and the associated entrainment of oxygen-depleted waters into the surface mixed layer. Patterns of the first empirical orthogonal functions (EOF) based on annual mean properties are almost identical for the simulated air-sea fluxes of O_2 and of heat and they show similar temporal behavior (r = 0.88). They reveal the typical tripole pattern for the North Atlantic Oscillation (NAO) and explain 28.4% of the interannual variability in simulated air-sea O_2 fluxes (17% for simulated heat fluxes). Regressing air-sea fluxes of O_2 to annual mean NAO-Index [*Hurrel*, 1995] clearly reflects this tripole pattern. During the years of a high NAO-Index O_2 uptake increases in the western subpolar North Atlantic due to increased deep convection, while in the eastern subtropics O_2 fluxes also increase due to the strengthening of the Azores high pressure area, which leads to intensified cooling by enhanced trade winds.
In contrast, the tropics (10°S to 15°N) show almost identical air-sea O_2 fluxes in the WIND and full forcing runs (r = 0.97 for outgassing of O_2). At the equatorial divergence water upwelled to the surface is warmed and becomes oversaturated with respect to O_2 and CO_2, which leads to permanent outgassing of both gases. Since the equatorial divergence is wind-stress driven, changes in outgassing of O_2 are strongly related to wind-stress variability.

Carbon

Simulated total uptake of pre-industrial carbon by the North Atlantic amounts to $0.3\,\mathrm{Pg\,C\,yr^{-1}}$ during our 54-year integration, with a temporal standard deviation of annual mean fluxes of $\pm 0.02\,\mathrm{Pg\,C\,yr^{-1}}$ (Figure 5a). This variability in oceanic carbon uptake is about ten times smaller than estimates from atmospheric inversions of *Bousquet et al.* [2000] but consistent with ocean model results of *McKinley et al.* [2004a], *Le Quéré et al.* [2000], *Wetzel et al.* [2005] and *Raynaud et al.* [2005] and novel atmospheric inversions [*Rödenbeck et al.,* 2003; *McKinley et al.,* 2004]. Our model results indicate that no single part of the model domain dominates the interannual variability of the North Atlantic carbon uptake. Generally smaller fluctuations of subtropical and equatorial CO_2 fluxes compared to those in the subpolar basin are compensated by the larger low latitude area. While the mean pre-industrial carbon uptake simulated by the model is located mainly in the subpolar North Atlantic ($0.22\pm0.01\,\mathrm{Pg\,C\,yr^{-1}}$), standard deviations of the annual air-sea fluxes are of the same magnitude ($0.10\pm0.01\,\mathrm{Pg\,C\,yr^{-1}}$) over the subtropical and the equatorial Atlantic.
Overall, the different model runs reveal a pronounced dominance of wind stress in driving the interannual variability of CO_2 fluxes in the entire Atlantic covered by our model (Figure 5e-g). In the subtropical and subpolar North Atlantic (north of 15°N), the magnitude of the variability of carbon uptake in the WIND run is about 70% of the full forcing run. Temporal changes in carbon uptake simulated by the WIND and the full forcing runs are correlated with r = 0.81 . In the equatorial Atlantic, the relatively shal-

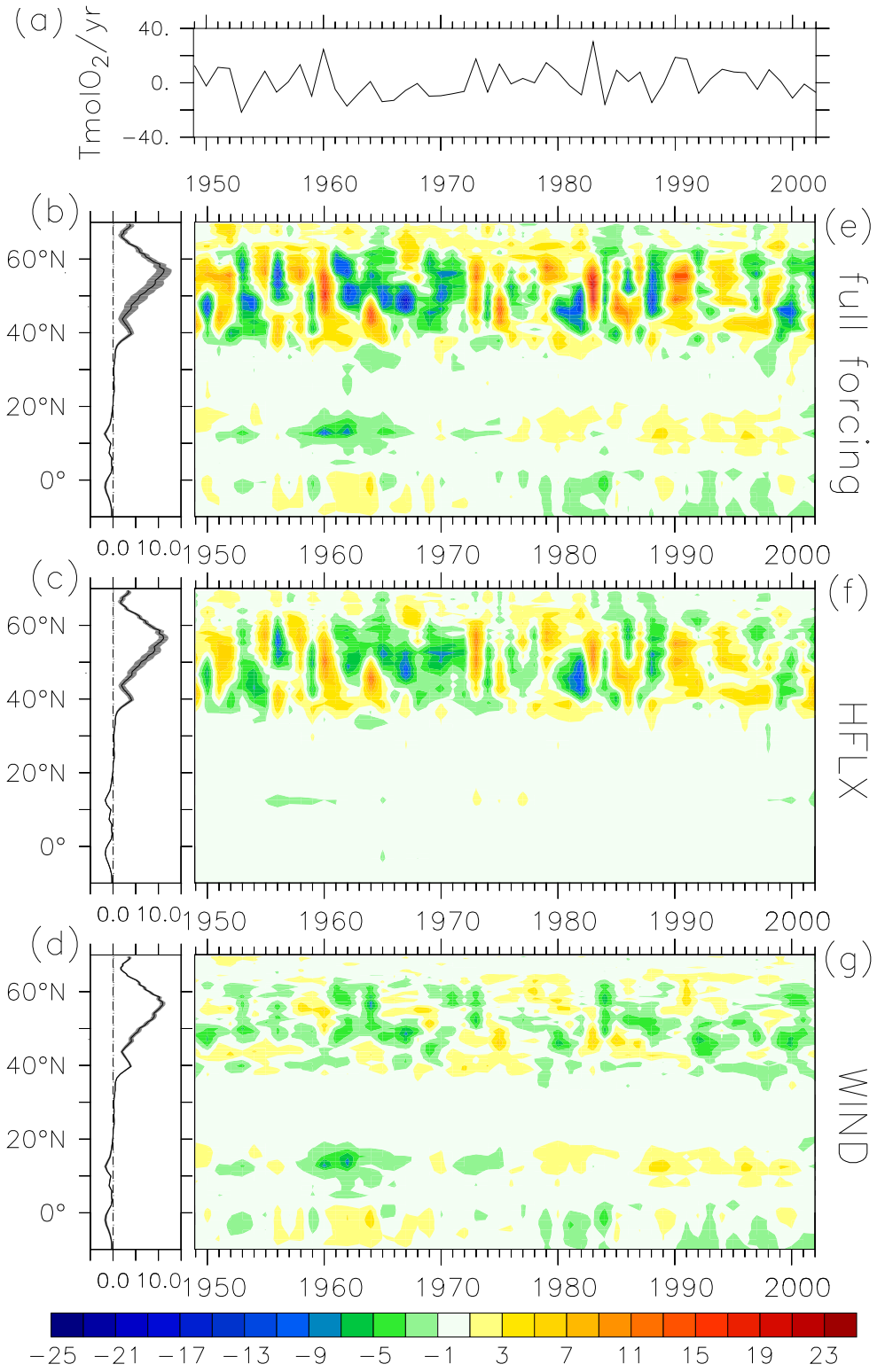

Figure 4: (a): Variability of detrended annual mean values of simulated O_2 uptake 0°N to 70°N in Tmol O_2 yr^{-1}. (b)-(d): Zonally averaged, annual mean, air-to-sea O_2 flux and standard deviation (shaded) of the interannual variability (mol m^{-2} yr^{-1}) for each of the full forcing, HFLX, and WIND simulations. (e)-(g): The corresponding zonally integrated, detrended anomalies in Mmol O_2 m^{-1} yr^{-1}. Positive values denote fluxes into the ocean.

low mixed layers lead to faster equilibration times of CO_2. Thus, CO_2 fluxes can still respond to DIC changes on seasonal to annual time scales and accordingly show a response to forcing mechanisms similar to the O_2 fluxes discussed above. Consequently, interannual changes in air-sea fluxes of CO_2 and O_2 are closely correlated (r = 0.72) near the equator (5°S to 5°N average) in the wind-stress driven equatorial upwelling. The temporal evolution of equatorial outgassing of CO_2 in the the WIND run shows the same high correlation with the full forcing run (r = 0.81) as in the area north of 15°N. Further sensitivity studies show that the dominant impact of variations in wind stress on variations in air-sea CO_2 exchange cannot only be explained by the wind-speed dependence of the piston velocity in the gas exchange formulation. An additional run with climatological piston velocity (run PISTON of Table 1) still yields 75% of the basinwide CO_2 flux variance of the full forcing run. In comparison, the WIND and HFLX runs explain 45% and 27% of the basinwide CO_2 flux variance, respectively.

Chemical buffering of changes in CO_2 concentrations by the large pools of carbonate and bicarbonate ions implies that the CO_2 equilibration time is about ten-times longer than for O_2. Thus rapid changes in DIC concentrations or in surface heat fluxes cannot be balanced by air-sea CO_2 fluxes on seasonal time scales in regions of deep wintertime mixed layers [*Broecker and Peng*, 1974]. This is consistent with our model results showing that simulated variability in the subpolar mixed layer's carbon inventory due to entrainment of DIC-rich waters by convection is more than ten times as important as are the variations due to air-sea CO_2 fluxes. In consequence, a geographical separation into heat-flux and wind-stress driven regions similar to those for O_2 cannot readily be established for the air-sea exchange of CO_2 over the North Atlantic. In the subtropical gyre our model results show a temporal correlation of 0.87 between $\frac{\partial pCO_2}{\partial SST}\frac{dSST}{dt}$ and $\frac{dpCO_2}{dt}$. This indicates an increased CO_2 uptake in this region by intensified cooling, e.g., associated with a negative NAO-phase and driven by changes in both enhanced wind stirring and reduced surface heat input. Our results confirm earlier findigs of *McKinley et al.* [2004a] and *Le Quéré et al.* [2003] suggesting that SST is the primary control of interannual variability of pCO_2 in the subtropical gyre.

Summary

Sensitivity tests with a coupled physical-biogeochemical model have allowed us to discern the major atmospheric driving mechanisms of interannual variability of air-sea O_2 exchange in the North and equatorial Atlantic. In the model simulations, the subpolar basin contributes most of the North Atlantic O_2 uptake and most of its interannual variability. The latter is driven by changes in wintertime convection controlled predominantly by the NAO through fluctuations in surface heat fluxes. On the other hand, outgassing

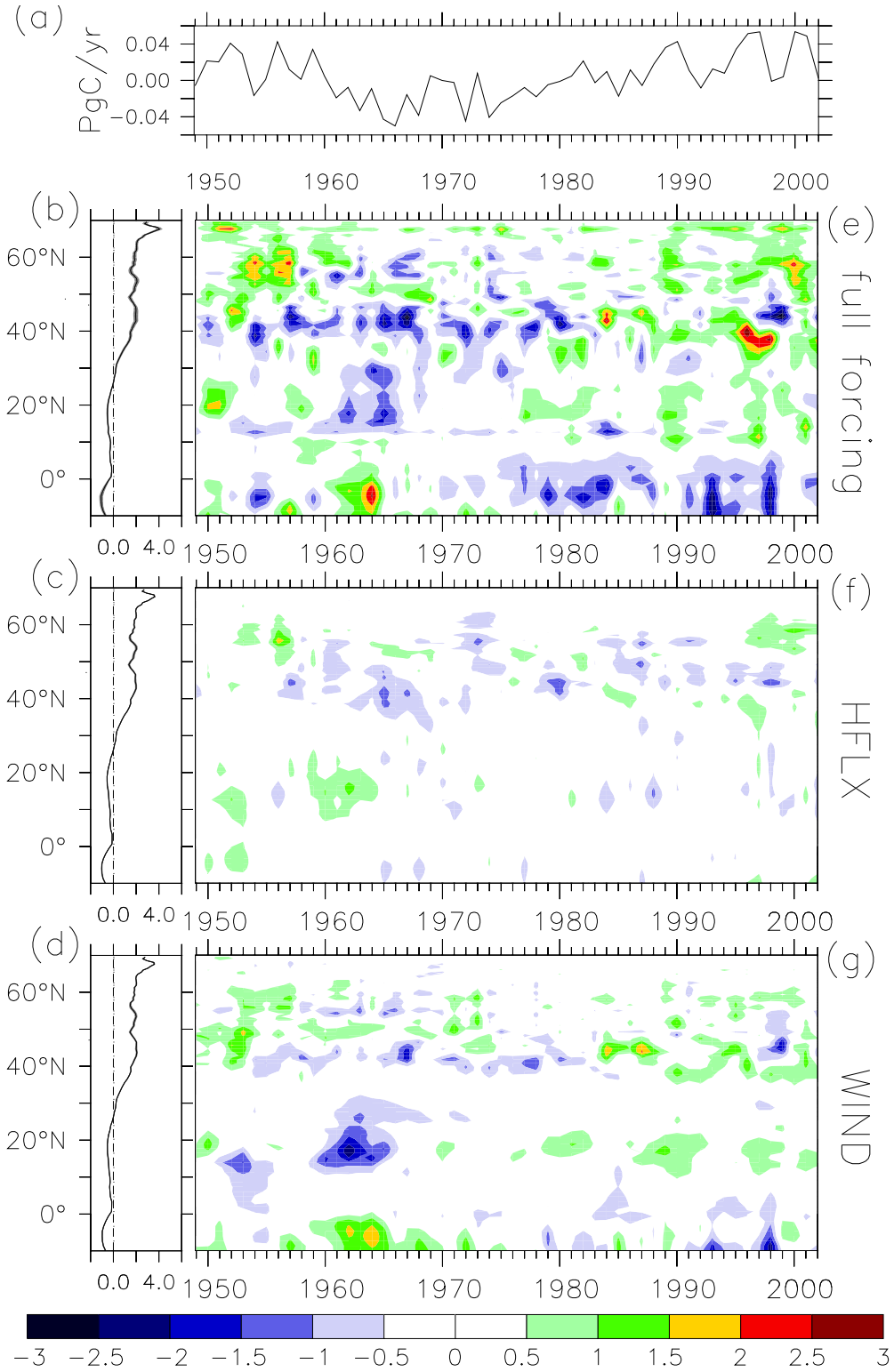

Figure 5: (a): Variability of detrended annual mean values of simulated CO_2 uptake 0°N to 70°N in Pg C yr^{-1}. (b)-(d): Zonally averaged, annual mean, air-to-sea CO_2 flux and standard deviation (shaded) of the interannual variability (mol m^{-2} yr^{-1}) for each of the full forcing, HFLX, and WIND simulations. (e)-(g): The corresponding zonally integrated, detrended anomalies in Mmol C m^{-1} yr^{-1}. Positive values denote fluxes into the ocean.

of O_2 is strongly related to changes in wind stress in the equatorial Atlantic. Interannual variability of CO_2 fluxes is driven mainly by wind stress variability over the entire model domain; only minor changes derive from variability in piston velocity and heat fluxes. Although the simulated mean carbon uptake is mostly located in the subpolar region, year-to-year fluctuations of the oceanic carbon uptake are of the same magnitude in the subpolar region, the subtropics and the equatorial Atlantic.

Acknowledgments
This work was partially funded by the DFG via SFB 460 and by the European Union via the IPs NOCES EVK2-2001-00115 and CARBOOCEAN 511176 (GOCE). We thank the two anonymous reviewers for their constructive comments that have helped to improve the paper and Niki Gruber for providing us with the Bermuda pCO_2 data.

References

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.

Bates, N. R., A. C. Pequignet, R. J. Johnson and N. Gruber (1999), A short term sink for atmospheric CO_2, *Nature*, *420*, 489–493.

Bopp, L., C. Le Quéré, M. Heimann and A. C. Manning (2002), Climate induced oceanic oxygen fluxes: Implications for the contemporary carbon budget, *Global Biogeochem. Cycles*, *16*, doi:10.1029/2001GB001445.

Boyer, T. P. and S. Levitus (1997), Objective analyses of temperature and salinity for the world ocean on a 1/4 degree grid. Technical report, NOAA Atlas NESDIS 11, U.S. Gov. Printing Office, Washington, D.C.

Bousquet, P., P. Peylin, P. Ciais, C. Le Quéré, P. Friedlingstein and P. Tans (2000), Regional changes in carbon dioxide fluxes of land and ocean since 1980, *Science*, *290*, 1342–1345.

Broecker, W. S. and T. H. Peng (1974), Gas exchange rates between air and sea, *Tellus*, *26*, 21–35.

Conkright, M. E., R. A. Locarnini, H. E. Garcia, T. D. O'Brien, T. P. Boyer, C. Stephens and J. I. Antonov (2002), World Ocean Atlas 2001: Objective analyses, data statistics, and figures, CD-ROM documentation. National Oceanographic Data Center, Silver Spring, MD, 17 pp.

Eden, C. and T. Jung (2000), North Atlantic interdecadal variability: Oceanic response to the North Atlantic oscillation (1865-1997), *J. Climate*, *14*, 676–691.

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.

Eden, C. and J. Willebrand (2001), Mechanisms of interannual to decadal variability of the North Atlantic circulation, *J. Climate*, *14*, 2266–2280.

Gruber, N., C. D. Keeling and N. R. Bates (2002), Interannual variability in the North Atlantic ocean carbon sink, *Science*, *298*, 2374–2378.

Hurrell, J. W. (1995), Decadal trends in the North Atlantic oscillation: Regional temperatures and precipitation, *Science*, *269*, 676–679.

Kalnay, E. et al. (1996), The NCEP/NCAR 40 years reanalysis project, *Bull. Amer. Meteor. Soc.*, *77*, 437–471.

Keeling, R. F. and S. R. Shertz (1992), Seasonal and interannual variations in atmospheric oxygen and implications for the global carbon cycle, *Nature*, *358*, 723–727.

Körtzinger, A., J. Schimanski, U. Send and D. Wallace (2004), The ocean takes a deep breath, *Science*, *306*, 1337.

Lafore, J.-P. et al. (1998), The Meso-NH Atmospheric simulation system. Part I: Adiabatic formulation and control simulations, *Annales Geophysicae*, *16*, 90–109.

Leonard, B. P. (1992), A stable and accurate convective modelling procedure based on quadratic upstream interpolation, *Comp. Meth. Appl. Mech. Eng.*, *19*, 59–98.

