

Variability of the Global Ocean Carbon Sink (1998 through 2011)

A thesis submitted to the School of Environmental Sciences of the
University of East Anglia in partial fulfilment of the requirements for the
degree of Doctor of Philosophy

By Peter Landschützer

April 2014

© This copy of the thesis has been supplied on condition that anyone who consults it is understood to recognise that its copyright rests with the author and that no quotation from the thesis, nor any information derived therefrom, may be published without the author's prior, written consent.

© Copyright 2014

by

Peter Landschützer

”We are at the very beginning of time for the human race. It is not unreasonable that we grapple with problems. But there are tens of thousands of years in the future. Our responsibility is to do what we can, learn what we can, improve the solutions, and pass them on.” (Richard P. Feynman)

Abstract

In this thesis a newly developed 2-step neural network approach is used to reconstruct basin-wide monthly maps of the sea surface partial pressure of CO₂ ($p\text{CO}_2$) at a resolution of $1^\circ \times 1^\circ$ for both the Atlantic Ocean from 1998 through 2007 and the global ocean from 1998 through 2011. From those, air-sea CO₂ flux maps are computed using a standard gas exchange parameterization and high-resolution wind speeds.

Observations form the basis of the studies conducted in this thesis. The neural network estimates benefit from a continuous improvement of the observations, i.e., the Surface Ocean CO₂ Atlas (SOCAT) database. Additionally, bottle samples were collected along the UK-Caribbean line to investigate the variability of the sea surface $p\text{CO}_2$ and its drivers.

The neural network derived $p\text{CO}_2$ estimates fit the observed $p\text{CO}_2$ data with a root mean square error (RMSE) of about 10 μatm in the Atlantic Ocean from 1998 through 2007 and about 12 μatm in the global ocean from 1998 through 2011, with almost no bias in both studies. A check against independent $p\text{CO}_2$ data reveals a larger RMSE, in particular in regions with strong $p\text{CO}_2$ variability and gradients.

Temporal mean contemporary flux estimates for the Atlantic Ocean ($-0.45 \pm 0.15 \text{ Pg C} \cdot \text{yr}^{-1}$) and the global ocean ($-1.54 \pm 0.65 \text{ Pg C} \cdot \text{yr}^{-1}$) agree well with recent studies. Trends and variabilities within the considered time periods are strongly influenced by climate modes. The global results from 1998 through 2011 reveal the strongest variability of the air-sea CO₂ fluxes in the Equatorial Pacific ($\pm 0.12 \text{ Pg C} \cdot \text{yr}^{-1}$, $\pm 1\sigma$), mainly driven by the El Niño Southern Oscillation (ENSO) climate mode. Trends towards a strengthening of the Southern Ocean carbon sink ($-0.36 \pm 0.07 \text{ Pg C} \cdot \text{yr}^{-1} \cdot \text{decade}^{-1}$) from 1998 through 2011 are potentially linked to the recent weakening of the Southern Annular Mode (SAM) index.

Acknowledgements

”Knowledge is in the end based on acknowledgement.”

(Ludwig Wittgenstein)

I cannot be grateful enough for the support from my wife, Leen. Whenever I was down you helped me up and whenever I needed you, you were there. You always believed in me and that is the most valuable gift one can ever receive.

I would like to express my gratitude to my supervisors Dorothee Bakker, Nicolas Gruber and Ute Schuster for your support, your helpful comments and your open ears for new ideas. Without your help, guidance and knowledge I would have been lost so many times.

I would like to thank my colleagues here at UEA, in particular Clare Ostle, Steve Jones and Oliver Legge. You have provided helpful advice, discussions and knowledge but most important of it all, you were always up for a laugh.

I would further like to thank all scientists and bright minds with whom I had so many helpful discussions at conferences and meetings, in particular Maciej Telszewski, who’s work introduced me to the world of neural networks, as well as Tristan Sasse, Shin-ichiro Nakaoka and Jiye Zeng, with whom I had fruitful and inspiring discussions about neural networks.

At last I would like to thank the SOCAT team and all scientists involved for their effort in creating the surface ocean $f\text{CO}_2$ database, but also captains, officers, and crew of all ships on which measurements have been made.

Contents

| | |
|---|-----------|
| Abstract | iv |
| Acknowledgements | v |
| 1 Context and background | 1 |
| 1.1 The Earth's climate system and its natural variability | 2 |
| 1.1.1 CO ₂ and the anthropogenic alteration of the climate system | 4 |
| 1.1.2 Interaction between the spheres: The global carbon budget | 6 |
| 1.2 Ocean–atmosphere interaction | 7 |
| 1.2.1 Air–sea gas flux measurements and methods | 7 |
| 1.2.2 Bulk formulation of the air–sea gas exchange | 9 |
| 1.2.3 Gas exchange rate and transfer velocity | 11 |
| 1.3 Sea surface partial pressure of CO ₂ | 12 |
| 1.3.1 Inorganic carbon cycle | 15 |
| 1.3.2 Ocean circulation and <i>p</i> CO ₂ distribution | 18 |
| 1.3.3 Sea surface <i>p</i> CO ₂ measurements | 19 |
| 1.4 Air–sea CO ₂ flux estimates | 22 |
| 1.4.1 Long–term mean air–sea CO ₂ flux estimates | 22 |
| 1.4.2 Variability and trends of the Atlantic Ocean sea surface <i>p</i> CO ₂ and air–sea flux | 24 |
| 1.4.3 Variability and trends of the global sea surface <i>p</i> CO ₂ and air–sea flux | 26 |
| 1.5 Climate variability and CO ₂ fluxes | 27 |
| 1.6 Current limitations on data interpolation methods | 30 |
| 1.7 Aim of the research | 32 |

| | | |
|----------|--|-----------|
| 2 | Methods | 34 |
| 2.1 | Using neural networks to estimate the sea surface $p\text{CO}_2$ | 36 |
| 2.1.1 | Neural networks | 37 |
| 2.1.2 | Data choice | 40 |
| 2.1.3 | Data preparation and validation data | 45 |
| 2.1.4 | Dividing the global ocean into biogeochemical provinces using a self-organizing map | 48 |
| 2.1.5 | Reconstructing the sea surface $p\text{CO}_2$ using a feed-forward network | 53 |
| 2.1.6 | Air-sea CO_2 flux calculation | 57 |
| 2.2 | Critical reflection on the method | 59 |
| 3 | $p\text{CO}_2$ sampling and processes in the North Atlantic | 61 |
| 3.1 | Underway measurements of the sea surface $p\text{CO}_2$ | 62 |
| 3.1.1 | Measurement system and maintenance | 64 |
| 3.1.2 | Quality control | 66 |
| 3.2 | Collection of discrete bottle samples along the UK-Caribbean line | 67 |
| 3.2.1 | Analysis of discrete bottle samples | 68 |
| 3.2.2 | Quality control | 70 |
| 3.2.3 | Comparison with independent data products | 71 |
| 3.2.4 | Variability of the sea surface DIC and TALK along the UK-Caribbean line | 74 |
| 3.3 | Sea surface $p\text{CO}_2$ along the UK-Caribbean line | 79 |
| 3.3.1 | Sea surface $p\text{CO}_2$ calculation | 79 |
| 3.3.2 | Validation of the calculated sea surface $p\text{CO}_2$ | 79 |
| 3.3.3 | Variability and drivers of the sea surface $p\text{CO}_2$ | 80 |
| 3.4 | Summary and conclusion | 86 |
| 4 | Seasonal to inter-annual variability of the Atlantic Ocean carbon sink | 88 |
| 4.1 | Residuals and validation | 90 |
| 4.2 | Validation with independent observations | 94 |
| 4.3 | Uncertainty of the air-sea flux | 95 |
| 4.4 | Decadal mean $p\text{CO}_2$ and air-sea CO_2 flux | 99 |

| | | |
|----------|--|------------|
| 4.5 | Seasonality | 103 |
| 4.6 | CO ₂ trends and inter-annual variability | 106 |
| 4.7 | Summary and conclusion | 113 |
| 5 | Variability of the global ocean carbon sink | 115 |
| 5.1 | Residuals and validation | 116 |
| 5.2 | Validation with independent observations | 117 |
| 5.3 | Uncertainty of the air-sea CO ₂ flux | 119 |
| 5.4 | Long term mean <i>p</i> CO ₂ and air-sea flux in the global ocean | 123 |
| 5.5 | Seasonality | 126 |
| 5.6 | Inter-annual variability and trends | 128 |
| 5.6.1 | The Pacific Ocean | 130 |
| 5.6.2 | The Atlantic Ocean | 130 |
| 5.6.3 | The Southern Ocean | 132 |
| 5.6.4 | The Indian Ocean | 136 |
| 5.6.5 | The global ocean | 136 |
| 5.7 | A global carbon budget (1998 through 2011) | 138 |
| 5.8 | Summary and conclusion | 141 |
| 6 | General discussions, conclusions and future research | 143 |
| 6.1 | Summary of the main findings | 144 |
| 6.2 | General discussion and conclusions | 145 |
| 6.3 | Future research | 150 |
| A | List of acronyms | 152 |
| B | List of symbols | 155 |
| | References | 159 |

List of tables

| | | |
|-----|--|-----|
| 2.1 | Datasets and their original resolution used in this thesis | 43 |
| 2.2 | Input and target vector elements of the neural network datasets | 45 |
| 2.3 | Arrangement of the neural network input vector elements | 47 |
| 4.1 | Statistical comparison between the Atlantic Ocean $p\text{CO}_2$ estimates and the SOCAT v1.5 data | 91 |
| 4.2 | Comparison between the neural network estimates and the results from the RECCAP Atlantic Ocean study | 101 |
| 4.3 | Results of the sensitivity runs | 103 |
| 5.1 | Statistical comparison between the global ocean $p\text{CO}_2$ estimates and the SOCAT v2 data | 117 |
| 5.2 | Statistical comparison between the global ocean $p\text{CO}_2$ estimates and independent timeseries stations | 119 |
| 5.3 | Temporal mean fluxes, trends and IAV for the major ocean basins and the global ocean | 126 |
| 6.1 | Comparison of the contemporary and the anthropogenic flux between the SOM–FFN results and recent studies | 147 |
| 6.2 | Comparison between basin-wide SOM–FFN results and recent estimates . | 149 |
| A.1 | List of acronyms | 152 |
| B.1 | Variables, symbols and description | 155 |

List of figures

| | | |
|------|--|----|
| 1.1 | Schematic of the global climate system and its interconnections | 4 |
| 1.2 | Radiative forcing of human activities and natural processes | 5 |
| 1.3 | Climatological mean $\Delta p\text{CO}_2$ distribution | 13 |
| 1.4 | Schematic of the ocean conveyor circulation | 18 |
| 1.5 | Schematic of the ocean surface currents | 20 |
| 1.6 | Number of $f\text{CO}_2$ observations per decade within SOCAT v1.5 | 21 |
| 1.7 | Thermal driven minus non-thermal driven seasonal $p\text{CO}_2$ amplitude . . . | 25 |
| 1.8 | Linear trends within 16 biomes in the global ocean | 27 |
| 1.9 | MEI, NAO and SAM climate indices | 28 |
| 1.10 | Number of months where $f\text{CO}_2$ measurements are available for each gridbox in SOCAT v1.5 | 31 |
| 2.1 | Schematic of the 2-step neural network approach | 38 |
| 2.2 | Schematic of the neural network processing chain | 39 |
| 2.3 | Schematic of the neuron connections within a multi-layer network | 40 |
| 2.4 | Temporal mean of the gridded SOCAT data used in this thesis | 42 |
| 2.5 | Flow chart of the different stages within the 2-step neural network method | 46 |
| 2.6 | Schematic of the neuron response towards an input vector within a SOM . | 49 |
| 2.7 | Biogeochemical provinces in the Atlantic Ocean derived from a SOM . . . | 52 |
| 2.8 | Biogeochemical provinces in the global ocean derived from a SOM | 54 |
| 3.1 | Schematic of the underway measurement system | 63 |
| 3.2 | Map of the ship tracks of the UK-Caribbean line where bottle samples were taken | 68 |
| 3.3 | Comparison of the sampled DIC and TALK data with climatologies | 73 |
| 3.4 | DIC and TALK data points along the UK-Caribbean line | 75 |

| | | |
|------|---|-----|
| 3.5 | SST and SSS data points along the UK–Caribbean line | 76 |
| 3.6 | DIC/TALK ratio and property plots | 77 |
| 3.7 | Comparison of the calculated $p\text{CO}_2$ data with underway observations and results from the SOM–FFN approach | 81 |
| 3.8 | Polynomial fit to the calculated $p\text{CO}_2$ data | 82 |
| 3.9 | Seasonal Hovmöller plot of the data along the UK–Caribbean line | 83 |
| 3.10 | Changes in $p\text{CO}_2$ along the UK–Caribbean line by drivers | 85 |
| 4.1 | Maps of the residuals between the Atlantic Ocean estimates and the SO- CAT v1.5 gridded data | 91 |
| 4.2 | Residuals as a function of the input parameters and the estimated $p\text{CO}_2$. | 92 |
| 4.3 | Variability of the $p\text{CO}_2$ estimates compared to timeseries stations in the subtropical Atlantic Ocean | 93 |
| 4.4 | Maps of the residuals between the Atlantic Ocean estimates and indepen- dent SOCAT v2 data | 96 |
| 4.5 | Empirical semi–variograms of randomly chosen residual ensembles in the Atlantic Ocean | 98 |
| 4.6 | Decadal mean surface ocean $p\text{CO}_2$ and CO_2 flux density in the Atlantic Ocean | 99 |
| 4.7 | Spatial $p\text{CO}_2$ difference compared to climatological data in the Atlantic Ocean | 103 |
| 4.8 | Zonally averaged seasonal cycle of the sea surface $p\text{CO}_2$ and the CO_2 flux in the Atlantic Ocean | 104 |
| 4.9 | Seasonal difference between the estimates and climatological data in the Atlantic Ocean | 104 |
| 4.10 | Thermally and non–thermally driven seasonal cycles in the Atlantic Ocean | 105 |
| 4.11 | Linear $p\text{CO}_2$ and flux trends in the Atlantic Ocean | 107 |
| 4.12 | Seasonal and annual mean CO_2 fluxes in the Atlantic Ocean | 110 |
| 4.13 | Annual mean fluxes along the UK–Caribbean line | 111 |
| 4.14 | Comparison between subtropical and subpolar data with the NAO index . | 112 |
| 5.1 | Maps of the residuals between the global ocean estimates and the SOCAT v2 gridded data | 118 |

| | | |
|------|---|-----|
| 5.2 | Variability of the $p\text{CO}_2$ estimates compared to timeseries stations in the subtropical Atlantic and Pacific Ocean | 120 |
| 5.3 | Map of the 11 global Ocean Inversion regions | 121 |
| 5.4 | Empirical semi-variograms of randomly chosen residual ensembles in the global ocean | 122 |
| 5.5 | Mean surface ocean $p\text{CO}_2$ and CO_2 flux density in the global ocean | 124 |
| 5.6 | Spatial $p\text{CO}_2$ difference between SOM-FFN estimates and climatological data in the global ocean | 126 |
| 5.7 | Seasonal cycles of the air-sea carbon flux in the major ocean basins | 127 |
| 5.8 | Annual mean integrated carbon flux from 1998 through 2011 | 128 |
| 5.9 | Linear $p\text{CO}_2$ trends in the global ocean | 129 |
| 5.10 | Seasonal and annual mean carbon flux in the major ocean basins | 131 |
| 5.11 | Southern Ocean $\Delta p\text{CO}_2$ compared to CHL and MLD | 133 |
| 5.12 | Southern Ocean $\Delta p\text{CO}_2$ response to SAM and ENSO | 134 |
| 5.13 | Spatial distribution of the leading EOF of the global surface ocean $p\text{CO}_2$ variability | 137 |
| 5.14 | Temporal distribution of the leading EOF of the surface ocean $p\text{CO}_2$ variability | 138 |
| 5.15 | Sources and sinks of the global ocean carbon budget from 1998 through 2011 | 139 |

Chapter 1

Context and background

”It is not knowledge, but the act of learning,
not possession but the act of getting there,
which grants the greatest enjoyment.”

(Carl Friedrich Gauss)

The aim of this thesis is to investigate the variability of the sea surface partial pressure of CO₂ ($p\text{CO}_2$) and the air–sea exchange of this gas, based on observations. This chapter provides an introduction on the relevance of this topic, what is already known and where the research of this thesis fits into the bigger context. It will provide an overview to the Earth’s climate system and anthropogenic alterations of the natural greenhouse gas effect and the role of the global ocean within the climate system. Roughly 50% of human emitted carbon remains in the atmosphere with the global ocean to–date being a sink for carbon dioxide (CO₂). The processes that control the sea surface $p\text{CO}_2$, the air–sea exchange of CO₂, as well as recently derived CO₂ sink estimates of the Atlantic Ocean and the global ocean will further be reviewed.

1.1 The Earth’s climate system and its natural variability

The Earth’s climate is a complex coupled system, which it is not simply driven by the interaction of its components, but also by external forcings. It receives its energy input from the sun, with a total amount of roughly $1370 \text{ W} \cdot \text{m}^{-2}$ (Kiehl and Trenberth, 1997). Most of this incoming radiation is reflected back to space, e.g. by aerosol particles and the natural albedo effect, and never reaches the Earth’s surface. In theory, Earth can be considered a black body, hence it acts as an ideal emitter of the incoming radiation and its absorbed energy equals the emitted radiation:

$$E_{\text{abs}} = E_{\text{em}} \quad (1.1)$$

Therefore, the emitted energy must equal the energy received from the sun according to the Stefan–Boltzmann law:

$$4 \cdot \pi \cdot R_S^2 \cdot \sigma \cdot T_S^4 \cdot \frac{\pi \cdot R_E^2}{4 \cdot \pi \cdot D^2} \cdot (1 - \alpha) = 4 \cdot \pi \cdot R_E^2 \cdot \sigma \cdot T_E^4 \quad (1.2)$$

The left hand side of the equation describes the emitted black body radiation by the sun, calculated via the radius (R_S) and the surface temperature (T_S) of the sun multiplied by the the amount of energy received by the Earth surface, calculated from the Earth radius

(R_E) and the distance to the sun (D). The variable σ describes the Stefan–Boltzman constant and α describes the amount of energy reflected due to the albedo effect. The right hand side describes the black body energy emitted by Earth, as a function of the radius (R_E) and the surface temperature (T_E). Solving this equation for the temperature of the Earth leads to:

$$T_E = T_S \cdot \sqrt{\frac{R_S \cdot \sqrt{1 - \alpha}}{2 \cdot D}} \quad (1.3)$$

On average, the calculated surface temperature of the Earth (T_E) is about -19°C , hence much lower than the actual observed temperature of the Earth surface. This temperature difference is linked to an important natural process. The emitted radiation by the planet is long wave radiation which gets absorbed by infra–red absorbing gases (Tyndall, 1861), such as carbon dioxide (CO_2), nitrous oxide (N_2O), ozone (O_3), water vapour (H_2O) and methane (CH_4). This effect, called greenhouse effect, alters the emitted Energy (E_{em}) and thereby the temperature of the lower atmosphere keeping it $\sim 30^\circ\text{C}$ warmer on average than without a greenhouse effect (Kiehl and Trenberth, 1997).

Atmospheric temperature is only one measure of the state of the climate system. Long term averages (usually over a period of 30 years) of wind, precipitation, humidity and other meteorological parameters are commonly used to define the climate of a certain region (Thorntwaite, 1948; Kottek *et al.*, 2006; Le Treut *et al.*, 2007). However, regional climates are only the result of the state of the global climate system. The global climate system consists of components or spheres, namely the atmosphere, the hydrosphere, the cryosphere, the lithosphere and the biosphere, which are interacting with each other. These interactions are illustrated in figure 1.1 from Le Treut *et al.* (2007) and include natural as well as anthropogenic processes linked to positive and negative climate feedbacks, such as the exchange of heat between the atmosphere and the hydrosphere, the exchange of gases between the atmosphere and the biosphere or the exchange of water between the atmosphere and the cryosphere.

Internal and inter–connected processes within the spheres as well as external forcing cycles that effect the incoming solar radiation of the sun (see figure 1.1) lead to natural climate variabilities. These variabilities range from timescales of millions of years like

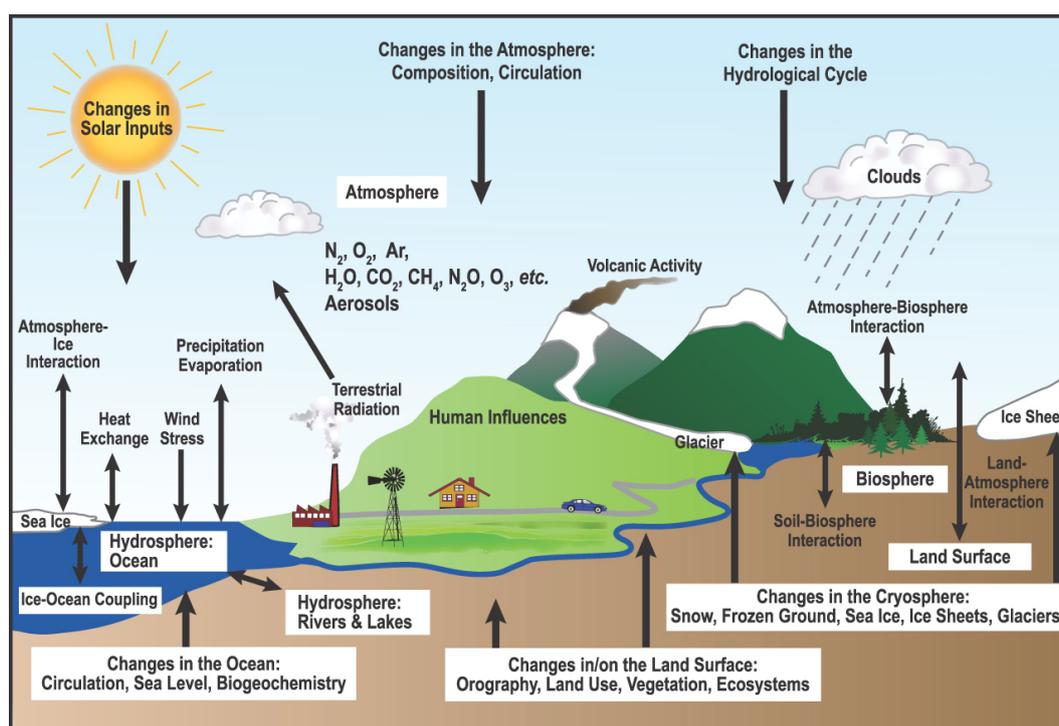


Figure 1.1: Schematic view of the components of the climate system, their processes and interactions from Le Treut *et al.* (2007).

e.g. the movement of the continents driven by the Earth's plate tectonics (Wegener, 1966), tens to hundreds of thousands of years, like e.g. the Milankovich cycles (Milankovitch, 1920, 1930) effecting the incoming solar radiation, i.e. altering the amount of energy received from the sun, to only a few years like e.g. major volcanic eruptions (Robock, 2000) which increase the amount of back-scatter particles in the atmosphere. Climate records, such as ice and sediment cores, reveal strong natural variabilities of the global climate in the past, which help to gain insight into the natural processes involved.

1.1.1 CO₂ and the anthropogenic alteration of the climate system

Past records show a variety of natural climate variabilities from ice-ages to warmer periods, but since the beginning of the industrial revolution, the Earth's climate faces a series of new humankind induced changes. Fossil fuel burning, land use change and cement production led to increasing concentrations of greenhouse gases in the atmosphere, altering the surface temperature of planet Earth. Ice-core records show that since the industrial revolution the increasing CO₂ concentrations continue to rise (Etheridge *et al.*, 1996), exceeding levels from the past 800,000 years (Lüthi *et al.*, 2008).

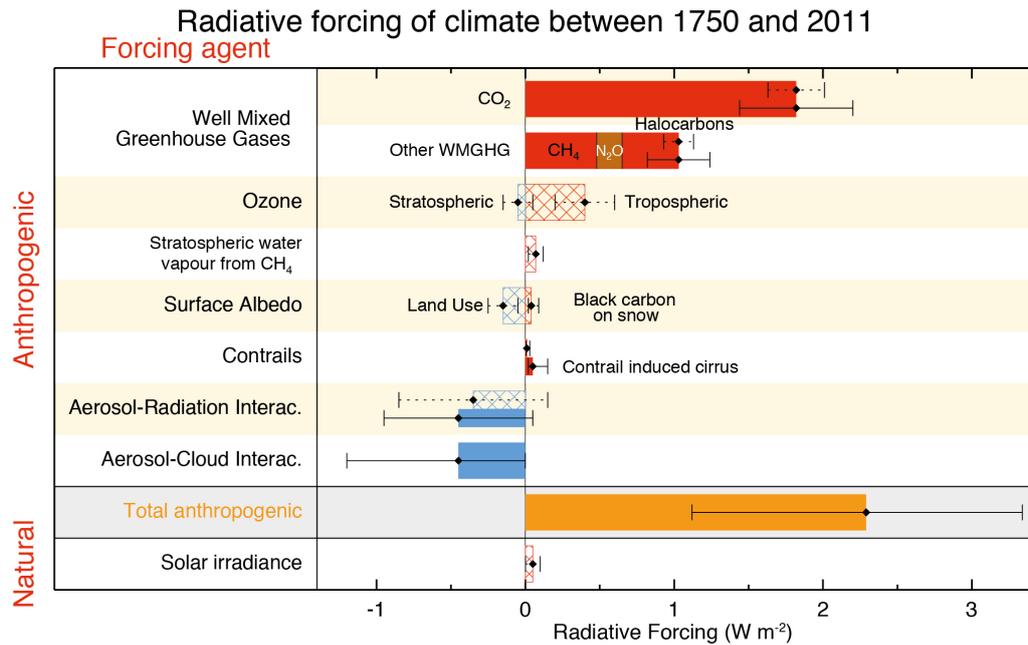


Figure 1.2: Summary of the principal components of the radiative forcing of climate change from Myhre *et al.* (2013). Forcing estimates are derived from observed abundance, model simulations in combination with observations, or purely model based (see Myhre *et al.* (2013)). Solid bars illustrate novel effective radiative forcing estimates, whereas hatched bars illustrate radiative forcing estimates as they were presented previously in IPCC reports (Myhre *et al.*, 2013). Uncertainties are illustrated by the whiskers of each bar.

Figure 1.2 from the fifth assessment report (Myhre *et al.*, 2013) of the Intergovernmental Panel on Climate Change (IPCC) quantifies the increase in radiative forcing (the imbalance between received radiation from the sun and emitted radiation from planet Earth) since the start of the industrial revolution in 1750 up to 2011, linked to human activities. It clearly shows the strongest contribution from greenhouse gases, especially CO₂. It further illustrates the relatively small effect of natural processes compared to human activities in observed changes in radiative forcing. Overall the estimated increase in anthropogenic-induced radiative forcing is estimated to be $\sim 2.3 \text{ W} \cdot \text{m}^{-2}$ (Myhre *et al.*, 2013), altering the imbalance between absorbed and emitted energy (equation 1.1).

The global marine and terrestrial surface temperature record HadCRUT4, as presented in Morice *et al.* (2012) from 1850 through 2012, reveals that surface temperatures are increasing on our planet. Morice *et al.* (2012) find the warmest temperatures in the 21st century, with 2010 being the warmest year since the beginning of the record.

Since the beginning of the industrial revolution, humankind has continuously altered the climate system. Future projections, as summarized in Table SPM.1 in Stocker *et al.*

(2013), show that most land areas are virtually certain to experience more frequent hot days by the end of the 21st century, driven by the increase in radiative forcing caused by the emission of greenhouse gases. These projections further suggest very likely increased incidence and/or magnitude of extreme high sea level.

1.1.2 Interaction between the spheres: The global carbon budget

Processes associated with the ocean and land carbon reservoirs and their exchange are known to substantially reduce the amount of CO₂ in the atmosphere. While carbon dioxide measurements are highly precise in the atmosphere, air–sea and especially air–land flux estimates do not yet have the same accuracy, due to temporal and spatial heterogeneity, and therefore provide large uncertainties in estimating the different pathways of emitted carbon (Le Quéré *et al.*, 2009).

In order to improve the projections of climate change in the future, it is essential to understand and quantify the amount of carbon removed from the atmosphere by the ocean and terrestrial systems. For example, the Global Carbon Project (<http://www.globalcarbonproject.org/>), established in 2001, aims to quantify the different pathways of CO₂. Emitted carbon remains in the atmosphere or is redistributed from the atmosphere to the land or ocean. Recent carbon budgets as presented in Le Quéré *et al.* (2013) and Sarmiento *et al.* (2010), consider these four major terms to close the budget following:

$$Em_{ff+luc} = F_{land} + F_{ocean} + \frac{dG_{atm}}{dt} \quad (1.4)$$

where the variable Em describes sources of CO₂ emission, dG_{atm}/dt is the atmospheric accumulation rate of CO₂ and F_{land} and F_{ocean} are land and ocean sinks, respectively. While Le Quéré *et al.* (2013) consider fossil fuel combustion (Em_{ff}) and land use change (Em_{luc}) for their emission term, Sarmiento *et al.* (2010) use fossil fuel emissions only. While fossil fuel emissions are fairly well known, e.g from energy statistics (Marland *et al.* (2005); http://cdiac.ornl.gov/trends/emis/meth_reg.html), and the atmospheric accumulation rate (dG_{atm}/dt) is very accurately measured in the atmosphere (Ballantyne *et al.* (2012); Ed Dlugokencky and Pieter Tans, NOAA/ESRL, www.esrl.noaa.gov/gmd/ccgg/trends/), the ocean flux term (F_{ocean}) in the budget to date is mainly based on model outputs

(Sarmiento *et al.*, 2010) or a combination of observations with trends and variabilities from ocean models (Le Quéré *et al.*, 2013). The last term in the budget, the global land flux (F_{land}), can be calculated as the residual between the emission and the other sink terms, hence the Sarmiento *et al.* (2010) sink term includes the net land sink (change in land use change are considered in this term, to avoid introducing additional uncertainties), whereas Le Quéré *et al.* (2013) report the residual land sink.

Budget estimates suggest that less than 50% of the emitted carbon remains in the atmosphere, e.g. Le Quéré *et al.* (2009) estimate 43% for the time period 1959 to 2008 and Le Quéré *et al.* (2013) estimate 46% for the time period 2002 to 2011. These budget estimates suggest that the remainder of the emitted carbon is being fairly equally spread between the land and ocean. However, all these budgets analyse various time periods and consist of different estimates, e.g. other models are used to estimate the net ocean flux.

1.2 Ocean–atmosphere interaction

1.2.1 Air–sea gas flux measurements and methods

There are several ways to quantify the air–sea gas flux of CO_2 , including direct flux measurements and air–sea flux parametrisations. One way to measure the air–sea flux of CO_2 is by using eddy correlation techniques (Jones and Smith, 1977). This technique correlates CO_2 fluctuations with turbulent vertical velocity fluctuations in the atmospheric surface layer (McGillis *et al.*, 2001), hence the flux measurements can be used to quantify the CO_2 flux over a wide range of wind speed and other environmental parameter (Jones and Smith, 1977; McGillis *et al.*, 2001). Eddy correlation measurements are highly precise and can be used to investigate fluctuations in the CO_2 flux within hour timescales (McGillis *et al.*, 2001). However, these measurements are point measurements and are therefore not suitable to quantify the global ocean carbon sink over longer time periods.