Le Quéré, C., J. C. Orr, P. Monfray and O. Aumont (2000), Interannual variability of the oceanic sink of CO_2 from 1979 through 1997, *Global Biogeochem. Cycles*, *14*, 1247–1265.

Le Quéré, C., O. Aumont, P. Monfray and J. C. Orr (2003), Propagation of climatic events on ocean stratification, marine biology and CO2: case studies over the 1979-1999 period, *J. Geophys. Res.*, *108* (C12), 3375, doi:10.1029/2001JC000920.

McKinley, G. A., M. J. Follows and J. Marshall (2000), Interannual variability of the air-sea flux of oxygen in the North Atlantic, *Geophys. Res. Lett.*, *27*, 2933–2936, doi:10.1029/2000GL011492.

McKinley, G. A., M. J. Follows, J. Marshall and S. Fan (2003), Interannual variability of the air-sea flux of O_2 fluxes and the determination of CO_2 sinks using atmospheric O_2/N_2, *Geophys. Res. Lett.*, *30*, 1101, doi:10.1029/2002GL016044.

McKinley, G. A., M. J. Follows and J. Marshall (2004), Mechanisms of air-sea CO_2 flux variability in the equatorial Pacific and the North Atlantic, *Global Biogeochem. Cycles*, *18*, GB2011, doi:10.1029/2003GB002179.

McKinley, G. A., C. Rödenbeck, M. Gloor, S. Houweling and M. Heimann (2004), Pacific dominance to global air-sea CO2 flux variablility: A novel atmospheric inversion agress with ocean models, *Geophys. Res. Lett.*, *31*, L22308, doi:10.1029/2004GL021069.

Oschlies, A. and V. Garçon (1999), An eddy-permitting coupled physical-biological model of the North Atlantic. Part 1: Sensitivity to advection numerics and mixed layer physics, *Global Biogeochem. Cycles*, *13*, 135–160, doi:10.1029/98GB02811.

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

Raynaud S., O. Aumont, K. B. Rodgers, P. Yiou and J. C. Orr (2005), Interannual-to-decadal variability of North Atlantic air-sea CO_2 fluxes, *Ocean Science Discussion*, *2*, 437–472.

Rödenbeck, C., S. Houweling, M. Gloor, M. Heimann (2003), CO_2 flux history 1982-2001 inferred from atmospheric data using a global inversion of atmospheric transport, *Atmos. Chem. Phys.*, *3*, 1919–1964.

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.

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

Wetzel, P., A. Winguth, E. Maier-Reimer (2005), Sea-to-air CO_2 flux from 1948 to 2003: A model study, *Global Biogeochem. Cycles*, *19*, GB2005, doi:10.1029/2004GB002339 .

Introduction to neural-network based mapping of pCO_2

Estimating pCO_2 from related parameters requires a formulation which takes into account the dependencies between these parameters and pCO_2. Unfortunately, a stringent relation in a mathematical sense between for example SST and pCO_2 does not exist, even though simple regression techniques may work well for data restricted in time and space. This fact is nicely illustrated by two studies of *Watson et al.* [1991] and *Lefèvre et al.* [2002] carried out in the subpolar and subtropical North Atlantic respectively, who found promising linear relations between pCO_2 and SST with coefficients being of the same magnitude in both studies but of opposite sign.
In our model study, the availability of data over the course of a year on an eddy-resolving grid disclosed the opportunity to take a look at physical and biological relationships between pCO_2 and for example SST or Chl respectively. The results of *Watson et al.* [1991] and *Lefèvre et al.* [2002] were relatively well reproduced when applying a multiple linear regression to our simulated data of SST, Chl and pCO_2 (Figure 6).

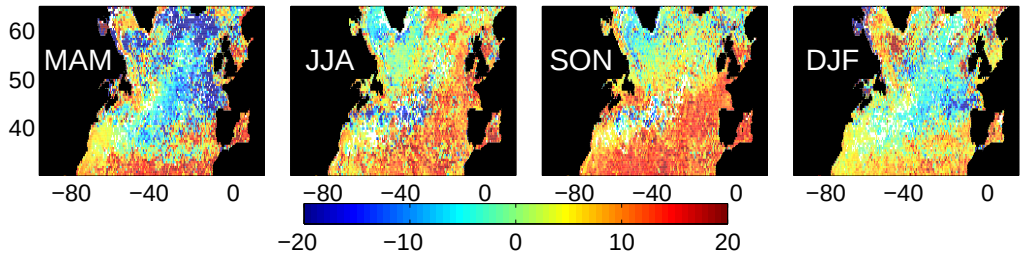

Figure 6: Seasonal mean regression coefficients of pCO_2 versus SST in ppm/K for a multiple linear regression applied monthly to the model data on a $1/2^\circ \times 1/2^\circ$ grid. White areas indicate an explained variance $< 50\%$.

The regression coefficients for SST shown in Figure 6 are subject to a considerable variability in both time and space. In spring (MAM) they are of opposite sign to that expected if warming of surface waters was mainly responsible. As an increasing solar radiation causes a warming in spring and a spring bloom, it is the biological drawdown of CO_2 which has the major impact in this season. In winter (DJF) a cooling of surface waters by strong cold winds leads to an enhanced vertical mixing of the water column. This is associated with a reduction of the vertical gradient in dissolved inorganic carbon (DIC) and an entrainment of DIC-rich waters and thus with an increase of surface pCO_2. Again, the sign of the temperature dependency is opposite to that expected from a simple cooling of water.
Overall, the significant fluctuations of the SST coefficient over relatively short

scales in time and space found in this model-based investigation indicates that a multiple linear regression is not a suitable technique for the interpolation of sparse data.

In the present study a neural network approach using a self-organizing map (SOM) was accomplished in order to estimate pCO_2 from related parameters. SOMs were introduced to science by Teuvo Kohonen (thus often referred to as Kohonen Feature Maps, KFM) in 1982 [*Kohonen*, 1982]. SOMs can recognize relationships in the data during the training process and are able to associate a parameter vector (e.g. [SST, Chl]) with a target value (e.g. pCO_2) in their application. The biological motivation, the training process and the functioning will be discussed in the following.
It is a well known fact in neurobiology that many structures in the brain are of planar topology [*Holmes*, 1918; *Glickstein*, 1988]. Since, for example, sensory experiences, are multidimensional, the question is how a planar structure can process multidimensional input.

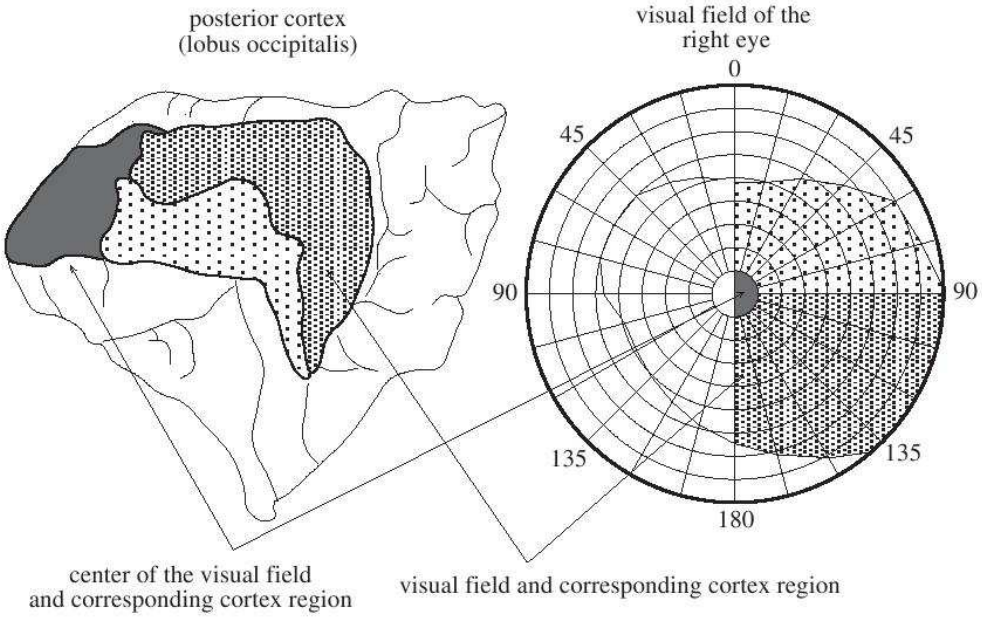

Figure 7: Mapping of the visual field on the cortex. Figure taken from *Neural Networks* by Raùl Rojas.

In the visual cortex exterior spatial relations contained in the visual information are mapped as a two-dimensional projection in the way that neighbouring regions in the visual field are processed by neighbouring regions in the cortex and signals from the center of the visual field are processed with higher resolution (Figure 7). Observations like these motivated scientists to mimic the brain's capability to create a low-dimensional representation of

a high-dimensional input with artificial computing units. Teuvo Kohonen discovered a principle which facilitates an automatical formation of a topological signal representation (self-organization). On the basis of a simplified example using VOS-line observations of SST, Chl and pCO_2 this process of self-organization shall be explained in the subsequent paragraph.

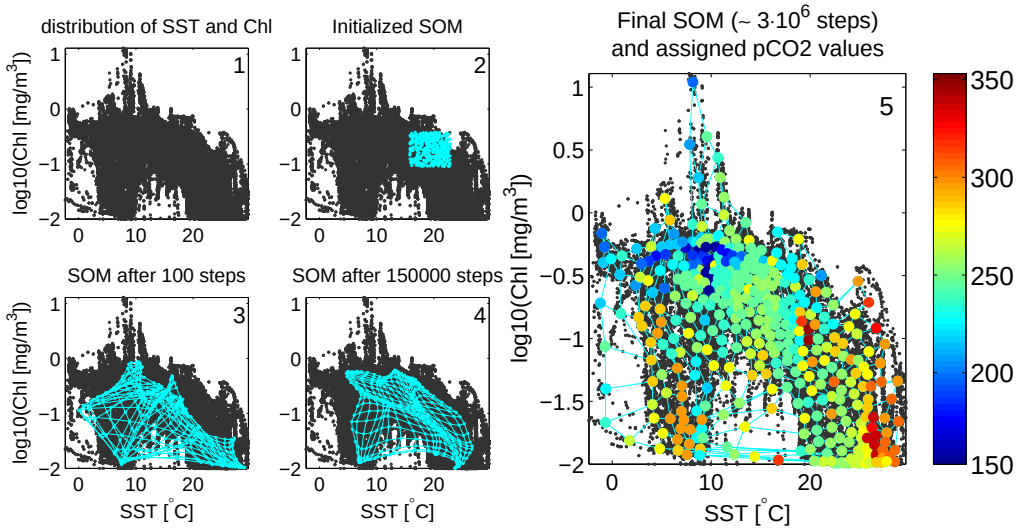

Figure 8: Formation of a SOM. Dimension and size of the SOM were reduced for illustration. See text for details.

First of all, samples of SST and Chl are normalized by their mean and standard deviation:

$$SST' = \frac{SST - \overline{SST}}{STD(SST)} \tag{1}$$

$$Chl' = \frac{log(Chl + 0.01) - \overline{log(Chl + 0.01)}}{STD(log(Chl + 0.01))} \tag{2}$$

For Chl a log-normal distribution is assumed following *Campbell* [1995]. A value of 0.01 $\frac{mgChl}{m^3}$ is added to avoid a large impact of data uncertainties at very low Chl concentrations.
The distribution of data is shown in Figure 8, panel 1. Data were renormalized for clarity.
At the beginning of the training process a network (in this case 20×20 2-dimensional vectors) of small random values between 0 and 1 is initialized (Figure 8, panel 2, vectors were renormalized). The initial neighbourhood radius (N(0)) is set to be approximately half the network radius (N(0)=6 for present example) to avoid a topological defect (a "knot") in the SOM. The

initial adaptation rate ($\eta(0)$) is set between 0 and 1 (e.g. ~ 0.9).

The iterative training process consists of 5 steps:

1. A randomly chosen vector from the normalized training data set x(index)=[SST', Chl'] is presented as a stimulus to the network and the euclidian distance between the stimulus vector and each vector of the network is calculated.

2. The best matching vector v_{ij} of the network (the "winner") is determined by a minimum distance criterion.

3. The "winner" is updated. Size and direction of v_{ij} are modified towards the stimulus vector:

$$v_{ij}(index+1) = v_{ij}(index) + \eta(index) \cdot (x(index) - v_{ij}(index)) \quad (3)$$

   All vectors in the neighbourhood N(index) are updated in the same way, but the adaptation rate η(index) decreases exponentially with distance from the "winner" (see equation 6 and Figure 9c).

4. Neighbourhood N(index) and adaptation rate η(index) decrease exponentially with increasing training index (see equations 4, 5 and Figure 9a, b).

5. Go back to 1. until convergence.

The inflation of the SOM during the training process is illustrated by Figure 8. In the initial phase, the updated neighbourhood of the "winner" and the adaptation rate are large and the shape of the SOM is strongly affected by the current stimulus vector (panel 3). With ongoing training the updated neighbourhood and the adaptation rate decrease and the SOM converges in a coverage of the training data distribution, in which areas with higher training data density are accounted for with better resolution of the SOM. Thus the SOM can be regarded as a lower dimensional topology preserving parameter space representation.

To achieve a convergence, the number of iterations during the training process (index_{max}) was set to the 28-fold length of the training data set after testing different values. Figure 9 indicates the exponential decrease of the updated neighbourhood and adaptation rate with training index (in multiples of training data set length) as well as the exponential decrease of the adaptation rate with distance from the "winner" (for index=1). The underlying equations are:

$$N(index) = N(0) \cdot e^{-(\frac{2 \cdot index}{index_{max}})^2} \quad (4)$$

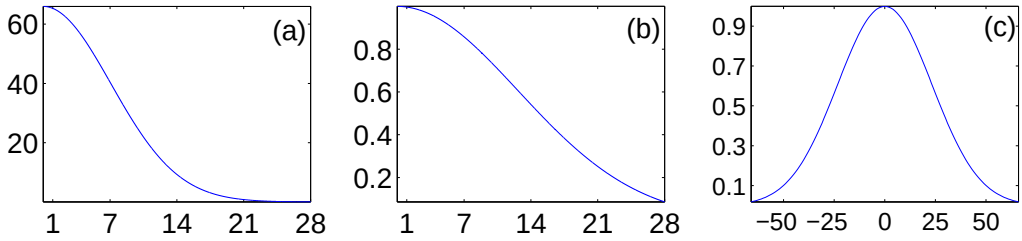

Figure 9: (a): Exponential decrease of updated neighbourhood with training index in multiples of training data set length. (b): Exponential decrease of adaptation rate with training index in multiples of training data set length. (c): Exponential decrease of adaptation rate with distance from best matching vector for an initial neighbourhood N(0)=66.

$$\eta(index) = \eta(0) \cdot e^{-5 \cdot (\frac{index}{2 * index_{max}})^2} \tag{5}$$

$$\eta(distance) = \eta(index) \cdot e^{-(\frac{distance}{N(index)/2})^2} \tag{6}$$

Finally, pCO_2 values need to be assigned to the SOM. Several techniques are possible here. First of all, in addition to the network representing the parameter space, a pCO_2 map with random values (scattered around the mean pCO_2 value) can be initialized at the beginning of the training process. During the training the value of the pCO_2 map associated with the coordinates (i,j) of the best matching vector v_{ij} of the network is updated in exactly the same way as described by step 3. The modification is done towards the pCO_2 value (pCO_2(index)) belonging to the samples of x(index)=[SST', Chl']. In this case the formation of the pCO_2 map is achieved during the training process which normally saves calculating time.
Alternatively the assigning can be done subsequent to the training process. For each vector of the SOM the euclidian distances between this vector and all training data vectors are calculated. The minimum distance criterion again determines the best matching vector of the training data set. The pCO_2 value belonging to this best matching vector is then assigned to the considered SOM vector.
This procedure can also be done the other way round. In this case the euclidian distances between each vector of the training data set and all SOM vectors are calculated and the pCO_2 value belonging to the considered training data vector is assigned to the best matching SOM vector. Here, SOM vectors can remain unassigned and the pCO_2 value needs to be determined by an interpolation between neighbourhood values.
In the present study, after comparing the mapping accuracy for all three methods, the third procedure was used for assigning pCO_2 values to the

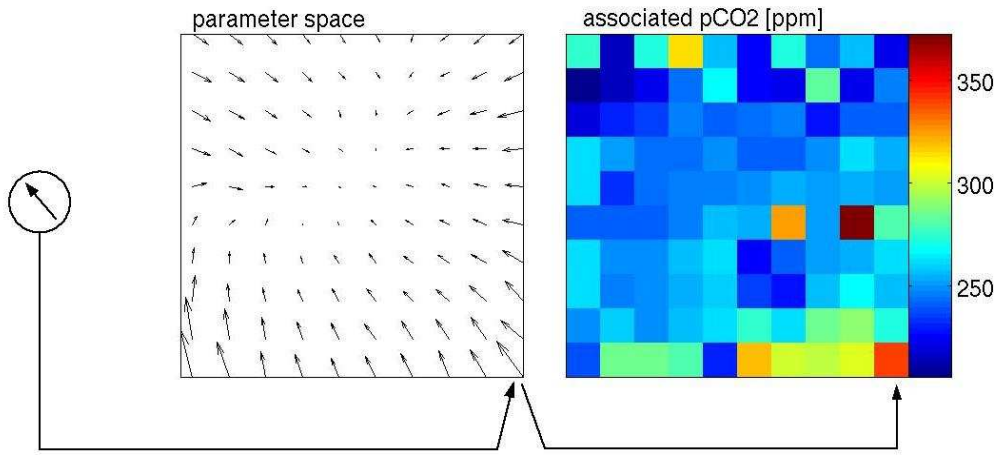

Figure 10: Schematic mode of operation of a trained KFM. An input vector (left) is associated with the pCO_2 value (right) of the best matching vector (middle) of the parameter space representation. Note: For clarity, the dimension of the KFM was reduced to 10×10 and the input (and parameter) space was chosen to be two-dimensional (SST and Chl only).

trained SOM. Values for unassigned vectors were determined by a linear interpolation in which pCO_2 values of the first two neighbourhood squares surrounding the unassigned vector were considered and direct neighbours of the unassigned vector were weighted twice.

The application of a trained SOM is similar to the assigning of pCO_2 values. A vector consisting of the same parameters (e.g. remotely sensed SST and Chl) which have to be normalized in the same way is presented to the SOM (Figure 10). The parameter vector is then associated with the pCO_2 value of the best matching vector of the SOM determined by the minimum distance criterion.

Please note that the example presented above was simplified for illustration. An attempt to map pCO_2 exclusively from simulated VOS-line observations of SST and Chl did not succeed. The additional use of the sampling position and time (day in the year) as training data was inevitable. In this case the input data were normalized as follows:

$$Day' = cos\frac{Day \cdot 2\pi}{365} \tag{7}$$

$$Lat' = \frac{latitude - 15°N}{65°N - 15°N} \tag{8}$$

$$Lon' = \frac{longitude - 88°W}{11°E - 88°W} \tag{9}$$

$$SST' = \frac{SST - \overline{SST}}{STD(SST)} \tag{10}$$

$$Chl' = \frac{log(Chl + 0.01) - \overline{log(Chl + 0.01)}}{STD(log(Chl + 0.01))} \tag{11}$$

Latitude and Longitude were normalized by their maxima and minima. Here, only data between 15°N and 65°N were considered to ensure a monthly VOS-line coverage. The input for "Day" was normalized sinusoidally, following *Lefèvre et al.* [2005] and emphasizing the seasonal cycle and allowing January to be close to both February and December.
In order to find an optimal size for the SOM, 16 KFMs (20×20, 40×40, ..., 320×320 vectors) were trained with the synthetic VOS-line observations of the reference year 2005. The basinwide RMS-error of the estimated pCO_2 as a function of the map size showed a substantial decrease of the mapping error on increasing the number of vectors from 20×20 to 220×220, and a slightly increasing RMS-error thereafter. Thus, a SOM size of 220×220 vectors and an initial neighbourhood size N(0)=66 was chosen.

The technical literature on SOMs offers a broad and detailed insight to a theoretical understanding but is usually lacking in practical advice. Most of the equations and values determining the training procedures described above were found by combining theoretical considerations and practical trial-and-error optimizations. Simply testing mapping errors for different SOM configurations remained the most powerful tool.

Alternative methods are conceivable as well. For SOM-based pCO_2 mapping it is also possible to obtain a topological representation of the parameter space by a training with remotely sensed data. In this case, VOS-line samples are only used to assign pCO_2 values to the trained map. So far it is not clear whether this is a beneficial strategy. It impedes the additional use of position in time and space as training parameters. Furthermore, remote sensing errors can contaminate the training data and result in a deterioration of the parameter space topology. Hopefully, this will be subject of future studies.
Backpropagation neural networks provide the option of a non-linear regression between pCO_2 and related parameters. Their potential for mapping pCO_2 has been analyzed in a preliminary study. Unfortunately, a suitable configuration ensuring an accurate reproduction of data not contained in the training has not been found.

In the present study no use has been made of available neural network software. In contrast to simple and mostly self-explanatory regression tech-

niques, the functioning of neural networks can easily remain a black box. Thus, backpropagation neural networks and SOMs were developed by the author to ensure a comprehension as extensive as possible.

References

Campbell, J. W. (1995), The lognormal distribution as a model for bio-optical variability in the sea, *J. Geophys. Res.*, *100*(C7), 13237-13254, doi:10.1029/95JC00458.

Glickstein, M. (1988), The discovery of the visual cortex, *Scientific American*, *259*, 84–91.

Holmes, G. (1918), Disturbances of the visual orientation, *The British Journal of Ophthalmology*, *2(1918)*, 449–468.

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

Lefévre, N., A. Taylor (2002), Estimating pCO_2 from sea surface temperatures in the Atlantic gyres, *Deep Sea Res., Part I*, *49*, 539–554.

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

Watson, A. J., C. Robinson, J. E. Robinson, P. J. le B. Williams, M. J. R. Fasham (1991), Spatial variability in the sink for atmospheric carbon dioxide in the North Atlantic, *Letters to Nature*, *350*, 50–53.

Helpful literature

Neural Networks, Raùl Rojas, 1996.
(free download at: http://page.mi.fu-berlin.de/rojas/neural)

Theorie der neuronalen Netze : eine systematische Einführung, Raùl Rojas, 1993.

Self-Organizing Maps, Teuvo Kohonen, 2001.

Neuronale Netze : eine Einführung in die Neuroinformatik selbstorganisierender Netzwerke, Helge Ritter, Thomas Martinetz, Klaus Schulten, 1990.

Ein kleiner Überblick über Neuronale Netze, David Kriesel, 2007.
(free download at: http://www.dkriesel.com)
An english version is coming soon.

Neuronale Netze : eine Einführung in die Neuroinformatik, Rüdiger Brause, 1991.

Programmierung neuronaler Netze, Hilger Kruse, 1991.

Praktische Einführung in neuronale Netze, Alessandro Mazzetti, 1992.

Neural-network based estimates of North Atlantic surface pCO_2 from satellite data - a methodological study

–*Journal of Geophysical Research*, accepted–

Tobias Friedrich[1]
(corresponding author)

Andreas Oschlies[1]

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

Abstract

A new method is proposed to estimate ocean-surface pCO_2 from remotely sensed surface temperature and chlorophyll data. The method is applied to synthetic observations provided by an eddy-resolving biogeochemical model of the North Atlantic. The same model also provides a perfectly known simulated pCO_2 "ground truth" used to quantitatively assess the success of the estimation method. Model output is first sampled according to realistic Voluntary Observing Ship (VOS) and satellite coverage. The model-generated VOS "observations" are then used to train a self-organizing neural network that is subsequently applied to model-generated "satellite data" of surface temperature and surface chlorophyll in order to derive basinwide monthly maps of surface pCO_2. The accuracy of the estimated pCO_2 maps is analyzed with respect to the "true" surface pCO_2 fields simulated by the biogeochemical circulation model. We also investigate the accuracy of the estimated pCO_2 maps as a function of VOS-line coverage, remote sensing errors, and the interpolation of missing remote sensing data due to cloud cover and low solar irradiation in winter. For a simulated "sampling" corresponding to VOS lines and patterns of optical satellite coverage of the year 2005, the neural net can successfully reproduce pCO_2 from model-generated "remote sensing data" of SST and Chl. Basinwide RMS-errors amount to 19.0 μatm for a hypothetical perfect interpolation scheme for remote-sensing data gaps, and 21.1 μatm when climatological surface temperature and chlorophyll values are used to fill in areas lacking optical satellite coverage.

Introduction

With respect to the Earth's radiation balance, CO_2 is the most important anthropogenically affected greenhouse gas, whose content in the atmosphere has increased by 30% since the beginning of the industrial revolution. Still, the atmospheric increase constitutes only about half of the accumulated anthropogenic emissions. The ocean is thought to have absorbed some 40% of these emissions so far [*Sabine et al.,* 2004]. However, estimates of regional patterns and temporal fluctuations of the ocean's CO_2 uptake are prone to large uncertainties. Reducing the uncertainties in air-sea CO_2 fluxes on synoptic scales is also relevant for improved estimates of regional terrestrial carbon sources and sinks, which eventually have to be accounted for in suitable schemes to monitor and control anthropogenic CO_2 emissions.
A direct calculation of the air-to-sea CO_2 flux requires detailed knowledge of the gas-transfer velocity and the partial pressure of CO_2 (pCO_2) in the ocean's surface waters. In the Atlantic Ocean a pCO_2 monitoring network was initialized by the European project CAVASSOO (Carbon Variability Studies by Ships of Opportunity) in 2001. It is currently extended with better coverage and higher measurement accuracy by the European CARBOOCEAN project in cooperation with US-American partners (Table 2).
In contrast to CO_2 concentrations in the well mixed atmosphere, seawater pCO_2 is subject to high spatial and temporal variability at the ocean's surface as a result of changes in solubility, mainly driven by temperature. Furthermore, marine biology alters the total dissolved inorganic carbon (DIC) content of the water which affects pCO_2. Voluntary observing ship (VOS) measurements analyzed by *Lüger et al.* [2004] exhibited a pCO_2 variance of 23.8 μatm with a peak-to-peak amplitude of about 200 μatm. In the presence of such a high pCO_2 variability one cannot exclusively rely on simple interpolation between isolated local measurements. Alternative techniques need to be investigated to obtain reliable estimates of basinwide pCO_2 on synoptic time scales. When targeting the monitoring of anthropogenic CO_2 fluxes, the accuracy of any pCO_2 estimate has to be quite high. For example, *Watson et al.* [1991] noted that a bias of 1 μatm in a global ΔpCO_2 (atmospheric pCO_2 minus sea-surface pCO_2) estimate would result in an uncertainty in the calculated ocean carbon uptake of $\sim$ 0.2 Pg/yr which is about 10 % of the total uptake estimated by *Takahashi et al.* [2002].
Ocean-surface pCO_2 depends on sea level pressure, DIC, total alkalinity, sea surface temperature (SST) and salinity, of which currently only SST is observable by satellite-based remote sensing on a basinwide scale. Thus a straightforward calculation of basinwide pCO_2 is impeded by the limited data availability. In addition to SST, remote sensing of ocean color, from which one can estimate concentrations of surface chlorophyll (Chl), are typically delivered daily with, apart from cloudy and high-latitude winter scenes, essentially global coverage. These two remotely sensed data sets contain valuable information about the two main physical and biological drivers of the above

variables determining pCO_2. Although approximative regressions may work well in some circumstances [*Watson et al.*, 1991; *Bates et al.*, 1995; *Lüger et al.*, 2004; *Lefèvre et al.*, 2002, 2005], a stringent algebraic relation between SST, Chl and pCO_2, does not exist. For SST, the simple temperature dependency of surface pCO_2 leading to a decrease in pCO_2 with decreasing SST is often superimposed by entrainment of DIC-rich water due to vertical mixing caused by a cooling of the ocean surface. The Chl-pCO_2 relation depends strongly on season. The spring-bloom increase in Chl concentrations usually causes a significant decrease in pCO_2, whereas variations in summer Chl concentrations can be totally decoupled from pCO_2. Furthermore, the relationship between satellite-derived surface Chl and phytoplankton biomass is complicated and depends, among others, on the actual physiological state of the phytoplankton.

In the present study we therefore explore the potential of non-algebraically combining local VOS measurements of pCO_2 and related variables (SST, Chl) with global satellite data of SST and Chl (but not pCO_2) for generating high-resolution monthly maps of surface pCO_2 estimates for an entire ocean basin.

VOS-line based sampling of pCO_2, SST and Chl is simulated in the framework of a high resolution biogeochemical circulation model of the North Atlantic. A neural network (Kohonen Feature Map, KFM) is trained with these synthetic "observations" to reproduce model-generated "measurements" of pCO_2 from SST and Chl along the modeled VOS lines. The KFM trained by the model-generated VOS-line data is then applied on the basin scale to model-generated remote sensing data of SST and Chl. The accuracy of the resulting basinwide pCO_2 maps is analyzed with regard to missing data due to clouds and low solar irradiation in winter, remote sensing errors, and spatio-temporal coverage of the VOS-line data used to train the KFM.

The paper is organized as follows: In the subsequent section we present the high-resolution biogeochemical circulation model. The procedure used to simulate VOS-line sampling and remote sensing is described in section 3. Section 4 deals with the neural network and its configuration. The presentation and discussion of our results (section 5) is split into a simulation to test different scenarios without the interference of remote sensing data gaps, the handling of missing data in the remote sensing of SST and Chl, CO_2 fluxes calculated from pCO_2 estimates, and an evaluation of our method. Section 6 summarizes our results.

Model and Forcing

A pelagic nitrogen-based nutrient-phytoplankton-zooplankton-detritus ecosystem model [*Oschlies and Garçon*, 1999] is coupled to a 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 [*Pakanowski*, 1995]. The model domain spans the Atlantic Ocean from 18°S to 70°N at a horizontal resolution of 1/12° × 1/12° 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, a Haney-type heat flux condition as given by *Barnier et al.* [1995] and a restoring condition for sea surface salinity. A sensitivity experiment was run with daily ECMWF forcing as described in *Eden and Jung* [2006]. 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].

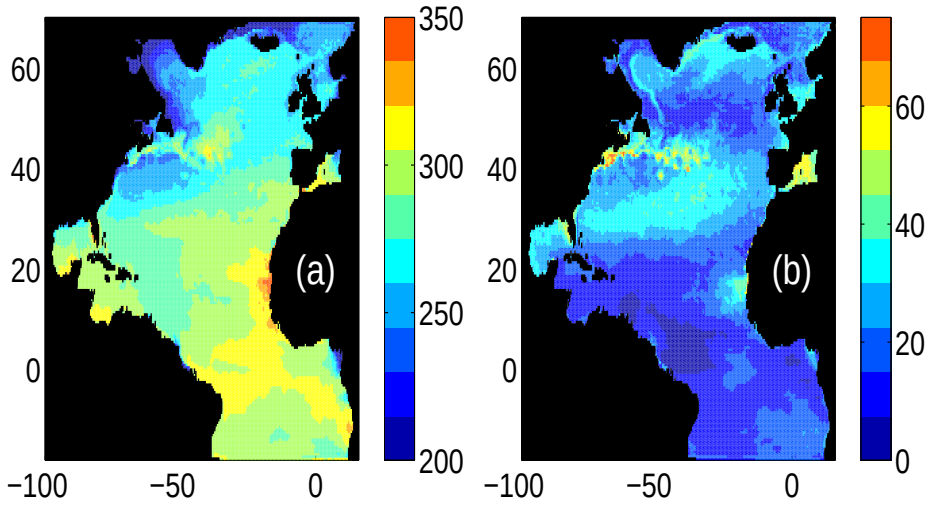

Figure 11: Mean (a) and variance (b) of modeled surface pCO_2 in μatm subsampled on 1/2° × 1/2° grid. Note the pre-industrial atmospheric pCO_2 setup of the model.

Initial conditions for nitrate are taken from *Conkright et al.* [2002], and for dissolved inorganic carbon from the pre-industrial estimate of the GLODAP data set [*Key et al.*, 2004].The atmospheric pCO_2 remains on a pre-industrial level, but varies seasonally and latitudinally according to a nonlinear fit to observational estimates by *Conway et al.* [1994]. The coupled biogeochemical-physical model is integrated over a 10 year spin-up period, and model data used here for simulating VOS-line based observations and remote sensing are taken from the model year 11. The annual mean and variance of simulated surface pCO_2 is shown in Figure 11.

Data

The VOS-line coverage of the North Atlantic in the year 2005 was, until then, the most comprehensive one in the VOS pCO_2 sampling history. Approximately 740,000 measurements had been gathered in the area 10°S to 70°N ($\sim$ our model domain). The observational coverage maintained by the European project CARBOOCEAN and US-American partner projects, includes a monthly monitoring along the route UK–Caribbean, and also between the North Sea and the southern tip of Greenland (Figure 12, Table 2). Because of the lack of monthly coverage further south, we restricted the area considered in our study to the region north of 15°N. We also note that the VOS data of the year 2005 set does not contain any data points in the Labrador Sea. Consequently, no training data was available for the mapping of pCO_2 in this region of particular importance for the oceanic CO_2 uptake [*Sabine et al.*, 2004].
Infrared and optical satellite data coverage is limited by clouds and, in the case of Chl, low solar irradiation at high latitudes in winter. Figure 12 indicates the number of days per month during which satellite data for Chl could be obtained by SeaWiFS for the year 2005. The data availability (for Chl) is typically less than 50% at every grid point for any month of the year, whereas almost no data is available for Chl north of 50°N in winter between November and February.

In the following, we will sample the output of our high-resolution biogeochemical-circulation model with a sampling schedule that exactly follows the coverage of VOS lines and remote sensing data of the year 2005. Thus, we do not use the actual observations but only the associated meta-information.
To assess the reliability of our methodological study, we first check whether the temporal and spatial variability of the modeled pCO_2 is representative of real pCO_2 data. To this extent, we compare model-generated and observed pCO_2 variability along the VOS cruise tracks conducted by the carcarrier M/V Falstaff in the years 2002/2003 [*Lüger et al.*, 2004] encompassing a full annual cycle on 15 transits with more than 78,000 samples in the area between UK/Northern Spain and North America.

The amplitude of the variations in pCO_2 is comparable for synthetic and actual measurements (Figure 13a). There is no phase relationship between the fluctuations since the stochastic nature of the variations overweighs the seasonal influences on this scale of sampling ($\leq$10 km). The sampling scale is typically smaller than the model grid spacing of $1/12^\circ \times 1/12^\circ$ cos(latitude). Thus, for the analysis of along-track spatial scales, all observations falling into the same model grid box are averaged. Subsequently, model-generated and actual observations are linearly detrended and linearly interpolated onto the same grid of 7×7 km^2 corresponding to the model resolution at the mean latitude (40°N) of the M/V Falstaff cruises. On this grid, the pCO_2

CARBOOCEAN:		
Ship	**PIs / Chief Scientists**	**Reference for data portal**[1]
Nuka Arctica	T. Johannessen A. Olsen	26na2005
Hespérides	A. F. Rios	29he2005
Poseidon	T. Steinhoff D. W. R. Wallace A. Körtzinger	06po2005
G. O. Sars	T. Johannessen A. Olsen	58gs2004
Trans Carrier	A. Omar	58tc2005
Santa Maria	U. Schuster A. Watson	64sa2005
Falstaff	T. Steinhoff D. W. R. Wallace A. Körtzinger	77ff2005
Quima	M. Gonzalez-Davila J. M. Santana-Casiano I. Rodriguez Ucha	quima2005
US-american partners:		
Ship	**PIs / Chief Scientists**	**Reference**
R. H. Brown	R. Castle J. Shannahoff R. Wanninkhof	www.aoml.noaa.gov/ocd/gcc/index.php
Explorer	B. Huss R. Wanninkhof	www.aoml.noaa.gov/ocd/gcc/index.php
Skogafoss	K. Sullivan D. Pierrot R. Wanninkhof	www.aoml.noaa.gov/ocd/gcc/index.php

Table 2: Vessels, Principal Investigators and reference websites for year 2005 VOS-line data used for our study.