The air–sea gas flux has been further estimated using ocean interior data and an ocean inversion approach (see e.g. Gloor *et al.*, 2003). This approach is based on the definition of a tracer (C^*) which has no sources or sinks in the interior ocean and is represented by the sum of pre–industrial air–sea gas exchange of CO_2 and the anthropogenic invasion of CO_2 (Gruber *et al.*, 1996). The tracer is redistributed in the ocean by transport and

mixing, hence if these transport processes are inverted, gas fluxes can be implied from this method (Gloor *et al.*, 2003; Gruber *et al.*, 2009). The advantage of this method is, that it does not need a gas exchange formulation between atmosphere and ocean and it provides estimates for both the pre–industrial and the recent exchange of CO₂. The disadvantage of this method is that its quality relies on interior carbonate system quantities with large uncertainties attached. Furthermore, inter–annual to decadal variabilities of the upper ocean carbon cycle may be reflected differently in the inversion fluxes.

At last, the air–sea flux of carbon dioxide can be estimated via a simple bulk formulation, which describes the air–sea flux of CO₂ as a function of the partial pressure difference of $p\text{CO}_2$ ($\Delta p\text{CO}_2$) between the atmosphere and the ocean and a kinetic term, describing the wind influence on the gas exchange (see section 1.2.2). The advantage of this method is that the mole fraction of CO₂ can be measured at sea (see chapter 3 for more detail) and in the atmosphere with low uncertainties attached. The disadvantage of the $\Delta p\text{CO}_2$ method is that it strongly depends on the formulation of a wind-driven kinetic term. Furthermore, large areas of the global ocean remain poorly sampled, leading to the necessity of data interpolation methods in order to estimate the air–sea gas flux on a global scale from observations (e.g. Takahashi *et al.*, 2009; Sasse *et al.*, 2013; Rödenbeck *et al.*, 2013). Within this thesis, the $\Delta p\text{CO}_2$ method will be applied to estimate the air–sea gas flux of the Atlantic Ocean (chapter 4) and the global ocean (chapter 5), hence it will be discussed in more detail in the following section.

Other regional and global flux estimates stem from atmospheric inversions (see e.g. Gurney *et al.*, 2008), which use atmospheric transport models and CO₂ data measured in the atmosphere to assess the sources and sinks of the contemporary air–sea and air–land gas flux. While the faster atmospheric transport and mixing compared to the ocean circulation leads to coarse output resolution in space, atmospheric inversion estimates have a high temporal resolution in comparison to ocean inversions (Wanninkhof *et al.*, 2013b).

Ocean general circulation models (e.g. Le Quéré *et al.*, 2007; Doney *et al.*, 2009; Graven *et al.*, 2012) also use the bulk formulation approach to quantify the air–sea CO₂ gas exchange, however the sea surface $p\text{CO}_2$ is computed from the state variables of the carbonate system for each timestep, using the information from the previous timestep, hence the resulting air–sea flux is less dependent on the gas transfer formulation than in

observation–based estimates (Wanninkhof *et al.*, 2013b). Ocean general circulation models provide estimates of the anthropogenic flux, as their pre-industrial ocean is saturated. They are, however, not constrained by observations, hence the sea surface $p\text{CO}_2$ is a process based estimate and therefore relies on the quantification of all carbonate system processes.

1.2.2 Bulk formulation of the air–sea gas exchange

The expression for the gas flux can be derived by applying a simple model where the transport of mass through the interface is controlled by turbulence and molecular diffusion (Jähne *et al.*, 1987).

Air–sea gas exchange is driven by a concentration difference of a gas to its equilibrium concentration ($[\text{CO}_{2,eq}]$). According to Henry’s Law the equilibrium concentration in a liquid equals the solubility of a gas times the partial pressure of the gas above the liquid with which it is in thermodynamic equilibrium, hence:

$$[\text{CO}_{2,eq}] = K_0 \cdot p\text{CO}_2 \quad (1.5)$$

where K_0 describes the solubility of CO_2 . Following Weiss (1974) the solubility of CO_2 is derived from:

$$\begin{aligned} \ln(K_0) = & A_1 + A_2 \cdot (100/T) + A_3 \cdot \ln(T/100) \\ & + s \cdot [B_1 + B_2 \cdot (T/100) + B_3 \cdot (T/100)^2] \end{aligned} \quad (1.6)$$

where T denotes the absolute temperature of the ocean surface water, s the salinity and A_{1-3} and B_{1-3} are empirically derived constants and can be obtained from Weiss (1974). The air–sea flux is directly proportional to the partial pressure difference of CO_2 or the concentration difference between the oceanic concentration and the equilibrium concentration (Sarmiento and Gruber, 2006):

$$\Delta p\text{CO}_2 = p\text{CO}_{2,\text{atm}} - p\text{CO}_{2,\text{w}} \quad (1.7)$$

$$\Delta[\text{CO}_2] = [\text{CO}_{2,\text{atm}}] - [\text{CO}_{2,\text{w}}] \quad (1.8)$$

Hence the ocean is either saturated or it acts as a source or a sink of CO_2 . In a simple 2–layer model (Broecker and Peng, 1974; Sarmiento and Gruber, 2006), where ocean and atmosphere are expressed as 2 molecular layers, gas transport through each of these layers is driven by diffusion and can be expressed by Fick’s first law:

$$F_{\text{CO}_2} = -\epsilon \cdot \frac{\partial[\text{CO}_2]}{\partial z} \quad (1.9)$$

In this simple model only the vertical component z is taken into account and the flux through the layers can further be represented by finite differences (Sarmiento and Gruber, 2006):

$$F_{\text{CO}_2} = -\epsilon \cdot \frac{\Delta[\text{CO}_2]}{\Delta z} \quad (1.10)$$

The diffusion coefficient ϵ can be expressed as a function of the film layer thickness Δz and a kinetic term called the gas transfer velocity k :

$$\epsilon = k \cdot \Delta z \quad (1.11)$$

Liss and Slater (1974) proposed that the gas transfer velocity of CO_2 is orders of magnitude larger in the atmosphere, hence for the transport of mass through the interface, only the water sided velocity (k_w) need to be accounted for. By further assuming that there are no sink and source terms at the interface, the flux of CO_2 can be expressed via:

$$F_{\text{CO}_2} = -k_w \cdot ([\text{CO}_{2,\text{atm}}] - [\text{CO}_{2,\text{w}}]) \quad (1.12)$$

or in terms of partial pressure:

$$F_{\text{CO}_2} = -k_w \cdot K_0 \cdot (p\text{CO}_{2,\text{atm}} - p\text{CO}_{2,\text{w}}) \quad (1.13)$$

Equation 1.13 is the most commonly used bulk–formulations for the calculation of the air–sea CO₂ flux. While the partial pressure difference term in equation 1.13 describes the direction and the driver of the flux, the air–sea gas flux magnitude strongly depends on the kinetic gas transfer velocity k_w , which varies as a function of the wind speed and the sea surface temperature.

1.2.3 Gas exchange rate and transfer velocity

Gas exchange is influenced by breaking waves, bubbles, temperature and humidity gradients, which are linked to wind speed (Woolf, 1997), hence the formulation of the gas transfer velocity strongly depends on the wind speed, which links these dependencies to the air–sea gas flux (e.g. Nightingale *et al.*, 2000).

Although there was an improvement of field measurement techniques and laboratory studies, like wind tunnel studies, in the last decades, there is still an uncertainty concerning the transfer velocity. Jähne *et al.* (1987) first proposed the following relationship for the transfer velocity:

$$k_w = \kappa^{-1} \cdot \text{Sc}^{-n} \cdot u^* \quad (1.14)$$

The first term κ describes a dimensionless constant. It is the so-called transfer resistance factor (Jähne *et al.*, 1987). Of particular importance for the transfer velocity are the second term, the Schmidt number (Sc) and its exponent (n), as well as the third term, the friction velocity (u^*).

The Schmidt number is given following Jähne *et al.* (1987) as:

$$Sc = \frac{\nu}{\epsilon} \quad (1.15)$$

where ν is the kinematic viscosity of water and ϵ is the molecular diffusivity of the gas. The kinematic viscosity influences the thickness of the layer, where the gas exchange takes place. The Schmidt number depends on the salinity and the temperature of the seawater. For temperatures between 0 and 30°C, freshwater (salinity of 0) and seawater (salinity of 35), the Schmidt number can be calculated using the empirical relationship and empirically derived coefficients for CO₂ from Wanninkhof (1992) and Keeling *et al.* (1998):

$$Sc = A - B \cdot T + C \cdot T^2 - D \cdot T^3 \quad (1.16)$$

For the Schmidt number exponent n (see equation 1.14), different values are taken depending on the model, which is used to determine the gas transfer velocity. Experimentally, the number of the exponent is determined from multi tracer measurements (Jähne *et al.*, 1987).

During the last decades more and more gas transfer velocity models were developed using the latest measurement techniques and wind tunnel studies. All models commonly use the wind speed at a height of 10 meters above sea level rather than the friction velocity, as the former can be directly measured. The wind speed dependency in transfer velocity models ranges from linear (e.g. Liss and Merlivat, 1986) over quadratic (e.g. Wanninkhof, 1992) to cubic (e.g. Wanninkhof and McGillis, 1999), and combinations of linear and quadratic (e.g. Nightingale *et al.*, 2000).

1.3 Sea surface partial pressure of CO₂

The previous section illustrated how the flux of CO₂ between the atmosphere and the ocean depends on the gas transfer velocity and the $p\text{CO}_2$ difference ($\Delta p\text{CO}_2$). While atmospheric mixing ratios are fairly homogeneous in space, the sea surface $p\text{CO}_2$ shows more regional variability (Takahashi *et al.*, 2002), hence the $\Delta p\text{CO}_2$ is largely driven by

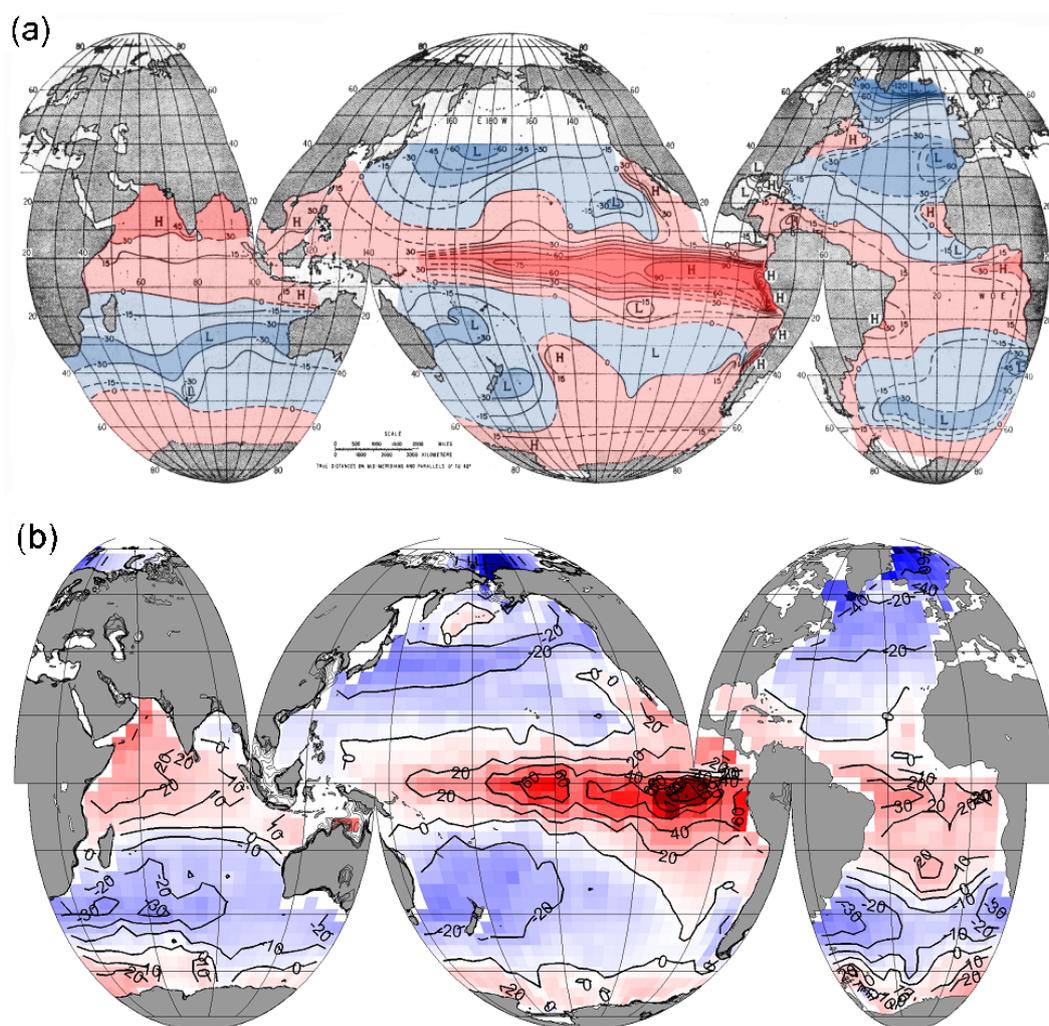


Figure 1.3: Climatological mean $\Delta p\text{CO}_2$ (ocean minus atmosphere in μatm) distribution from (a) the study of Keeling (1968) (colors added by N. Gruber) and (b) the climatology of Takahashi *et al.* (2009). Positive or red areas indicate supersaturation of CO₂ in the ocean, negative or blue areas indicate undersaturation.

the sea surface $p\text{CO}_2$.

Over the past decades, several studies have been conducted to investigate the distribution of the sea surface $p\text{CO}_2$ and the $\Delta p\text{CO}_2$ using available shipboard observations. Figure 1.3 illustrates one of the oldest ((a) Keeling, 1968) and a recent ((b) Takahashi *et al.*, 2009) global $\Delta p\text{CO}_2$ estimate. While results of Keeling (1968) are derived from data collected on 14 vessels between 1900 and 1967, the dataset used by Takahashi *et al.* (2009) includes ~ 3 million data points collected on vessels between 1970 through 2007 (Takahashi *et al.*, 2008) and has been normalized to the reference year 2000.

Although ~ 40 years apart, there is remarkable agreement between these two estimates. Figure 1.3 (a) and (b) illustrate the general pattern of high sea surface partial

pressures and supersaturation of CO₂ in the tropical oceans, lower partial pressure and undersaturation in the temperate mid latitude oceans, supersaturation in the high latitude Southern Ocean and the lowest partial pressures with strong undersaturation in the high latitude North Atlantic.

The sea surface partial pressure of CO₂ is strongly influenced by physical, chemical and biological processes, explaining the pattern illustrated in figure 1.3. In their work, Takahashi *et al.* (1993) propose that changes in the surface $p\text{CO}_2$ can mainly be described as a function of the sea surface temperature (T), the dissolved inorganic carbon (DIC), the total alkalinity (TALK) and the salinity (s) of the seawater. Hence the changes in the $p\text{CO}_2$ can be derived from:

$$d p\text{CO}_2 = \frac{\partial p\text{CO}_2}{\partial T} \cdot dT + \frac{\partial p\text{CO}_2}{\partial \text{DIC}} \cdot d\text{DIC} + \frac{\partial p\text{CO}_2}{\partial \text{TALK}} \cdot d\text{TALK} + \frac{\partial p\text{CO}_2}{\partial s} \cdot ds \quad (1.17)$$

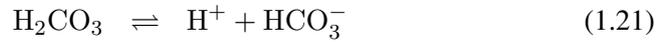
The relationships between the non-thermal drivers of equation 1.17 and the sea surface $p\text{CO}_2$ is yet not fully quantified. Takahashi *et al.* (1993) found experimentally, when DIC, TALK and salinity were kept constant, i.e. by heating a closed parcel of water, the relationship between $p\text{CO}_2$ and temperature is accurately described by:

$$\frac{\partial \ln p\text{CO}_2}{\partial T} = 4.23\% \cdot ^\circ \text{C}^{-1} \quad (1.18)$$

Hence the natural logarithm of the sea surface $p\text{CO}_2$ increases by about 4% when temperature increases by 1°C. This temperature relationship is largely related to the solubility effect quantified in equation 1.6. Takahashi *et al.* (1993) note that for surface waters, the effect of temperature and DIC changes have the largest effect on the sea surface $p\text{CO}_2$. Processes effecting the sea surface $p\text{CO}_2$ distribution are further investigated in the following subsections.

1.3.1 Inorganic carbon cycle

Gaseous CO₂ (CO_{2(gas)}) that dissolves in seawater gets hydrated to form aqueous CO₂ (CO_{2(w)}). CO_{2(w)} further reacts with water to form carbonic acid (H₂CO₃), which dissociates in two steps to form bicarbonate (HCO₃⁻) and carbonate ions (CO₃²⁻). These reactions are summarized following Dickson *et al.* (2007):



The carbonic acid concentration and the concentration of CO₂ in seawater can be combined to:

$$[\text{H}_2\text{CO}_3^*] = [\text{CO}_{2(\text{w})}] + [\text{H}_2\text{CO}_3] \quad (1.23)$$

$$(1.24)$$

And the equilibrium relationships of these reactions are given by:

$$K_0 = \frac{[\text{H}_2\text{CO}_3^*]}{p\text{CO}_2} \quad (1.25)$$

$$K_1 = \frac{[\text{HCO}_3^-][\text{H}^+]}{[\text{H}_2\text{CO}_3^*]} \quad (1.26)$$

$$K_2 = \frac{[\text{CO}_3^{2-}][\text{H}^+]}{[\text{HCO}_3^-]} \quad (1.27)$$

The sum of the products formed by these reactions is the total dissolved inorganic carbon concentration:

$$[\text{DIC}] = [\text{H}_2\text{CO}_3^*] + [\text{HCO}_3^-] + [\text{CO}_3^{2-}] \quad (1.28)$$

In order to determine the CO₂ system in the ocean, TALK has to be included. TALK of seawater is a measure of the proton deficit of the solution relative to a zero level and is defined by Dickson (1981) as:

$$\begin{aligned} [\text{TALK}] = & [\text{HCO}_3^-] + 2 \cdot [\text{CO}_3^{2-}] + [\text{B}(\text{OH})_4^-] + [\text{OH}^-] - [\text{H}^+] \quad (1.29) \\ & + [\text{HPO}_4^{2-}] + 2 \cdot [\text{PO}_4^{3-}] + [\text{SiO}(\text{OH})_3^-] + [\text{NH}_3] \\ & + [\text{HS}^-] - [\text{HSO}_4^-] - [\text{HF}] - [\text{H}_3\text{PO}_4] \end{aligned}$$

There are however different definitions in the literature, e.g. Peng *et al.* (1987) and Takahashi *et al.* (1982).

The sea surface $p\text{CO}_2$ can be determined by reactions 1.20 to 1.22 as a function of the reaction products and their equilibrium constants, as well as approximations of the DIC and TALK concentrations (see e.g. Sarmiento and Gruber, 2006):

$$p\text{CO}_{2(w)} = \frac{K_2}{K_0 \cdot K_1} \cdot \frac{[\text{HCO}_3^-]^2}{[\text{CO}_3^{2-}]} \approx \frac{K_2}{K_0 \cdot K_1} \cdot \frac{(2 \cdot [\text{DIC}] - [\text{TALK}])^2}{[\text{TALK}] - [\text{DIC}]} \quad (1.30)$$

where:

$$[\text{HCO}_3^-] \approx 2 \cdot [\text{DIC}] - [\text{TALK}] \quad (1.31)$$

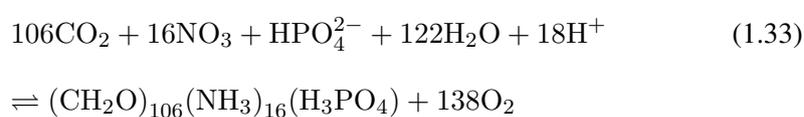
$$[\text{CO}_3^{2-}] \approx [\text{TALK}] - [\text{DIC}] \quad (1.32)$$

Hence, the surface $p\text{CO}_2$ is affected by the ratio of the equilibrium constants, the DIC and the TALK concentration of the seawater. While the solubility of seawater and the other equilibrium constants are determined by the sea surface temperature and salinity, the DIC concentration is influenced by the exchange of CO₂ with the atmosphere and both DIC and TALK are influenced by biological processes and mixing.

Equation 1.18 already illustrated the temperature dependence of the sea surface $p\text{CO}_2$ when salinity, DIC and TALK are kept constant. Without biological activity, mixing and

the exchange of gases with the atmosphere, the surface ocean $p\text{CO}_2$ would strictly follow this temperature dependence, leading to high partial pressures at the Equator and low partial pressures towards the poles, following the example of Broecker and Peng (1982). In contrast, very rapid gas exchange would result in the sea surface $p\text{CO}_2$ being in equilibrium with the atmospheric CO₂ whereas the DIC concentration would be at minimum at the Equator increasing towards the poles due to its temperature dependence. However, figure 1.3 reveals that the sea surface $p\text{CO}_2$ distribution is more complex than explained by temperature alone and that gas exchange is not sufficiently fast for the surface ocean $p\text{CO}_2$ being in equilibrium with the atmospheric $p\text{CO}_2$.

Photosynthesis and remineralisation are the most important biological processes considering the distribution of sea surface carbon dioxide. This process removes CO₂ and nutrients from the surface water to form organic matter and to produce oxygen. Redfield *et al.* (1963) proposed a constant stoichiometric ratio between carbon dioxide, nutrients and oxygen linked to photosynthetic production:



Biological production decreases the sea surface $p\text{CO}_2$, the DIC concentration and the concentration of free protons, increasing the TALK concentration. The global net primary production by phytoplankton is estimated to remove 45–50 Pg C · yr⁻¹ from the inorganic sea surface carbon pool (Longhurst *et al.*, 1995) in the form of organic carbon. The majority of organic matter produced at the ocean euphotic zone is recycled in the ocean surface and the remainder is further exported into the deeper layers of the ocean (export production) by sinking particles (Sarmiento and Gruber, 2006). The contrary process, the remineralisation of organic matter in the deeper ocean layers, in combination with deep wintertime mixing and upwelling leads to the re-introduction of DIC, increasing surface $p\text{CO}_2$ concentrations (Sarmiento and Gruber, 2006). Biological production is limited by the available amount of light and nutrients and therefore shows strong regional and seasonal variability.

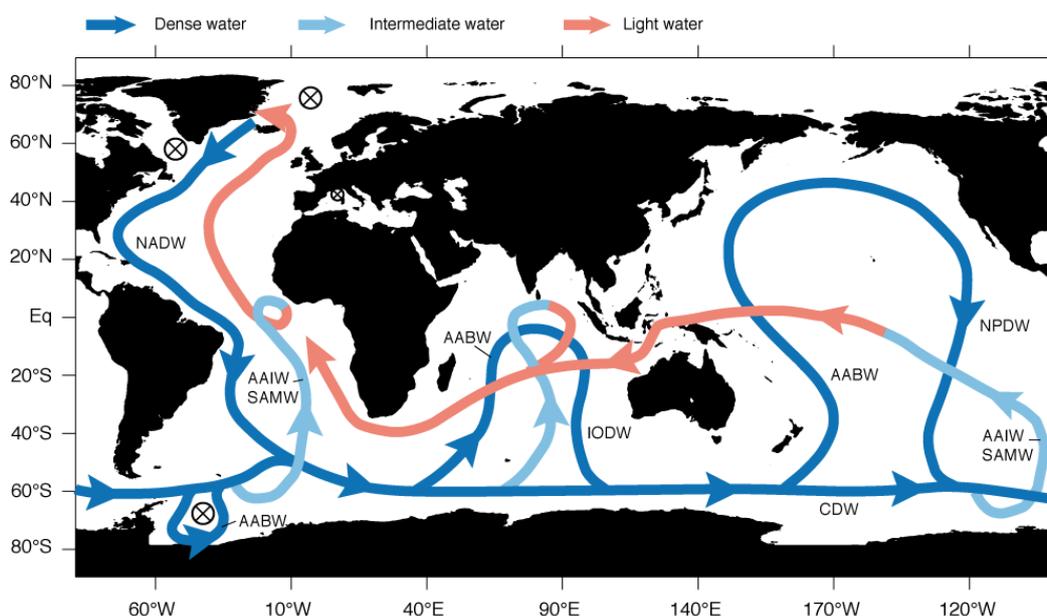


Figure 1.4: Schematic view of the ocean conveyor circulation and the deep water formation zones (marked with circles) from Gnanadesikan and Hallber (2002) as illustrated in Sarmiento and Gruber (2006). Dark blue lines and arrows illustrate the pathway of dense or deep water, light blue lines and arrows illustrate the pathway of intermediate and mode waters and red illustrates the pathway of light waters. Abbreviations, i.e. NADW (North Atlantic Deep Water), AAIW (Antarctic Intermediate Water) SAMW (Subantarctic Mode Water), AABW (Antarctic Bottom Water), CDW (Circumpolar Deep Water), IODW (Indian Ocean Deep Water) and NPDW (North Pacific Deep Water), refer to the relevant water masses.

1.3.2 Ocean circulation and $p\text{CO}_2$ distribution

Figure 1.3 illustrates the non-uniform distribution of the $\Delta p\text{CO}_2$. The dependency of the sea surface $p\text{CO}_2$ on temperature, gas exchange and biology have been discussed in the previous section. The ocean however is not static, hence there are several features in figure 1.3 that are influenced by the large scale thermohaline ocean circulation and the smaller scale wind driven circulation of the ocean surface waters.

The salinity-driven ocean conveyor circulation with its deep water formation zones affects the distribution of sea surface $p\text{CO}_2$. Figure 1.4 from Gnanadesikan and Hallber (2002) as illustrated in Sarmiento and Gruber (2006) shows the main circulation pathways and the main heat exchange areas within the global ocean. Deep water forms in the high latitude Atlantic Ocean and the Southern Ocean as North Atlantic Deep Water (NADW) and Antarctic Bottom Water (AABW), respectively. The conveyor circulation was first proposed by Robinson and Stommel (1959) and has been updated over the years (Gnanadesikan and Hallber, 2002; Lozier, 2010).

Deep water formation provides a pathway for carbon from the ocean surface into the

interior ocean. In particular the formation of NADW leads to a strong interior transport of surface ocean carbon, while the sea surface $p\text{CO}_2$ is low due to the cooling of the northwards movement of the water mass (see figure 1.3). Large quantities of anthropogenic carbon, subducted into the deep ocean, are stored in the high latitude North Atlantic (Sabine *et al.*, 2004; Khatiwala *et al.*, 2013).

Besides the thermohaline global ocean conveyor circulation, Ekman transport, driven by the wind stress and the Coriolis force, is one of the major driving factors affecting the motion of the surface waters and the resulting $p\text{CO}_2$ distribution (Sarmiento and Gruber, 2006).

Ekman transport replaces surface water masses with water masses from deeper layers of the ocean. This results in coastal upwelling and warming of carbon rich water masses. In the tropical zones of both hemispheres the dominant wind patterns are directed eastwards (Easterlies), resulting in a surface water Ekman transport northwards in the northern hemisphere and southwards in the southern hemisphere. This leads to a water mass upwelling at the Equator, which brings water masses, rich in remineralized carbon, back into the surface layer. The surface transport of carbon enriched water mass in combination with warming results in high $p\text{CO}_2$ waters along the Equator, especially seen in the eastern Equatorial Pacific (see figure 1.3).

The sea surface $p\text{CO}_2$ distribution is further influenced by the surface ocean current system, illustrated in figure 1.5. Warm water masses (originated from the low latitudes) transport carbon rich waters towards the higher latitudes where water masses start to cool, increasing the CO₂ solubility. In contrast, cool water masses originating from higher latitudes move towards the low latitudes where they get warmed, thereby reducing the gas solubility, leading to an increase in the sea surface $p\text{CO}_2$.

1.3.3 Sea surface $p\text{CO}_2$ measurements

Since the earliest regional estimates of the $\Delta p\text{CO}_2$ from Wattenberg (1933) in the central and South Atlantic based on the Meteor expedition from 1925 to 1927, there has been a strong increase in ocean carbon data over the last century. Projects like the Geochemical Ocean Sections program (GEOSECS) from 1972 to 1978 and the World Ocean Circulation Experiment (WOCE) in the 1990s helped to better understand the ocean carbon

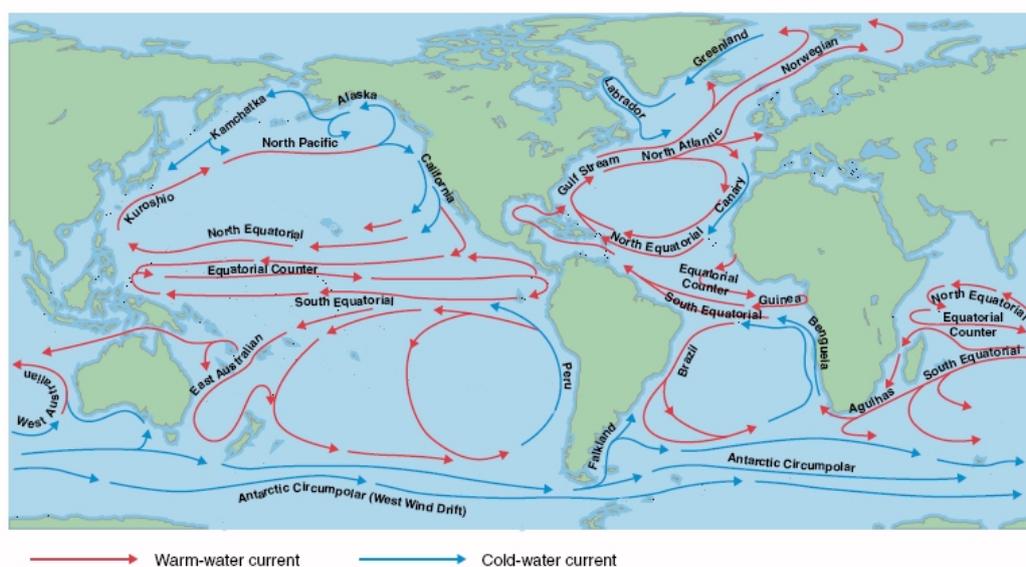


Figure 1.5: Schematic of the ocean surface warm and cold water currents as illustrated at <http://oceanmotion.org/html/impact/conveyor.htm>. It illustrates the polewards transport of warm waters and the equatorwards transport of cold waters from higher latitudes, as well as the gyre circulation in both the North and South Atlantic and Pacific.

system (e.g. Broecker and Peng, 1982) and to develop new quality standards (e.g. Dickson *et al.*, 2007) to reduce the uncertainty of the data. Ship-time is expensive, therefore, within the last decades, there has been an increase in $p\text{CO}_2$ measurements on-board Voluntary Observing Ships (VOS) (e.g. Cooper *et al.*, 1998; Schuster and Watson, 2007), which are commercial vessels, participating in the program to obtain a large number of observations on a repeated ship route.