[1] http://dataportal.carboocean.org/front_content.php?idcat=142

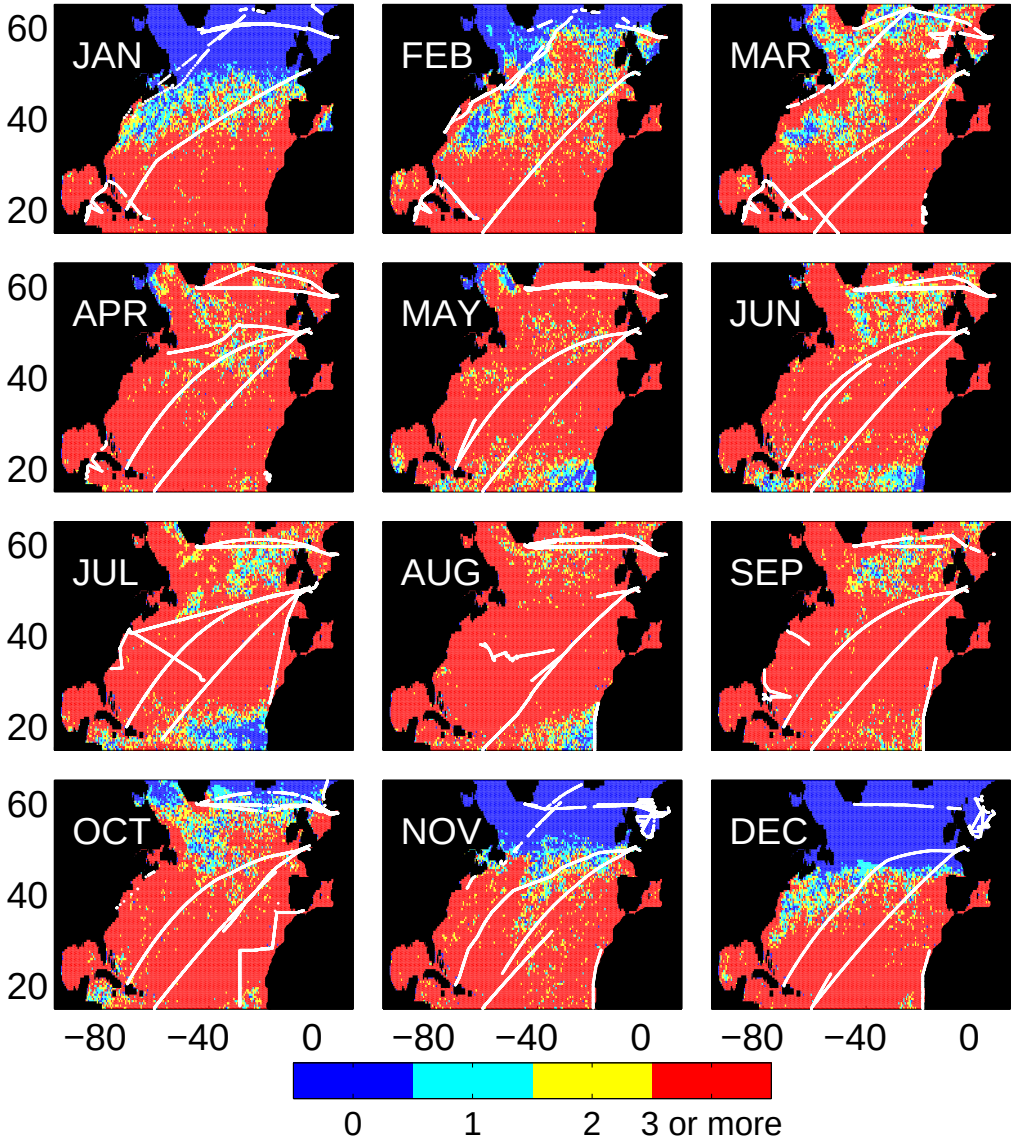

Figure 12: Availability of remote sensing data for chlorophyll (number of days per month) and monthly tracks of VOS-lines for the year 2005 (white lines).

variance amounts to 23.8 μatm for the actual observations and to 32.5 μatm for the model-generated "data", respectively.

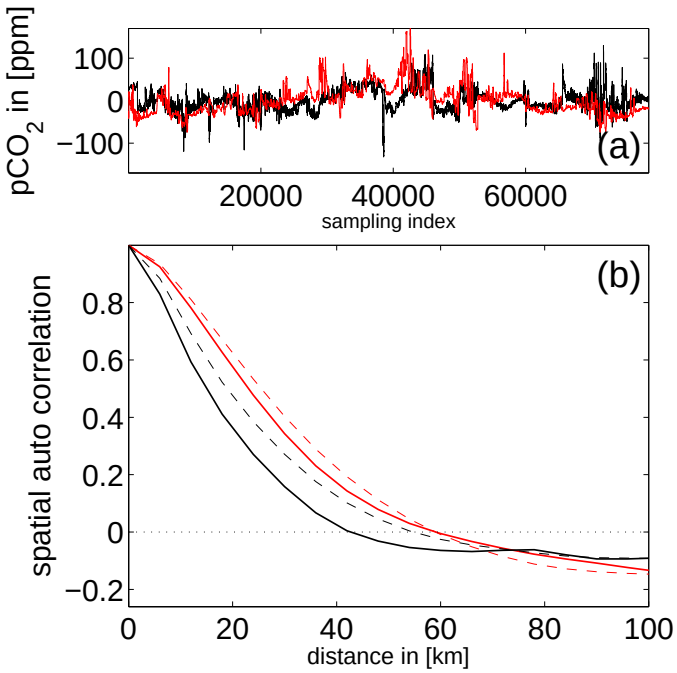

Figure 13: **(a):** Temporal pCO_2 variability for actual and model-generated measurements of carcarrier M/V Falstaff [*Lüger et al.*, 2004] black: observed, red: model-generated. Because of the pre-industrial atmospheric pCO_2 model setup the mean was removed from each curve. **(b):** Corresponding mean along-track spatial autocorrelation scales in km for observed (black solid), model-generated (red solid) pCO_2 and observed (black dashed), model-generated (red dashed) SST.

As expected, the mean spatial scales of the observed SST are larger than those of the observed pCO_2. This is relatively well reproduced by the model-generated data (Figure 13b). However, for both variables the mean spatial scales of the model-generated data in our model are larger than those of the real data. The spatial auto-correlation function crosses zero at 43 km for the real and at 59 km for the synthetic data. A possible reason for the too large scales in the model might be the absence of a diurnal cycle and the use of climatological forcing in the model, leading to less smallscale variability for the model-generated data. Overall, the calculated scales are significantly shorter than those presented by *Murphy et al.* [2001] who analyzed pCO_2 auto-correlation scales in the Gulf of Alaska and *Li et al.* [2005] who analyzed mean global auto-correlation scales of pCO_2 for the period 1970-2003. Compared to *Murphy et al.* [2001] the shorter scales found in our analysis can be explained by the higher eddy activity in the North Atlantic [*Chelton*

et al., 2007]. The different time scales considered may explain the differences with respect to the study by *Li et al.* [2005].

Methods

Estimating basinwide pCO_2 from remotely sensed SST and Chl requires a formulation which converts the two observed quantities (SST, Chl) into pCO_2. Finding such a relationship between pCO_2 and SST and/or Chl has been subject of previous studies [*Watson et al.*, 1991, *Bates et al.*, 1995, *Lüger et al.*, 2004, *Lefèvre et al.*, 2002, 2005]. Nevertheless, an exact mathematical relationship $\overline{pCO_2}$ = f(SST, Chl) or $\overline{pCO_2}$ = f(SST, Chl, Position, Time) does not exist as pCO_2 depends on other factors than local SST and Chl. Surface waters with identical SST and Chl can well have different pCO_2 levels. For the context of simple linear functional relationships, this can be illustrated by applying a multiple linear regression to the model-generated data:

$$pCO_2 = \mathrm{A} \cdot SST + \mathrm{B} \cdot Chl + \mathrm{C} \tag{12}$$

with the coefficients A,B and C being functions of space and time.

For computational reasons, the model data were subsampled on a grid of $1/2^\circ \times 1/2^\circ$ resolution. The regression coefficients A(Lat,Lon,Month) (Figure 14) and B(Lat,Lon,Month) (not shown) exhibit considerable variability both in time and in space. Spatial gradients in the SST coefficient A(Lat,Lon,Month) can reach values up to $60\,\mu$atm/°C/100 km whereas temporal fluctuations easily exceed $\pm 20\,\mu$atm/°C in a month. This makes it difficult to envisage some division of the North Atlantic into seasonal or regional biogeophysical regimes with constant regression coefficients.

In contrast to standard regression techniques, Kohonen Feature Maps (KFMs) [*Kohonen*, 1982] work as an associative memory that can be used to estimate pCO_2 from SST and Chl (and position in time and space) without fitting a predefined mathematical function. The concept of KFMs is motivated by neurobiological findings about the topological organization of sensory experiences in certain areas of the brain. The KFM consists of a topological arrangement of neurons that associate an input vector (here SST, Chl, position, time) with an output (here pCO_2). Similar to the brain, a KFM has to learn how to associate input and output. This is achieved by confronting the KFM with "stimulus" pairs of input and output variables in a training process. A detailed description of the formation of topological maps during the training process can be found in *Kohonen* [1982] and *Lefévre et al.* [2005]. Figure 15 illustrates a simple scheme of the functioning. In the training process the mean topological features of the observed combinations [SST, Chl, pCO_2] are memorized in a map of (normalized) vectors and corresponding

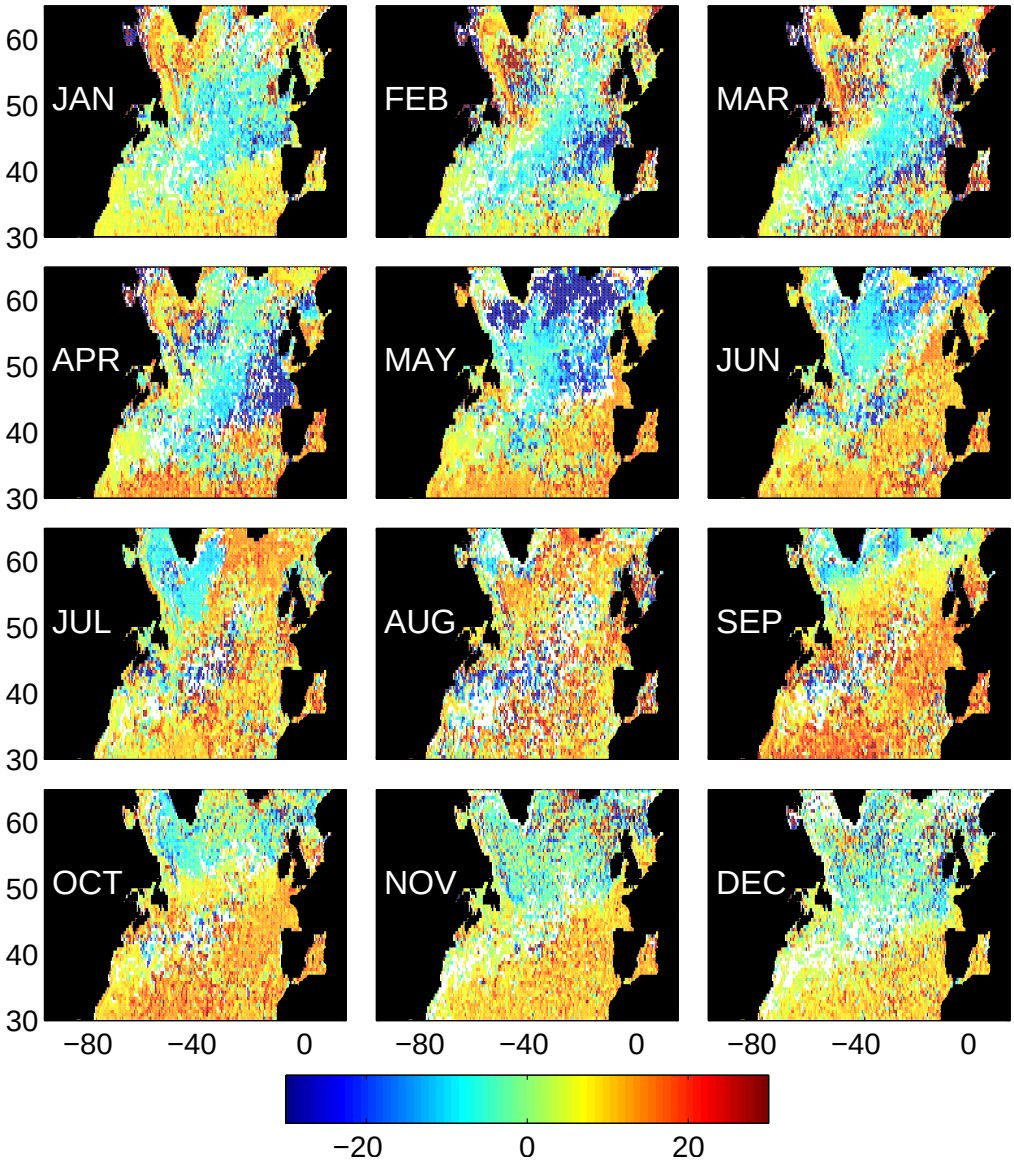

Figure 14: Regression coefficients of model-generated pCO_2 versus model-generated SST in μatm/°C for a multiple linear regression applied monthly to the model data on a 1/2° × 1/2° grid according to equation 12. White areas indicate regions where the multiple linear regression accounts for less than 50 % of the model-generated pCO_2 variance.

pCO_2 values. In the application process, an input vector (normalized SST and Chl in the example of Figure 15) is associated with the pCO_2 value belonging to the best fitting vector of the parameter space representation.
In our case, model-generated VOS-line samples of SST and Chl and corresponding Latitude, Longitude, Day (in the year) are used to form a 5-dimensional parameter space representation in the training process. In the subsequent application phase the KFM can associate simulated remote sensing data of SST and Chl (and the respective position and time) with a pCO_2 value. Input data was normalized as follows:

$$Day' = cos\frac{Day \cdot 2\pi}{365} \tag{13}$$

$$Lat' = \frac{latitude - 15°N}{65°N - 15°N} \tag{14}$$

(only data between 15° N and 65° N were considered)

$$Lon' = \frac{longitude - 88°W}{11°E - 88°W} \tag{15}$$

$$SST' = \frac{SST - \overline{SST}}{STD(SST)} \tag{16}$$

$$Chl' = \frac{log(Chl + 0.01) - \overline{log(Chl + 0.01)}}{STD(log(Chl + 0.01))} \tag{17}$$

Overlined variables refer to the spatial-temporal average.
The input for 'Day' was normalized sinusoidally, following *Lefévre et al.* [2005] and emphasizing the seasonal cycle and allowing January to be close to both February and December. A log-normal distribution was assumed for model-generated Chl following *Campbell* [1995]. A value of 0.01 $\frac{mgChl}{m^3}$ has been added to avoid a large impact of data uncertainties at very low Chl concentrations.

Since a KFM works as an associative allocation between parameter vectors and target values, the number of combinations [pCO_2,SST,Chl,Day,Lat,Lon] that are set during the training process and later used as "look-up table" to estimate pCO_2 from the input variables [SST,Chl,Day,Lat,Lon] is restricted by the size of the map. By choosing a map that is too small (i.e., a too small number of neurons), it would not be possible to adequately represent all features of the training data. The map would already be occupied by the dominating features, leaving no room for more infrequently sampled regions or events, such as the West African upwelling. In order to find an optimal size for the topological map, we trained 16 KFMs (20×20, 40×40, ..., 320×320 neurons) with the synthetic observations (SST, Chl and corresponding Day, Lat, Lon) of the model-generated VOS-lines and reconstructed surface pCO_2

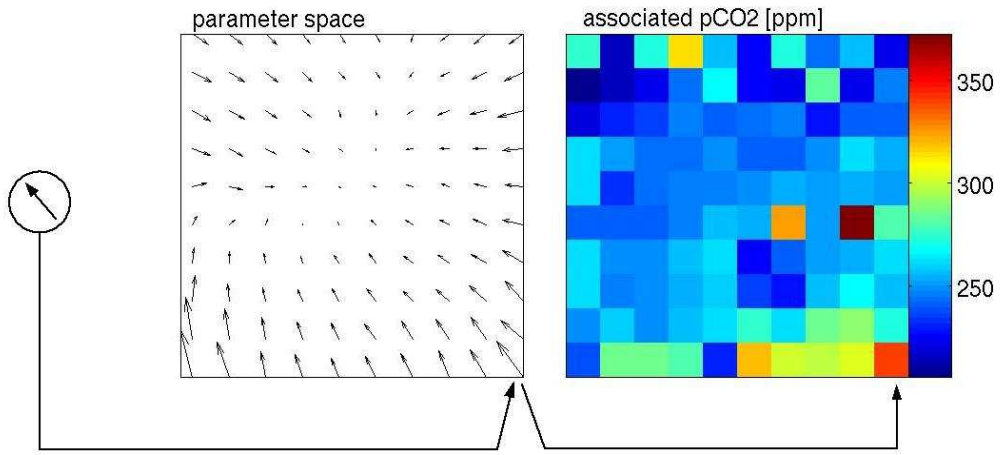

Figure 15: Schematic mode of operation of a trained KFM. An input vector (left) is associated with the pCO_2 value (right) of the best matching vector (middle) of the parameter space representation. X- and y-axis denote (dimensionless) width of KFM. Note: For clarity, the dimension of the KFM was reduced to a 10×10 array of neurons, and the input (and parameter) space was chosen to be two-dimensional only (e.g., SST and Chl).

from model-generated remote sensing data of SST and Chl (and corresponding Day, Lat, Lon). Missing data due to a lack in optical satellite coverage were not considered in this case. The basinwide RMS-error of the estimated pCO_2 as a function of the map size shows a substantial decrease of the mapping error on increasing the number of neurons from 20×20 to 220×220, and a slightly increasing RMS-error thereafter.
Nowadays, remote sensing of SST and Chl is normally available at a spatial resolution of approximately 9 km (SeaWiFS). For computational reasons, we subsampled the model-generated remote sensing data on a somewhat coarser $1/2° \times 1/2°$ grid. The impact of remote sensing errors on the pCO_2 estimates was investigated by adding Gaussian noise with σ=0.5 K to the model-generated SST data, and Gaussian nose with a variance of σ=30 % to the actual model-generated Chl data.