In order to investigate the temporal variability of the sea surface $p\text{CO}_2$ and other related carbon system parameters in the ocean, several timeseries stations and moorings were launched (see e.g. <http://cdiac.ornl.gov/oceans/Moorings/>). Three of the oldest stations include, the Hawaiian Ocean Timeseries (HOT) station (Dore *et al.*, 2009), the Bermuda Atlantic Timeseries Station (BATS) Hydrostation "S" (Gruber *et al.*, 2002; Bates, 2007) and the European Station for Timeseries in the Ocean (ESTOC) (González-Dávila *et al.*, 2007), which were launched in 1988, 1954 and 1995, respectively.

This increasing community measurement effort is reflected in the Lamont–Doherty Earth Observatory (LDEO) database, which was the first established database gathering available sea surface $p\text{CO}_2$ measurements on a global scale. While in the mid-1990s, Takahashi *et al.* (1997) had a total of 250,000 data available for their work, this number continued to increase to 940,000 in the early 2000's (Takahashi *et al.*, 2002) up to roughly

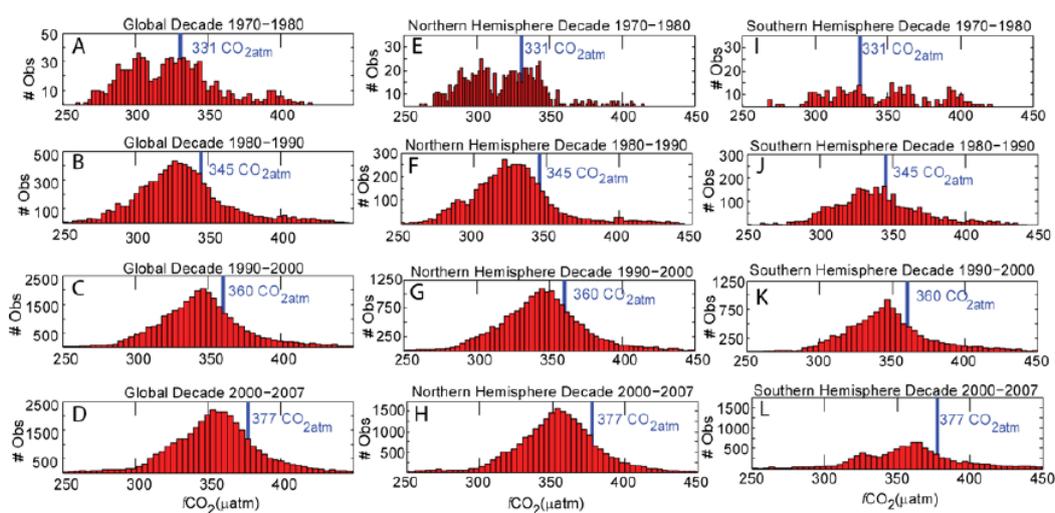


Figure 1.6: Number of $f\text{CO}_2$ observations and their distribution per decade within SOCAT v1.5 data base as illustrated in Sabine *et al.* (2013), for (A)–(D) the global ocean, (E)–(H) the northern hemisphere and (I)–(L) the southern hemisphere. Blue lines indicate the decadal mean atmospheric CO₂ molar fraction in ppm.

3,000,000 towards the end of the last decade (Takahashi *et al.*, 2008, 2009). The current version includes a total of roughly 6.7 million measurements (Takahashi *et al.*, 2013).

More recently, the community effort led to the creation of the up-to-date largest unified quality-controlled database: the Surface Ocean Carbon Atlas (SOCAT) (Pfeil *et al.*, 2013; Bakker *et al.*, 2014). Initiated in 2007 and first published in 2011, the aim of the database was to provide a publicly available and regularly updated set of surface $f\text{CO}_2$ observations (Pfeil *et al.*, 2013), which initially included 6.3 million data points within the period 1970 through 2007. Updated in 2013 (Bakker *et al.*, 2014), the current version includes over 10 million data points up to the year 2011.

Figure 1.6 shows the amount of data included in the SOCAT v2 databased as illustrated in Sabine *et al.* (2013) for each decade from 1970 through 2007. It clearly indicates the increase in surface CO₂ observations in time, but further illustrates the unequal distribution of measurements in space. Within each decade, the majority of measurements were obtained in the northern hemisphere.

1.4 Air–sea CO₂ flux estimates

1.4.1 Long–term mean air–sea CO₂ flux estimates

The previous sections emphasised the role of sea surface $p\text{CO}_2$ measurements and the importance of air–sea gas exchange, regarding the concentration of greenhouse gases in the atmosphere. A recent estimate by Wanninkhof *et al.* (2013b) suggests a mean anthropogenic CO₂ uptake from 1990 to 2010 by the global ocean of $-2.0 \text{ Pg C} \cdot \text{yr}^{-1}$ for the last two decades. This estimate is based on the compilation of different approaches within a recent Regional Carbon Cycle Assessment and Processes (RECCAP) synthesis, including estimates derived (i) using an empirical update of the Takahashi *et al.* (2009) climatology (Park *et al.*, 2010), (ii) from ocean general circulation models that include a full representation of the oceanic carbon cycle (e.g. Le Quéré *et al.*, 2007; Doney *et al.*, 2009; Graven *et al.*, 2012), (iii) from inversions of ocean interior carbon measurements (e.g. Gruber *et al.*, 2009) and (iv) from inversions of atmospheric CO₂ (e.g. Gurney *et al.*, 2008). These methodologically very different methods show a good agreement regarding the 20-year mean flux ranging from $-1.9 \pm 0.3 \text{ Pg C} \cdot \text{yr}^{-1}$ to $-2.4 \pm 0.3 \text{ Pg C} \cdot \text{yr}^{-1}$ (Wanninkhof *et al.*, 2013b).

Recently, more observation–based estimates of the contemporary mean air–sea CO₂ flux were derived using the increasing amount of available observations. Takahashi *et al.* (2009) estimates a global net flux, referenced to the year 2000, of $-1.6 \pm 0.9 \text{ Pg C} \cdot \text{yr}^{-1}$, using available surface ocean $p\text{CO}_2$ measurements from the LDEO database (Takahashi *et al.*, 2008) and an advection–based interpolation method. A climatological mean estimate from Sasse *et al.* (2013), using carbon bottle data within the surface mixed layer in combination with a combined clustering–regression technique estimates a net flux of $-1.55 \pm 0.35 \text{ Pg C} \cdot \text{yr}^{-1}$. Both observation–based examples are roughly $0.4\text{--}0.5 \text{ Pg C} \cdot \text{yr}^{-1}$ lower than suggested by Wanninkhof *et al.* (2013b).

The difference between the observation–based estimates and the estimates included in Wanninkhof *et al.* (2013b) can be linked to riverine–derived carbon. The contemporary air–sea flux is a combination of a natural and an anthropogenic driven flux component. The natural air–sea flux of CO₂ is estimated to be balanced on global scale (Gruber *et al.*, 2009), i.e. the pre–industrial ocean was on average neither a sink, nor a source of atmospheric CO₂. However, when taking into account riverine–derived carbon (Jacobson *et al.*,

2007), the pre–industrial ocean describes a source of CO₂ to the atmosphere of 0.45 Pg C · yr⁻¹.

A special focus within this thesis is on the Atlantic Ocean. Over the last two decades, the Atlantic Ocean (44°S to 79°N and west of 19°E) has taken up about -0.49 ± 0.11 Pg C · yr⁻¹ on average (Schuster *et al.*, 2013), with about half of it being driven by the uptake of anthropogenic CO₂, while the other half represents an uptake flux of natural CO₂ (Gruber *et al.*, 2009; Schuster *et al.*, 2013). This makes the Atlantic Ocean one of the most important sinks per area for atmospheric CO₂, and especially for anthropogenic CO₂ (Sabine *et al.*, 2004; Mikaloff Fletcher *et al.*, 2006, 2007; Gruber *et al.*, 2009).

This sink estimate was taken from the Atlantic Ocean RECCAP study by Schuster *et al.* (2013) where the authors reviewed different methodologies to estimate the air–sea CO₂ fluxes and provided a "best" estimate. Furthermore, RECCAP–based estimates exist for different ocean basins (e.g. Lenton *et al.*, 2013), the global ocean (Wanninkhof *et al.*, 2013b) as well as for the air–land flux (e.g. Haverd *et al.*, 2013). While the Atlantic Ocean RECCAP study includes (I) the *p*CO₂ climatology of Takahashi *et al.* (2009) and the Tier 1 methodologies described in Schuster *et al.* (2013) which include (II) Ocean Inversions (Gruber *et al.*, 2009), (III) Atmospheric Inversions (e.g. Peylin *et al.*, 2013), (IV) Ocean Biogeochemical Models (e.g. Le Quéré *et al.*, 2007; Doney *et al.*, 2009; Graven *et al.*, 2012), as well as observation–based results including (V) a SOCAT v1.5 based multi parameter regression (Schuster *et al.*, 2013) and (VI) an estimate based on the *p*CO₂ database of Takahashi *et al.* (2009) updated by McKinley *et al.* (2011), the best estimate is derived from two observation–based sources, namely (I) using ocean surface partial pressure of CO₂ (*p*CO₂) measurements (Takahashi *et al.*, 2009) and (II) from inversions of ocean interior carbon measurements (e.g. Gruber *et al.*, 2009). These estimates, however, differ in the way that model and inversion-based estimates report the anthropogenic carbon flux, i.e. include riverine-derived carbon, whereas observation–based estimates report the contemporary flux only.

1.4.2 Variability and trends of the Atlantic Ocean sea surface pCO₂ and air–sea flux

The Atlantic Ocean sink varies substantially by season, which is in part driven by the seasonal variations in surface ocean pCO₂ linked to temperature and biological activity, but also related to circulation (Takahashi *et al.*, 1993, 2002; Schuster *et al.*, 2013). The sink strength is further affected by seasonal variations in surface ocean winds and atmospheric CO₂. Surface ocean pCO₂ varies over a wide range above and below the atmospheric pCO₂, with much of the seasonal amplitude dominated by temperature in the subtropical regions in both hemispheres (Gruber *et al.*, 2002; Takahashi *et al.*, 2002; Sarmiento and Gruber, 2006), explaining the summer maximum in surface ocean pCO₂. In contrast, biological processes acting in synergy with ocean mixing and circulation dominate the seasonal pCO₂ cycle in Equatorial and high latitude regions (poleward of 40°N and 40°S) (Takahashi *et al.*, 1993; Bennington *et al.*, 2009), explaining the summer minimum in oceanic pCO₂. Due to their opposite phasing, thermal and non-thermal drivers cancel each other out along the regime boundaries at around 40°S and 40°N (Takahashi *et al.*, 2002), leading to a minimum in the seasonal amplitude at those latitudes. This is illustrated in figure 1.7 for the global ocean. In their study, Schuster *et al.* (2013) identified a broad agreement among independent seasonal flux estimates in the temperature-driven subtropics, but not elsewhere.

Long term trends and inter–annual variability of the Atlantic carbon sink represent a source of substantial disagreement between different methodologies and studies (Schuster *et al.*, 2013). Using surface ocean pCO₂ observations, Schuster and Watson (2007) argued for decrease in the North Atlantic carbon sink of $\sim 0.24 \text{ Pg C} \cdot \text{yr}^{-1}$ and a reduction in the seasonal pCO₂ amplitude in both the subtropical and temperate North Atlantic from the mid–1990s to the mid–2000s. Further support for a decreasing North Atlantic sink comes from Lefèvre *et al.* (2004), Lüger *et al.* (2006), Olsen *et al.* (2006), and Schuster *et al.* (2009b), although each study analysed different regions and periods and also used different methods to determine trends. Schuster *et al.* (2009b), for example, analysed data from 1990 until 2006 in the eastern subpolar gyre and throughout most of the central North Atlantic, while Olsen *et al.* (2006) focused on the Nordic Seas, but looked at a more extended period, i.e., from 1981 until 2002/2003.

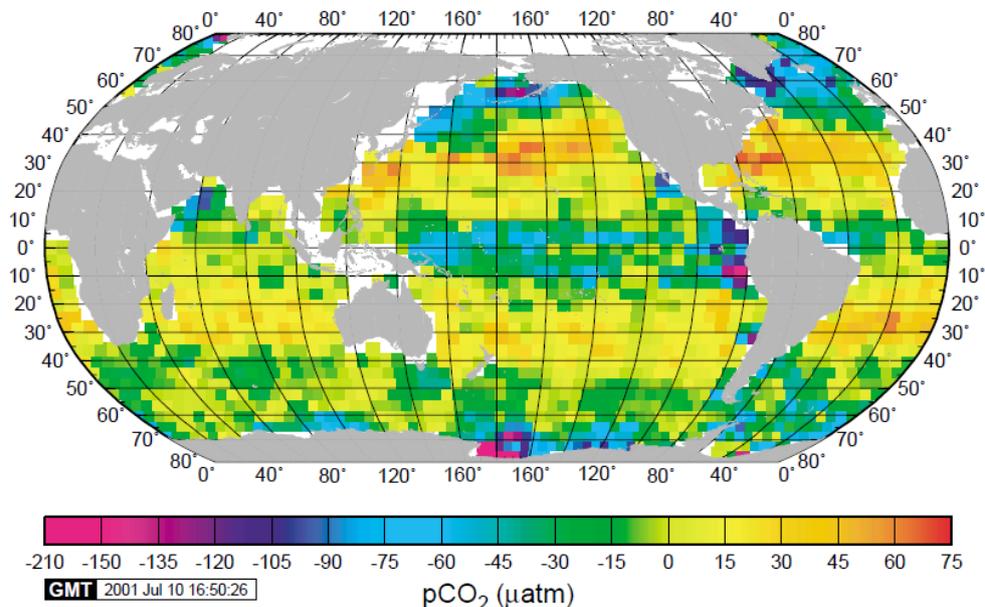


Figure 1.7: Thermal driven minus non-thermal driven seasonal amplitude as illustrated in Takahashi *et al.* (2002). Positive values indicate where the thermal effects exceed the non-thermal effects and negative values indicate where non-thermal effects exceed the thermal effects.

Results from the RECCAP study (Schuster *et al.*, 2013) show only limited inter-annual variability within the Atlantic and Arctic Ocean from 1990 to 2010. The standard deviations of the inter-annual variability (deseasonalized and detrended estimates from various methods included in RECCAP) are ranging from $\pm 0.015 \text{ Pg C} \cdot \text{yr}^{-1}$ derived from a multi parameter regression method to $\pm 0.055 \text{ Pg C} \cdot \text{yr}^{-1}$ derived from the median of the ocean inversion studies.

Trends from 1995–2009 reported in the Atlantic Ocean RECCAP study of Schuster *et al.* (2013) are derived from the available ocean biogeochemical models, atmospheric inversion studies and the estimate based on the $p\text{CO}_2$ database of Takahashi *et al.* (2009) updated by McKinley *et al.* (2011). These trends were derived from the slope of the linear fit of the deseasonalized flux data. Atmospheric inversions suggest a small but positive flux trend, i.e. a trend towards a decreasing carbon sink, for the Atlantic Ocean including the Arctic Ocean of $17 \pm 26 \text{ Tg C} \cdot \text{yr}^{-1} \cdot \text{decade}^{-1}$, while model and database trends indicate a negative flux trend of $-34 \pm 14 \text{ Tg C} \cdot \text{yr}^{-1} \cdot \text{decade}^{-1}$ and $-290 \pm 7.4 \text{ Tg C} \cdot \text{yr}^{-1} \cdot \text{decade}^{-1}$, respectively for the same region.

The initial year and period of data analysed for trends are crucial aspects to consider when answering different questions (Gruber, 2009). McKinley *et al.* (2011) pointed out

that when the surface ocean trends in $p\text{CO}_2$ are analysed over more than 25 years, all regions in the North Atlantic exhibit trends that are not statistically different from the trend in atmospheric CO₂, implying no change in the sink/source strength. However, when the periods of analyses were shortened to 10 years and the beginning and end years shifted, substantial trends emerged, largely reflecting inter–annual to decadal timescale variability.

1.4.3 Variability and trends of the global sea surface $p\text{CO}_2$ and air–sea flux

Wanninkhof *et al.* (2013b) investigates the inter–annual variability of the global ocean carbon flux from 1990 to 2010, based on ocean biogeochemical models, atmospheric inversions and an empirical update of the Takahashi *et al.* (2009) climatology (Park *et al.*, 2010). These authors found a “best” estimate of the IAV of $\pm 0.2 \text{ Pg C} \cdot \text{yr}^{-1}$ (1σ), which ranges from $\pm 0.16 \text{ Pg C} \cdot \text{yr}^{-1}$ (1σ) for the median of the available ocean biogeochemical models to $\pm 0.40 \text{ Pg C} \cdot \text{yr}^{-1}$ (1σ) for the atmospheric inversion model median. Wanninkhof *et al.* (2013b) further note that inter–annual variabilities from individual models are larger ($\sim \pm 0.25 \text{ Pg C} \cdot \text{yr}^{-1}$), indicating that the inter–annual variability of the different models is not coherent in time.

Trends for the period 1990 to 2010 as illustrated in Wanninkhof *et al.* (2013b) show strong differences between the different methodologies. The smallest trend signal of the air–sea CO₂ flux of $-0.13 \text{ Pg C} \cdot \text{yr}^{-1} \cdot \text{decade}^{-1}$ stems from the median of the available atmospheric inversion models. In contrast the median of the ocean inversion fluxes shows a substantially larger trend of $-0.5 \text{ Pg C} \cdot \text{yr}^{-1} \cdot \text{decade}^{-1}$ (Wanninkhof *et al.*, 2013b)

While natural climate–driven variabilities have their largest effect on the natural CO₂ fluxes, trends in the contemporary fluxes are mainly driven by the anthropogenic increase of CO₂ in the atmosphere. Fay and McKinley (2013) analysed 30 year global trends over large scale ocean biomes and concluded that for most regions, trends in the sea surface $p\text{CO}_2$ are parallel or slightly smaller than atmospheric trends implying hardly any change in the ocean carbon sink/source strength over such a large timescale. However, on shorter timescales or when changing the start and end year, trends emerge in response to climate mode signals, similar to the study of McKinley *et al.* (2011). The effect on trends compared for different timeperiods is shown in figure 1.8 following the illustration in Fay

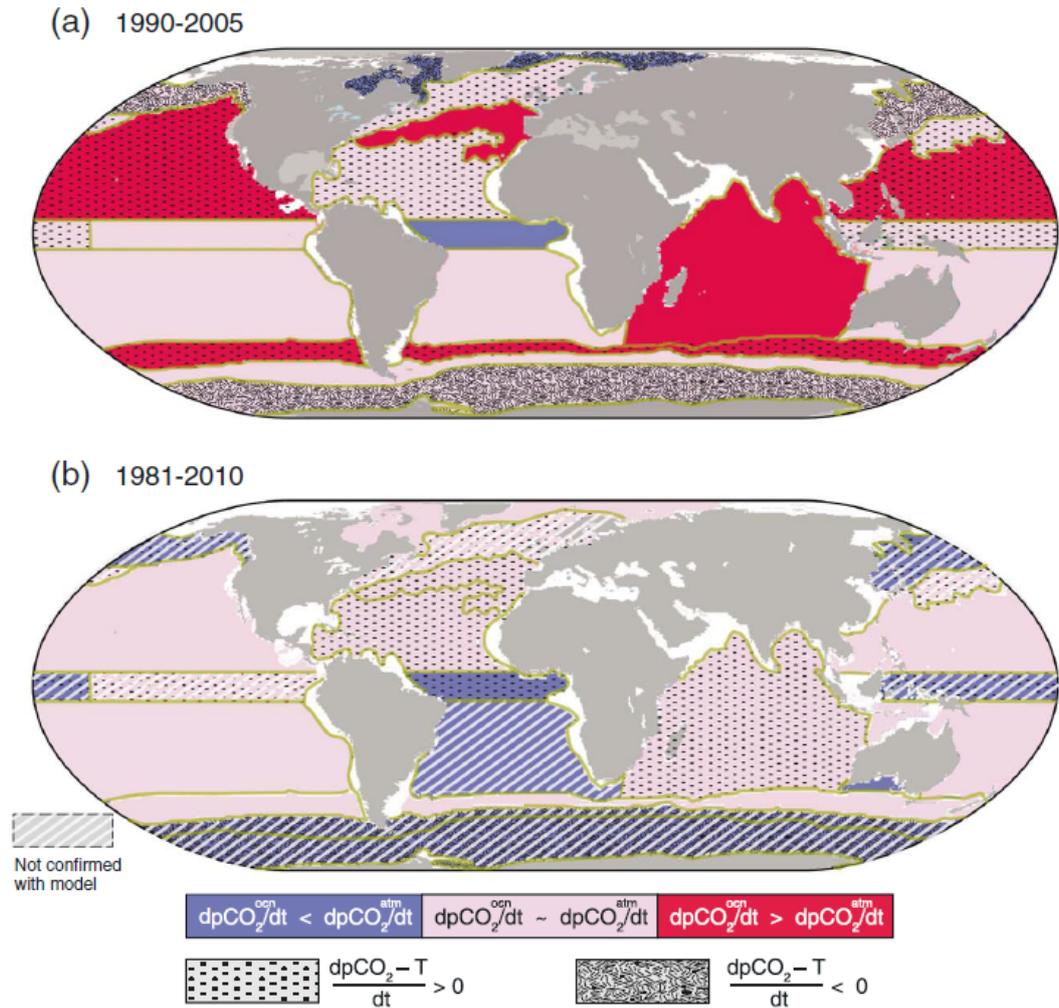


Figure 1.8: Linear trends calculated for 16 biome regions as illustrated Fay and McKinley (2013) for (a) the 1990 to 2005 period and (b) the 1981–2010 period. Blue biome areas indicate a shallower trend in sea surface pCO_2 whereas red biome areas indicate a steeper sea surface pCO_2 trend. Pink biome areas indicate where the ocean pCO_2 trend is parallel to the atmospheric pCO_2 trend

and McKinley (2013).

1.5 Climate variability and CO₂ fluxes

Inter-annual variabilities, strongly influenced by climate, are largely affecting the natural CO₂ flux (Le Quéré *et al.*, 2010; Wanninkhof *et al.*, 2013b; Fay and McKinley, 2013). Previous studies linked strong regional to basin-wide variabilities of the contemporary air–sea CO₂ flux to different climate modes. Three of the most commonly discussed climate indices are illustrated in figure 1.9, namely the Multivariate ENSO (El Niño Southern Oscillation) index (MEI), the North Atlantic Oscillation winter index (NAO) and the

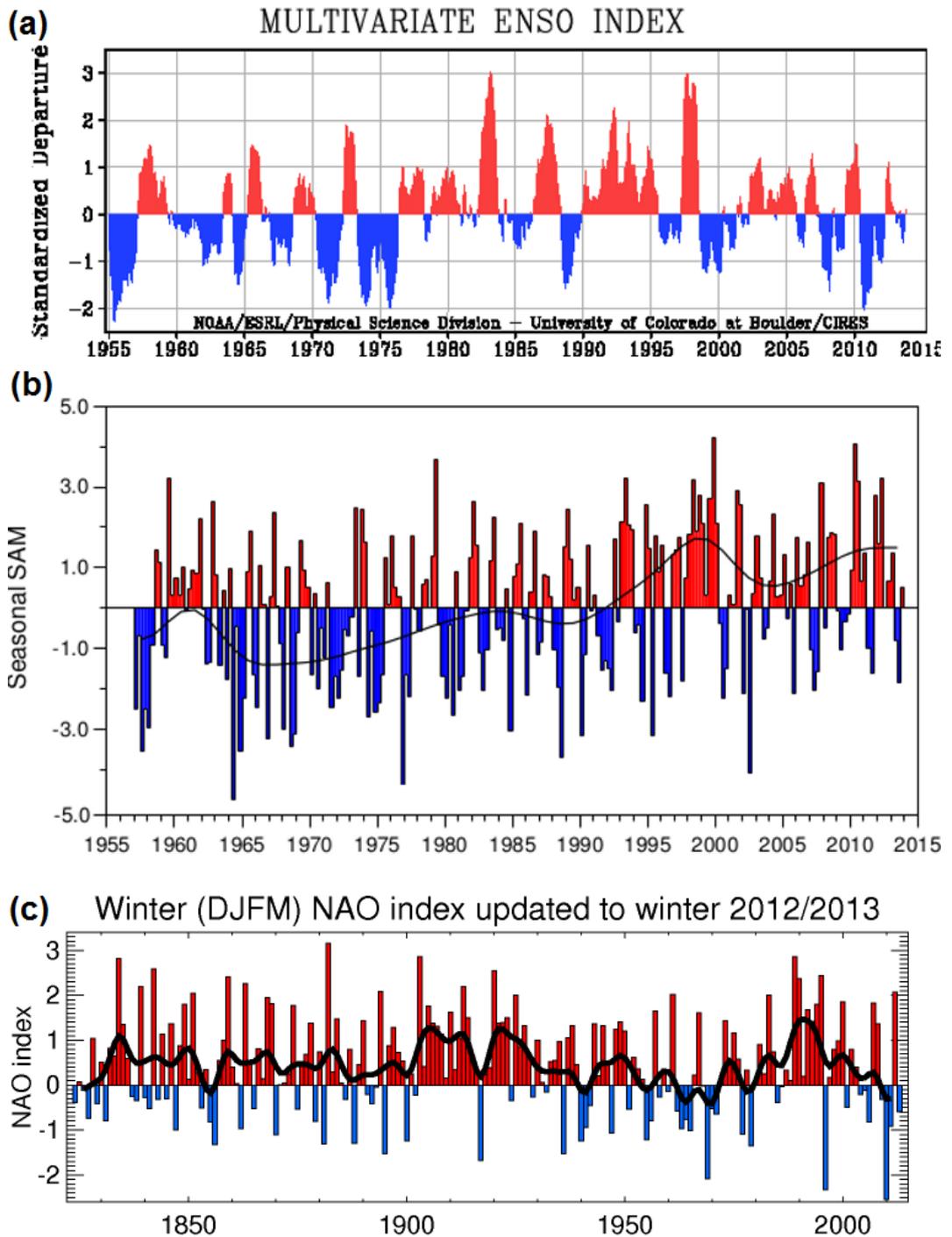


Figure 1.9: Timeseries of (a) the Multivariate ENSO index (MEI) as illustrated in Wolter and Timlin (2011), (b) the Southern Annular Mode (SAM) index as illustrated in Marshall (2003) and (c) the North Atlantic Oscillation (NAO) index as illustrated in Osborn (2011) for an extended time period.

Southern Annular Mode index (SAM). These climate modes and their influence on the air–sea flux of CO₂ are introduced in this section.

ENSO describes the most important coupled ocean–atmosphere phenomenon to cause global climate variability on inter–annual time scales, and refers to strong periodically

occurring variations in sea level pressure and sea surface temperatures in the Equatorial Pacific, which are reflected in the sea surface $p\text{CO}_2$ and the exchange of CO₂ with the atmosphere (Feely *et al.*, 2006). ENSO events appear in two phases, namely (i) the El Niño phase with warm sea surface temperatures and high surface level pressures in the Equatorial Pacific followed by (ii) La Niña with cold surface water temperatures and low sea level pressure in the (see e.g. Trenberth *et al.*, 2007).

The MEI (figure 1.9 (a)) is based on six observed climate variables over the Tropical Pacific, namely (i) pressure, (ii) zonal and (iii) meridional components of the surface wind, (iv) sea surface temperature, (v) surface air temperature, and (vi) total cloudiness fraction of the sky (Wolter, 1987). From those, the MEI is further calculated as the first unrotated Principal Component (PC) of all six observed fields combined (Wolter, 1987).

A study of Feely *et al.* (2006) in the Equatorial Pacific over a 23 year period from 1981–2004 indicates a strong inter-annual signal of air-sea CO₂ flux, driven by the ENSO climate mode, and an additional seasonal variability in the eastern Equatorial Pacific. El Niño periods are characterized by a reduction in the trade wind intensity, reducing the upwelling effect along the Equatorial Pacific (McPhaden *et al.*, 1998). This leads to a reduction in the amount of re-mineralized carbon reaching the ocean surface layer while surface temperatures increase. While the effect of ENSO in the Pacific Ocean is thoroughly discussed in the literature, the global and intra-basin wide contemporary CO₂ flux response to ENSO has barely been addressed in the literature.

The SAM is associated with high and pressure anomalies between mid and high latitudes in the southern hemisphere, hence it reflects changes in the main westerly wind belt (Trenberth *et al.*, 2007). Previous studies suggest, that the trend towards a positive SAM phase in the last decades from ~ 1965 onwards (see figure 1.9 (b)) has enhanced westerly winds and shifted them towards the poles, leading to enhanced upwelling of neutral carbon and a reduction in the Southern Ocean carbon sink (e.g. Le Quéré *et al.*, 2007). More recently, Fay and McKinley (2013) argue that the response to the recent weak negative trend (from 1998 onwards in figure 1.9 (c)) in the Southern Annular Mode (SAM) led to increasing CO₂ flux trends in the Southern Ocean.

The NAO is the dominant large-scale climate mode in the Atlantic Ocean (e.g. Hurrell, 1995) and impacts sea-surface $p\text{CO}_2$ via changes in the driving parameters. During

positive NAO phases (see figure 1.9 (c)), sea surface temperature shows a tripole pattern with cold anomalies in the subpolar region and warm anomalies in the mid latitudes and corresponding changes in vertical mixing and nutrient supply (Marshall *et al.*, 2001).

Using surface ocean $p\text{CO}_2$ observations, Schuster and Watson (2007) argued for decrease in the North Atlantic carbon sink of $\sim 0.24 \text{ Pg C} \cdot \text{yr}^{-1}$ and a reduction in the seasonal $p\text{CO}_2$ amplitude in both the subtropical and temperate North Atlantic from the mid-1990s to the mid-2000s. The authors linked this reduction to the large changes that occurred in the climate mode of the North Atlantic over this period, i.e. a shift of the North Atlantic Oscillation (NAO) from very positive phases in the early 1990s to negative and near-zero phases in the mid-2000s (illustrated in figure 1.9 (c)). Corbière *et al.* (2007) supported this conclusion on the basis of their observations from the subpolar gyre over the 1993 to 2003 period, pointing out that the larger than expected increase in the observed $p\text{CO}_2$ is mainly a result of rapid warming.

Based on the results of a global ocean biogeochemistry model, Thomas *et al.* (2008) argued that the trend toward a smaller North Atlantic sink identified by Schuster and Watson (2007) is transitory and is expected to rebound in the near future, i.e. that this decrease is part of a "natural" fluctuation and should not be interpreted as a signal of anthropogenic climate change. They interpreted the decline in the sink strength to be the result of a NAO-driven reduction in the transport of water by the North Atlantic Current into the eastern subpolar gyre. In a contrasting modeling study, Ullman *et al.* (2009) argued that the North Atlantic carbon sink actually increased from the mid-1990s to the mid-2000s. They proposed that the declining trend in the NAO from the early to mid-1990s until the mid-2000s led to reduced convective mixing in the subpolar gyre, counteracting the impact of warming. Ullman *et al.* (2009) argue, that Thomas *et al.* (2008) do not analyze changes in vertical mixing, which Ullman *et al.* (2009) found to dominate the large scale trends in the subpolar gyre.

1.6 Current limitations on data interpolation methods

A major challenge in detecting trends and variabilities in the global ocean carbon sink from observations is due to the highly heterogeneous distribution of the surface ocean $p\text{CO}_2$ measurements in time and space, as illustrated in figure 1.10. Different approaches

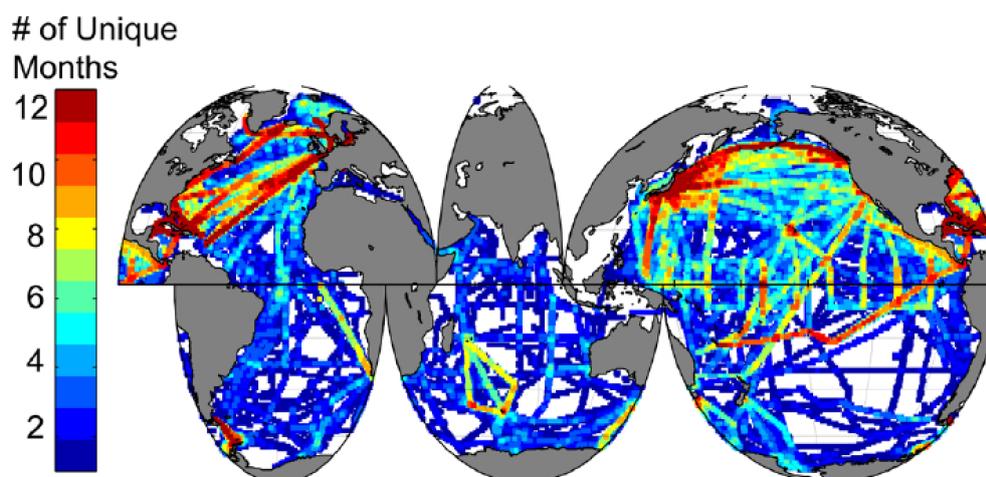


Figure 1.10: Number of unique months where $f\text{CO}_2$ measurements are available for each gridbox in SOCAT v1.5, as illustrated in Sabine *et al.* (2013)

have been employed to overcome this limitation and to create basin wide estimates. These include the binning of data to $4^\circ \times 5^\circ$ bins in latitude and longitude followed by an advection-based interpolation method (Takahashi *et al.*, 1999, 2003, 2009), binning of data to large-scale biogeochemical provinces (McKinley *et al.*, 2011), multi-linear regression models (e.g. Chierici *et al.*, 2009; Peng and Wanninkhof, 2010), neural network-based methods (e.g. Lefèvre *et al.*, 2005; Friedrich and Oschlies, 2009; Telszewski *et al.*, 2009; Sasse *et al.*, 2013) and a novel diagnostic model of ocean mixed layer biogeochemistry (Rödenbeck *et al.*, 2013). Each of these approaches has its strengths and weaknesses.

For example, the binning and interpolation scheme employed by Takahashi *et al.* (1999) is well suited for constraining monthly climatologies. However, its coarse resolution tends to smooth out small-scale features. Nevertheless, the method is not sensitive to outliers, due to the large scale smoothing. The binning to large-scale biogeochemical provinces works well to determine long-term trends (McKinley *et al.*, 2011), but its resolution is even more coarse. The multi-linear regression models allow very finely resolved estimates. However, the explained variance in these statistical models is often relatively low, causing substantial uncertainties in the estimated fields. More recently, self-organizing map (SOM) based neural network techniques have been applied to estimate the sea surface $p\text{CO}_2$. Lefèvre *et al.* (2005) and Telszewski *et al.* (2009) have applied this technique to cluster available observations based on their relationship towards relevant input parameters that are known to affect the sea surface $p\text{CO}_2$ distribution. These authors

then used the average $p\text{CO}_2$ from each cluster to estimate the sea surface $p\text{CO}_2$ where no observations exist and similar input data relations apply. The disadvantage of this method however is, that only a discrete set of clustered $p\text{CO}_2$ data are available to estimate the sea surface $p\text{CO}_2$ where no observations exists, hence it is favourable to use large amounts of clusters. However, data sparsity can leads to clusters without corresponding observations, limiting the extrapolation power of this method. To overcome this limitation Sasse *et al.* (2013) applied a 2-step approach, where the authors use a multi-linear regression technique, rather than the cluster average to obtain a continuous relationship between input data and $p\text{CO}_2$. This approach however, assumes linearity between the sea surface $p\text{CO}_2$ and its predictors.

1.7 Aim of the research

In order to investigate the variability of the sea surface $p\text{CO}_2$ and the resulting air-sea carbon flux, a novel neural network based approach that overcomes most of the aforementioned issues which have limited previous studies, is introduced in chapter 2. This method is capable of capturing a large amount of variability in $p\text{CO}_2$ due to the non-linear predictor-observation relationship on a fine $1^\circ \times 1^\circ$ spatial grid. The method determines the non-linear relationships between the surface ocean $p\text{CO}_2$ observations and a set of independent observations to produce global sea surface maps of $p\text{CO}_2$ on a monthly basis. The network gathers information from similar ocean biogeochemical provinces and provides regional and global $p\text{CO}_2$ estimates.

The newly developed neural network method relies on the assumption that the ocean carbon sink and its variability can be estimated as a function of proxy variables, which are subjectively chosen. Furthermore, it relies on ocean carbon measurements in order to establish a correct relationship. Therefore, the work presented in this thesis benefits from the recent publication of the SOCAT database, which provides the to-date largest global data set of surface ocean fugacity of CO_2 (similar to the $p\text{CO}_2$, but accounting for the non-ideal behaviour of CO_2) observations (Pfeil *et al.*, 2013; Sabine *et al.*, 2013; Bakker *et al.*, 2014). Chapter 3 gives an introduction on the $p\text{CO}_2$ measurement techniques used in the North Atlantic Ocean, similar to those included in SOCAT. Furthermore, bottle samples collected on-board a VOS ship are used to investigate the variability of the major

carbonate system parameter, namely DIC, TALK and the sea surface $p\text{CO}_2$, along the UK–Caribbean shipping line.

The neural network method introduced in chapter 2 will be used to investigate the mean state, as well as the variability of the sea surface $p\text{CO}_2$ and the air–sea gas flux of CO_2 in the Atlantic Ocean (chapter 4) and the global ocean (chapter 5). The estimated sea surface $p\text{CO}_2$ will first be validated with independent data from timeseries stations and moorings, before the results are analysed. The global results in chapter 5 provide an estimate of the ocean sided pathway of the globally emitted carbon and therefore provide the basis to close the global carbon budget based on surface ocean CO_2 observations. Finally, the results and findings of this thesis will be discussed in chapter 6.

Chapter 2

Methods

”Borders. I have never seen one.
But I have heard they exist in the
minds of some people.”

(Thor Heyerdahl)

Methods of the Atlantic Ocean study in the following chapter have been published in:

**A neural network–based estimate of the seasonal to
inter–annual variability of the Atlantic Ocean carbon sink**

P. Landschützer¹, N. Gruber², D.C.E. Bakker¹, U. Schuster^{1,6}, S. Nakaoka³, M.R.
Payne^{2,4}, T.P. Sasse⁵, J. Zeng³

¹School of Environmental Sciences, University of East Anglia, Norwich Research Park, Norwich,
UK

²Institute of Biogeochemistry and Pollutant Dynamics, ETH Zürich, Zürich, Switzerland

³Center for Global Environmental Research, National Institute for Environmental Studies, Tsukuba,
Ibaraki, Japan

⁴Section for Oceanography and Climate, National Institute for Aquatic Resources, Technical Uni-
versity of Denmark, Charlottenlund, Denmark

⁵Climate Change Research Centre, University of New South Wales, Sydney, Australia

⁶Now at: College of Life and Environmental Sciences, Hatherly Laboratories, University of Ex-
eter, Exeter, UK

Biogeosciences, 2013, 10, 7793–7815, doi: 10.5194/bg-10-7793-2013

*The work and analysis presented in this chapter was undertaken by P. Landschützer.
Co–authors on this publication provided guidance and suggestions regarding the analysis
and the results to help and address the interests of the wider scientific community.*

This chapter will provide an introduction to neural networks, the basic principle of how they can be used to estimate a certain target value from dependent input data and how they are used in particular for the studies presented in this thesis. A novel 2-step approach will be introduced where the global ocean is firstly divided into biogeochemical provinces, and secondly the input-target relationship is reconstructed for each province separately. This non-linear relationship is then used to estimate sea surface $p\text{CO}_2$ targets where no direct observations are available.

2.1 Using neural networks to estimate the sea surface $p\text{CO}_2$

In order to estimate the sea surface partial pressure and the air-sea gas flux of CO_2 , two neural network methods are used to reconstruct the relationship between input variables and target measurements of the sea surface $p\text{CO}_2$ on a monthly $1^\circ \times 1^\circ$ resolution. This novel 2-step approach is used to establish numerical relationships between the sea surface $p\text{CO}_2$ and a suite of input data that are known to drive its variability. In the first stage, a neural network clustering algorithm (SOM) is used to define a discrete set of biogeochemical provinces that share a common relationship between the independent inputs and the target $p\text{CO}_2$. In the second stage, for each biogeochemical province a non-linear and continuous relationship is derived between $p\text{CO}_2$ and the input parameters on the basis of a feed-forward network (FFN) method. This input-target relationship is then used to estimate surface ocean $p\text{CO}_2$ for each month and each pixel.

The resulting surface ocean $p\text{CO}_2$ distribution is then combined with corresponding atmospheric $p\text{CO}_2$ data and wind-speed based estimates of the gas transfer velocity to construct the mean and variability of the Atlantic Ocean carbon sink (chapter 4) from 1998 through 2007 using the SOCAT v1.5 dataset (Pfeil *et al.*, 2013; Sabine *et al.*, 2013) and of the global ocean carbon sink (chapter 5) from 1998 through 2011 using the SOCAT v2 dataset (Bakker *et al.*, 2014). In both stages, for the Atlantic study and the global ocean study, global input and target data sets are used, taking advantage of the fact that biogeochemical provinces with limited coverage in a particular ocean basin can learn from observations in the same biogeochemical province in another ocean basin.

Prior to the decision of combining the two neural network methods, both of them were tested separately. The SOM technique, as it was previously applied (e.g. Telszewski *et al.*,

2009) and tested here, can only reconstruct a discrete number of estimates, hence in order to capture the range of globally observed $p\text{CO}_2$ values, and further be able to capture small variabilities, over a time-span of ten years or more, a large amount of processing units (or neurons) is required. This results in three major issues: I) A large number of neurons requires large computational performance, II) data sparsity leads to empty clusters when large numbers of neurons are applied, hence the method partly loses its extrapolation power and III) The non-uniform distribution of both input variables and observations require extensive pre-processing of the data, in the form of data normalization, to not bias the cluster result towards one particular input.

The stand-alone FFN results provide a continuous range of $p\text{CO}_2$ estimates and data normalization is not required, however, yet again, the non-uniform observation distribution provides a major challenge for this method. The aim of the method is to minimize the mean squared error between estimates and observations, hence where the majority of the observations exist in time (e.g. 2005 through 2007) and in space (e.g. the temperate North Atlantic and North Pacific) the method is able to reproduce the data with virtually no bias, however data poor regions receive less weight, and the method overfits the temperate regions, but loses the extrapolation power over large data poor areas of the global ocean. This is reflected in high variance within the estimated fields, which is not in line with available observations.

The combination of both methods aims to overcome most of this issues, i.e., to produce a continuous number of estimates within certain pre-defined province. This province definition assures that data poor regions receive more weight and are not overpowered by data rich regions.

Figure 2.1 shows a schematic of the method, from the choice of input variables via the neural network steps to the final product. Neural networks, their applications in science, as well as each step individually are explained in more detail in the following subsections.

2.1.1 Neural networks

Neural networks have been previously applied to solve complex problems in many scientific disciplines, such as atmospheric sciences (e.g. Kolehmainen *et al.*, 2001; Marzban

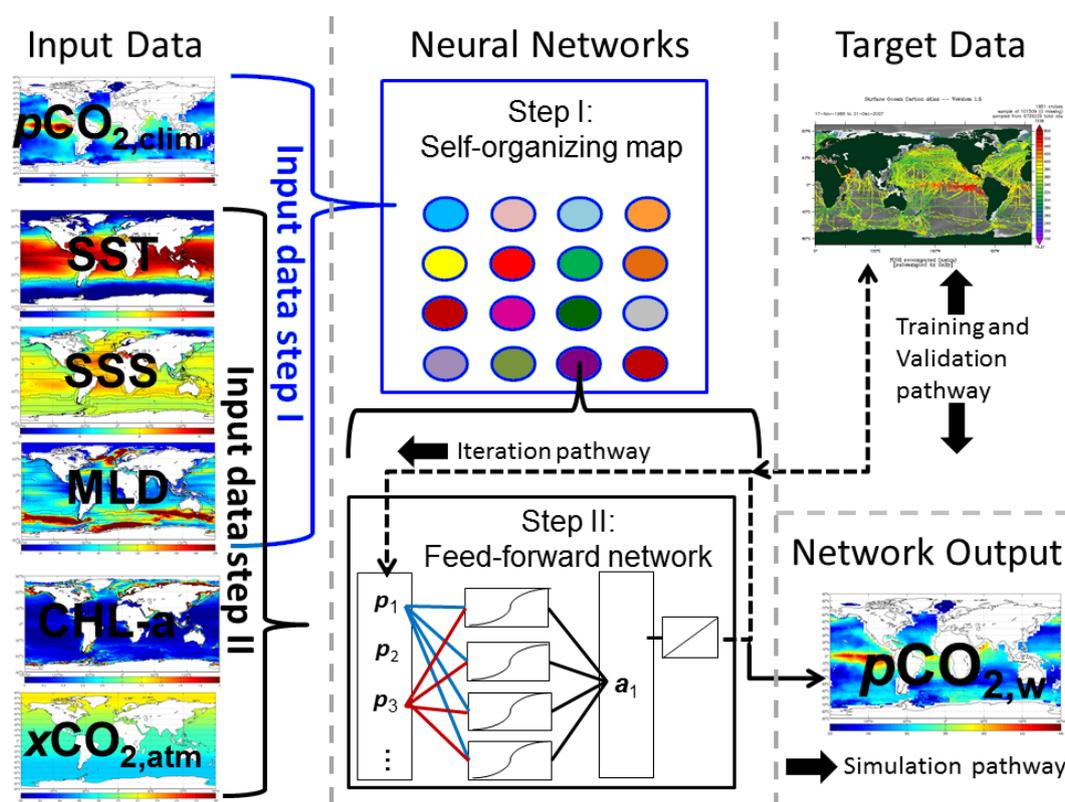


Figure 2.1: Schematic of the neural network method, including the input data, the two neural network approaches (step 1 is coloured in blue and step 2 is coloured in red) including their different pathways, the target observations and the final $p\text{CO}_2$ fields.

and Stumpf, 1996), climate science (e.g. Knutti *et al.*, 2003), terrestrial science (e.g. Lundstedt, 1992; Maier and Dandy, 2000) and ocean science (e.g. Tsai *et al.*, 2002; Tangang *et al.*, 1997). More recently, neural network techniques made their way into the field of ocean biogeochemistry. Lefèvre *et al.* (2005), Friedrich and Oschlies (2009), Telszewski *et al.* (2009), Sasse *et al.* (2013) and Nakaoka *et al.* (2013) have recently applied neural network techniques to create maps of the sea surface $p\text{CO}_2$. However, neural network based maps were previously limited to certain ocean basins (e.g. Nakaoka *et al.*, 2013) or climatologies of the global ocean (e.g. Sasse *et al.*, 2013).

The term neural network (or artificial neural network) arises from the biological processing unit called "neuron" of the human brain. Although neural networks as they are used in science are purely mathematical, they are inspired by the strong inter-connection of the biological neurons and the unique way these inter-connected units solve specific problems by learning from previous examples. There are several different groups of neural networks, designed to solve specific problems, such as pattern recognition, identification,

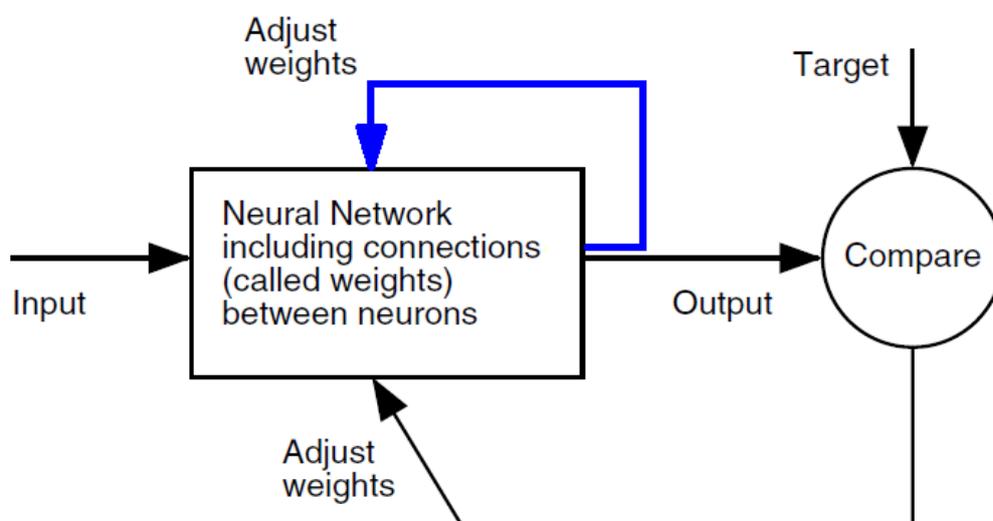


Figure 2.2: Schematic of the typical neural network processing chain from Demuth *et al.* (2008) represented by the black lines. Inputs are processed in the neural network and the network outputs are compared to targets. Iteratively, the network weights are adjusted to minimize the output–target misfit. The blue line represents the typical processing chain of a SOM, where no output is produced and the weights are adjusted according to the presented input data.

classification, speech, vision, and control systems (Demuth *et al.*, 2008). Despite the large variety of neural networks and their applications, they are typically all trained in a way that a particular input leads to a specific target output. The network training is an iterative process, where the coefficients of each neuron, called weights of a neuron, are adjusted after every iteration so the network output best matches a target value (Demuth *et al.*, 2008). Figure 2.2 shows this general process chain of neural networks following Demuth *et al.* (2008). Inputs are provided to the neural network, they are processed and an output is generated. The output is then validated against targets and the weights of the neurons are adjusted in response.

Neural networks are commonly differentiated in static and dynamic categories and by the number of neuron layers. Static neural networks, like e.g. the feed–forward neural network used here in step 2, compute the network output and adjust the weight of the neurons in response to a single input, whereas dynamic networks compute their output and adjust the neuron weights in response to current and previous network inputs and outputs combined. Neural networks are commonly organized in network layers (often referred to as hidden layers and output layer), where each layer consists of inter–connected neurons. The inter–connections between inputs and layer neurons is shown in figure 2.3. Input data

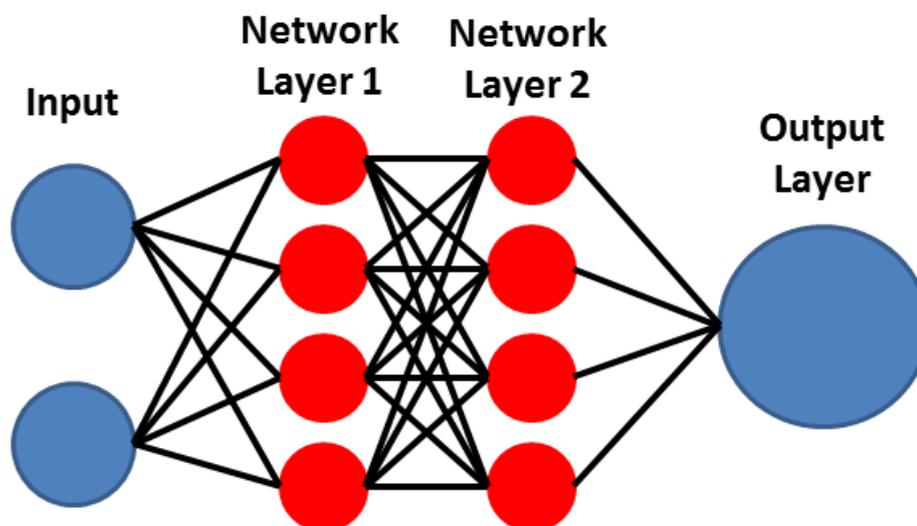


Figure 2.3: Schematic of the connections of a three layer neural network, including two network layers and one output layer. The connections illustrate the link between inputs, neurons and targets illustrated in figure 2.2.

are processed in these network layers via their system of weighted connections.

2.1.2 Data choice

In order to train a neural network, sets of input and target data have to be provided. The $f\text{CO}_2$ observations from the SOCAT v1.5 (Sabine *et al.*, 2013) and v2 (Bakker *et al.*, 2014) database, gridded onto monthly $1^\circ \times 1^\circ$ fields, form the basis of the computations. The cruise-weighted average was used here, where first all data collected from the same cruise within a grid box were averaged and in a second step, the grid box averages from each cruise separately were further averaged within the same box to avoid biasing the final value towards cruises where high resolution measurements were obtained (Sabine *et al.*, 2013). These SOCAT data sets includes global observations over the period 1970 to 2007 in version 1.5 and over the period from 1970 to 2011 in version 2 and underwent an extensive series of automatic and manual secondary quality controls (Pfeil *et al.*, 2013). For both the SOCAT v1.5 and the SOCAT v2 database, the reported $f\text{CO}_2$ estimates were converted to $p\text{CO}_2$ using the formulation (see e.g. Weiss, 1974; Körtzinger, 1999):

$$p\text{CO}_2 = f\text{CO}_2 \cdot \exp\left(P_{atm}^{surf} \frac{B + 2 \cdot \delta}{R \cdot T}\right)^{-1} \quad (2.1)$$

where P_{atm}^{surf} is the total atmospheric surface pressure, B and δ are virial coefficients (Weiss, 1974), R is the universal gas constant and T is the absolute temperature (Reynolds *et al.*, 2002). The monthly mean sea level pressure of the National Centers for Environmental Prediction (NCEP) was used for P_{atm}^{surf} (Kalnay *et al.*, 1996). The correction from $f\text{CO}_2$ to $p\text{CO}_2$ describes only minimal changes in absolute terms, resulting in a global mean difference of only $\sim 1.3 \mu\text{atm}$ between the two fields.

Figure 2.4 (a) shows the distribution of the converted gridded $p\text{CO}_2$ observations of the SOCAT v1.5 database from 1998 through 2007 and figure 2.4 (b) shows the distribution of the SOCAT v2 database from 1998 through 2011. It demonstrates the heterogeneous distribution of observations in the global ocean. While the northern hemisphere is spatially well covered, large areas have little data coverage in the southern hemisphere.

In the Atlantic Ocean the $f\text{CO}_2$ (and converted $p\text{CO}_2$) data distribution of the SOCAT v1.5 dataset has relatively good spatial coverage, but yet are highly skewed in time. The number of $1^\circ \times 1^\circ$ pixels with $f\text{CO}_2$ measurements within the chosen study area varies from as low as ~ 180 ($\sim 1\%$ of all data) per year in 1999 and 2000, to over 4000 ($\sim 20\%$ of all data) per year in 2006 and 2007, with the latter two years accounting for 40% of all measurements. This provides a challenge for conventional regression models to not introduce a systematic bias towards regions and years, where the majority of observations exist. The new 2-step neural network approach compensates for the limited coverage within a province in a certain ocean basin by learning from observations in the same biogeochemical province in another ocean basin. The global data on the other hand are more homogeneously distributed in time. In contrast, the Atlantic has a good spatial coverage, while this is not the case for many of the other ocean basins.

A crucial choice concerns the selection of the independent input variables used for the training of the networks. Takahashi *et al.* (1993) suggested that changes in the sea surface $p\text{CO}_2$ are determined by four main drivers, namely the sea surface temperature (SST), sea surface salinity (SSS) the concentration of total dissolved carbon (DIC) and the total alkalinity (TALK) of the surface waters, representing the ocean chemistry and

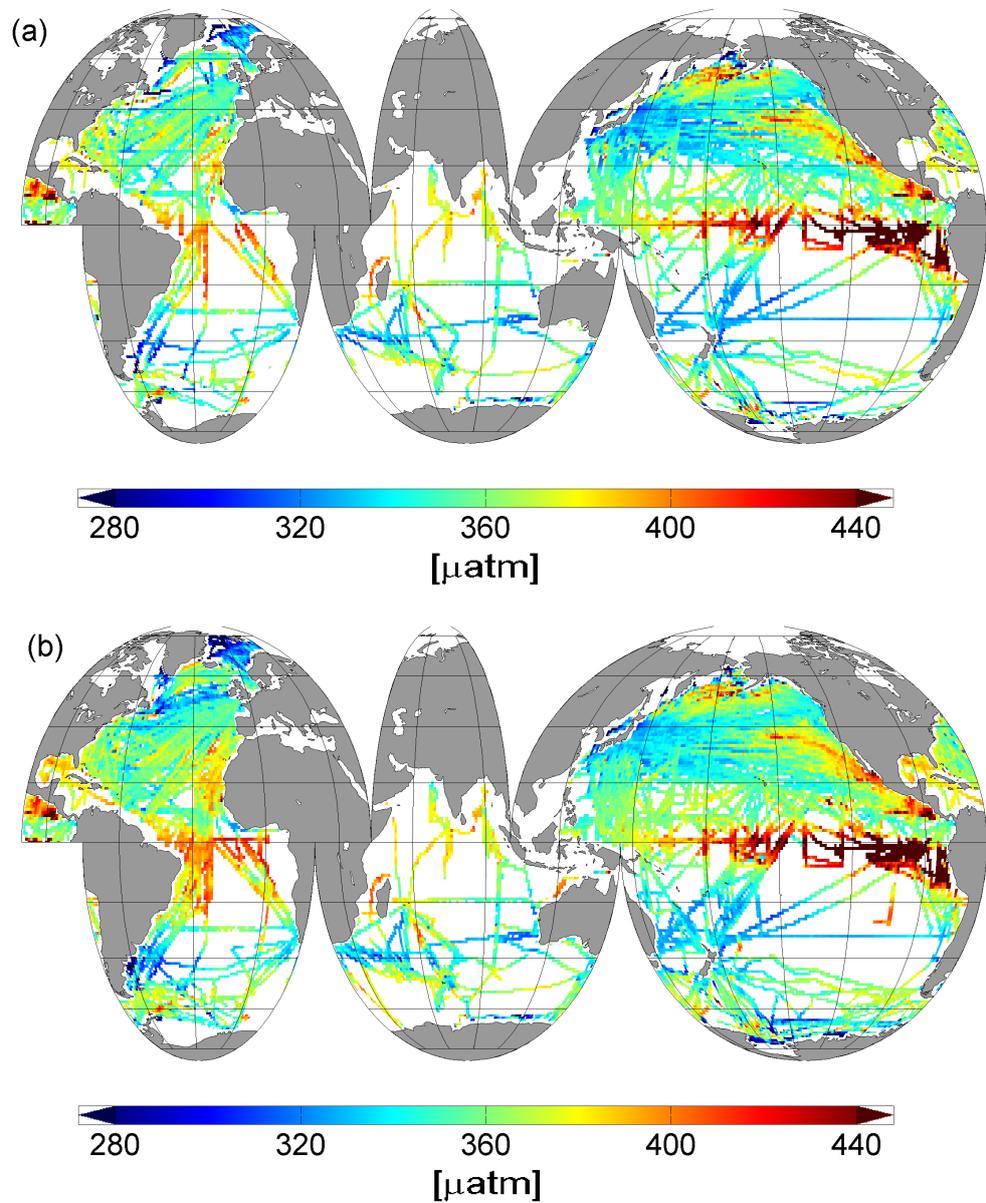


Figure 2.4: (a) Map of the gridded observations within the SOCAT v1.5 database (Pfeil *et al.*, 2013; Sabine *et al.*, 2013) from 1998 through 2007 and (b) map of the gridded SOCAT observations included in version 2 (Bakker *et al.*, 2014) from 1998 through 2011.

physics. The choice of SST and SSS was made as these parameters represent proxies as suggested by Takahashi *et al.* (1993), and are available globally on a monthly timescale, spanning the entire study period. However, to-date TALK and DIC products are limited to annual and monthly climatologies and are therefore not suited to reconstruct year-to-year variabilities.

Hence, in addition to SST and SSS, the method requires additional information about

Table 2.1: Input datasets and their original resolution used to estimate the sea surface $p\text{CO}_2$ within the Atlantic Ocean study (chapter 4) and the global ocean study (chapter 5).

| input variables | original resolution | Atlantic Ocean study | global ocean study |
|----------------------------|--|----------------------|--------------------|
| Reynolds SST | $1^\circ \times 1^\circ \times \text{monthly}$ | X | X |
| SeaWiFS CHL | $9\text{km} \times 9\text{km} \times \text{monthly}$ | X | - |
| Globcolour CHL | $1^\circ \times 1^\circ \times \text{monthly}$ | - | X |
| ECCO2 MLD | $1/4^\circ \times 1/4^\circ \times \text{daily}$ | X | X |
| ECCO2 SSS | $1/4^\circ \times 1/4^\circ \times \text{daily}$ | - | X |
| SODA SSS | $1/2^\circ \times 1/2^\circ \times \text{monthly}$ | X | - |
| Globalview $x\text{CO}_2$ | $0.05 \sin(\text{rad}) \times \text{monthly}$ | X | X |
| $p\text{CO}_2$ climatology | $72^\circ \times 45^\circ \times \text{monthly}$ | X | X |

physical, chemical and biological processes which alter the sea surface DIC and TALK concentrations and therefore effect the sea surface $p\text{CO}_2$. In section 1.3.1 it was noted, that biological production decreases the surface DIC concentration and increases the TALK concentration, hence a proxy is required to establish the link between the effect of biological production and sea surface $p\text{CO}_2$. Since the start of the SeaWiFS satellite mission in 1997 (SeaWiFSProject,

<http://oceancolor.gsfc.nasa.gov/cgi/l3>), high resolution monthly and inter-annually varying chlorophyll-a (CHL) data are available and are therefore added as a simple but unique proxy to represent the biological effect on the sea surface $p\text{CO}_2$ distribution.

Furthermore, vertical mixing allows the re-introduction of DIC from deeper layers of the ocean (see section 1.3.1), increasing the partial pressure of CO_2 at the sea surface. Hence the ocean mixed layer depth (MLD) was added as an additional physical parameter.

At last, trends in sea surface $p\text{CO}_2$ are attributed to the increase in atmospheric CO_2 (see e.g. McKinley *et al.*, 2011), hence the atmospheric mole fraction of CO_2 ($x\text{CO}_{2,atm}$) was added to represent the main driver for the anthropogenic induced trend in sea surface $p\text{CO}_2$. The same product is further used for the calculation of the air-sea gas flux (section 2.1.6).

There are however more inter-annually varying proxies that were tested within the decision making process, e.g. sea surface height and the surface ocean winds. The final decision to use SST, SSS, MLD, CHL and $x\text{CO}_2$ was based on error analysis, i.e., which combination of variables provides the best fit compared to SOCAT as well as independent timeseries data, which will be introduced in the next subsection.