Results and Discussion

With respect to possible efforts to monitor CO_2 sources and sinks, a monthly resolution of basinwide pCO_2 maps seems to be a reasonable goal. In the following, our assessment will therefore focus on the accuracy of monthly maps of pCO_2 estimated from model-generated remote-sensing data of SST and Chl via a KFM trained by model-generated VOS data. Because we restrict ourselves to synthetic observations, we can reliably assess the quality

of our pCO_2 estimates with respect to the known "true" pCO_2 simulated by the biogeochemical circulation model. This would not be possible for pCO_2 estimates from real satellite data as, in that case, the true pCO_2 is not known apart from a few data points (e.g., VOS lines) that would generally be used to train the KFM.
Our methodological study and the analysis of our results focuses on the following questions:

- What is the accuracy of basinwide maps of surface pCO_2 for present-day VOS-line coverage and how sensitive is this accuracy to changes in VOS coverage?
- What is the impact of errors in the remotely sensed SST and Chl data?
- What is the impact of incomplete remote sensing coverage and how can data gaps be interpolated?
- What is the accuracy of CO_2 fluxes calculated from estimated pCO_2 maps?

A first issue addresses the KFM mapping frequency needed to generate robust monthly-mean pCO_2 estimates. Since remotely sensed SST and Chl are available on a daily basis, at least two alternative ways are conceivable of how to apply the data to the KFM. As a first option, daily values of SST and Chl can be converted by the KFM into daily values of pCO_2, which are subsequently averaged into monthly means. As a second option, monthly means of remotely-sensed SST and Chl can be applied to the KFM to directly estimate monthly-mean pCO_2 maps.

Figure 16 shows a comparison of the seasonally averaged RMS-error in the estimated pCO_2 maps for the two methods. For the upper panels (a-d), pCO_2 was calculated from model-generated daily SST and Chl, and for the lower panels (e-h), pCO_2 was estimated from model-generated monthly-mean SST and Chl satellite data. For all seasons the mapping error turned out to be larger when model-generated daily SST and Chl data were used. Initially, this result seems counterintuitive: The KFM was trained to reproduce simulated in-situ pCO_2 from model-generated instantaneous in-situ SST and Chl data, and the mapping process is a nonlinear one: $pCO_2(\overline{SST},\overline{Chl}) \neq \overline{pCO_2}$. Two features help to understand this counterintuitive behavior. The pCO_2 training data of the model-generated VOS-lines as well as the model-generated "true" pCO_2 in the considered domain are distributed nearly

normally (Figure 17a, dashed and solid lines), with most values concentrated in the interval of 230-330 μatm (note the pre-industrial setup of the model). In this central pCO_2 range, the KFM resolves features most precisely, whereas it tends to overestimate (underestimate) lower (higher) pCO_2 values (Figure 17b). Largest mapping errors occur for $pCO_2 < 200\mu$atm and for $>$

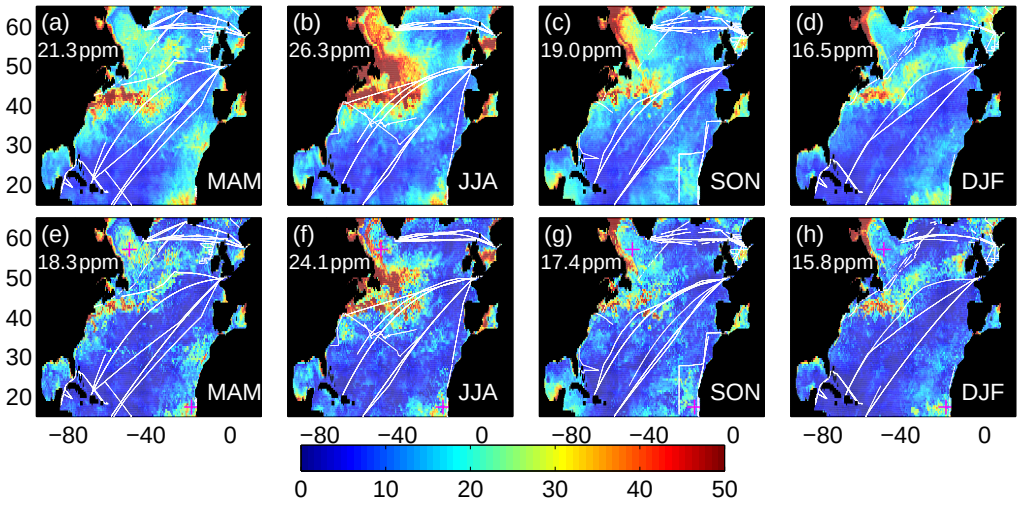

Figure 16: Seasonal mean RMS-error in μatm for mapping based on model-generated VOS-lines of the year 2005 (white lines) and daily (a-d) or monthly mean (e-h) model-generated remote sensing data of SST and Chl. Magenta crosses in the Labrador Sea and off West Africa in (e-h) denote positions of two simulated stationary monitoring buoys (see text for details).

400μatm (Figure 17a/b, dots). Particularly in areas of large gradients, e.g. in the North Atlantic Current (NAC) region close to the Grand Banks, where high pCO_2 values of the NAC encounter fairly low values of the Labrador Current, the KFM is not able to resolve such a high spatial variability. If the input data for the mapping consists of monthly mean values, these gradients are weaker.

The remote sensing errors of SST and Chl provide the second reason, leading to a deviation when the KFM is applied. As these errors are assumed to be of stochastic nature they tend to average out over the course of one month, resulting in more reliable monthly mean values. Significantly larger errors of the daily mapping in the Labrador Sea and Grand Banks region in summer (Figure 16b and f) can be attributed to the remote sensing errors for model-generated Chl. A more detailed analysis reveals a marginal and consistently smaller mapping error over the year for a perfect remote sensing of SST (Figure 18a, magenta dotted). With regard to a perfect remote sensing of Chl (Figure 18a, magenta dashed) a substantial improvement is primarily found for the months May to August.

In a number of sensitivity experiments we also investigated the importance of the individual data sets for accurately estimating surface pCO_2. Even though a straightforward relation between Chl concentrations and pCO_2 does not exist, disregarding model-generated Chl observations (both in the training and

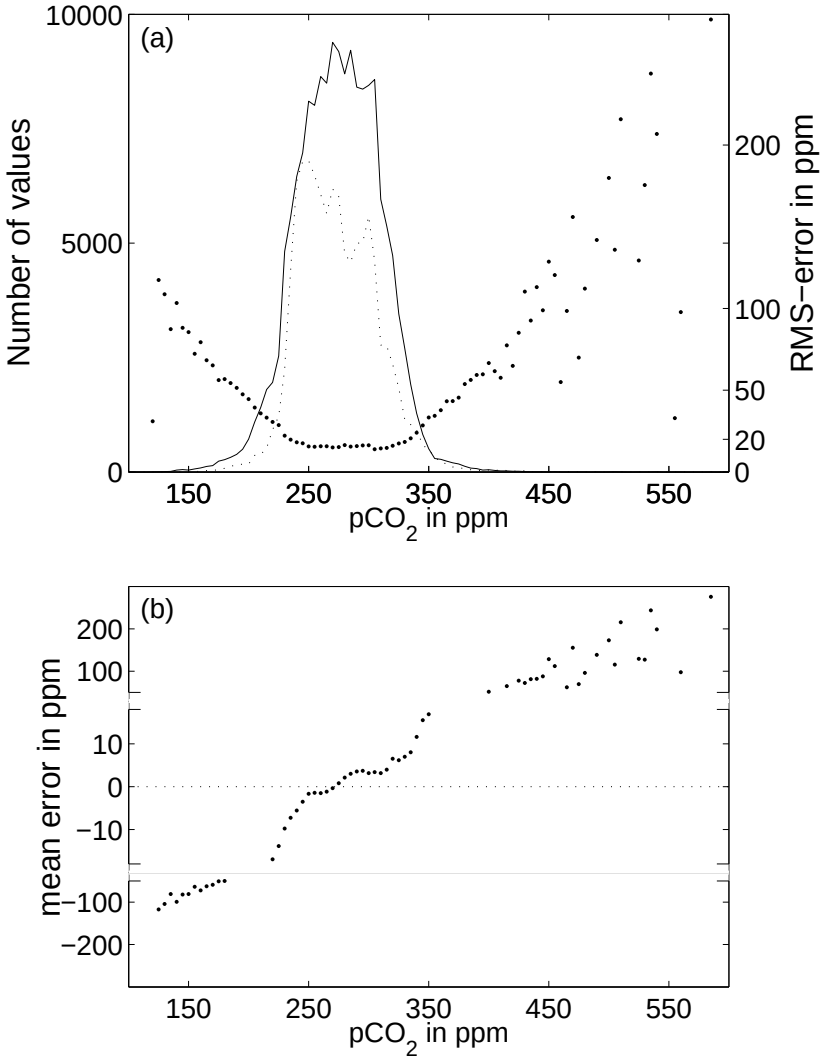

Figure 17: **(a):** Number of values in basinwide modeled pCO_2 (solid line) and model-generated VOS-line data of the year 2005 (dotted) and RMS-error in μatm (dots) vs. modeled pCO_2 in μatm (5 μatm intervals). **(b):** Mean error of estimated pCO_2 in μatm (dots) vs. modeled pCO_2 in μatm (5 μatm intervals). Mapping based on simulated VOS-lines of the year 2005 and monthly mean model-generated remote sensing data of SST and Chl. Note the split y-axis in (b) and the pre-industrial atmospheric pCO_2 setup of the model.

in the application of the KFM) leads to considerably larger mapping errors, in particular during the spring bloom between February and June (Figure 18a, blue line).

The same holds for neglecting position in time and space. Use of these coordinates in the mapping procedure appears unfounded at first sight and impedimentary for a theoretical insight into the dependance of pCO_2 on

physical and biological parameters. However, an attempt to use only model-generated SST and Chl as input variables was not successful (Figure 18, cyan line). Adding simulated mixed layer depth to the training and KFM input data (neglecting that there is no currently known method available to accurately determine mixed layer depth remotely) could not improve this result substantially.
Figure 18a (green line) additionally shows the importance of measuring Chl onboard along the VOS-lines instead of gathering training data through remote sensing. When remotely sensed Chl data with their relatively large errors are used in the training process, the topology of the KFM is deteriorated by the remote sensing errors which leads to larger mapping errors compared to more accurate in-situ training data.

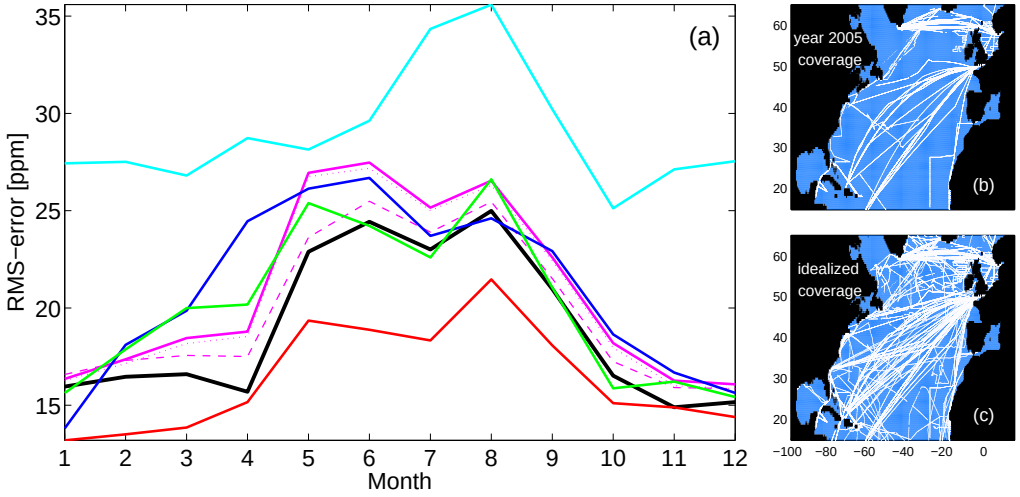

Figure 18: **(a):** Annual cycle of RMS-error in μatm for basinwide (15°N to 65°N) mapping of pCO_2. **Magenta:** Mapping based on simulated VOS-lines of the year 2005 and daily remote sensing of SST and Chl when perfect satellite coverage is assumed (solid), as well as neglecting remote sensing errors in SST (dotted) and Chl (dashed). **Black:** Mapping based on simulated VOS-lines of the year 2005 and monthly mean remote sensing data of SST and Chl disregarding missing data (reference simulation). **Green:** Similar to black, but remote sensing of Chl for training instead of onboard sampling. **Cyan:** Similar to black, but using only SST and Chl and not position and day. **Blue:** Similar to black, but using SST and Position only. **Red:** Training and mapping based on an simulated idealized VOS-line coverage. **(b)** and **(c):** Annual VOS-lines coverage of the year 2005 and idealized coverage.

Even for the case with model-generated in-situ Chl training data and monthly-mean remotely sensed SST and Chl, estimated pCO_2 still shows large errors in various regions (Figure 16e-h and Figure 18a black line). Beside the fact

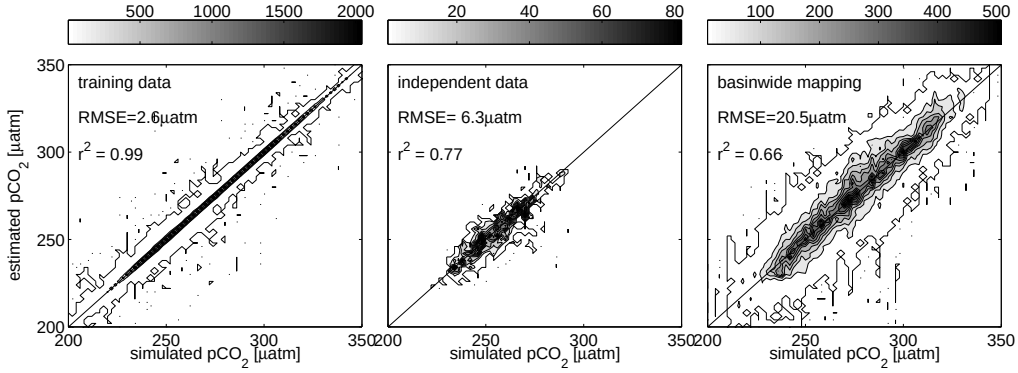

Figure 19: Comparison of model-generated "true" and estimated pCO_2 in numbers of data for (left) reproduction of training data, (middle) reproduction of an model-generated independent data set and (right) basinwide (15°N to 65°N) mapping corresponding to Figure 16 (e-h) and Figure 17. Note that given basin-mean RMS-error in the right panel is non-area weighted. The associated area-weighted basin-mean RMS-error amounts to 19 μatm.

that unreproducible large gradients can still occur in monthly mean values of pCO_2, these mapping errors arise from the VOS-line coverage of the year 2005 which, notwithstanding all endeavors, still has to be regarded as sparse. Large mapping errors in the Labrador Sea and around the Grand Banks (particularly in summer) do not necessarily indicate a deficiency of the method but can be seen as a consequence of the lack of training data in this area.
A more detailed analysis of the errors in the estimated pCO_2 maps confirms that these are largest in areas far away from any training data set. As shown by Figure 19a, the KFM can accurately reproduce the training data set with an RMS-error of 2.6 μatm, while the errors are already more than twice as large along individual VOS-lines excluded from the training data set (Figure 19 middle). For the entire basin, however, the RMS-error of the estimated pCO_2 field amounts to 20.5 μatm (area-weighted: 19 μatm). In contrast to our model study, in which the "true" pCO_2 of the model is known, the latter error would be difficult to estimate for real applications of the methods using real data. The conventional error assessment along individual VOS-lines not contained in the training data is not necessarily indicative of the basinwide error (Figure 19 middle and right panel).

Impact of spatial resolution

In order to add training data in regions with particularly large errors in the estimated pCO_2, two stationary monitoring buoys were additionally simulated in the model at the positions of 17°N / 20°W close to the West African upwelling and at 57°N / 55°W in the Labrador Sea (magenta crosses in Fig-

ure 16e-h). A weekly sampling of local SST and pCO_2 together with remote sensing of Chl was assumed and added to the VOS-line data for the training of the KFM. A clear regional improvement of the pCO_2 estimates was observed for the Labrador Sea buoy leading to a local 20% decrease of the RMS-error, though having no significant effect in the basinwide RMS-error. For the simulated West African buoy the regional decrease in the error of the estimated pCO_2 was smaller. It was also limited to a significantly smaller area and even accompanied by an error increase in the surrounding region. This slight increase was caused by the fact that the observed pCO_2 values and biological and physical properties vary on small scales. Still, in the absence of sufficient surrounding observations, simulated local buoy data are taken as the best guess for the whole area by the KFM. The effect of both simulated buoys on the KFM estimate of pCO_2 could be greatly enhanced by an associated spatial variability study.
In order to assess the potential of enhanced VOS sampling, the KFM was trained with a simulated idealized VOS-line coverage. It was assumed that in addition to VOS-lines of the year 2005, all freely accessible lines [3] would have been occupied in the same year. The area-weighted, basinwide annual mean RMS-error can be reduced to 16.4 μatm in this case, compared to 19 μatm for the actual lines of the year 2005 (Figure 18a, red line).

Impact of optical satellite coverage

As discussed previously in section 3, remote sensing of ocean color and SST is limited by cloud cover and, in the case of ocean color, by low solar irradiation in winter. Hence, the treatment of missing satellite data needs to be investigated. Basically, missing data can be replaced in two ways: Climatological values of SST and Chl might be used as input data for the KFM, or climatological pCO_2 could complete a mapping which is applied only to optically covered pixels. Both alternatives have been analyzed. Climatological pCO_2, SST and Chl were simulated by a model run with slightly different forcing (daily forcing for the considered year instead of monthly mean forcing) to account for possible biases in our climatologies. Simulated SST and Chl climatologies were assumed to be available on a $1/2^\circ \times 1/2^\circ$ grid. For the simulated pCO_2 climatology we have chosen the Takahashi [*Takahashi et al.*, 1997] resolution of $4^\circ \times 5^\circ$.