The time period of the analysis is restricted from January 1998 to December 2007 for the first study concentrating on the Atlantic Ocean basin (chapter 4) and January 1998 to December 2011 for the second study covering the global ocean study (chapter 5), due to the temporal limitations of the data that were chosen for these studies (CHL, SOCAT v1.5 and SOCAT v2)

For the Atlantic study, presented in chapter 4, the National Oceanic and Atmospheric Administration (NOAA) Optimum Interpolation (OI) sea surface temperature v.2 (SST) (Reynolds *et al.*, 2002), the SeaWiFS mapped CHL (SeaWiFSProject, <http://oceancolor.gsfc.nasa.gov/cgi/l3>), the MLD data from the Estimating the Circulation and Climate of the Ocean, Phase II (ECCO2) project (Menemenlis *et al.*, 2008), the Simple Ocean Data Assimilation (SODA) SSS data (Carton and Giese, 2008) and the monthly atmospheric $x\text{CO}_{2,atm}$ from GLOBALVIEW-CO2 (2011) are used. Due to the additional years from 2008 through 2011 of the global study, presented in chapter 5, the SSS and CHL datasets are replaced by the ECCO2 project SSS (Menemenlis *et al.*, 2008) and the Globcolour CHL (<http://www.globcolour.info>) datasets, respectively. In addition, the monthly $p\text{CO}_2$ climatology, with the reference year 2000, of Takahashi *et al.* (2009) is used in both studies as an additional input parameter for defining biogeochemical provinces. The datasets used within the different studies are further summarized in table 2.1. In view of their strongly skewed distribution, MLD and CHL have been log-transformed before used as input data.

The chosen input data illustrated in table 2.1 represent a variety of observation-based interpolation products ($p\text{CO}_2$ climatology, Reynolds SST, Globalview $x\text{CO}_2$), data assimilation based products (SODA SSS), satellite derived products (SeaWiFS CHL, Globcolour CHL, Reynolds SST), and model estimates constraint by in-situ (e.g. ARGO) data (ECCO2 MLD, ECCO2 SSS). While satellite based products have the advantage of near global direct coverage, other products rely on the spatial and temporal coverage of available observations. While e.g. the SST product has small uncertainties attached, due to the satellite-observation bias correction applied (Reynolds *et al.*, 2002), MLD and SSS products are indirect model estimate constrained by temperature and salinity profiles (Menemenlis *et al.*, 2008) and are therefore assumed to have larger uncertainties attached.

Table 2.2: Input and target vector elements for each subset used within the method. The subscript ds describes de-seasonalised data, which are computed by subtracting the long-term mean seasonal cycle from the original dataset. Additionally, $\log(\text{CHL})$ was excluded from sets FINP, FITR, FIVAL and FINP2 to estimate $p\text{CO}_2$ where no satellite chlorophyll-a is available due to cloud cover.

| Set name | Elements n of the j^{th} input vectors \mathbf{p}_n^j | Targets (t^j) |
|---|---|------------------------|
| SINP (SOM Input) | SST, $\log(\text{MLD})$, SSS, $p\text{CO}_{2,clim}$ | - |
| FINP, FITR, FIVAL (FFN Input, Training and Validation) | SST, $\log(\text{CHL})$, $\log(\text{MLD})$, SSS, $x\text{CO}_{2,atm}$, SST $_{ds}$, CHL $_{ds}$, MLD $_{ds}$, SSS $_{ds}$, $x\text{CO}_{2,atm,ds}$ | $p\text{CO}_{2,SOCAT}$ |
| FINP2 (FFN Input 2) | SST, $\log(\text{CHL})$, $\log(\text{MLD})$, SSS, $x\text{CO}_{2,atm}$, SST $_{ds}$, CHL $_{ds}$, MLD $_{ds}$, SSS $_{ds}$, $x\text{CO}_{2,atm,ds}$ | - |

2.1.3 Data preparation and validation data

Data from the previous section with an original resolution finer than the required $1^\circ \times 1^\circ$ were averaged onto the desired grid, whereas input data with a coarser resolution were interpolated using a bilinear interpolation. Further, monthly averages were taken of all inputs with a finer temporal resolution.

Seasonal variabilities are much larger than trends and year-to-year variabilities within the input data for most parts of the ocean, hence, to highlight anomalies and year-to-year trends within the data sets, de-seasonalised sets of the input variables were further produced by removing their long-term mean seasonal cycle from the original dataset. This approach aims to assure that small signals are not hidden by much larger signals.

In the next step the monthly $1^\circ \times 1^\circ$ input data are rearranged into three major data sets. Each of these data sets consists of input vectors (\mathbf{p}_n) where the input data are organized as row vector elements, for example SST, $\log(\text{MLD})$, SSS, and $p\text{CO}_{2,clim}$ for the self-organizing map input (SINP) dataset (table 2.2). Table 2.3 as an example shows how the input vectors are arranged for the Atlantic Ocean study from 1998 through 2007, namely by sampling each input dataset at the same space-time point. Every input vector \mathbf{p}_n consists of data points at the same geographical location on a $1^\circ \times 1^\circ$ grid at the same point in time, as illustrated in table 2.3. Two of the main data sets, SINP and the feed-forward network input 2 set (FINP2) do not have a corresponding target dataset (table 2.2) as they are used to produce basin-wide outputs. Input vectors with empty vector

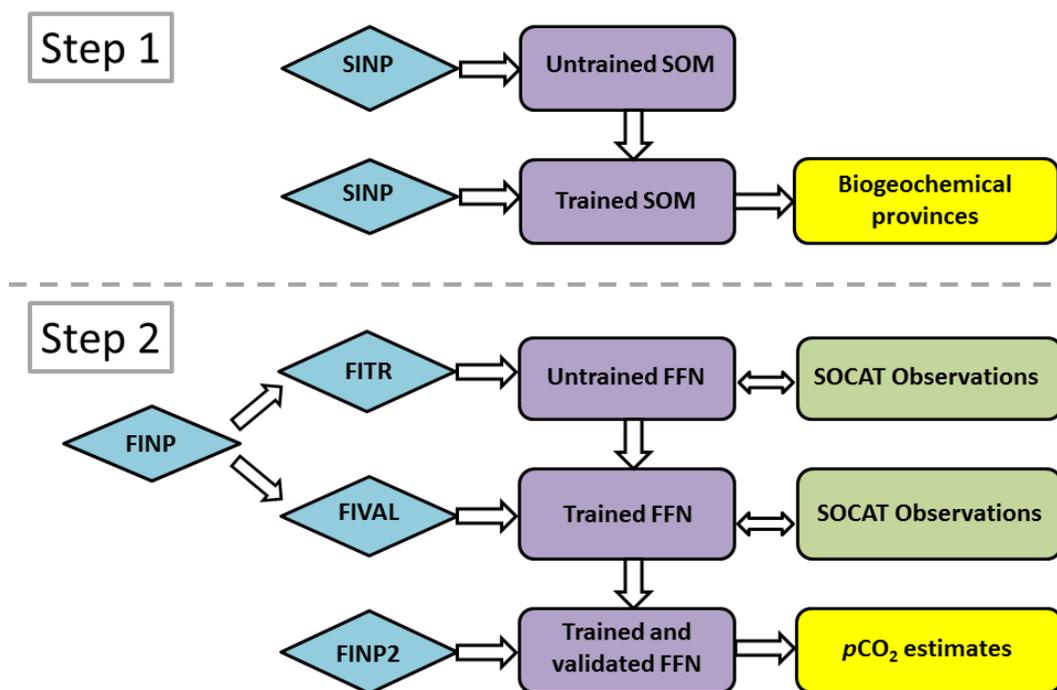


Figure 2.5: Flow chart of the different stages within the neural network methods and the datasets used to train and validate the neural networks.

elements, e.g. where no salinity data are available, were further removed from these data sets. The third major set, the feed–forward network input set (FINP), consists only of input vectors where corresponding SOCAT observations, or targets (t), are available, i.e., they are further subsampled in time at the locations where observations are available to estimate the input–target relationship. In order to train the FFN, two sub sets of the FINP set are created, namely the actual training (FITR) set and a validation (FIVAL) set (table 2.2). The different stages of the 2–step neural network method, as well as the datasets used are illustrated in figure 2.5.

Where no chlorophyll–a satellite data are available, due e.g., to cloud cover or lack of sufficient light at high latitudes in winter time, the sea surface $p\text{CO}_2$ is estimated with the remaining input parameters. This applies to about 22% of all pixels in the Atlantic Ocean from 1998 through 2007 and about 18% of all estimated global pixels from 1998 through 2011 and mainly concerns the high latitude oceans in winter. Since chlorophyll concentrations tend to be low and unvarying during these months, it is expected that this choice has a relatively small influence on the results.

To evaluate the results of the Atlantic Ocean study (chapter 4), data from two sources

Table 2.3: Arrangement of r input vectors with n vector elements from the global input data from 1998 to 2007

| | Point in time and space | Input vector elements |
|---------|---------------------------|---|
| $j=1$ | Jan 1998; 89.5°S; 179.5°W | $\mathbf{p}_1^1, \mathbf{p}_2^1, \dots, \mathbf{p}_n^1$ |
| $j=2$ | Jan 1998; 89.5°S; 178.5°W | $\mathbf{p}_1^2, \mathbf{p}_2^2, \dots, \mathbf{p}_n^2$ |
| : | : | : |
| $j=p-1$ | Jan 1998; 89.5°S; 178.5°E | $\mathbf{p}_1^{p-1}, \mathbf{p}_2^{p-1}, \dots, \mathbf{p}_n^{p-1}$ |
| $j=p$ | Jan 1998; 89.5°S; 179.5°E | $\mathbf{p}_1^p, \mathbf{p}_2^p, \dots, \mathbf{p}_n^p$ |
| $j=p+1$ | Jan 1998; 88.5°S; 179.5°W | $\mathbf{p}_1^{p+1}, \mathbf{p}_2^{p+1}, \dots, \mathbf{p}_n^{p+1}$ |
| : | : | : |
| $j=q-1$ | Jan 1998; 89.5°N; 178.5°E | $\mathbf{p}_1^{q-1}, \mathbf{p}_2^{q-1}, \dots, \mathbf{p}_n^{q-1}$ |
| $j=q$ | Jan 1998; 89.5°N; 179.5°E | $\mathbf{p}_1^q, \mathbf{p}_2^q, \dots, \mathbf{p}_n^q$ |
| $j=q+1$ | Feb 1998; 89.5°S; 179.5°W | $\mathbf{p}_1^{q+1}, \mathbf{p}_2^{q+1}, \dots, \mathbf{p}_n^{q+1}$ |
| : | : | : |
| $j=r$ | Dec 2007; 89.5°N; 179.5°E | $\mathbf{p}_1^r, \mathbf{p}_2^r, \dots, \mathbf{p}_n^r$ |

that are not included in SOCAT v1.5 are used. Firstly, timeseries data are extracted from the combined record from BATS (Bermuda Atlantic Time Series Station) and Hydrostation "S" (Bates, 2007; Gruber *et al.*, 2002) and the European Station for Time Series in the Ocean (ESTOC) (e.g. González-Dávila *et al.*, 2007). Sea surface $p\text{CO}_2$ data from BATS are calculate from the actual recorded TALK and DIC observations, as further explained in chapter 3. Secondly, a total of 3065 additional data points within the Atlantic basin from 1998 through 2007 are used from the updated SOCAT v2 database, which were not included in version 1.5 and therefore constitute independent data.

Results of the global study (chapter 5) are validated with the extended temporal record of the timeseries above, as well as further independent timeseries data not included in SOCAT v2 and available from the Carbon Dioxide Information Analysis Center (CDIAC) CO_2 timeseries and mooring project (<http://cdiac.ornl.gov/oceans/Moorings/>). Data from timeseries stations that do not report the sea surface $p\text{CO}_2$ directly, have been calculated from the actual recorded TALK and DIC observations using the method explained in chapter 3.

To evaluate the sensitivity of the results with regard to the chosen input data, 4 sensitivity runs were performed within the Atlantic Ocean (chapter 4) study, namely

- SR1 (sensitivity run 1), where the SODA sea surface salinity was replaced with the World Ocean Atlas 2009 (Antonov *et al.*, 2010) sea surface salinity climatology
- SR2, where the ECCO2 MLD product was replaced with the de Boyer Montegut

(de Boyer Montegut *et al.*, 2004) MLD climatology

- SR3, where the SST product was replaced with the SODA (Carton and Giese, 2008) sea surface temperature product
- SR4, where chlorophyll-a was excluded as an input parameter

2.1.4 Dividing the global ocean into biogeochemical provinces using a self-organizing map

In a first neural network step, SOM method (Kohonen, 1987, 2001) is used to partition the global ocean into 16 biogeochemical provinces. Such a biogeochemical province is characterized by each of its locations having a similar relationship among all considered input data. Based on trial and error, the choice of climatological $p\text{CO}_2$ from Takahashi *et al.* (2009), SST, $\log(\text{MLD})$ and SSS as the input data for the SOM (SINP data set in table 2.2) was made. Chlorophyll-a, i.e., $\log(\text{CHL})$, was excluded due to its many missing values. This section provides more detail to the first step outlined in figure 2.5.

A SOM is a neural network based cluster algorithm that can detect regularities within the provided input data and then learns to group similar input data together (Demuth *et al.*, 2008). It is a one layer neural network, hence the network consists of only one network layer (see figure 2.3). It is capable of identifying similar input data, arranged as input vectors via their Euclidean distance towards the neurons of the network which then physically respond by moving towards the input vectors. As a one layer network, a self-organizing map does not have an output layer, hence does not compute outputs. It is the adjustment of the neurons weights, i.e. the response of the neurons and their adjustment towards the inputs, that makes this algorithm suitable to cluster the global ocean. Figure 2.6 shows the adjustments of neurons (their physical movement) in response to input data.

The choice of 16 provinces represents a subjectively determined optimum between too many provinces with too little data but a high degree of correlation between the provinces, and too few provinces with a lot of data, but too high variance in the data. Self organizing maps have previously been used to estimate the sea surface $p\text{CO}_2$ directly (Lefèvre *et al.*, 2005; Friedrich and Oschlies, 2009; Telszewski *et al.*, 2009; Nakaoka *et al.*, 2013) by assigning a set of discrete $p\text{CO}_2$ values to a large amount of trained neurons. Here, a similar approach is used to train the neurons, but rather than assigning a $p\text{CO}_2$ value to

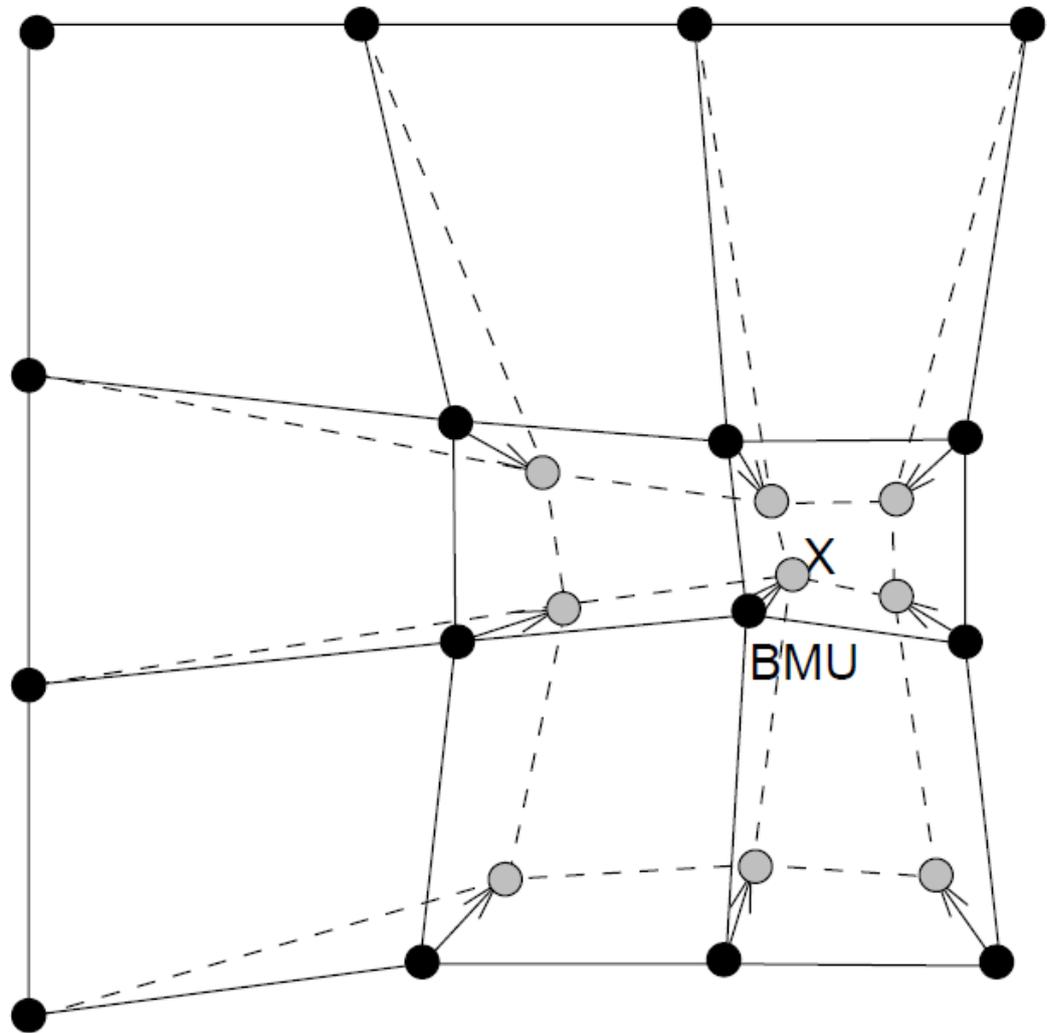


Figure 2.6: Neuron response to input data marked as X following Vesanto *et al.* (2000). Neurons and their neighbours organized on a regular grid, are physically drawn towards the input. BMU stands for "best matching unit" and refers to the closest, or winning neuron.

each neuron, a second neural network is used (step 2) to estimate a continuous set of $p\text{CO}_2$ values for each province.

The chosen map was initialized with 16 neurons, organized on a 2-dimensional 4×4 point hexagonal grid. This means that the input data are clustered into 16 neurons, which represent the 16 biogeochemical provinces. In this case each weight of a neuron consists of a four element vector (SST, $\log(\text{MLD})$, SSS, $p\text{CO}_{2,clim}$), representing its coordinates. The distance between a neuron and an input vector is calculated via a distance function. These processing units are initially spread over a 2-dimensional field, in this case in a hexagonal formation, forming a single layer of neurons. Experience has shown, however,

that the choice of neuron topology does not have a significant effect on the final province distribution. The use of neurons, their initialization and their distance relation describes the biggest difference towards other clustering algorithms, e.g. k-means clustering (e.g. Wagstaff *et al.*, 2001). For both studies, the Euclidean distance between a neuron's weight vector and the input vectors of the SINP dataset was used for the distance function. The weight matrix ($\mathbf{W}_{m=16,n=4}$), which is formed by the 16 neurons with their 4 weight vector elements, was randomly initialized.

After initialization, the training vectors are introduced to the SOM with each training parameter as one element of the vector. For the j^{th} input vector \mathbf{p}_n^j with length n , the Euclidean distances to each of the $i=1,\dots,16$ neurons represented by each row i of the weight matrix are calculated:

$$d_i^j = \sqrt{\sum_{l=1}^n (\mathbf{W}_{i,l} - \mathbf{p}_l^j)^2} \quad (2.2)$$

\mathbf{d}_m^j comprises a vector containing the Euclidean distances to each neuron i of the input vector \mathbf{p}_n^j . The smallest element of the distance vector, i.e. the shortest distance element, marks the distance towards the closest neuron, called the winning neuron. The neuron i is updated by moving towards the average position of all the input vectors it was identified as a winner, or a close neighbour of the winner, hence neurons physically move in response to the input vectors (Demuth *et al.*, 2008). This sort of training is called "batch training". This is done by adjusting the i^{th} row of the initial weights matrix after the iterative step q following Vesanto *et al.* (2000):

$$\mathbf{W}_{i,n}(q+1) = \frac{\sum_{j=1}^r \mathbf{S}(d_{neighbour} - d_i^j) \cdot \mathbf{p}_n^j}{\sum_{j=1}^r \mathbf{S}(d_{neighbour} - d_i^j)} \quad (2.3)$$

where $\mathbf{W}_{i,n}$ is the i^{th} row of the input weight matrix, d_i^j is the Euclidean distance between the neuron i and the presented j^{th} input vector and $d_{neighbour}$ is the neighbourhood radius. \mathbf{S} describes the step function. Neighbouring neurons will only be updated if $\mathbf{S}(d_{neighbour} - d_i^j) > 0$. During the training of the SOM $d_{neighbour}$ decreases in two phases from an initial coarse training phase where the surrounding nearest neighbours are updated that lie

within a neighbourhood radius of 3, which lasts for the first 100 iterations, to a final fine training phase where only the winning neuron is updated.

After $q=200$ iterations (presenting all input vectors to the SOM 200 times), the training is stopped and the input vectors are re-introduced, without updating the SOM. In return every vector receives the neuron number i of the winning neuron, where $d_i^j = \min(\mathbf{d}_m^j)$, until every training vector is labelled with a number between 1 and 16, representing the province the vector belongs to. Since every input vector has a geographical location (see table 2.3) the global ocean can now be divided into 16 biogeochemical provinces.

As the four vector elements of the weight vector and the input vectors \mathbf{p}_n^j represent the coordinates of the map, a normalized set of input data would assure equal weighting of each input parameter (when Euclidean distances are computed). However, here the relative weights of the input data are forced toward the climatological sea surface $p\text{CO}_2$ data, in order to minimize the variance of $p\text{CO}_2$ within each biogeochemical province for the entire period. To do so, input data are not normalized or altered, with the exception of MLD, which was log-transformed (table 2.2). As a consequence, the range between the lowest and highest value of sea surface $p\text{CO}_2$ is one order of magnitude larger than that for SST, and about another order of magnitude larger than that for the remaining input parameters (log(MLD), SSS). This was done to reduce the misfit between estimates and observations in the second stage of the fitting, i.e. in the FFN. As a consequence, the biogeochemical provinces follow the seasonal pattern of the sea surface $p\text{CO}_2$ climatology, as the Euclidean distances between neuron weights and input vectors are largely dominated by one element, namely sea surface $p\text{CO}_2$, meaning that the seasonality of sea surface $p\text{CO}_2$ climatology at any given location will be mostly determined by the seasonal changes of the biogeochemical provinces and to a lesser degree by the seasonal cycle of the input data in the second stage of the fitting. In addition, owing to the climatological nature of the used Takahashi *et al.* (2009) sea surface $p\text{CO}_2$ data, there are few inter-annual shifts in the distribution of the biogeochemical provinces. The dynamics, i.e. the change in shape of the provinces from one month to the next and the much smaller changes from one year to the next of the provinces, forms the largest difference to conventional provinces or biomes (e.g. Longhurst *et al.*, 1995; McKinley *et al.*, 2011; Fay

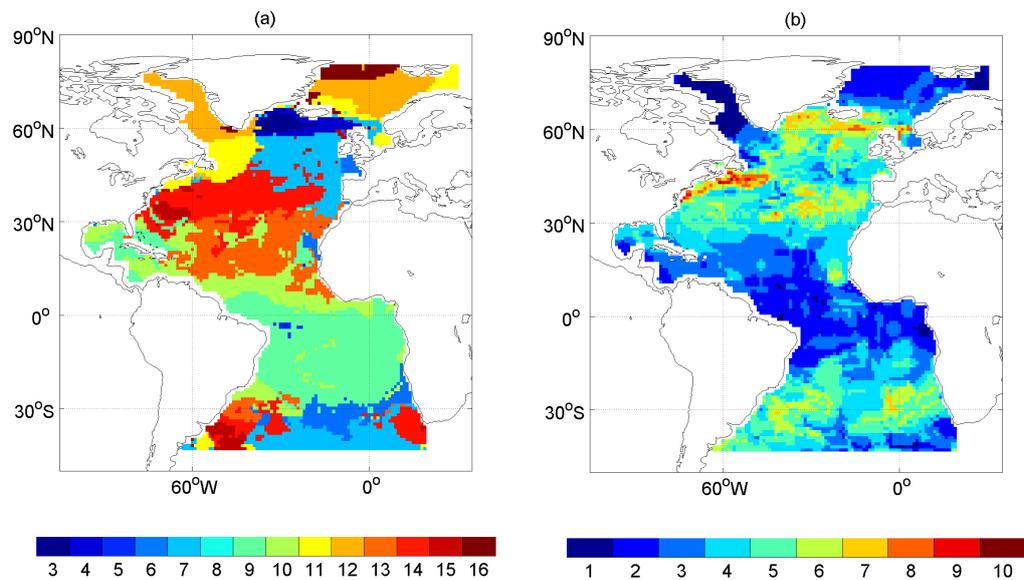


Figure 2.7: Map of the biogeochemical provinces in the Atlantic Ocean identified by the self-organizing map (SOM) method: (a) Province number of the mode (i.e., most frequent occurrence). (b) The number of provinces every pixel belongs to from 1998 to 2007.

and McKinley, 2013). No time or space information was introduced to the SOM, therefore these temporal variations emerge solely from the temporal variability of the monthly input data from 1998 through 2007 in the Atlantic Ocean and from 1998 through 2011 in the global ocean.

Figure 2.7 (a) shows the mode of the provinces of the Atlantic study (chapter 4), i.e., the province each pixel mainly belongs to from 1998–2007 and figure 2.7 (b) shows the number of shifting provinces per pixel. Despite their strong seasonal dynamics in space (figure 2.7 (a)) and time (figure 2.7 (b)), the estimated biogeochemical provinces exhibit a coherent large-scale behaviour, reflecting the well known oceanic structures such as the gyres, the equatorial regions, and the high-latitude North Atlantic. These provinces vary in time and space mainly in accordance with the variability of the climatological $p\text{CO}_2$. In the tropics, and the high latitude North Atlantic, the climatological $p\text{CO}_2$ vary little seasonally and therefore the provinces remain fairly steady, with only minimal province shifts. In contrast, the gyre regions of both hemispheres exhibit much larger seasonal variability, hence pixels there undergo many more province changes. The largest shifts occur along the Gulf Stream, where certain pixels change their province association up to 10 times.

Additionally, for the global ocean study (chapter 5), small "island" provinces with a surface area smaller than 10 connected pixels ($10 \times 1^\circ$ longitude $\times 1^\circ$ latitude grid boxes) of each month were removed and replaced by the mode of the surrounding provinces. Figure 2.8 again shows the mode of the resulting provinces and the number of shifting provinces for each pixel, during the full time period from 1998 through 2011 for the global ocean. Similar to the Atlantic Ocean regimes, regions with strong seasonal $p\text{CO}_2$ variability undergo stronger province shifts, resulting in larger numbers of provinces per pixel.

2.1.5 Reconstructing the sea surface $p\text{CO}_2$ using a feed–forward network

As a second step a feed–forward network (FFN) method is used to reconstruct the non–linear relationship between a set of input variables and corresponding targets, i.e., sea surface $p\text{CO}_2$, separately for each of the 16 biogeochemical provinces.

The FFN method is a type of backpropagation network method that is capable of approximating any function with a finite number of discontinuities (Demuth *et al.*, 2008). It consists of two layers, one network layer and one output layer (see figure 2.3). Similar to multi–linear regressions, a feed–forward network adjusts coefficients, i.e. the weights of every neuron in each network layer, to establish a relationship between inputs and targets. The adjustment of the weights follows an iterative process. The first iteration includes an initial guess, where the weights are randomly initialized, the estimates are computed and compared to the target observations (see figure 2.2). From there on the network goes backwards (hence the name backpropagation) and automatically re–adjusts the coefficients with the aim to reduce the mean squared error between estimates and targets. For each iteration, only a random subset of the data is used to train the network, while the remaining data are used for validation. The updating process of the coefficients is repeated until the network estimates derived from the validation set no longer improve significantly relative to the targets. The established relationship is then used to predict the $p\text{CO}_2$ for each point in time and space where no observations are available.

The feed–forward network was trained with the FINP dataset that included all input variables including their deseasonalised representation (see table 2.2). To this end the dataset was split into the 16 ocean provinces (FINP^k , $\text{FINP}2^k$, with $k=1, \dots, 16$) and each of

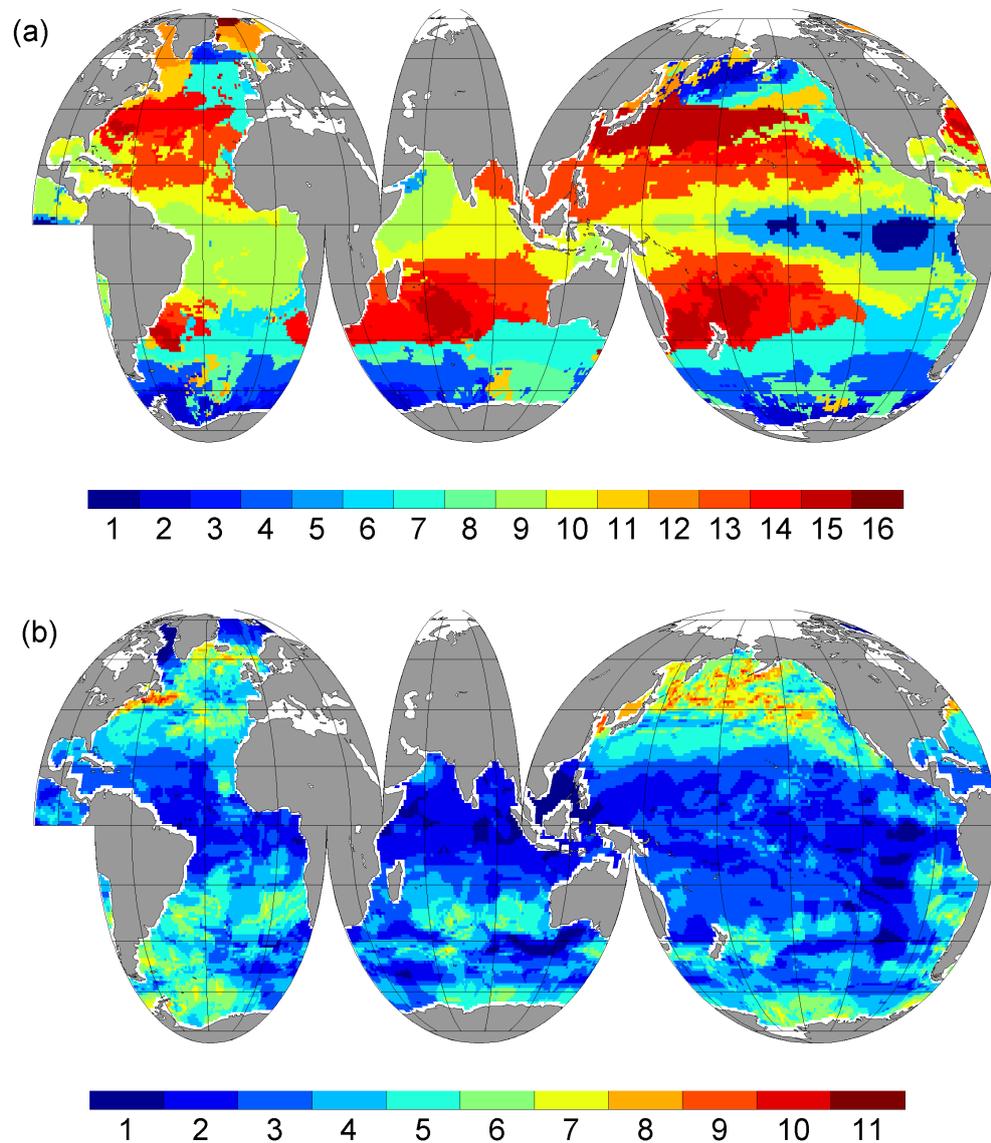


Figure 2.8: Map of the biogeochemical provinces in the global ocean identified by the Self-Organizing Map (SOM) method: (a) Province number of the mode (i.e., most frequent occurrence). (b) The number of provinces every pixel belongs to from 1998 to 2011.

them was processed separately. Due to the temporal and spatial variability of the provinces and the heterogeneous spatiotemporal distribution of the $p\text{CO}_2$ data, large differences exist in the number of observations within the different provinces. However, the neural network fit does not show degeneration as a function of the data density, as will be shown in chapter 4, since the temporal distribution and the spatial heterogeneity of the data does not lead to any major bias.

The feed–forward network uses two layers of neurons, one network layer of neurons (often referred to as “hidden” layer in the literature) using a tangents–sigmoid transfer function and one linear output layer. The response of the hidden layer to the input vector \mathbf{p}_n^j can be written as follows:

$$a_i = \frac{2}{1 + \exp(-2 \cdot (\mathbf{W}_{i,n} \cdot \mathbf{p}_n^j + b_i))} - 1 \quad (2.4)$$

Where a_i is the response of the hidden layer of i^{th} neuron. The response of all neurons m then forms the output vector of the hidden layer \mathbf{a}_m . This vector serves as input for the second layer of neurons, the linear output layer:

$$p\text{CO}_{2,est}^j = \mathbf{W}_m \cdot \mathbf{a}_m + b \quad (2.5)$$

Equation (2.4) and (2.5) show how the network calculates the scalar output $p\text{CO}_{2,est}^j$ for the j^{th} input vector \mathbf{p}_n^j in two steps, each step referring to one layer of the network. In the hidden layer the input vectors are multiplied with the weight matrix (weight vectors of each neuron organized in matrix form) of the hidden layer $\mathbf{W}_{m,n}$ and added to the layer bias vector \mathbf{b}_m . The output vector of the hidden layer \mathbf{a}_m is created using a tangents–sigmoid transfer function (equation 2.4), that computes elements of \mathbf{a}_m in the range from -1 to $+1$. Similar to the SOM, the size of $\mathbf{W}_{m,n}$ is determined by the size n of the input vector and the number m of neurons in the hidden layer. The length of \mathbf{b}_m and \mathbf{a}_m is as well determined by the number of neurons m . In the linear output layer \mathbf{a}_m , is processed the same way as \mathbf{p}_n^j in the hidden layer, with the exception that the output layer only consists of one neuron to produce one scalar $p\text{CO}_2$ estimate ($p\text{CO}_{2,est}^j$) for every input vector. Furthermore, the linear output layer allows $p\text{CO}_{2,est}^j$ to have any value between $-\textit{infinity}$ and $+\textit{infinity}$. During the training the weights and biases of each layer get iteratively adjusted to minimize the error between the network output $p\text{CO}_{2,est}^j$ and the scalar target element t^j from the corresponding SOCAT database that corresponds to \mathbf{p}_n^j . Therefore, the network can only be trained by those input vectors which do have co–located $p\text{CO}_2$ observations.

Before the training starts, the training vectors with corresponding observations from the FINP^k dataset are provided to initialize the network layer size. The size of the two network layers is determined by the number of neurons and the number of input vector elements n . The networks initial weights and biases are randomly generated. An important parameter that has to be provided before training starts is the number of neurons for the hidden layer. Too few neurons are not able to reproduce realistic results, whereas too many neurons decrease the computational performance and cause over-fitting and therefore the network is not able to generalize (Demuth *et al.*, 2008). Since the number of inputs and targets varies per province, one best number of neurons can not be provided to be used for all 16 provinces. Therefore a pre-training is performed, increasing the number of neurons in the hidden layer parabolically starting from two neurons up to a number where the ratio between of the training sample size (number of input vectors) to the number of weights does exceed 30. Amari *et al.* (1997) proposed this ratio to prevent artificial neural networks from over-fitting.

During every pre-training process the FINP^k set is divided into two independent subsets. The first (FITR^k) is used to train the network and the second (FIVAL^k) is used for validation within the method. Amari *et al.* (1997) suggested an optimal split (r_{opt}) between training and validation data as a function of the modifiable parameters h :

$$r_{opt} = \frac{1}{\sqrt{2 \cdot h}} \quad (2.6)$$

Modifiable parameter refers to the weights and biases of the network. During every pre-training process the FITR^k training vectors and the corresponding FITR^k targets are introduced to the network and the weights and biases are iteratively updated in the direction where the performance function, which is the mean squared error between network outputs $p\text{CO}_{2,est}^j$ and FITR^k targets t^j , decreases most rapidly. The FFN uses the Levenberg-Marquardt (Marquardt, 1963) algorithm, which is a quasi Newton method, to update weights and biases in every iteration step with the aim to reduce the mean squared error between outputs and targets. This can be written as follows:

$$\mathbf{x}_{q+1} = \mathbf{x}_q - [\mathbf{J}^T \cdot \mathbf{J} + \mu \cdot \mathbf{I}]^{-1} \cdot \mathbf{J}^T \cdot \mathbf{e} \quad (2.7)$$

The variable \mathbf{x}_q describes a vector of the current weights and biases, \mathbf{I} is the identity matrix, \mathbf{J} is the Jacobian matrix that contains first derivatives of the network errors with respect to the weights and biases and \mathbf{e} describes a vector containing the network errors. The scalar μ decreases after every successful step and increases after an increase of the performance function (Demuth *et al.*, 2008). The application of the algorithm in neural networks is described in more detail in (Hagan and Menhaj, 1994; Hagan *et al.*, 1996).

After every iteration of each pre-training, the network is validated by using the FIVAL^k sub-set. The updated weights and biases are used to simulate outputs from the FIVAL^k inputs and the mean squared error between these outputs and the FIVAL^k targets is calculated. Every pre-training of the network stops automatically when six consecutive iterations do not reduce the network's error on the FIVAL^k targets to prevent the network from over-fitting. After the pre-trainings with increasing number of neurons the one where the mean squared error of the validation data set FIVAL^k is a minimum is selected within each province and receives the optimal number of neurons for the actual training process.

During the actual training process the number of neurons is adjusted according to the best pre-training performance for each of the 16 provinces separately. Ten trainings are performed, where validation data are randomly picked, according to equation (2.6), out of the entire pool of observations to avoid over-fitting of the network output. After every training the trained network is used to simulate $p\text{CO}_{2,est}^j$ from the FINP^k dataset and the output is the average of the 10 training cycles, to end up with one monthly estimate for the time period 1998 through 2007 in the Atlantic Ocean study and 1998 through 2011 in the global ocean study for each province. After 16 FFN runs the results of the 16 provinces are combined to retrieve the $p\text{CO}_2$ estimates on a monthly global $1^\circ \times 1^\circ$ grid, which will be presented in chapter 4 and chapter 5.

2.1.6 Air-sea CO_2 flux calculation

The estimated $p\text{CO}_2$ fields are further used to calculate the air-sea flux density in $\text{mol C} \cdot \text{m}^{-2} \cdot \text{yr}^{-1}$ for each month and $1^\circ \times 1^\circ$ pixel from:

$$F_{\text{CO}_2} = -k_w \cdot K_0 \cdot (1 - f_{ice}) \cdot (p\text{CO}_{2,atm} - p\text{CO}_2) \quad (2.8)$$

where K_0 is the mainly temperature driven solubility of CO_2 (calculated in $\text{mol C} \cdot \text{m}^{-3} \cdot \mu\text{atm}^{-1}$) and k_w is the gas transfer velocity (calculated in $\text{m} \cdot \text{yr}^{-1}$) introduced in chapter 1 of this thesis. The air–sea flux results will be presented in chapter 4 for the Atlantic Ocean and in chapter 5 for the global ocean. Throughout this thesis the flux is defined positive upward, i.e., outgassing of CO_2 from the ocean to the atmosphere is positive, and ocean CO_2 uptake from the atmosphere is negative. f_{ice} refers to the percent of ice cover within a region derived from Rayner *et al.* (2003). For the gas transfer velocity (here calculated in $\text{cm} \cdot \text{hr}^{-1}$) the formulation of Wanninkhof (1992) is used with the scaling factor of Sweeney *et al.* (2007), i.e.:

$$k_w = 0.27 \cdot (\text{Sc}/660)^{-\frac{1}{2}} \cdot u^2 \quad (2.9)$$

where Sc the dimensionless Schmidt number and u the monthly mean Cross–Calibrated Multi–Platform (CCMP) wind speed (Atlas *et al.*, 2011) at a height of 10 meters above the sea surface.

The solubility of CO_2 is calculated according to Weiss (1974) and the Schmidt number according to Wanninkhof (1992) using the same SST and SSS input data used for the training of the neural network.

The partial pressure of atmospheric CO_2 , i.e., $p\text{CO}_{2,atm}$, is computed from the dry air mixing ratio $x\text{CO}_2$ of GLOBALVIEW-CO2 (2011), taking into account the water vapor correction according to Dickson *et al.* (2007):

$$p\text{CO}_{2,atm} = x\text{CO}_{2,atm} \cdot (P_{atm}^{surf} - P_{\text{H}_2\text{O}}) \quad (2.10)$$

where P_{atm}^{surf} is the sea–level pressure from NCEP (Kalnay *et al.*, 1996), and $P_{\text{H}_2\text{O}}$ describes the water vapor pressure.

2.2 Critical reflection on the method

Compared to conventional methods, the presented method in this chapter is capable of reconstructing small amounts of noise within the sea surface $p\text{CO}_2$ data, due to its non-linear input-target data relationship, hence it qualifies to investigate the inter-annual variability of the sea surface $p\text{CO}_2$ which is often buried under much larger signals, such as the seasonal cycle. The 2-step approach further provides estimates on sparse observation data regions, by applying relationships learned from similar ocean regions, where larger amounts of data exist.

However, like with any other method based on statistics, the neural network reconstructs a relationship based on the provided input data, which are subjectively chosen. This implies that the drivers for the seasonal and inter-annual variability, as well as trends need to be well known and only their exact relationships remain uncertain or unknown. For example, atmospheric CO_2 drives the trend signal in the global ocean. This atmospheric CO_2 signal is rather linear over the considered time periods in this thesis, hence any other linearly increasing or decreasing input parameter, e.g. calendar date or number of pirates in the global ocean, that would replace the atmospheric CO_2 as an input parameter, might result in the same sea surface $p\text{CO}_2$ estimate, although the established relationship is based on a completely wrong assumption.

Moreover, uncertainties exist in the chosen input data, in particular those derived from model estimates constraint by in-situ observations (e.g. MLD). These uncertainties, however, do not necessary reflect the uncertainty of the final $p\text{CO}_2$ estimate, as the variability of the $p\text{CO}_2$ is entirely determined by those input parameter that are correlated to the available target observations. The effect of different input data will be further tested in chapter 4 by sensitivity runs SR1-SR4.

Furthermore, many of the input data that drive the sea surface $p\text{CO}_2$ distribution and its variability are correlated with each other, hence occasionally, only a subset of input data are needed in order to establish a statistical relationship valid to reconstruct the sea surface $p\text{CO}_2$ observations within the SOCAT database, although this does not imply that the input data of the chosen subset are the ultimate drivers. This inter-correlation make it further difficult to quantify the contribution of the $p\text{CO}_2$ change of each input parameter, hence the method presented does not give a direct quantification of the involved processes.

Previous research has focused on the exact relationship between input data and the sea surface $p\text{CO}_2$ (Takahashi *et al.*, 1993). As a result, mechanistic understanding exists to-date regarding the $p\text{CO}_2$ –SST relationship (see section 1.3). The method presented here does not assume any prior knowledge between input data and target $p\text{CO}_2$, hence the method may be improved by implementing the well known $p\text{CO}_2$ –SST relationship (e.g. estimating changes in $p\text{CO}_2$ as a combination of the $p\text{CO}_2$ –SST relationship with a neural network based estimate of the non temperature driven changes) and is a suggestion for future work.

Data sparse regions are estimated from the relationship derived from the available observations within the same province. The similarity of the input–target relationship within one province is another assumption. The validation approach using the FIVAL dataset within the FFN provides a first control mechanism to help the network generalize a relationship and to not over–fit the available observations within a province. However, further validation of the neural network outputs it needed and a thorough assessment with independent data needs to be conducted where available.

In conclusion, the neural network approach is only as valid as the chosen input and target data and it can only be truly validated where observations exist, hence a large part of this thesis will be assigned to validating the final $p\text{CO}_2$ estimates in the Atlantic Ocean (chapter 4) and the global ocean (chapter 5).

Chapter 3

$p\text{CO}_2$ sampling and processes in the North Atlantic

”It doesn’t matter how beautiful your theory is,
it doesn’t matter how smart you are.
If it doesn’t agree with the experiment, it’s wrong.”

(Richard P. Feynman)

In this chapter underway measurements of the sea surface $p\text{CO}_2$ as well as discrete bottle samples obtained along the UK–Caribbean ship route are introduced and the carbonate system parameter along the line are investigated. This chapter further provides an overview on the measurement techniques and the quality control of in–situ data. Since 2002, underway measurements obtained along the line are used within the SOCAT database, which forms the basis of the neural network technique introduced in chapter 2. Discrete bottle samples, collected during four field campaigns, are used to investigate the variability and the drivers of the sea surface $p\text{CO}_2$ along the UK–Caribbean line.

The student's contribution to the work presented in this chapter includes the support of the data PI Ute Schuster in the regular maintenance of the underway measurement system on–board M/V Benguela Stream (section 3.1.1), the regular quality control of the data (section 3.1.2) as well as laboratory–based testing of the equipment. The planning and preparation process for four individual voyages, the sampling of discrete bottle data (section 3.2) on–board M/V Benguela Stream, their analysis in the laboratory (section 3.2.1), as well as the quality control of the data (section 3.2.2) was carried out as shared work by the student and Clare Ostle, with great support by Oliver Legge and the technical staff of the University of East Anglia. All the analysis of the data and the interpretation of the results, presented from section 3.2.3 onwards, was solely conducted by the student. Although the work presented here fits only one chapter of this entire thesis, it has to be noted that a much more significant amount of time spent within this PhD project was dedicated to this work.

3.1 Underway measurements of the sea surface $p\text{CO}_2$

Since 2002 the University of East Anglia (UEA) has been measuring the underway sea surface $p\text{CO}_2$ from Portsmouth, UK, to the Caribbean using an automated equilibration system (Schuster and Watson, 2007), developed at UEA, funded by the European Commission (EC) projects CAVASSOO, CARBOOCEAN and CARBOCHANGE. The instrument is currently mounted on the M/V Benguela Stream, a reefer vessel participating in the Voluntary Observing Ship (VOS) program, and is set up to measure sea surface carbon dioxide, marine air CO_2 , SST and SSS to obtain a high number of observations at a high accuracy. A similar system has been successfully operated along this route between June

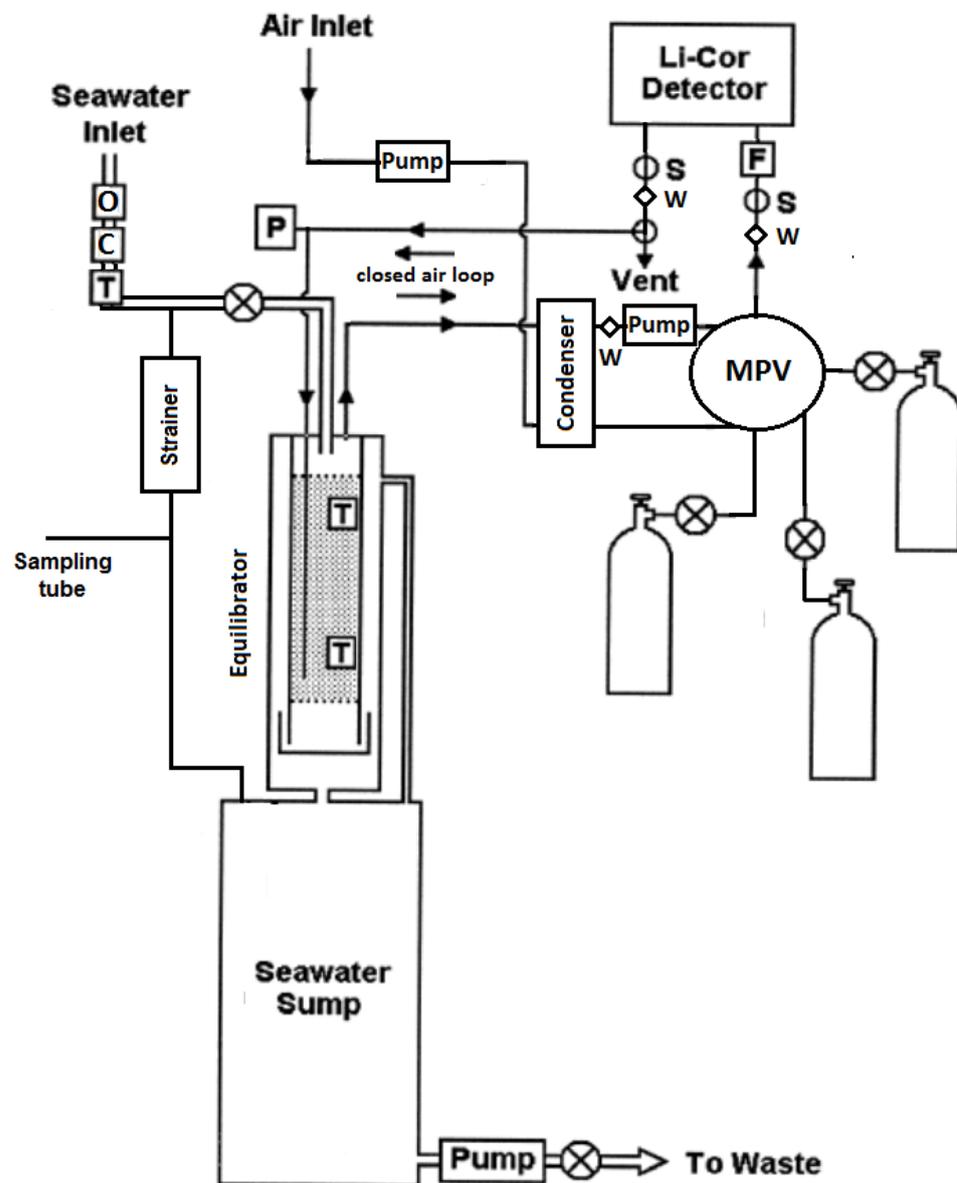


Figure 3.1: Schematic of the underway measurement system updated from Cooper *et al.* (1998). Sensors are labelled with T for temperature, C for conductivity, O for oxygen, P for pressure and F for mass flow controller. The labels MPV, W and S refer to multi position valve, water watchers and solenoid valves, respectively.

1994 through August 1995 (Cooper *et al.*, 1998) and from 2002 onwards a large number of underway sea surface $p\text{CO}_2$ observations in the North Atlantic subtropics are collected from vessels along this shipping route using a similar measurement set-up, with slight modifications over time.

3.1.1 Measurement system and maintenance

Figure 3.1 shows the schematic of the measurement system as illustrated in Cooper *et al.* (1998) but modified to the most recent set-up mounted on M/V Benguela Stream. Seawater carbon dioxide ($\text{CO}_{2,w}$) can not be directly measured, hence it has to be equilibrated with a gas first (see equation 1.19). This is done in a specifically designed equilibrator (Cooper *et al.*, 1998), which consists of a gas tight tube filled with raschig rings (hatched area in figure 3.1). The equilibrator is currently located below sea level, hence no pump is in use, as the regular water pressure is sufficient for a constant stream of seawater from the inlet. While seawater percolates from the top to the bottom of the equilibrator, the raschig rings provide a large surface area for a rapid exchange of CO_2 between the seawater and air, the latter circulated in a closed loop, passing the equilibrator from the bottom to the top. The air equilibrates with the seawater and is then circulated using a diaphragm pump through a LiCor 7000 infra-red gas analyser, where both the mole fraction of CO_2 ($x\text{CO}_{2(gas)}$) and moisture of the air is measured. Seawater that passes through the equilibrator ends up in a sump tank and gets further pumped back to sea.

Inside the equilibrator two Pt100 temperature sensors are mounted to measure both the temperature on the top and at the bottom of the equilibrator. Furthermore a pressure sensor is mounted on the top of the equilibrator. Both temperature and pressure are measured to convert $x\text{CO}_2$ to carbon dioxide in partial pressure units ($p\text{CO}_2$) and to correct for temperature changes between the seawater inlet and the equilibrator (see e.g. Dickson *et al.*, 2007).

At the seawater inlet of the ship, which leads to the equilibrator, digital Aanderaa sensors are located, namely a temperature sensor 4050, an oxygen optode sensor 3835 and a conductivity sensor 3919, mounted in a sensor housing. These digital sensors are connected to a data-logger next to the electronics cabinet, which converts the sensor reading into the desired units. Furthermore, marine air is pumped from the port side bridge wing of the ship to the LiCor 7000 system. Both marine air and air from the equilibrator pass a condenser. This is a control mechanism to avoid water vapour to condense in the LiCor 7000 gas analyser and therefore damage the unit. The instrument includes three water watchers which shut down the instrument in case of water approaching the analyser. Furthermore, a mass flow controller is positioned before the LiCor 7000 gas analyser which

regulates the air flow from the equilibrator, the marine air line and three gas standards to $100 \text{ ml} \cdot \text{min}^{-1}$.

To calibrate the LiCor, three 10 litre standard gases are currently used on-board. The standards contain a known mixing ratio of carbon dioxide, i.e., roughly 250 parts per million (ppm), 350 ppm and 450 ppm, respectively. These standards are pre- and post-calibrated in the laboratory, against highly accurate NOAA gas standards, to determine the correct mixing ratio before they are applied on-board. The aim of the standards is to calibrate the seawater derived $x\text{CO}_2$ measurements.

The LiCor 7000 unit, the Aanderaa data logger and the electronics cabinet are connected to an instrument computer. The measurement software on the computer regulates the measurement procedures and automatically records the readings of all sensors, and stores them in a data file. The on-board measurement system is fully automated, hence no scientist needs to operate the system. The measurement system works in a self-repeating 3-step cycle, repeatedly measuring air from the equilibrator, marine air and gas standards.

A complete round trip of the ship takes four weeks, which includes two Atlantic crossings, one southbound voyage from Le Havre, France ($49^\circ\text{N } 29'$ and $0^\circ\text{E } 6'$) leading usually south of the Azores through the subtropical gyre to the southern Caribbean islands ($\sim 10\text{--}12^\circ\text{N}$ and $\sim 60^\circ\text{W}$), and one northbound return voyage from Manzanillo in the Dominican Republic ($19^\circ\text{N } 42'$ and $71^\circ\text{W } 44'$) back to Portsmouth, UK ($50^\circ\text{N } 44'$ and $1^\circ\text{W } 5'$). However, ship routes depend on weather, current and sea-state conditions and Caribbean arrival ports vary occasionally.

After arrival of the ship in Portsmouth, an on-board maintenance procedure is carried out. Firstly, data from the last voyage are downloaded and analysed to see if there have been technical problems which need immediate attention. Further on, the standard gas cylinder pressures and the amount of water in the condenser are checked and a pressure leak test is carried out for the equilibrator air loop, to check if the line is gas tight. Afterwards, the Aanderaa sensors in the seawater housing and, if necessary, the equilibrator are cleaned. Next, the analogue pressure and temperature sensors are calibrated. The Pt100 sensors are calibrated against the Aanderaa temperature sensor in three different temperatures, namely in warm water, cool water and melting ice. At last, the standard measurement routine is started before the ship leaves for the next voyage.

3.1.2 Quality control

Before the final data product is derived, a series of quality checks need to be performed, to determine the robustness of the measurements. In a first step temperature sensor readings of a voyage are compared. The seawater line leading from the seawater inlet to the equilibrator is well insulated, hence a good calibration of the sensors will lead to fairly similar readings between the Aanderaa sensor at the seawater inlet and the two Pt100 sensors in the equilibrator. Therefore, differences between the seawater inlet and equilibrator readings might indicate a problem with the water flow.

Due to the mass flow controller, gas flow rates are supposed to be fairly constant at $100 \text{ ml} \cdot \text{min}^{-1}$. However, if flow rates of the equilibrator air loop or the marine air supply drop, this might indicate problems with the air pumps in either loop. Dropping gas standard flow rates might indicate a leak in the gas line or problems with the regulator settings.

The LiCor 7000 gas analyser further consists of a pressure sensor, whose readings are compared with those from the equilibrator pressure sensor. Furthermore, the ship crew reports meteorological data, measured on the bridge, including atmospheric pressure for weather forecasts (<http://www.sailwx.info/>), hence these data can be further used to check the precision of the pressure sensor readings.

In a next step the gas standard readings are analysed. Due to the strong ambient temperature difference at the instrument between the port of Portsmouth and the Caribbean of up to 30°C , the moisture content of the air passing through the LiCor 7000 changes for both marine air and air passing the equilibrator, but they should be zero for the three gas standards. Within roughly 30 minutes of gas standard measurement time, both moisture and CO_2 content are expected to be steady.

Change of water masses or biological activity can cause steep gradients in the seawater $x\text{CO}_2$ readings. Salinity (calculated from conductivity) and oxygen provide indicators if such a steep gradient is caused by water mass changes or by biological activity. Additionally, the crew on-board collects nutrient and salinity bottle samples roughly every four and 12 hours, respectively, while the ship is at sea.

Atmospheric CO_2 is in general fairly constant throughout an Atlantic Ocean crossing,

hence strong variability within a short time period might indicate wind driven contermination of marine air through the ship's stack. Meteorological data (<http://www.sailwx.info/>), as well as ship speed and direction are used to identify ship contermination.

The quality controlled data have been used in several studies (e.g. Cooper *et al.*, 1998; Schuster and Watson, 2007; Schuster *et al.*, 2009b; Watson *et al.*, 2009) and are included in the major sea surface $p\text{CO}_2$ and $f\text{CO}_2$ databases, namely the Lamont–Doherty Earth Observatory (LDEO) database (Takahashi *et al.*, 2013) and the Surface Ocean CO_2 Atlas (SOCAT) database (Pfeil *et al.*, 2013; Sabine *et al.*, 2013; Bakker *et al.*, 2014).

3.2 Collection of discrete bottle samples along the UK–Caribbean line

In addition to the automated measurements a total of four accompanied field campaigns were carried out in the North Atlantic Ocean on-board M/V Benguela Stream in April/May 2012 (BS56 – Peter Landschützer on-board), June/July 2012 (BS58 – Clare Ostle on-board), September/October 2012 (BS62 – Clare Ostle on-board) and January/February 2013 (BS66 – Peter Landschützer on-board), each voyage representing one of the four seasons. A total of 325 bottle samples were collected with, on average, 40 seawater bottle samples per crossing during the day in intervals of about 2 hours. Figure 3.2 shows the ship tracks for each voyage split up into southbound and northbound voyages. Samples were taken from a bypass line of the main seawater line towards the equilibrator, located right after the sensor housing (see figure 3.1). At this bypass line a t-piece splits the bypass line into two lines. The first, a water hose leading directly to the sump tank underneath the equilibrator, and the second, a roughly one meter long piece of tubing, which was used to fill up the bottles.

Right before the t-piece a strainer was located to filter small particles. This strainer was cleaned once a day to prevent organic material to be captured in the bottles. The first sample of the day was collected in a 500 ml bottle, used as replicate sample to later assess the quality of the analysis in the laboratory. All other samples were collected in 250 ml bottles.

Before taking the samples, all bubbles appearing in the sample tubing were removed and the bottles were properly rinsed. The bottles were filled by placing the end of the

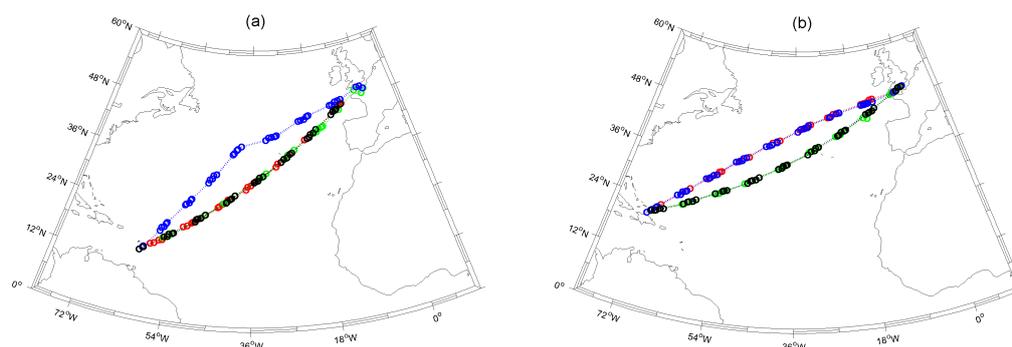


Figure 3.2: Map of the ship tracks of the (a) southbound and (b) northbound UK–Caribbean line in months where bottle samples were taken. Red colours refer to the Spring voyage in April/May 2012, Green colours refer to the Summer voyage in June/July 2012, Blue colours refer to the Autumn voyage in September/October 2012 and black colours refer to the Winter voyage in January/February 2013. Circles indicate where bottle samples were taken along the line.

tygon tubing at bottom of the bottles and then filled by overflowing them with about four times the bottle's volume, according to the standard operating procedure (SOP) given by Dickson *et al.* (2007).

After the bottles were filled, the stopper was placed on the bottles to avoid gas exchange and water contamination while moving from the sampling site to the fixing station. Immediately after collecting the samples, temperature and salinity of the seawater were recorded, as well as time, date and geographical position of the ship. The samples were then fixed on-board by first removing 1% by volume of the seawater bottle and then adding 50 μl of saturated mercuric (II) chloride (HgCl_2) in every 250 ml bottle and 100 μl in every 500 ml bottle, following the SOP by Dickson *et al.* (2007). Mercuric (II) chloride prevents changes of inorganic carbon to organic carbon and vice versa while the samples are stored. The stopper was wiped with a clean tissue and greased to seal the bottle to prevent the exchange of gases while being stored in a dark place on the ship.

3.2.1 Analysis of discrete bottle samples

The samples were analysed within two laboratory campaigns, one after the first two voyages and the second one at the end of the field campaign. Two VINDTA 3C (Versatile INstrument for the Determination of Total inorganic carbon and titration Alkalinity) instruments were used simultaneously to determine the TALK concentration via titration

and the DIC concentration via coulometry (Mintrop, 2011). The VINDTA 3C system operates at an accuracy of $\pm 1 \mu\text{mol} \cdot \text{kg}^{-1}$ for both TALK and DIC when analysing open ocean water samples (Mintrop *et al.*, 2000). Nutrients, for the calculation of TALK, were provided from the samples taken on-board and interpolated to the actual DIC and TALK measurement points.