Figure 20 illustrates the seasonally averaged RMS-error for a mapping that replaces missing data in the remote sensing with climatological SST and Chl. Mapping errors are similar to those obtained earlier for the idealized case of no missing data (Figure 16e-h). Comparing the annual cycle of the errors (Figure 21) with the missing data ratio (Figure 12) does not reveal a cor-

[3] http://cdiac.ornl.gov/oceans/pCO_2inv.html

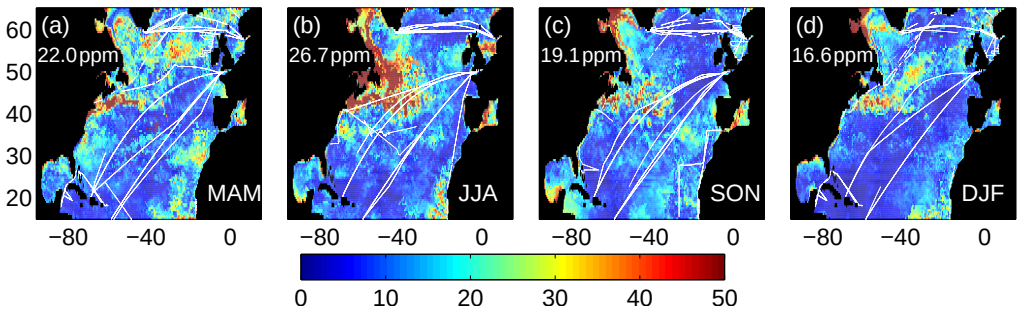

Figure 20: Seasonal mean RMS-error in μatm for mapping based on simulated VOS-lines and remote sensing of SST and Chl of the year 2005. Missing data caused by cloud cover or low solar irradiation were replaced by model-generated climatological values of SST and Chl. Seasonal VOS-lines of the year 2005 (white lines).

relation of the error size with optical coverage, which is lowest in winter. Instead, errors of the estimated pCO_2 fields are largest in summer, similar to the earlier simulation that assumed no missing data.
Overall, it becomes apparent that the mapping errors are dominated by the method's inability to reproduce areas of large pCO_2 gradients whereas the replacement of missing data is of secondary importance. In our simulation, substituting gaps in optical satellite coverage by model-generated climatological pCO_2 achieves better results than using model-generated climatological SST and Chl input variables to the KFM (Figure 21). However, it should be kept in mind that real-world climatological pCO_2 under increasing atmospheric CO_2 levels is still afflicted with considerable uncertainties, which was neglected in our pre-industrial simulation.

CO_2 fluxes

One major reason for estimating basinscale pCO_2 maps is the assessment of marine CO_2 uptake. Air-sea gas exchange is, besides the variation from the equilibrium pCO_2 value, a function of the gas transfer velocity (piston velocity). This piston velocity is here assumed to depend quadratically on wind speed, following *Wanninkhof* [1992]. As a result, temporal and spatial patterns of error in CO_2 fluxes calculated from estimated pCO_2 do not necessarily resemble RMS-error patterns shown in Figures 16 and 20. In fact, errors in estimated wintertime pCO_2 tend to cause larger errors in CO_2 fluxes whereas errors in the summer, which dominate in our simulations, will have a minor effect.

In our simulation the annual cycle of the model-generated "true" CO_2 fluxes

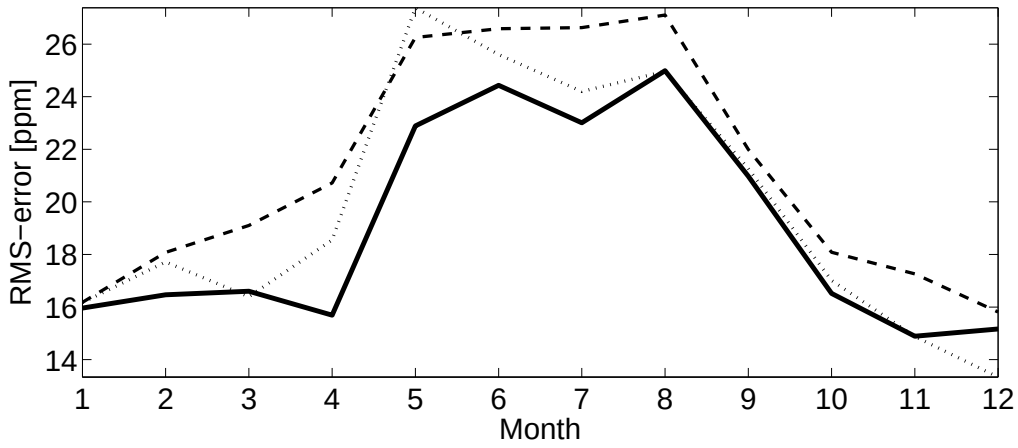

Figure 21: Annual cycle of RMS-error in μatm for basinwide (15°N to 65°N) mapping of pCO_2 for an assumed perfect satellite coverage (solid) and for a replacement of missing data by model-generated climatological pCO_2 (dotted) and model-generated climatological SST and Chl (dashed).

is very well reproduced in the considered region 15°N to 65°N when CO_2 fluxes are calculated from the mapped pCO_2 replacing missing data in the remote sensing with simulated climatological SST and Chl (Figure 22, top left panel). The RMS error of monthly and regionally averaged CO_2 fluxes amounts to 0.15 mol/m^2/yr. The amplitude of the seasonal cycle of the estimated flux amounts to 3.3 mol/m^2/yr (3.1 mol/m^2/yr when we assume perfect satellite coverage) and overestimates the model-generated "true" amplitude by 6.8 %.

Regionally, significant deviations from the model-generated "true" CO_2 flux (Figure 22) and CO_2 uptake (Table 3) can occur. Largest errors in CO_2 flux are found in the North Atlantic Current region from December to April (where the main uptake takes place in our model) and derive from errors in the estimated pCO_2 in this region of large pCO_2 gradients. Errors in CO_2 fluxes are also elevated by the replacement of missing data due to cloud cover and low solar irradiation in late fall and winter. In the regions 25°N to 35°N and north of 55°N, as well as in the basinwide mean, more than half of the RMS error of monthly and regionally averaged CO_2 fluxes derives from the substitution of missing data. In particular north of 55°N, CO_2 uptake is overestimated for the months March to May which is caused by an overestimate of the pCO_2 drawdown. This overestimate is not found in the simulation without missing data and is caused by too high synthetic climatological Chl values taken as a replacement for the missing data. Whether disregarding or replacing missing data in the remote sensing, errors in estimated CO_2 fluxes are not in phase with pCO_2 mapping errors but have a maximum in winter when errors in the estimated pCO_2 maps are comparatively low.

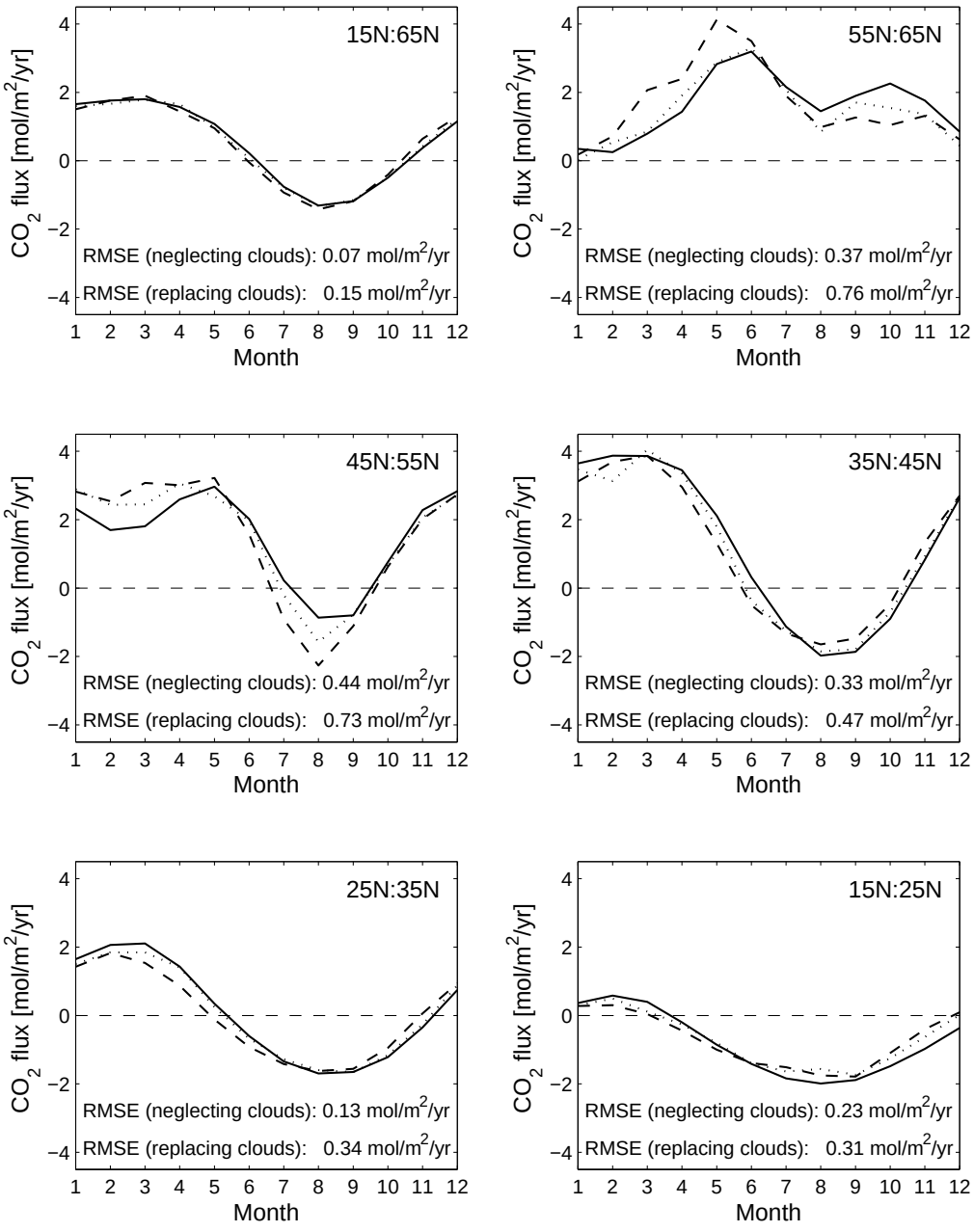

Figure 22: Annual cycle of model-generated "true" (solid) and estimated CO_2 fluxes, replacing missing data in the simulated remote sensing with model-generated climatological SST and Chl (dashed) and for assumed perfect satellite coverage (dotted) for given latitudes in mol/m^2/yr. CO_2 fluxes were calculated using the piston velocity formulation of *Wanninkhof* [1992]. Positive values denote fluxes into the ocean.

Region	**area-weighted mean RMS-error [μatm]**	**model-generated "true" CO_2 uptake [TgC/yr]**	**model-generated "true" CO_2 uptake - estimated CO_2 uptake [TgC/yr]**
15°N:65°N	21.1	190.6	+11.5
55°N:65°N	24.4	78.5	-3.6
45°N:55°N	25.8	81.7	+2.5
35°N:45°N	25.7	102.1	+8.9
25°N:35°N	15.2	13.0	+11.4
15°N:25°N	18.9	-84.6	-7.7

Table 3: Area-weighted RMS-error, model-generated "true" CO_2 uptake in TgC/yr and deviation from simulated "true" CO_2 uptake (model-generated "true" CO_2 uptake–estimated CO_2 uptake) in TgC/yr for pCO_2 mapping replacing missing data in the simulated remote sensing with model-generated climatological SST and Chl. CO_2 fluxes were calculated using the piston velocity formulation of *Wanninkhof* [1992].

Method evaluation

When evaluating the presented method it is important to distinguish between different objectives of a basinwide pCO_2 mapping and thus between local mapping errors and basin-mean errors. As local mapping errors are large in various regions it is not clear whether the presented method can reliably detect interannual fluctuations and trends in the considered areas and hence provide a benefit compared to a simple use of a pCO_2 climatology. For the basin-mean view, however, we do believe that there is additional benefit. *Lüger et al.* [2004] found climatological pCO_2 values [*Takahashi et al.*, 2002] to be on average lower than observed values by 13 μatm in the eastern basin (35°W to 0°W) of the North Atlantic and to be, on average, higher by 2 μatm for the western part (80°W to 35°W). For the eastern part the differences were dominated by deviations of up to 22 μatm between January and May. In comparison, for our method we find basin-mean differences (model-generated "true" pCO_2 - estimated pCO_2) of -2.1 μatm (1.8 μatm) for the western (eastern) part with maximum monthly deviations of 9.5 μatm.

The model-generated basin-mean pCO_2 is overestimated by 0.51 μatm by the KFM mapping technique with missing remote sensing data replaced by climatological SST and Chl values. Taking into account ΔpCO_2 values (atmospheric pCO_2 - sea surface pCO_2) of 26 μatm in the mid-1990s and 5 μatm in the early 2000s as reported by *Schuster and Watson* [2007] for the region 20°N to 65°, this estimated basin-mean error is still about one order of magnitude smaller than observed decadal changes. In our simulation the slight overestimate of basin-mean pCO_2 leads to an underestimate of basin-

wide CO_2 uptake by 6.0 % (Table 3) which is a promising value. However, it must be kept in mind that the accuracy of the presented method strongly depends on the relationships between SST, Chl and pCO_2. In particular for Chl which, in the model is simply a constant fraction of phytoplankton biomass, this relationship is likely to be more complex in the real ocean. Other aspects that have been neglected in the current model are effects of calcium-carbonate forming organisms on alkalinity and hence pCO_2, as well as variations in the carbon-to-nutrient ratios of organic matter.

Conclusions

By simulating present-day observational sampling schemes in the context of an eddy-resolving biogeochemical circulation model, we have demonstrated that VOS-line measurements and satellite data can be combined by a self-organizing neural network in order to generate monthly basinwide maps of pCO_2. Our results show that, besides the still insufficient VOS-line coverage and the frequent occurrence of missing data in the remote sensing of SST and Chl, the high spatial variability of pCO_2 in the NAC region is the most crucial limiting factor for the accuracy of the mapping procedure suggested here for the North Atlantic. The basinwide RMS error for a hypothetical perfect satellite coverage (or a hypothetical perfect scheme to interpolate missing data due to clouds and low solar irradiation in winter) amounts to 19.0 μatm, but reaches locally 50 μatm and more in the region of large pCO_2 gradients, even though training data and an independent data set can be reproduced with a much lower RMS error of 2.6 μatm 6.3 μatm respectively .

Compared to model-generated daily data of remotely sensed SST and Chl, synthetic monthly mean data of SST and Chl provide a more reliable and successful basis for producing basinwide pCO_2 charts since remote sensing errors and large pCO_2 gradients are reduced. Simulating the VOS-lines and actual optical satellite coverage of the year 2005, a basinwide mean RMS-error of 21.1 μatm is found when missing data in the remote sensing are replaced by model-generated climatological values of SST and Chl. The annual cycle of the model-generated "true" CO_2 fluxes can be well reproduced from the estimated pCO_2. The amplitude of the annual cycle is slightly overestimated by 6.8 % while the basinwide CO_2 uptake is underestimated by 6.0 %.

The extension of monitoring pCO_2 is highly recommended as our results indicate a significant improvement in the basinscale pCO_2 maps for an increased VOS-line coverage. For this purpose, morel observations would be particularly useful in the Labrador Sea and the NAC area. The additional use of float-based pCO_2 sampling might provide a substantial benefit.

Acknowledgments

Our thanks go first to the numerous scientists and personnel responsible for

the collection of VOS-line and SeaWiFS data that form the basis of our model simulation. We are grateful to C. Eden for running the biogeochemical ocean model. We also thank the anonymous reviewers for their helpful comments. This work was funded by the European Union via CARBOOCEAN 511176 (GOCE).

References

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.

Bates, N. R., A.F. Michaels, A.H. Knap (1995), Seasonal and interannual variability of oceanic carbon dioxide species at the U.S. JGOFS Bermuda Atlantic Time-series Study (BATS) site, *Deep Sea Res., Part II*, *43*, 347–383.

Bates, N. R., A. C. Pequignet, R. J. Johnson and N. Gruber (1999), A short term sink for atmospheric CO_2, *Nature*, *420*, 489–493.

Campbell, J. W. (1995), The lognormal distribution as a model for bio-optical variability in the sea, *J. Geophys. Res.*, *100*(C7), 13237-13254, doi:10.1029/95JC00458.

Chelton, D. B., M. G. Schlax, R. M. Samelson, and R. A. de Szoeke(2007), Global observations of large oceanic eddies, *Gephys. Res. Lett.*, *34*, doi:10.1029/2007GL030812.

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, Volume 4: Nutrients. S. Levitus, Ed., NOAA Atlas NESDIS 54, U.S. Government Printing Office, Wash., D.C., 392 pp.

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. 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.

Gaspar, P., Y. Gregoris, and J.-M. Lefevre, (1990), A simple 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*, 179–193.

Kalnay, E. et al. (1996), The NCEP/NCAR 40 years reanalysis project, *Bull. Amer. Meteor. Soc.*, *77*, 437–471.

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. Cycle*, *18*, GB4031, doi:10.1029/2004GB002247.

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

Lefévre, N., A. Taylor (2002), Estimating pCO_2 from sea surface temperatures in the Atlantic gyres, *Deep Sea Res., Part I*, *49*, 539–554.

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

Le Quéré, C., J. C. Orr, P. Monfray and O. Aumont (2000), Interannual variability of the oceanic sink of CO_2 from 1979 through 1997, *Global Biogeochem. Cycles*, *14*, 1247–1265.

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

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

Murphy, P. P., N. Yukihiro, D. E. Harrison, N. K. Larkin (2001), Scales of spatial variability for surface ocean pCO_2 in the Gulf of Alaska and Bering Sea: toward a sampling strategy, *Gephys. Res. Lett.*, *28*, 1047–1050.