To analyse the TALK content of the samples, the bottles are opened and maintained in a water bath at a temperature of $\sim 25^\circ\text{C}$. A known volume of a sample is then dispensed into a water-jacked cell, which is as well maintained at a temperature of $\sim 25^\circ\text{C}$. In the cell, 150 μl increments of hydrochloric acid are titrated into the sample using a piston burette.

The titration is stopped after the total amount of acid, added to the sample in the cell reaches a total of 4.2 ml. The titration process is monitored by a Titrimo pH meter (readable to 0.1 mV), which uses a proton sensitive electrode, located within the cell. The TALK calculation is done both graphically and mathematically (Schuster *et al.*, 2009a; Mintrop, 2011) by plotting the volume of acid against the electromotive force measured by the pH meter. The resulting titration curve shows two inflection points, namely the protonation of carbonate and bicarbonate, and the acid consumption up to the second point is equal to the titration alkalinity.

The DIC concentration is measured by dispensing a known volume into a stripping chamber, where the sample is acidified with 8.5% phosphoric acid. This is done to convert all carbonate species to CO_2 gas. Pure nitrogen (N_2) is bubbled through a fine frit at the bottom of the stripping chamber, to ensure that the solution is stripped of CO_2 . The nitrogen gas acts as a carrier gas for CO_2 , hence to remove all CO_2 before entering the stripping chamber, it passes through a CO_2 absorbent. The CO_2 enriched gas then passes a Peltier cooling system to condense water vapour before it reaches a coulometer cell.

The coulometer cell consists of a main body and a side arm. In the main body a cathode solution, consisting of a mixture of water, tetra-ethyl-ammonium bromide, ethanolamine, dimethylsulfoxide and thymolphthalein indicator, is absorbing CO_2 in the gas stream. The side arm contains of an anode solution consisting of a mixture of saturated potassium iodide in water and dimethylsulfoxide (Dickson *et al.*, 2007). In the main body, CO_2 reacts with ethanolamine to produce hydroxyethylcarbamic acid, which

causes a change in pH. As a result of this reaction, a color change occurs, which is measured via transmittance at ~ 610 nm (Dickson *et al.*, 2007). The measurement system aims to maintain the transmittance at a constant level of 29%, hence hydroxide ions are generated at a platinum cathode by electrolyzing water. The required electrons for this process are generated at a silver anode of the side arm. This generates a current which is related by the Faraday constant to the moles of CO_2 absorbed by the solution (Johnson *et al.*, 1993). When the coulometer readings drop below a set endpoint of 50 counts \cdot minute $^{-1}$ the measurement system stops, sample results are recorded and DIC can be calculated following Dickson *et al.* (2007) from:

$$\text{DIC} = \frac{N_S - b \cdot t - a}{c} \cdot \frac{1}{V_S \cdot \rho} \quad (3.1)$$

Where N_S is the coulometer reading in counts for the sea water sample, a is the acid blank in counts, b is the background level of the system in counts \cdot min $^{-1}$, c is the coulometer calibration factor in counts \cdot mol $^{-1}$, t is the time required to measure the water sample in minutes, V_S is the volume of the sea water sample at the temperature of use in dm 3 and ρ is the density of the seawater sample in g \cdot cm $^{-3}$.

For the most accurate work, two further minor corrections may need to be made to compute the total DIC in the original seawater sample, namely for the dilution by mercuric chloride when the sample was collected and for the exchange of CO_2 with the headspace in the sampling bottle.

3.2.2 Quality control

In a first stage, the quality of the VINDTA 3C procedure needs to be assessed. This is done in two ways. Firstly, the accuracy is assessed by using certified reference material (CRM) standards from the Scripps Oceanographic Institute, San Diego, USA, and secondly the precision is assessed by using the 500 ml bottles that were taken as the first sample every day at sea. CRMs are highly accurate 500 ml standards with a known concentration of DIC and TALK. During the laboratory analysis, CRM standards are analysed as the first and the last sample every day. After the laboratory analysis is complete, the measured DIC and TALK concentrations of the CRMs are compared to the certified concentration

of the standards and if necessary, for the TALK analysis, the recorded acid concentration is corrected to reduce the offset between the measured and the certified concentration of the CRMs. This acid correction is then applied to all TALK samples analysed with the same VINDTA 3C.

Secondly 500 ml sample bottles, or replicate bottles, are used to assess the quality of the measurements. For the VINDTA 3C analysis, a sample volume of 250 ml is needed to derive the DIC and TALK concentration, hence 500 ml bottles can be used for in-bottle replicates. The measured TALK concentrations are supposed to be within the uncertainty provided by the VINDTA 3C measurement system ($\pm 1 \mu\text{mol} \cdot \text{kg}^{-1}$), hence larger differences between replicates indicate bad quality of the data. In this case, no further samples are analysed and 500 ml seawater junk samples are used instead to assess when the replicate offset is back within the provided uncertainty.

After completing the analysis, the resulting DIC and TALK concentrations are quality controlled using the World Ocean Circulation Experiment (WOCE) flagging system. Every measurement receives a number, namely 2 for good data, 3 for probably bad data and 4 for bad data. Additionally, a flag 9 for missing data was introduced, which is not in line with the WOCE flagging system. Overall, six DIC data points with a flag 4 and one DIC data point with a flag 9 were assigned. Four TALK data points received a flag 4 and three more data points were assigned a flag 9. All other data points were assigned a flag 2 and were considered for further analysis.

Finally, the error of the DIC and TALK analysis procedure was assessed by the standard deviation of all CRM bottles used resulting in a DIC uncertainty of $\pm 2.55 \mu\text{mol} \cdot \text{kg}^{-1}$ and a TALK uncertainty of $\pm 1.46 \mu\text{mol} \cdot \text{kg}^{-1}$, which is slightly larger than the reported uncertainty of the VINDTA 3C measurement system.

3.2.3 Comparison with independent data products

To-date two databases provide the largest sets of DIC and TALK data measured in the Atlantic Ocean, namely the GLODAP v1.1 (Global Ocean Data Analysis Project, Key *et al.* (2004)) database and the CARINA (Carbon in the Atlantic Ocean, CARINA-Group (2009)) database. While the CARINA database includes DIC and TALK data, measured between 1977 and 2006, GLODAP v1.1 Atlantic Ocean data are available between 1972

and 1998. Both databases include measurements at several depths. Besides the temporal offset between the data within the databases and the bottle data sampled onboard M/V Benguela Stream, a comparison (not shown here) reveals, that there is not a single data point, measured within the same month (independent from the sample year) that has a spatial difference less than 1°latitude and 1°longitude with the Benguela Stream samples.

In order to compare all bottle data that have received a WOCE flag of 2, available climatological products are used, which are based on the above databases and the LDEO surface ocean carbonate parameter dataset. In particular, the recently released monthly gridded DIC and TALK climatologies, normalized to the year 2005, of Takahashi and Sutherland (2013), available from Biological and Chemical Oceanography Data Management Office (<http://www.bco-dmo.org/dataset/3961>) are used. The comparison is illustrated in figure 3.3.

The climatological product includes data from the global GLODAP database which consists of 122 WOCE, Joint Global Ocean Flux Study (JGOFS), NOAA Ocean – Atmosphere Exchange Study (OACES) and other international and historical cruises conducted in the 1990s. The aim of GLODAP was to create a high quality global dataset to determine the global distributions of both natural and anthropogenic inorganic carbon. GLODAP was made available for the public via the Carbon Dioxide Information Analysis Center (<http://cdiac.ornl.gov/oceans/glodap/>). The Takahashi and Sutherland (2013) climatology product further includes about 4800 DIC and TALK data from the CARINA database (CARINA-Group, 2009) and about 2600 $p\text{CO}_2$ data from the LDEO database. The monthly DIC and TALK climatologies are presented on the same $4^\circ \times 5^\circ$ grid as the Takahashi *et al.* (2009) $p\text{CO}_2$ climatology. For the comparison of the sea surface DIC and TALK in figure 3.3 (a) and (b), data points of the climatological surface layer were co-located to the nearest point where bottle samples were collected.

The mean difference between the sampled DIC data and the climatological data (figure 3.3 (a)) is $-3.44 \mu\text{mol} \cdot \text{kg}^{-1}$, with a RMSE of $14.77 \mu\text{mol} \cdot \text{kg}^{-1}$, with the regression line closely following the optimal red regression line. Figure 3.3 (b) shows a similar match between bottle samples and the climatological TALK, with the regression line closely following the optimal red regression line. The mean difference is $4.76 \mu\text{mol} \cdot \text{kg}^{-1}$ and the RMSE is $14.36 \mu\text{mol} \cdot \text{kg}^{-1}$ between the independent TALK products.

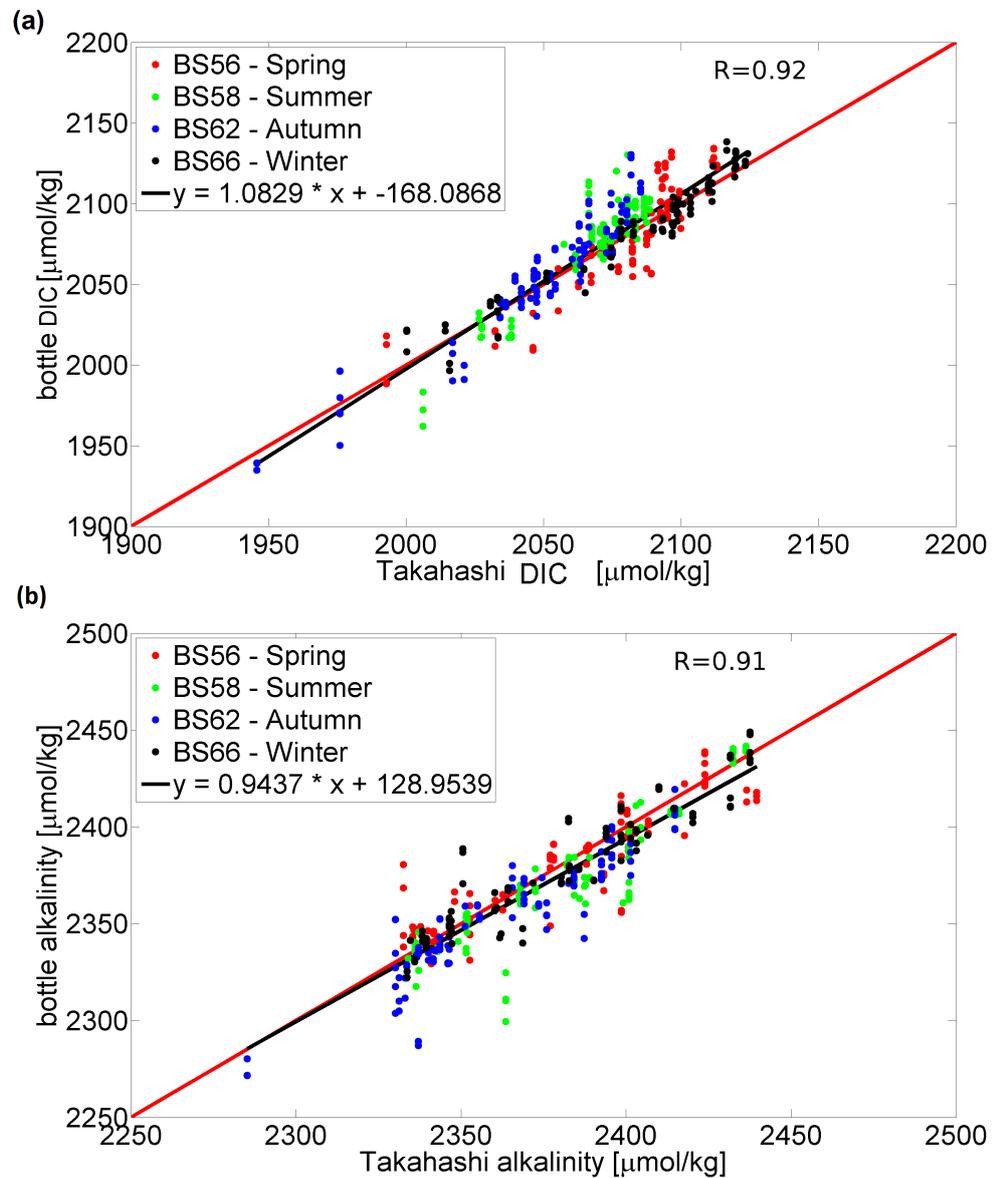


Figure 3.3: Analysed DIC and TALK data along the UK–Caribbean line that received a WOCE flag of 2 compared to climatologies of Takahashi and Sutherland (2013), available from Biological and Chemical Oceanography Data Management Office (<http://www.bco-dmo.org/dataset/3961>) for (a) DIC and (b) TALK. Colours refer to the different seasons when the measurements were taken, following the color code of figure 3.2, namely red for spring, green for summer, blue for autumn and black for winter. The black lines show the regression lines of all data points combined and the red lines show the optimal one–on–one regression line.

It needs to be noted here, that the comparison with climatologies presented in this section is far from ideal, given their temporal offset and their coarse resolution. An updated version of GLODAP database (GLODAPv2) is currently in progress, due to be published in February 2014.

3.2.4 Variability of the sea surface DIC and TALK along the UK–Caribbean line

Figure 3.4 shows the DIC and TALK data along the UK–Caribbean line that received a WOCE flag of 2 split up into the four seasons. Towards the most western longitudes, DIC drops to its lowest values along both transects (see figure 3.4 (a) and (b)). TALK, illustrated in figure 3.4 (c) and (d) increases from the east to the west throughout all seasons along both transects and peaks at $\sim 40\text{--}50^\circ\text{W}$ along the southbound transect and at $\sim 50\text{--}60^\circ\text{W}$ along the northbound transect. Similar to DIC, TALK concentrations drop towards the most western longitudes

TALK is mainly determined by freshwater addition through precipitation and loss through evaporation within the subtropical Atlantic Ocean, hence it is strongly related to SSS and only to a lesser degree to SST (see e.g. Lee *et al.*, 2006). In contrast, the DIC distribution is strongly dependent on the solubility of CO_2 (K_0), which is mainly driven by temperature and to a lesser degree by salinity (see e.g. Sarmiento and Gruber, 2006).

From the east to the west and from the cold to the warm seasons, temperatures, illustrated in figure 3.5 (a) and (b), increase and thereby reduce the solubility of seawater CO_2 . TALK and SSS are closely related, as expected, illustrated in figure 3.5 (c) and (d).

Figure 3.6 (a) shows that the DIC/TALK ratio decreases as a function of longitude along both transects. This illustration was chosen as it reveals three distinct water masses, the first east of 24°W , the second between $24\text{--}55^\circ\text{W}$ and the third west of 55°W along both transects, coloured in cyan, black and magenta, respectively. The most eastern water mass shows a medium decrease towards the west in the DIC/TALK ratio, which is stronger along the southbound transect. The mid longitude water mass is determined by a strong decrease towards the west in the DIC/TALK ratio which is almost equally pronounced along both transects. The most western water mass shows the weakest decrease in the ratio towards the west.

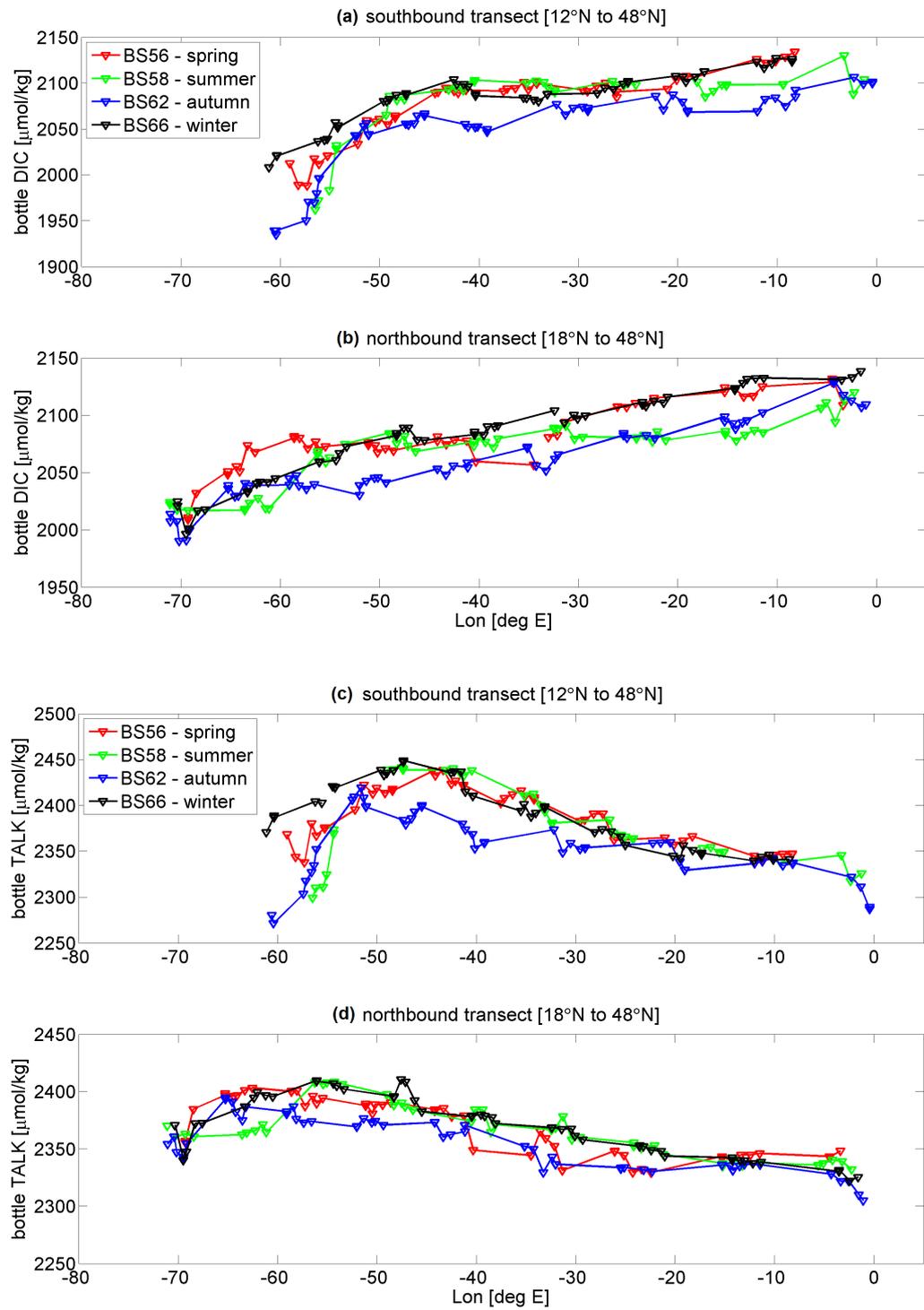


Figure 3.4: Analysed DIC and TALK data along the UK–Caribbean line that received a WOCE flag of 2 split up into the 4 seasons when data were sampled. (a) shows DIC data along the southbound voyages, (b) shows DIC data along the northbound return voyages, (c) shows TALK data along the southbound voyages and (d) shows TALK data along the northbound return voyages. Colours for the different seasons follow the color code from figure 3.2, namely red for spring, green for summer, blue for autumn and black for winter.

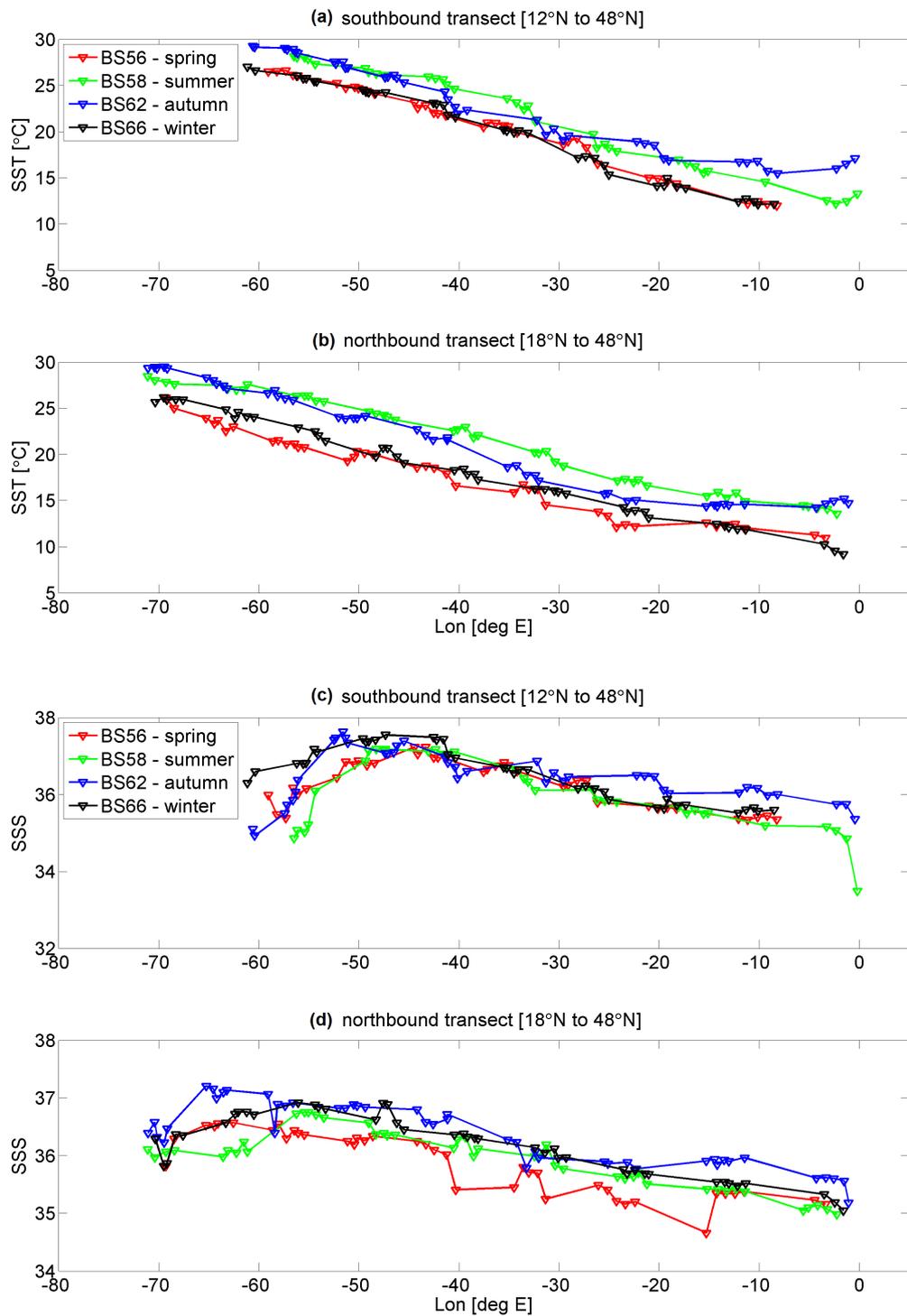


Figure 3.5: Underway SST and SSS data along the UK–Caribbean line measured by the Aanderaa sensors at the seawater inlet split up into the 4 seasons when data were sampled. (a) shows SST data along the southbound voyages, (b) shows the SST data along the northbound return voyages, (c) shows SSS data along the southbound voyages and (d) shows SSS data along the northbound return voyages. Colours for the different seasons follow the color code from figure 3.2, namely red for spring, green for summer, blue for autumn and black for winter.

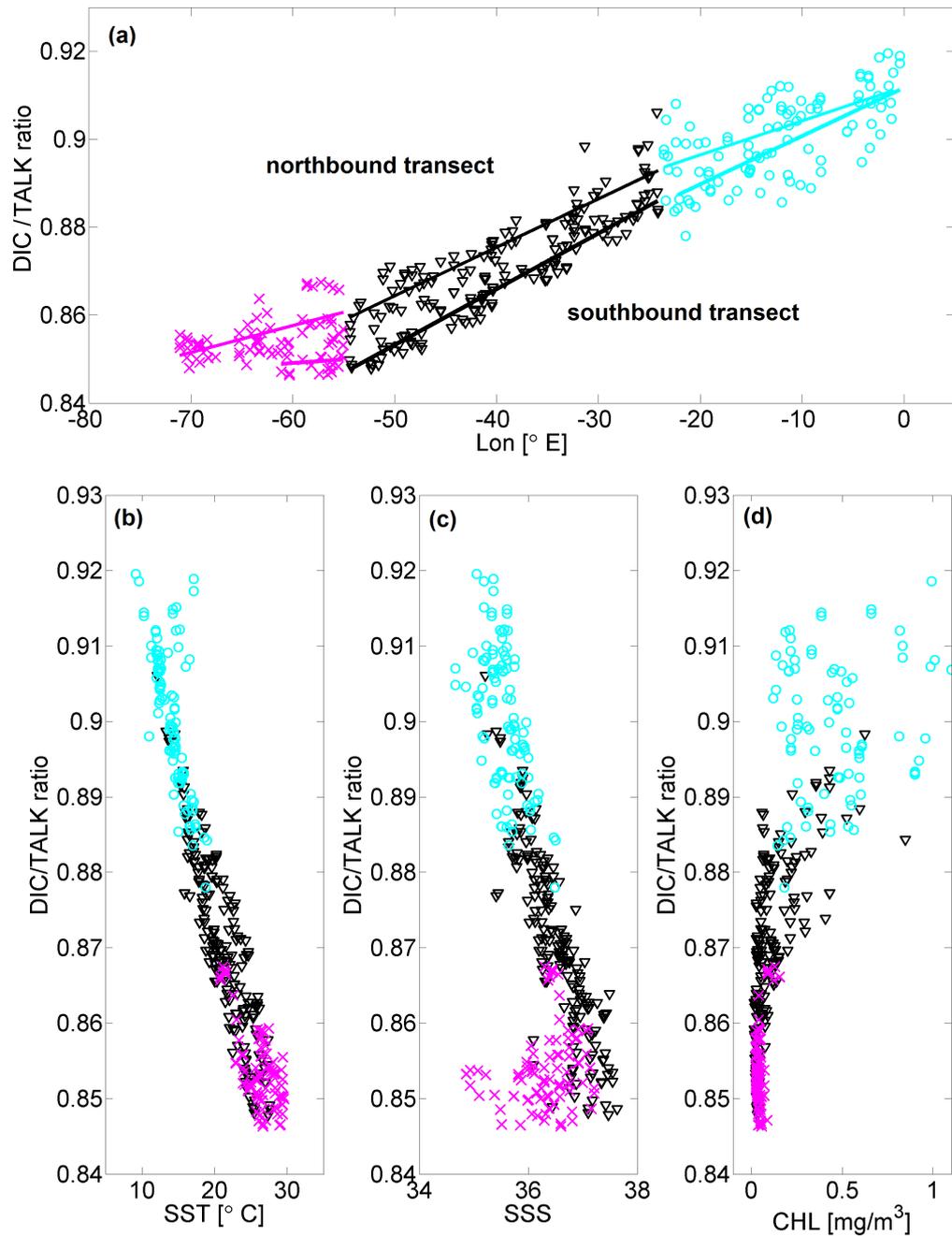


Figure 3.6: Property plots of (a) the DIC/TALK ratio as a function of longitude including linear regression lines for both northbound and southbound transects, (b)–(d) DIC/TALK ratio as a function of SST, SSS and CHL, respectively. Cyan circles mark measurements taken east of 24°W, black triangles mark measurements taken between 24°W–55°W and magenta crosses mark measurements taken west of 55°W, based on the changing DIC/TALK ratio in (a).

Figure 3.6 (b)–(d) illustrate the DIC/TALK ratio as a function of the main drivers, combined for both transects, namely SST, SSS and CHL, representing the photosynthetic activity. For CHL, the monthly Globcolour (<http://www.globcolour.info>) product (see chapter 2) for the years 2013/2013 was taken and the nearest satellite data point was used for the comparison to the actual observations.

The DIC/TALK ratio shows a linear relationship with SST in figure 3.6 (b) along all water masses. Within the subtropics, SSS explains roughly 80% of the TALK variability (Millero *et al.*, 1998), hence the SST effect on the TALK distribution is fairly small, but is much larger on DIC. However, SST can not be regarded as a stand-alone driver, since SSS, biological production and gas exchange are not constant. For example along the southbound transect within the mid section water mass (24°W–55°W), DIC barely changes as the ship moves west, hence the increase in TALK drives the DIC/TALK ratio in this region.

The DIC/TALK ratio in figure 3.6 (c) shows a linear decrease throughout most parts of the transect as function of salinity. Here, the increasing evaporation/transpiration ratio alters the SSS as well as the TALK concentration, whereas DIC is only influenced to a lesser degree by SSS changes. The water mass furthest west provides an exception to the linear decrease of the DIC/TALK ratio as a function of salinity. Here, TALK decreases with decreasing SSS, while DIC decreases at a similar rate, resulting in a barely changing DIC/TALK ratio. SST does not explain the drop in DIC (see figure 3.4 and figure 3.5), hence a plausible explanation for the almost steady ratio within the most western water mass can be provided by the entrainment of freshwater, causing low DIC concentrations, and further explaining a simultaneous drop in both SSS and TALK.

Figure 3.6 (d) illustrates the DIC/TALK ratio as a function of CHL. Within the water mass furthest to the east, biological production becomes more important. Photosynthesis decreases the amount of carbon in the surface ocean, leading to a decrease in DIC, whereas the formation of organic matter further decreases the amount of free protons, resulting in an increase in TALK, i.e., decreases the DIC/TALK ratio. Furthermore, the spread in the data might indicate seasonal variability of the biological effect.

3.3 Sea surface $p\text{CO}_2$ along the UK–Caribbean line

3.3.1 Sea surface $p\text{CO}_2$ calculation

The DIC and TALK data provide the two required carbonate system parameters for the calculation of the sea surface $p\text{CO}_2$, together with the directly measured SST, SSS and atmospheric pressure at sea level. Furthermore, nutrient data (silicate and phosphate) sampled on-board were co-located to the closest DIC and TALK sampling locations and used for the $p\text{CO}_2$ calculation. For the calculation the Matlab version of CO2SYS is used which is based on the original program for DOS and EXCEL (Lewis and Wallace, 1998). The program uses two out of the four parameters (DIC, TALK, either $f\text{CO}_2$ or $p\text{CO}_2$, and pH) of the carbonate system in seawater and their relationship to compute the remaining parameters, in this case the sea surface $p\text{CO}_2$ and pH.