Oschlies, A. and V. Garçon (1999), An eddy-permitting coupled physical-biological model of the North Atlantic. Part 1: Sensitivity to advection numerics and mixed layer physics, *Global Biogeochem. Cycles*, *13*, 135–160, doi:10.1029/98GB02811.

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

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.

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

Watson, A. J., C. Robinson, J. E. Robinson, P. J. le B. Williams, M. J. R. Fasham (1991), Spatial variability in the sink for atmospheric carbon dioxide in the North Atlantic, *Letters to Nature*, *350*, 50–53.

How to estimate North Atlantic surface pCO_2 from ARGO float data

–*Journal of Geophysical Research*, submitted–

Tobias Friedrich[1]
(corresponding author)

Andreas Oschlies[1]

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

Abstract

A novel method for mapping North Atlantic surface pCO_2 on a nearly basinwide scale is presented and analyzed. VOS-line coverage of the year 2005 is simulated using an eddy-resolving model and the synthetic "observations" of SST and SSS form a training data set for a self-organizing neural network that is subsequently applied locally to estimate pCO_2 from simulated year-2005 ARGO float data. The punctual pCO_2 estimates at the float positions are extrapolated using objective mapping. For ARGO float coverage of the year 2005, the resulting 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.

Introduction

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. Automated instruments on board of Voluntary Observing Ships (VOS) measuring partial pressure of CO_2 (pCO_2) at the ocean's surface have provided a substantial benefit for this quantification 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 and VOS-line coverage depends on commercial ship tracks (and funding) rather than on observational requirements. Therefore an interpolation between VOS-line observations remains difficult. On the other hand, there are currently more than 3100 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 parameters controlling pCO_2 variability at the ocean's surface without being subject to problems typically associated with remote sensing, such as the large fraction of missing data due to cloud cover and low solar irradiation at high latitudes in winter. In the present study we demonstrate in a model simulation how VOS-line observations of pCO_2, SST and SSS may be utilized to train a self-organizing neural network that is subsequently applied to ARGO float data in order to derive nearly basinwide monthly mean maps of surface pCO_2 for the North Atlantic.

Model Configuration

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 [*Pakanowski*, 1995]. The model domain spans the Atlantic Ocean from 18° S to 70° N at a horizontal resolution of $1/12° \times 1/12°$ 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, 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 climatology provided by *Levitus et al.* [1982] is used with a time scale of 30 days. In order to account for the presence of sea ice, the heat CO_2 and O_2 flux is 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 $2 \times 10^{10}\,m^4/s$ for viscosity) and a level-1.5 closure scheme for vertical turbulent mixing following *Gaspar et al.* [1990].
Initial conditions for nitrate are taken from *Conkright et al.* [2002] and for dissolved inorganic carbon from the pre-industrial 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 observational estimates by *Conway et al.* [1994].
The coupled biogeochemical-physical model is integrated over a 10 year spin-up period, and model data used here for simulating VOS-line based observations and remote sensing are taken from the model year 11.

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 23a). In a basin-wide comparison of the simulated annual cycle's amplitude to the Takahashi climatology [*Takahashi et al.*, 1997] the general patterns are well reproduced for the subpolar North Atlantic. In the subtropics and the North Atlantic Current region a significant larger amplitude in the annual cycle is found for simulated pCO_2.
Besides a general reproduction of pCO_2 observations, for a reliable methodological study variability of simulated pCO_2 and simulated related parameters must be representative of real data. In our previous study [*Friedrich and Oschlies*, 2007 (submitted)] it was shown that temporal variability along VOS cruise tracks conducted by the carcarrier M/V Falstaff in the years 2002/2003 [*Lüger et al.*, 2004] is similar for observed and simulated pCO_2. Autocorrelation scales for observed SST were found to be larger than those of observed pCO_2 which was well reproduced by 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 climatological forcing in the model which leads to less smallscale variability for the simulated data.
For the physical part of the model eddy energy spectra were calculated by *Eden* [2007]. In the subpolar North Atlantic eddy kinetic energy was found to have a maximum at a marginally larger scale in the observations. Overall the energy spectra and corresponding length scales and their lateral variations for the model simulation were found to be in remarkably good agreement with observations.

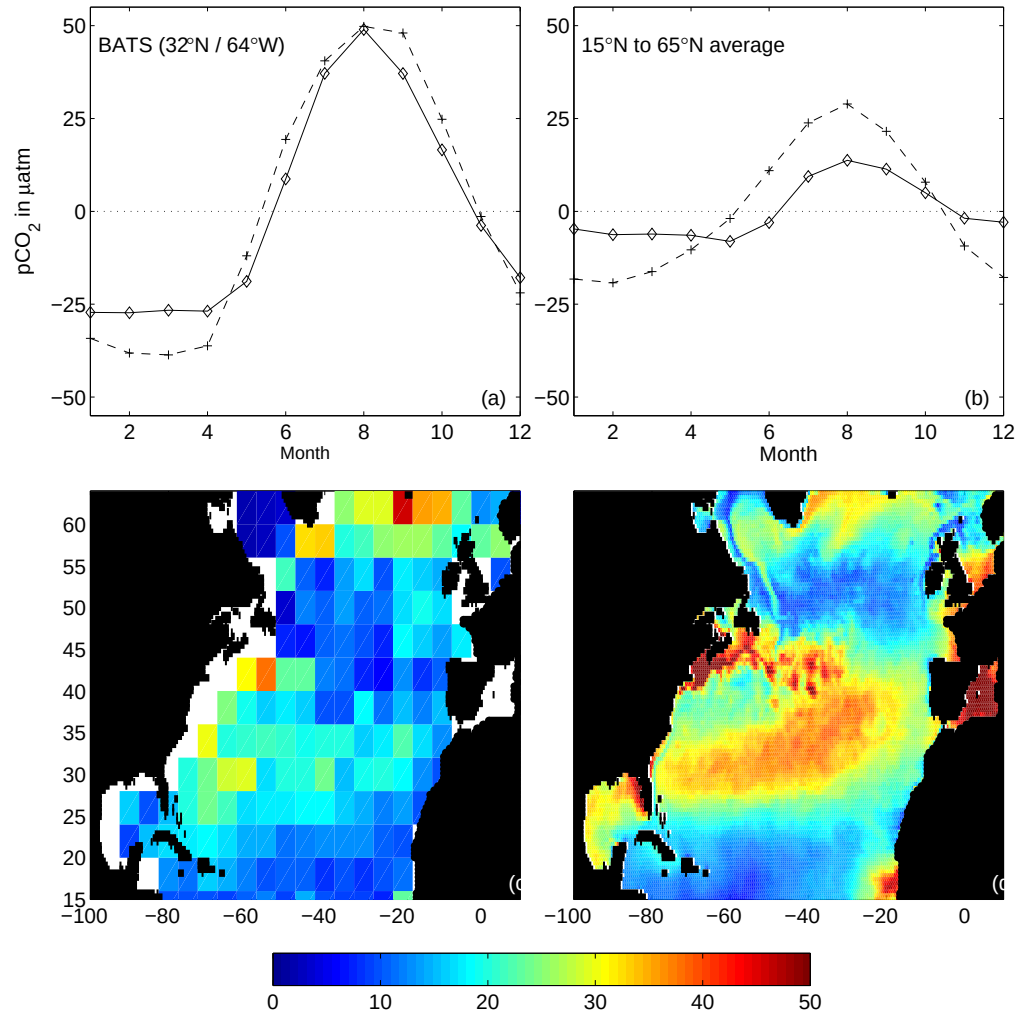

Figure 23: **(a):** Seasonal cycle (annual mean removed) of observed (solid) and simulated (dotted) oceanic pCO_2 at Bermuda. **(b):** Seasonal cycle (annual mean removed) of observed [*Takahashi et al.*, 2008] (solid) and simulated (dotted) oceanic pCO_2 averaged over 15°N to 65°N. Annual variance of observed [*Takahashi et al.*, 2008] (c) and simulated (d) pCO_2 in μatm. Bermuda pCO_2 data were kindly provided by Niki Gruber. Note the preindustrial setup of the model.

Data and Methods

Self-organizing neural network

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 ($\sim$ our model domain). As shown by *Lefèvre et al.* [2005] and *Friedrich and Oschlies* [2007, submitted] these observations can be used for the training of a self-organizing map (SOM) which, in turn, can be used to estimate pCO_2 from related pa-

rameters (e.g. SST, SSS) over the area represented by the training data. SOMs are a subtype of neural networks and work as an associative memory. During the training a lower dimensional representation of the training data (month, latitude, longitude, SST, SSS) is created in a self-organizing process (Figure 24a-d). Subsequently, 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 hyperplane over the training data in which the grade of discretization (and hence the resolution) depends on the training data density (Figure 24e). In the application of the SOM, an input vector 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 detailed description of the training process and the functioning of the SOM can be found in *Kohonen* [1982], *Lefèvre et al.* [2005] and *Friedrich and Oschlies* [2007, submitted].
In our case, pCO_2 is, besides its dependence on sea level pressure, a function of dissolved inorganic carbon (DIC), total alkalinity, SST and SSS. Because for any individual ocean basin total alkalinity can be calculated from SSS using a nonlinear empirical fit [e.g., *Eden and Oschlies*, 2006] ARGO SST and SSS data already provide substantial (though punctual) information about parameters that determine pCO_2. In addition to VOS-line samples of pCO_2, SST and SSS, we here use the position and the month of the samples to form the training data set. Even though their use appears unfounded since month, latitude and longitude have no direct impact on pCO_2, the SOM can regard them as information about an underlying background pCO_2 field which is then altered by SST and SSS. A limitation of the training process to SST and SSS would cause substantially larger mapping errors.

Objective mapping

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 seawater pCO_2 is subject to high spatial variability, an extrapolation to surrounding areas has to account for the typical scales of this variability.
In the present study, simulated VOS-line observations and ARGO-based estimates of pCO_2 were extrapolated monthly to a $1/2^\circ \times 1/2^\circ$ grid using objective mapping and a simple Gaussian weighting. A mean cut-off 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 in the North Atlantic. For the influence radius (R_{inf}) we use a value of $1/2^\circ$ which corresponds to a mean autocorrelation value of 0.5 [*Li et al.*, 2005]. We are aware of the problem that using a mean cut-off and influence radius neglects any regional fluctuations in spatial variability, but so far the sparseness of pCO_2 obser-

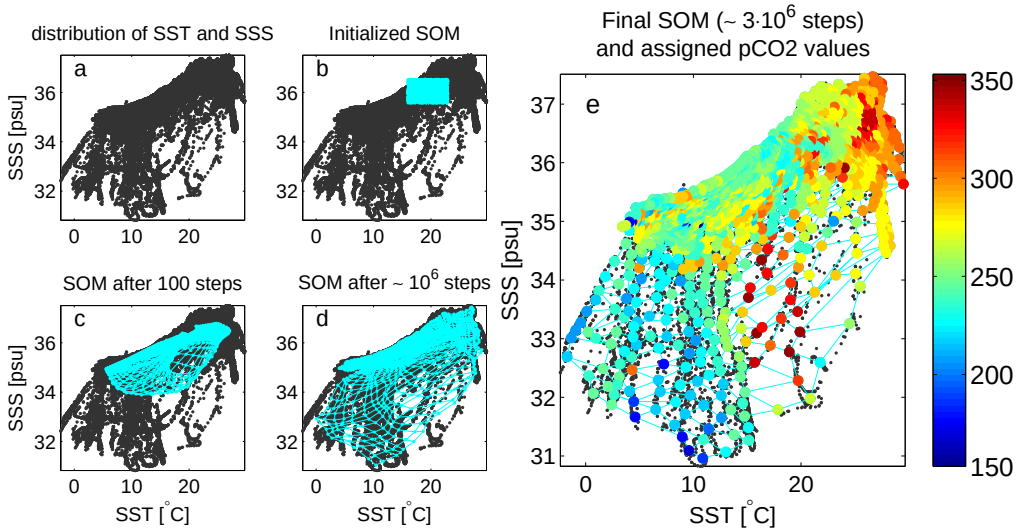

Figure 24: Formation of a SOM. **(a):** Distribution of SST and SSS in simulated VOS-line data. **(b):** Randomly initialized SOM-network. **(c)** and **(d):** Inflation during training process. **(e):** Final SOM-network with assigned pCO_2 values in μatm. Dimension of the SOM was reduced for illustration to SST and SSS only. Note the preindustrial setup of the model.

vations does not allow reliable regional autocorrelation-length estimates on a monthly time scale. The Gaussian weights for all data within the cut-off radius are then calculated by:

$$\kappa_i = \frac{e^{-(\frac{d_i}{R_{inf}})^2}}{\sqrt{2\pi}}. \tag{18}$$

with (d_i) being the distance between the grid point and the i-th data point (x_i) within the cut-off radius.
Calculating decorrelation lags in time on a monthly scale would require several months long uninterrupted timeseries of in-situ pCO_2 data. As reported by *Li et al.* [2005], not even the Bermuda timeseries (Hydrostation "S" and BATS) is sufficient for this analysis. Allthough temporal decorrelation scales (and hence temporal cut-off 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 calender month into monthly means.

Results and Discussion

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 on 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 synthetic "observations" using an objective mapping method with a 2° cut-off scale derived from *Li et al.* [2005] approximately 70 % of the region 15°N to 65°N remains uncovered over the course of the year. 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 4 and Figure 25). Thus, relying exclusively on (extrapolated) VOS-line pCO_2 data impedes reliable estimates of basinwide 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%
RMS-error of annual cycle of basin-mean pCO_2 (see Figure 25)	5.02 μatm		2.01 μatm	
Annual mean of simulated "true" pCO_2 – annual mean of estimated pCO_2	2.21 μatm		-0.03 μatm	
Mean RMS-error of pCO_2	14.4 μatm		15.9 μatm	

Table 4: Seasonal coverage, RMS-error in reproduction of basin-mean pCO_2 annual cycle, deviations from annual mean pCO_2 and mean RMS-error for extrapolated simulated VOS-line "observations" of the year 2005 and float-based mapping of pCO_2. RMS-errors were calculated over areas indicated in the first row (15°N:65°N).

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 $4° \times 5°$ 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 *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 climatological and the simulated "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 a about 10 μatm indicating that the quality of the pCO_2 fields used as missing data replacement is crucial even for the extensive VOS line coverage of the year 2005.

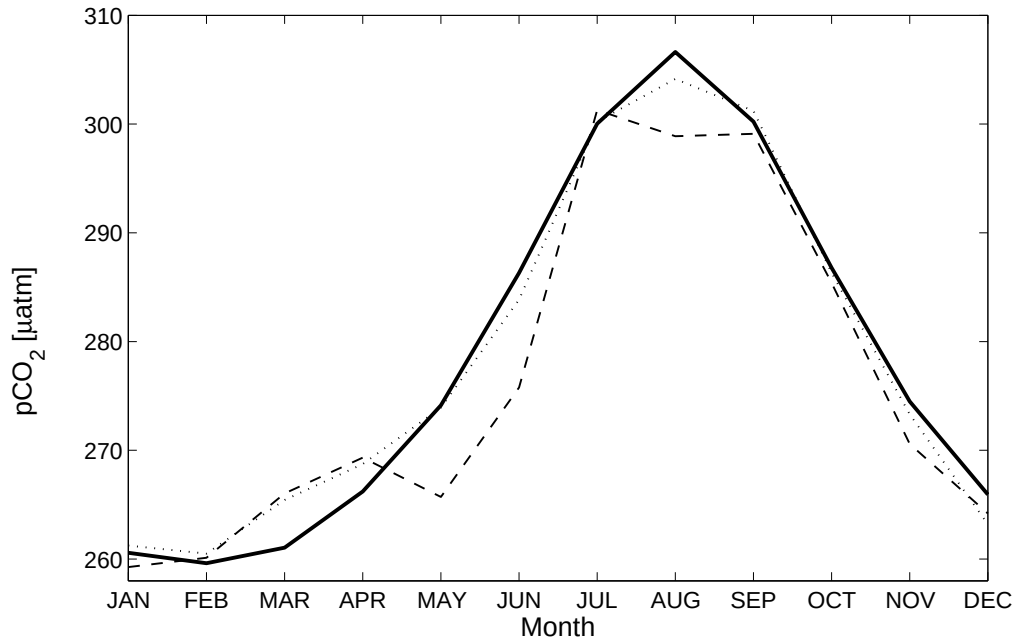

Figure 25: Annual cycle of basin-mean (15°N to 65°N) pCO_2 in μatm. **Solid:** Simulated "true" annual cycle of pCO_2. **Dashed:** Extrapolated VOS-line data. **Dotted:** Mapping based on ARGO float data of SST and SSS and extrapolated VOS-line data. Note the preindustrial setup of the model.

In addition to the regional high-density coverage of pCO_2 measurements achieved by VOS lines, ARGO floats provide a large number of approxi-

mately evenly distributed sampling points. For the considered area (15°N to 65°N) this number 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 simulated VOS-line "observations" of the year 2005 and applied to the simulated ARGO float data of the same year. The punctual pCO_2 estimates were extrapolated by objective mapping with a 2° cut-off and a 0.5° influence radius and subsequently used to fill in regions uncovered by VOS-lines. This combination of extrapolated synthetic "observations" and float-based estimates of pCO_2 increases the coverage of the region 15°N to 65°N to approximately 70 %.

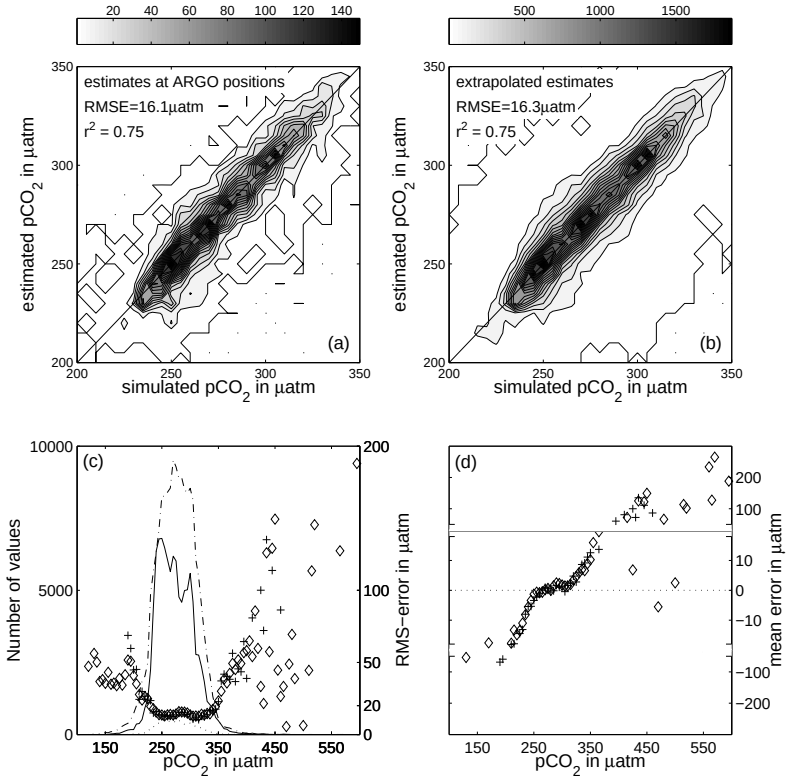

Figure 26: Comparison of simulated and estimated pCO_2 in numbers of data for (a) estimates at ARGO float positions and (b) extrapolated estimates. **(c):** Number of values in modeled pCO_2 (dash-dot), simulated year 2005 VOS-lines (solid) and simulated ARGO floats (dotted). RMS-error in μatm vs. modeled pCO_2 in μatm (5 μatm intervals) for estimates at ARGO float positions (crosses) and extrapolated estimates (diamonds). **(d):** Mean error in μatm vs. 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 (d) and the preindustrial setup of the model.

Figure 25 and Table 4 respectively compare the observational coverage and accuracy in the reproduction of the pCO_2 annual cycle and annual mean pCO_2 for extrapolated VOS-line data and extrapolated float-based estimates. Even though extrapolated VOS-line observations give the most precise estimates for the area covered by the objective analysis, their spatial and temporal coverage leads to a lack of representativeness of these estimates for the entire basin. ARGO float-based estimates assure a much higher accuracy in reproduction of both the annual mean pCO_2 cycle and the annual mean. 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 for the region considered.

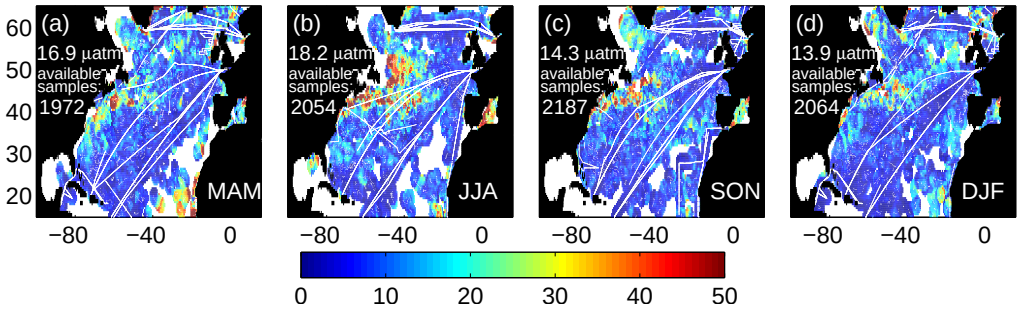

Figure 27: Seasonal mean RMS-error in μatm and number of available ARGO-float samples for mapping based on year 2005 VOS-lines and ARGO float coverage. White lines and dots: corresponding seasonal VOS-lines and ARGO float positions of the year 2005.

The distributions of estimated and simulated "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 26. The mapping results in a RMS-error of 16.1 μatm which is almost not altered by the subsequent extrapolation. For simulated VOS-lines (Figure 26, solid line), basin wide modeled pCO_2 (dash-dot) and simulated ARGO floats (dotted) most pCO_2 values are concentrated in the interval of 230-330 μatm (note the pre-industrial setup of the model). In this range, due to the associated most precise resolution of the SOM, RMS-errors are found to be minimum. Largest RMS-errors occur for $pCO_2 < 200\mu$atm and for $> 400\mu$atm, which are not well represented by the training data set of simulated VOS data, whereas lower (higher) pCO_2 values are overestimated (underestimated).
Figure 27 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 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 in *Friedrich and Oschlies* [2007, submitted] 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.

Comparison to remote-sensing based mapping

The methodological study by *Friedrich and Oschlies* [2007, submitted] showed that a basinwide view on pCO_2 can be obtained by combining VOS-line observations and remote sensing data of SST and chlorophyll (Chl) 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 at most a value of 50 % of the days per month in only a few parts of the basin for any month of the year. This uneven distribution of data gaps necessitates a replacement of missing data, e.g. with climatological SST and Chl values which unfortunately results in biased monthly means forming then the basis of the mapping.

Figure 28 compares the annual cycles of CO_2 fluxes [4] derived from remote-sensing based pCO_2 estimates in which lacks in optical satellite coverage were replaced by climatological values of SST and Chl [*Friedrich and Oschlies*, 2007 (submitted)] and from float-based mapping 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 reproducing 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%, allthough 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 remote sensing-based mapping benefits from a fortuitous canceling of errors over the basin.
It must be emphasized that remote sensing-based mapping as presented in *Friedrich and Oschlies* [2007, submitted] and the ARGO-based method presented here have a similar potential for estimating pCO_2 and CO_2 fluxes in all

[4] CO_2 fluxes were calculated from estimated and simulated "true" pCO_2 fields using the piston velocity formulation of *Wanninkhof* [1992].

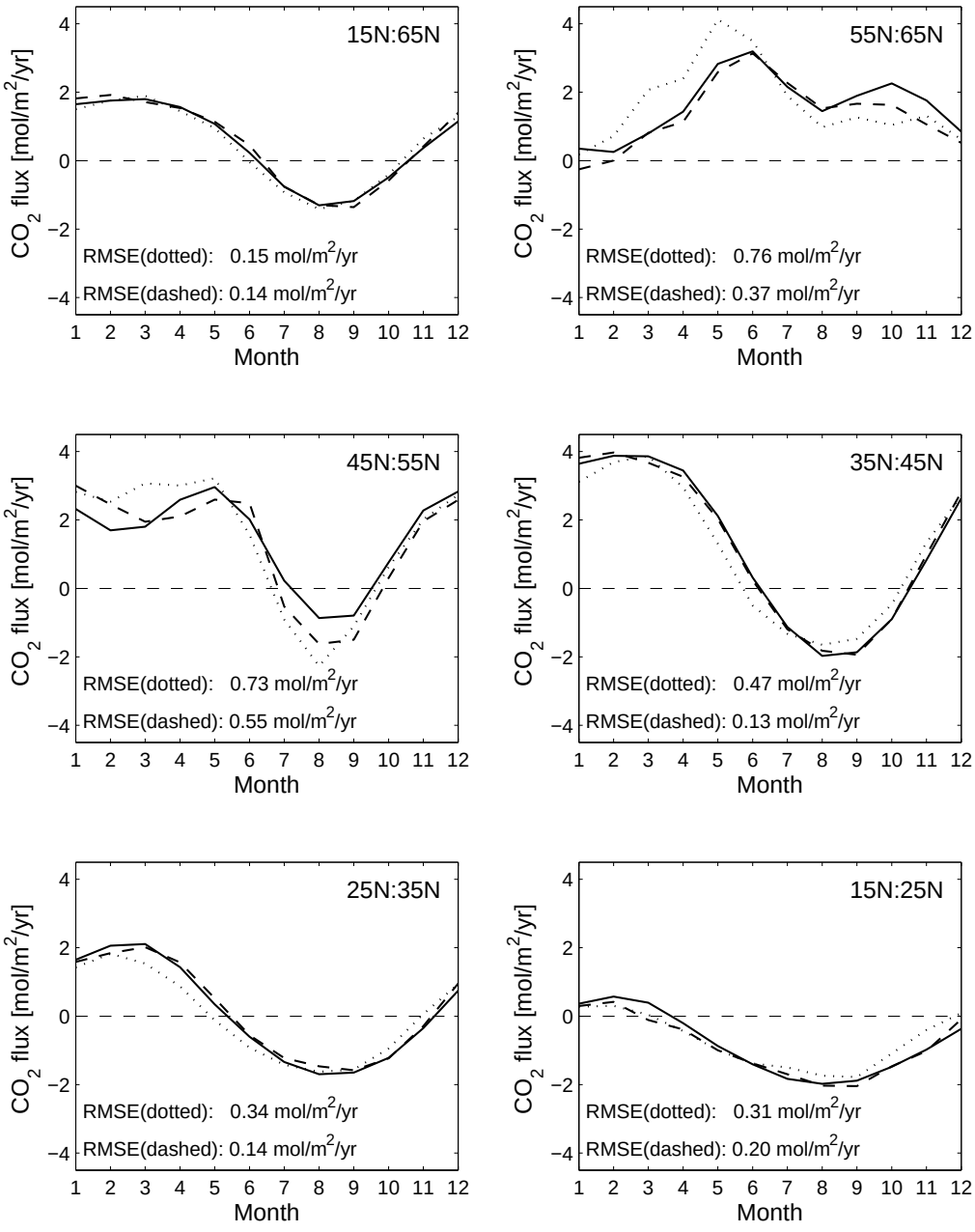

Figure 28: Annual cycle of simulated "true" (solid) and estimated CO_2 fluxes averaged over given latitude ranges in mol/m^2/yr. Underlying pCO_2 fields correspond to Figure 27 for dashed line or were mapped from remote sensing of SST and Chl replacing missing data with climatological values for SST and Chl (dotted line), respectively. Positive values denote fluxes into the ocean. Note the preindustrial setup of the model.

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 render unnecessary a dealing with missing data.
Furthermore, 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.

Mixed layer depth and chlorophyll

In addition to the relatively evenly distributed horizontal sampling points, ARGO floats also provide information about the vertical. Depth profiles of SST and SSS can be used to calculate mixed layer depth (MLD). The use of MLD or Chl as training parameters for the SOM has also been simulated in our methodological study. Chl cannot yet be easily measured with floats and has to be remotely sensed at their positions for the application of the SOM. Thus, the presented method would lose its major advantage.
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 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 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.
The gain, however, turns out to be small. In this hypothetical scenario, even if in situ MLD were accurately available along the VOS lines (to train the SOM), using in situ MLD or Chl additionally to SST and SSS reduced the RMS-error in reproducing the annual cycle of pCO_2 by about 10 % and 12 % respectively.

Conclusions

We have developed and analyzed a novel method that combines VOS-line and ARGO float data by a self-organizing neural network in order to gener-

ate nearly basinwide monthly maps of pCO_2. In contrast to remote sensing, 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 with an RMS-error of 15.9 μatm. This good coverage allows for reliable pCO_2 estimates representative for the basin mean. The spatial patterns reveal large errors in the North Atlantic Current and the West African upwelling region. 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%, respectively, compared to a remote sensing-based mapping.

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 C. Eden for running the biogeochemical ocean model. This work was funded by the European Union via CARBOOCEAN 511176 (GOCE).

References

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, Volume 4: Nutrients. S. Levitus, Ed., NOAA Atlas NESDIS 54, U.S. Government Printing Office, Wash., D.C., 392 pp.

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. (2006), 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 (2007), Neural-network based estimates of North Atlantic surface pCO_2 from satellite data - a methodological study, *J. Geophys. Res.*, submitted.

Gaspar, P., Y. Gregoris, and J.-M. Lefevre, (1990), A simple 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*, 179–193.

Kalnay, E. et al. (1996), The NCEP/NCAR 40 years reanalysis project, *Bull. Amer. Meteor. Soc.*, *77*, 437–471.

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. Cycle*, *18*, GB4031, doi:10.1029/2004GB002247.

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

Lefévre, N., A. J. Watson, 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, US Department of Commerce, NOAA, US Government Printing Oce, Washington, DC.

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

Oschlies, A. and V. Garçon (1999), An eddy-permitting coupled physical-biological model of the North Atlantic. Part 1: Sensitivity to advection numerics and mixed layer physics, *Global Biogeochem. Cycles*, *13*, 135–160, doi:10.1029/98GB02811.

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

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.

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

Conclusions

Two closely connected aims have been pursued in my studies: Understanding driving mechanisms of surface pCO_2 variability in the North Atlantic and developing and testing methods to map pCO_2 on a basinwide scale.
The sensitivity tests carried out by using different combinations of climatological and interannually varying wind stress and heat flux anomaly forcing fields showed the benefit of a model-based assessment of the processes causing temporal pCO_2 fluctuations. The analysis indicated a decoupling of CO_2 and O_2 fluxes in the subtropical and subpolar North Atlantic. For the variability of O_2 fluxes a pronounced separation into a heat-flux driven subtropical and subpolar and wind driven equatorial basin was revealed. Conversely, CO_2 flux variability was found to be wind-stress driven in the entire model domain. The simulated mean carbon uptake is mostly located in the subpolar region. However, year-to-year fluctuations of the oceanic carbon uptake are of the same magnitude in the subpolar region, the subtropics and the equatorial Atlantic. The consequence for pCO_2 monitoring aiming to reliably detect interannual variability and trends is that all parts of the North Atlantic must be included in an observing system.
Basinwide and accurate estimates of pCO_2 remain a difficult challenge. It was demonstrated that VOS-line observations and remote sensing or ARGO float data can be combined by a self-organizing neural network in order to generate monthly basinwide or nearly basinwide maps of pCO_2. For the reference year 2005 a basinwide mean RMS-error of 15.9 ppm is found for ARGO-float based estimates which cover about 70% of the considered area. Mapping based on remote sensing of SST and Chl resulted in a basinwide mean RMS-error of 19.0 ppm for an idealized remote sensing without data gaps and 21.1 ppm if this missing data is replaced by climatological values of SST and Chl.
The most crucial limiting factor for the accuracy of both mapping procedures is high spatial variability of pCO_2 in the North Atlantic Current and the West African upwelling region. Additionally, large and important areas such as the Labrador Sea are still uncovered by VOS-lines reducing the amount and the representativeness of training data samples.
For remote-sensing based pCO_2 estimates several aspects were found to be pivotal for the accuracy. First of all, miss-mapping is large when carried out on a daily basis mostly due to remote sensing errors in Chl. Since these errors are of stochastical nature, monthly mean data of SST and Chl provide a more reliable basis for producing basinwide pCO_2 charts.
Furthermore, taking sampling position in time and space as additional parameters for a successful mapping turned out to be inevitable if the formation of the topological representation in the SOM is done with VOS-line observations. Even though their use appears unfounded since day, latitude and longitude have no direct impact on pCO_2, the SOM can regard them as information about an underlying background pCO_2 field which is then altered

by physical and biological parameters. A limitation of the training process to these physical and biological parameters such as SST, Chl and mixed layer depth causes substantially larger errors. An alternative approach using for example remotely sensed maps of these parameters as training data will be analyzed by the author in future studies.
Finally, the severely restricted availability of remote sensing data in particular at high latitudes in winter has a large impact on the mapping accuracy. More than half of the RMS-error in the annual cycle of estimated basinwide CO_2 fluxes must be attributed to missing data in the remote sensing which needed to be replaced by climatological values. Especially for the region north of 55°N, CO_2 uptake were significantly overestimated in the spring bloom caused by too high climatological Chl values taken as a replacement. The fact that a coincidence of major spring bloom carbon uptake and hardly detectable Chl concentrations is virtually guaranteed at high latitudes remains unsatisfactory.
ARGO-float based pCO_2 estimates can be a substantial improvement in this context. They provide evenly distributed information about SST and SSS independent from satellite coverage. Although magnitude and patterns of mapping errors are similar to those found for remote-sensing based estimates, annual cycles of pCO_2 and CO_2 fluxes are more accurately reproduced in all parts of the basin. This leads to more reliable and profound estimates compared to remote-sensing based mapping which partly profits from a cancellation of errors.
The sustainable benefit of the present study for reducing uncertainties in the quantification of the North Alantic carbon sink is offered by the model-based methodological study on the combination of observational data. The availability of a simulated "ground truth" enabled an analysis of the basinwide performance of either mapping procedures as well as an accuracy assessment for state-of-the-art mapping techniques. Simulating the common practice of validating the mapping against an independent data set revealed that basinscale errors can be about threefold larger than those expected from this conventional validation. In addition the eddy-resolving model allowed to take a global view on the spatial and temporal variability of the relationships between pCO_2 and for example SST or Chl respectively. Although approximative regressions may succeed in some regions, the assumption of an inherent division of the North Atlantic into seasonal-spatial biophysical regimes with regionally constant regression coefficients can not be supported. Future approaches aiming to deliver accurate basinwide pCO_2 estimates through parameterizations or neural-network based mapping should therefore always consider the significant temporal and spatial variability of physical and biological dependencies of pCO_2 which might not be sufficiently reflected by VOS-line data. Hence, testing the basinwide performance of promising mapping techniques in a suitable biogeochemical model should be regarded as a necessity for a reliable validation.

In concluding, I would like to express my hope that this study is regarded as another key example showing the benefit of combining observational and modeling approaches and that the terms "observer" and "modeler" are no longer seen as a dichotomy but just as two opposite ends of the same continuum.

Danksagung

Mein ganz besonderer Dank gilt Andreas Oschlies für seine hervorragende und geduldige Betreuung und seine stets herzliche und motivierende Art, sich für meine Fragen Zeit zu nehmen.

Der Deutschen Forschungsgemeinschaft und der Europäischen Union verdanke ich die Finanzierung meiner Promotion in den Projekten SFB 460 und CARBOOCEAN.

Im Weiteren möchte ich Carsten Eden, Doug Wallace und Arne Körtzinger hervorheben und für ihre Unterstützung und für zahlreiche anregende Diskussionen danken.

Bedanken möchte ich mich auch bei allen Personen, die im Rahmen der VOS-line, ARGO und SeaWiFS Programme verantwortlich sind für das Sammeln und Auswerten von Daten, die zum grossen Teil die Basis meiner Studien darstellen.

Danken möchte ich auch meinen Eltern, die von kleinst auf meine kindliche Neugier geweckt und gefördert haben.

Außerdem danke ich allen, die für mich da sind und wissen, dass sie hier gemeint sind.