In a first step, using the TALK definition by Dickson (1981) and the dissociation constants K_1 and K_2 obtained by Mehrbach *et al.* (1973) and refitted by Dickson and Millero (1987), the sea surface pH is calculated from DIC and TALK. The program iteratively calculates the pH via Newtons method from an initial start guess to a final pH value with zero residual. Finally, the sea surface fugacity ($f\text{CO}_2$) is calculated from the DIC concentration and the free protons following:

$$f\text{CO}_2 = \frac{[\text{DIC}]}{K_0} \cdot \frac{[\text{H}^+] \cdot [\text{H}^+]}{[\text{H}^+] \cdot [\text{H}^+] + K_1 \cdot [\text{H}^+] + K_1 \cdot K_2} \quad (3.2)$$

Finally, $f\text{CO}_2$ is converted to $p\text{CO}_2$ using equation 2.1

3.3.2 Validation of the calculated sea surface $p\text{CO}_2$

In order to validate the calculated $p\text{CO}_2$ data, they are in a first step directly compared to the measurements derived from the underway system on-board the ship. After the quality control of the underway CO_2 data, a total of 147 data points, at the same measurement point in time and space, remain for comparison with the bottle sample derived $p\text{CO}_2$. Unfortunately, due to technical problems, no underway data are available in summer and only a limited amount of data remain for the autumn and winter voyage. Therefore, for additional validation within all seasons, the gridded $1^\circ \times 1^\circ$ SOM–FFN estimates, derived

from the method presented in chapter 2, are used. The monthly $1^\circ \times 1^\circ$ SOM–FFN $p\text{CO}_2$ estimates from 1998 through 2011 were averaged in time for each month and further corrected to the actual sampling date by adding $1.5 \mu\text{atm}$ per year. The global SOM–FFN $p\text{CO}_2$ validation and results will be presented in chapter 5. Similar to the DIC and TALK comparison in figure 3.3, the SOM–FFN monthly data were co–located to nearest sampling point in space, within at least 1° longitude and 1° latitude, within the same month.

Figure 3.7 (a) shows the comparison between the calculated $p\text{CO}_2$ data, derived from the discrete bottle samples with the quality controlled underway data, while figure 3.7 (b) shows the comparison with the co–located $p\text{CO}_2$ data from the SOM–FFN. The regression in figure 3.7 (a) reveals a close match between the two independently determined sea surface $p\text{CO}_2$ data. The majority of all $p\text{CO}_2$ measurements lies within 300 to 400 μatm , however, a few data measured close to the shore of England and France exceed partial pressures of 500 μatm . Overall, the comparison shows a small mean difference between the 147 data points of 0.42 μatm and a RMSE of 10.09 μatm . The slope of the regression reveals a close match of all data along the entire range of partial pressures measured.

Figure 3.7 (b) reveals a larger spread between the calculated $p\text{CO}_2$ data and the SOM–FFN estimates. The mean difference between all data points is 0.12 μatm with a RMSE of 13.64 μatm . The larger spread between the data is likely linked to the co–location of the $1^\circ \times 1^\circ$ gridded data to the actual sampling point. Furthermore, figure 3.7 (b) shows that winter data lie in general below the red one–on–one regression line, whereas summer data lie above, indicating that (i) the 1.5 μatm trend correction does not accurately represent the actual trend within all seasons and (ii) the effect of inter–annual $p\text{CO}_2$ variabilities can not be reconstructed by a simple temporal correction.

3.3.3 Variability and drivers of the sea surface $p\text{CO}_2$

The comparison of the calculated sea surface $p\text{CO}_2$ data with underway measurements shows that both agree well with each other. Hence, there is confidence that these data provide an accurate basis to investigate the seasonal variability of the sea surface $p\text{CO}_2$ along the UK–Caribbean line. However, samples were taken during daytime only and the sampling locations are dependent on the ship speed as well as the ship route and vary

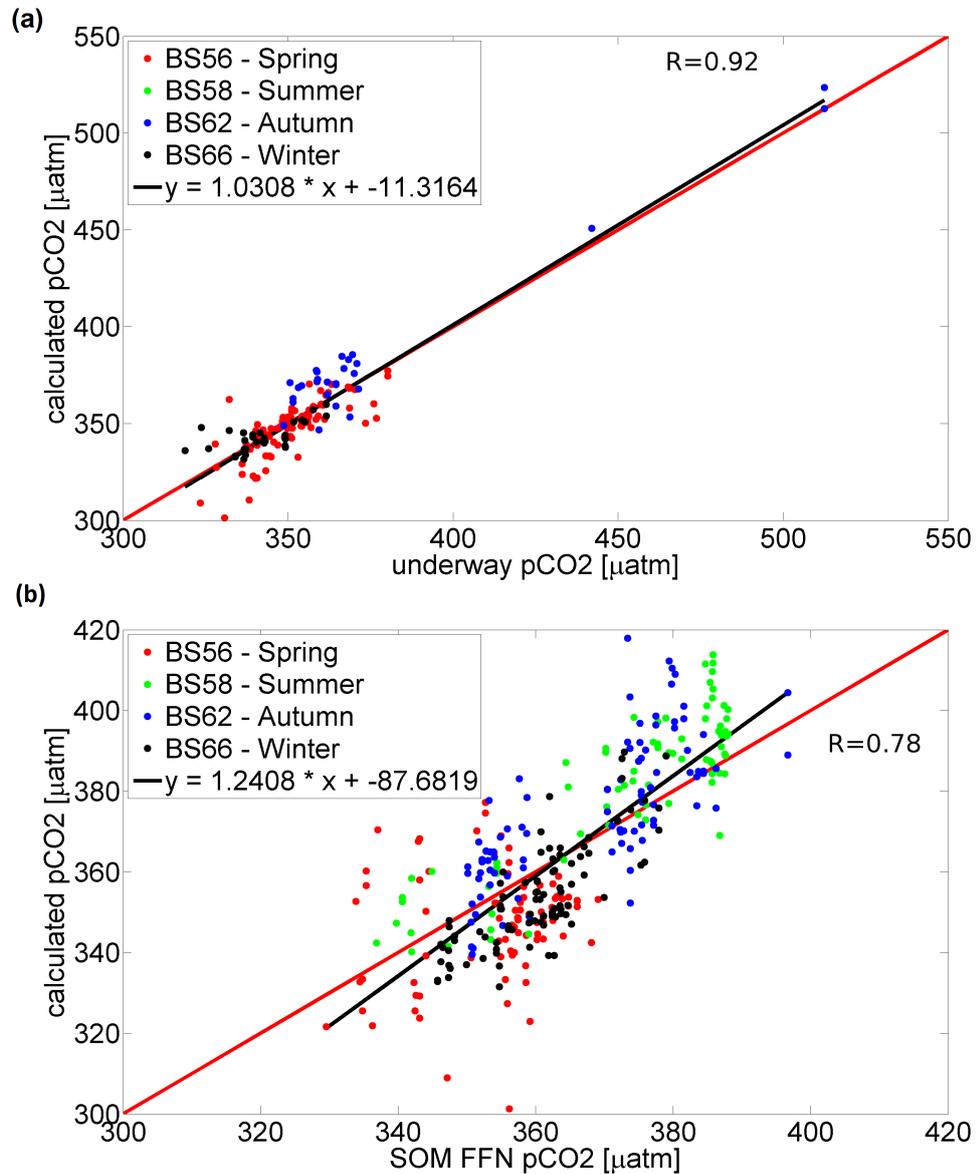


Figure 3.7: Calculated $p\text{CO}_2$ data along the UK–Caribbean line compared to (a) the quality controlled underway measurement derived from the infra–red absorption method and (b) the SOM–FFN derived with the method presented in chapter 2. Colours refer to the different seasons when the measurements were taken, following the color code of figure 3.2, namely red for spring, green for summer, blue for autumn and black for winter. The black lines show the regression lines of all data points combined and the red lines show the optimal one–on–one regression line.

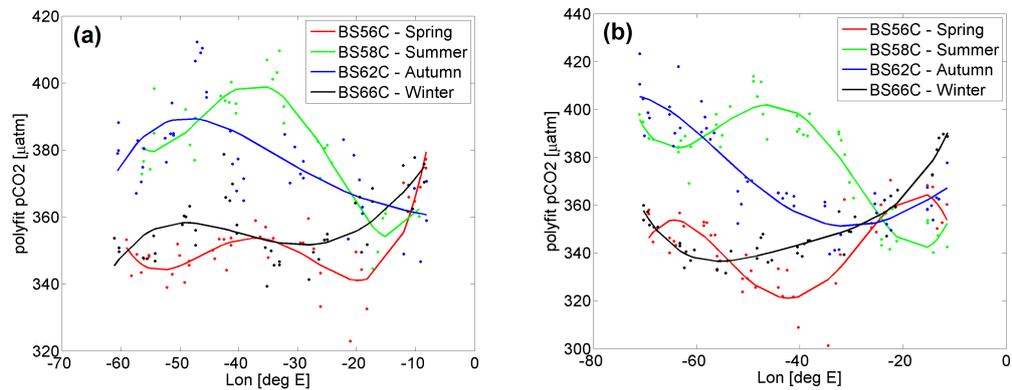


Figure 3.8: Seasonal $p\text{CO}_2$ data points and their polynomial fits as a function of longitude for (a) the southbound transect and (b) the northbound transect. Colours refer to the different seasons when the measurements were made, following the color code of figure 3.2, namely red for spring, green for summer, blue for autumn and black for winter.

between the different seasons as illustrated in figure 3.2. This circumstance makes a direct sample-to-sample comparison between different seasons impossible. As a consequence a different approach is used to investigate the seasonal variability of the sea surface $p\text{CO}_2$.

In a first step the individual voyages were split up into northbound and southbound transect and the four different seasons. Data points for each of the 8 individual transects were then plotted against their longitude co-ordinates and a 4th order polynomial was fit to the data:

$$y = a \cdot x^4 + b \cdot x^3 + c \cdot x^2 + d \cdot x + e \quad (3.3)$$

This is illustrated in figure 3.8. A fourth order polynomial was chosen as it shows the smallest bias compared to the actual calculated $p\text{CO}_2$. The variables a – e are the coefficients of the least squares fit. Data points east of 8°W were excluded, firstly because not every season was present along these longitudes and secondly because steep gradients close to shore strongly influence the polynomial fit. Hence, in the following, the seasonal variability is only discussed for the open ocean. Furthermore, differences between ship tracks are not accounted for in the following analysis.

Figure 3.8 provides an insight into the seasonal variability of the sea surface $p\text{CO}_2$ in the temperate North Atlantic Ocean. The maximum seasonal amplitude along both

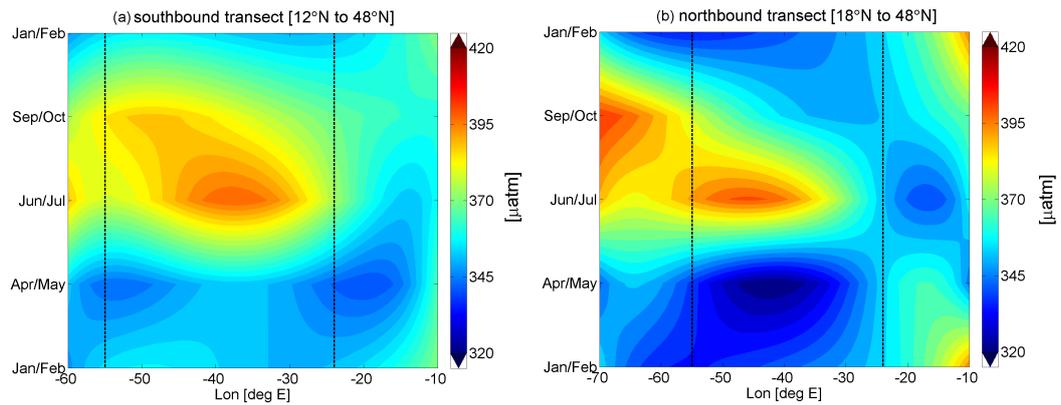


Figure 3.9: Hovmöller plot of the seasonal $p\text{CO}_2$ cycle along the UK–Caribbean line for (a) the southbound transect and (b) the northbound transect.

transects is mainly determined by the difference between summer (green line) and spring (red line), but substantially differs between both transects. While the northbound transect (figure 3.8 (b)) shows a maximum seasonal amplitude around 40°W of $\sim 80 \mu\text{atm}$, the maximum seasonal amplitude of the southbound transect (figure 3.8 (a)) does not exceed $50 \mu\text{atm}$. However, it can be argued here, that the amplitude difference along the northbound section is linked to the different ship tracks (see figure 3.2) between spring (further north) and summer (further south), which are furthest apart around 40°W, exactly where the strongest seasonal difference occurs.

The seasonal $p\text{CO}_2$ variability between east and west is further shown in figure 3.9. The Hovmöller contour plots in figure 3.9 were created by plotting the $p\text{CO}_2$ of every 1° longitude step from each polynomial derived from equation 3.3. Figure 3.9 illustrates the seasonal variability and the variability along the line from the west to the east following the example shown by Cooper *et al.* (1998). Along both transects, the North Atlantic subtropical gyre region ($\sim 24^\circ\text{W}$ to 55°W) shows a clear temperature driven seasonal pattern with high partial pressures in summer, whereas the most eastern longitudes show an anti-phased seasonal cycle, as clearly seen along the northbound transect. The seasonal maximum of the most western longitudes is shifted from summer at the southbound transect to autumn at the northbound transect. However, there is $\sim 6^\circ$ in latitude and $\sim 10^\circ$ in longitude difference between the port of arrival and the port of departure in the Caribbean. The borders, where these changes in the seasonality occur, coincide with the changes in the DIC/TALK ratio investigated in section 3.2.4.

In order to investigate the drivers of the $p\text{CO}_2$ variability along the UK–Caribbean line, the changes in the sea surface $p\text{CO}_2$ were split up into components as suggested by Takahashi *et al.* (1993) (see equation 1.17). East to west modifications in the $p\text{CO}_2$ resulting from temperature changes are computed using equation 1.18 as reported by Takahashi *et al.* (1993). DIC and TALK as well as salinity normalized TALK (sTALK) and DIC (sDIC), to compute the effect of freshwater forcing (FRESH), are derived from the approximations of Lovenduski *et al.* (2007):

$$\frac{\partial p\text{CO}_2}{\partial \text{DIC}} \cdot \Delta \text{DIC} = \frac{\text{sDIC}}{s_0} \cdot \frac{\partial p\text{CO}_2}{\partial \text{DIC}} \cdot \Delta s + \frac{s}{s_0} \cdot \frac{\partial p\text{CO}_2}{\partial \text{DIC}} \cdot \Delta \text{sDIC} \quad (3.4)$$

$$\frac{\partial p\text{CO}_2}{\partial \text{TALK}} \cdot \Delta \text{TALK} = \frac{\text{sTALK}}{s_0} \cdot \frac{\partial p\text{CO}_2}{\partial \text{TALK}} \cdot \Delta s + \frac{s}{s_0} \cdot \frac{\partial p\text{CO}_2}{\partial \text{TALK}} \cdot \Delta \text{sTALK} \quad (3.5)$$

$$\frac{\partial p\text{CO}_2}{\partial \text{FRESH}} \cdot \Delta \text{FRESH} = \frac{\text{sDIC}}{s_0} \cdot \frac{\partial p\text{CO}_2}{\partial \text{DIC}} \cdot \Delta s + \frac{\text{sTALK}}{s_0} \cdot \frac{\partial p\text{CO}_2}{\partial \text{TALK}} \cdot \Delta s \quad (3.6)$$

where:

$$\frac{\partial p\text{CO}_2}{\partial \text{DIC}} = \frac{p\text{CO}_2}{\text{DIC}} \cdot \gamma_{\text{DIC}} \quad (3.7)$$

$$\frac{\partial p\text{CO}_2}{\partial \text{TALK}} = \frac{p\text{CO}_2}{\text{TALK}} \cdot \gamma_{\text{TALK}} \quad (3.8)$$

with:

$$\gamma_{\text{DIC}} \approx \frac{3 \cdot \text{TALK} \cdot \text{DIC} - 2 \cdot \text{DIC}^2}{(2 \cdot \text{DIC} - \text{TALK}) \cdot (\text{TALK} - \text{DIC})} \quad (3.9)$$

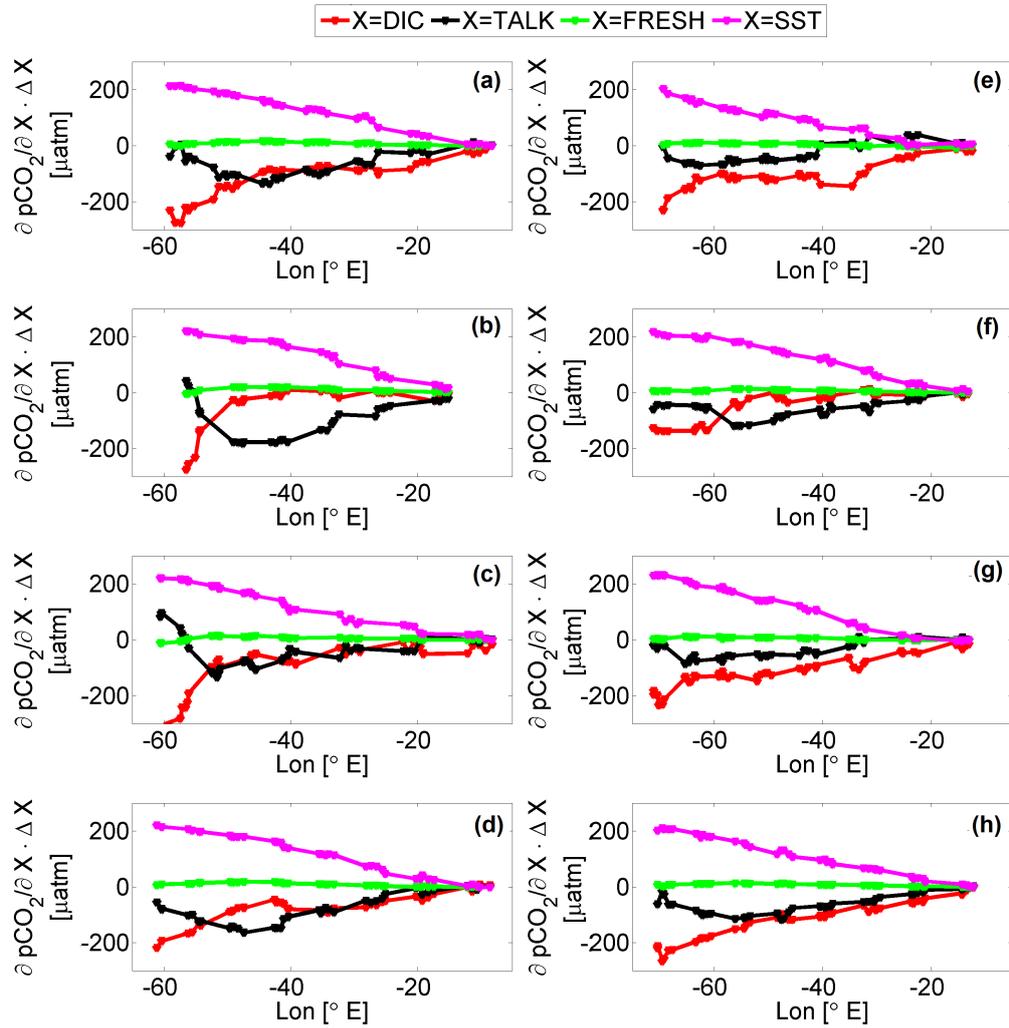


Figure 3.10: Sea surface $p\text{CO}_2$ changes plotted against longitude for the southbound (a)–(d) and the northbound (e)–(h) transects for each season from spring (top) to winter (bottom). The $p\text{CO}_2$ changes along the lines are split up into changes of the driving components, namely DIC (red), TALK (black), freshwater fluxes (FRESH – green) and SST (magenta). All changes are plotted relative to the first measurement in the east

$$\gamma_{\text{TALK}} \approx - \frac{\text{TALK}^2}{(2 \cdot \text{DIC} - \text{TALK}) \cdot (\text{TALK} - \text{DIC})} \quad (3.10)$$

The change in the sea surface $p\text{CO}_2$ along the line from the east to the west for each season and both southbound and northbound voyage are illustrated in figure 3.10. Takahashi *et al.* (1993) proposed a 4% change in the sea surface $p\text{CO}_2$ per $^{\circ}\text{C}$ warming of the surface water. This is reflected in a $p\text{CO}_2$ change $> 200 \mu\text{atm}$ from the east to the west in all seasons and along both transects. In contrast, freshwater changes only have a minor effect

on the sea surface $p\text{CO}_2$.

The temperature effect on the $p\text{CO}_2$ appears to be largely balanced from the east to the west by the DIC contribution along the northbound transect. Similar to SST, variations in DIC lead to $p\text{CO}_2$ changes of $< -200 \mu\text{atm}$ between the east and the west. These DIC shifts are however not linear, particularly along the southbound transect. The steady DIC concentration in the subtropical gyre is linked to the long residence time of carbon and the limited effect of mixing and biology. Hence, the contribution of DIC on the $p\text{CO}_2$ change is small in the mid longitudes, in particular along the southbound transect.

Increasing TALK concentrations from the east to the west, driven by the evaporation/transpiration ratio, resemble a reduction of free protons, resulting in a decrease in the sea surface $p\text{CO}_2$ along the line. Throughout all seasons, the TALK effect on sea surface $p\text{CO}_2$ changes reaches magnitudes similar to those driven by DIC, but is particularly strong in the subtropical gyre. In summer, along both transects, the TALK effect exceeds the DIC effect, counteracting the temperature driven increase in sea surface $p\text{CO}_2$.

The water mass furthest to the west (west of 55°W) has been linked to freshwater input in the previous sections. The water balance effect changes DIC proportionally to TALK, resulting in a low net freshwater effect on the sea surface $p\text{CO}_2$, which is in agreement with Takahashi *et al.* (1993).

3.4 Summary and conclusion

A total of 325 bottle samples was collected on-board M/V Benguela Stream during four field campaigns between April/May 2012 and January/February 2013, with each voyage representing one of the four seasons. TALK and DIC content of the bottle samples were analysed during two laboratory campaigns using a VINDTA 3C analysing system.

The DIC/TALK ratio reveals three distinct water masses along the UK–Caribbean line. The DIC/TALK ratio decreases linearly within all water masses as a function of temperature, indicating its influence on the DIC concentration throughout large parts of the crossings. Changes in salinity, linked to evaporation/transpiration changes, are expected to likewise change the TALK concentration, whereas DIC is expected to show only a minor response to salinity changes. This is well pronounced in water masses east of 55°W , but not the water mass furthest west, where DIC and TALK decrease equally,

linked to the entrainment of freshwaters. Biological production only plays a role in the water mass furtherst to the east (east of 24°W)

The analysed DIC and TALK data further form the basis to calculate the sea surface $p\text{CO}_2$. The calculated $p\text{CO}_2$ values agree well with the available quality controlled data obtained from the underway system installed on M/V Benguela Stream. For each crossing, i.e. southbound and northbound, the $p\text{CO}_2$ data again clearly show the three distinct regions. The first, west of 55°W, shows an autumn maximum of sea surface $p\text{CO}_2$. The second, from 24°W to 55°W, shows a temperature driven gyre seasonality with the lowest partial pressures in the cooler winter and spring months and $p\text{CO}_2$ maximum in summer. Here, the seasonal amplitude reaches $\sim 80 \mu\text{atm}$. The third region, east of 24°W (north of $\sim 40^\circ\text{N}$), shows the reversed seasonal cycle of the mid longitude water mass with high partial pressures in winter and spring and lower partial pressures in summer and autumn, as illustrated in figure 3.9. This region is part of the transition zone between the temperature dominated subtropical waters and the biologically driven subpolar waters (Takahashi *et al.*, 2002). DIC and TALK changes counteract the temperature driven $p\text{CO}_2$ increase along the line, with contributions to the $p\text{CO}_2$ change $< -200 \mu\text{atm}$ (compared to temperature driven changes $> 200 \mu\text{atm}$).

Bottle measurements along the UK–Caribbean line illustrate the close match between calculated sea surface $p\text{CO}_2$ and those measured via infra–red absorption, confirming the robustness of current measurement standards. The bottle samples provide the basis for a quantitative assessment on the variability of the sea surface $p\text{CO}_2$ along the ship route regarding its temperature, salinity, TALK and DIC dependency. Hence, discrete bottle samples form an accurate basis to investigate the sea surface $p\text{CO}_2$, its variability and drivers along the UK–Caribbean line.

Chapter 4

Seasonal to inter–annual variability of the Atlantic Ocean carbon sink

”Science cannot solve the ultimate mystery of nature.
And that is because, in the last analysis, we ourselves
are a part of the mystery that we are trying to solve.

(Max Planck)

The results presented in the following chapter have been published in:

A neural network–based estimate of the seasonal to inter–annual variability of the Atlantic Ocean carbon sink

P. Landschützer¹, N. Gruber², D.C.E. Bakker¹, U. Schuster^{1,6}, S. Nakaoka³, M.R. Payne^{2,4}, T.P. Sasse⁵, J. Zeng³

¹School of Environmental Sciences, University of East Anglia, Norwich Research Park, Norwich, UK

²Institute of Biogeochemistry and Pollutant Dynamics, ETH Zürich, Zürich, Switzerland

³Center for Global Environmental Research, National Institute for Environmental Studies, Tsukuba, Ibaraki, Japan

⁴Section for Oceanography and Climate, National Institute for Aquatic Resources, Technical University of Denmark, Charlottenlund, Denmark

⁵Climate Change Research Centre, University of New South Wales, Sydney, Australia

⁶Now at: College of Life and Environmental Sciences, Hatherly Laboratories, University of Exeter, Exeter, UK

Biogeosciences, 2013, 10, 7793–7815, doi: 10.5194/bg–10–7793–2013

The work and analysis presented in this chapter was undertaken by P. Landschützer. Co–authors on this publication provided guidance and suggestions regarding the analysis and the results to help and address the interests of the wider scientific community.

In this chapter the $1^\circ \times 1^\circ$ monthly estimated $p\text{CO}_2$ and air–sea flux fields derived from the 2–step neural network method (SOM–FFN) introduced in chapter 2 are presented for the Atlantic Ocean from 1998 through 2007. At first the quality of the $p\text{CO}_2$ product is determined by examination of the $p\text{CO}_2$ residuals, defined as the difference of the SOM–FFN $p\text{CO}_2$ estimates and gridded observations of the SOCAT v1.5 database (Sabine *et al.*, 2013), in both time and space. In a second, more robust test the SOM–FFN estimates are validated with independent timeseries products and data from the updated version of the SOCAT database (version 2). The residuals further form the basis for an in–depth uncertainty analysis of the air–sea gas flux. The decadal and seasonal mean results, computed from the $p\text{CO}_2$ estimates and the CO_2 flux are illustrated, and it is shown that they are consistent with recent studies. Finally, a $p\text{CO}_2$ and CO_2 flux trend and inter–annual variability analysis is conducted in the Atlantic Ocean, the North Atlantic and the South Atlantic sub–basins, within the considered time period.

4.1 Residuals and validation

A first check to test the quality of the neural network estimates is to look at the residuals. The combined SOM–FFN method obtains good fits with an overall mean r^2 between the fitted $p\text{CO}_2$ and the gridded Atlantic Ocean SOCAT v1.5 data of 0.87 and a root mean squared error (RMSE) of about $10 \mu\text{atm}$ (table 4.1). The overall bias is small ($-0.10 \mu\text{atm}$). Similar results apply to each year individually, indicating that the temporally inhomogeneous data distribution does not have a measurable effect on the estimates for each year.

The residuals are not entirely randomly distributed in space. As shown in figure 4.1 (a), the temporal mean residuals in each pixel show generally low values in the open ocean region, but tend to increase towards the fronts. The strongest model–observation discrepancies occur in the Equatorial Atlantic, along the Gulf Stream and North Atlantic Current as well as in the Norwegian, Greenland and the North Seas (see figure 1.5), i.e., mostly in regions with relatively strong horizontal gradients in surface ocean $p\text{CO}_2$.

The standard deviation of the residuals (figure 4.1 (b)) shows that the highest temporal errors occur again in the high latitudes of the North Atlantic, in particular the Norwegian and North Sea, as well as along the North American coastline and in the eastern South

Table 4.1: Statistical measures of the comparison of the neural network–based estimates of $p\text{CO}_2$ with the SOCAT v1.5 gridded observations (Sabine *et al.*, 2013) in the Atlantic Ocean from 44°S to 79°N and west of 30°E

| Period | r^2 | RMSE [μatm] | bias [μatm] | # data |
|-----------|-------|--------------------------|--------------------------|--------|
| 1998-2007 | 0.87 | 9.89 | -0.10 | 20003 |
| 1998 | 0.93 | 7.15 | -0.18 | 583 |
| 1999 | 0.89 | 9.35 | 1.62 | 186 |
| 2000 | 0.78 | 11.14 | -0.19 | 178 |
| 2001 | 0.83 | 11.48 | -0.87 | 566 |
| 2002 | 0.87 | 8.98 | 0.22 | 1941 |
| 2003 | 0.87 | 7.47 | -0.11 | 1963 |
| 2004 | 0.87 | 8.13 | -0.14 | 2729 |
| 2005 | 0.88 | 9.52 | -0.43 | 3575 |
| 2006 | 0.85 | 11.51 | -0.10 | 4280 |
| 2007 | 0.87 | 10.89 | 0.12 | 4002 |

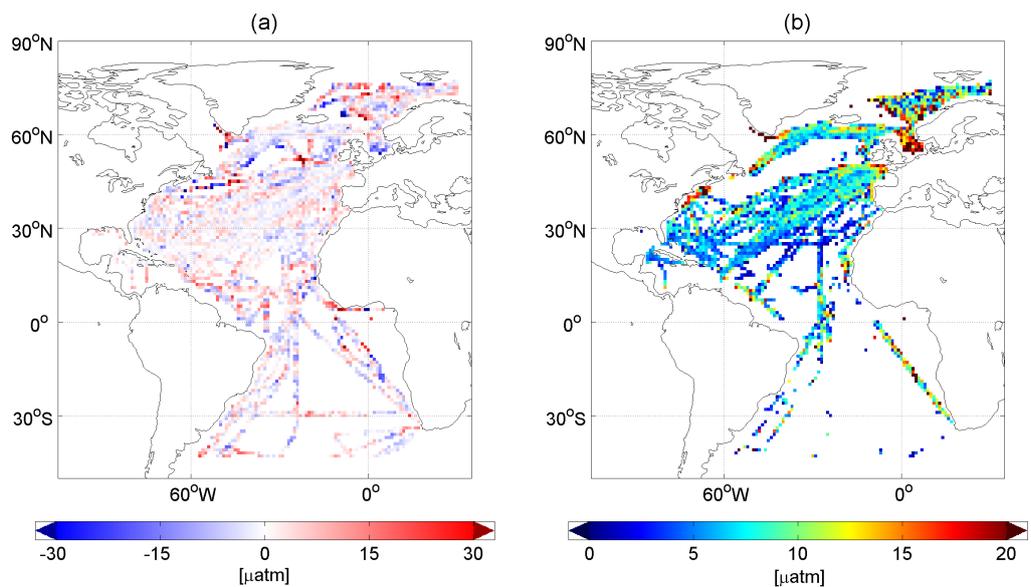


Figure 4.1: (a) Temporal mean residuals and (b) standard deviation of the residuals in μatm between neural network estimates and SOCAT v1.5 gridded observations (Sabine *et al.*, 2013) for the period from 1998–2007. Pixels that have a value in (a) but not in (b) indicate where only 1 observation in time is available.

Atlantic between 0 and 30°S . This indicates that the model input parameters are not able to predict all the temporal variability occurring in these regions with known biogeochemical complexity.

To test the impact of the inhomogeneous distribution of the neural network input data and $p\text{CO}_2$ observations, the residuals are further shown as a function of the input variables (figure 4.2). It is shown that residuals are independent of the magnitude of the

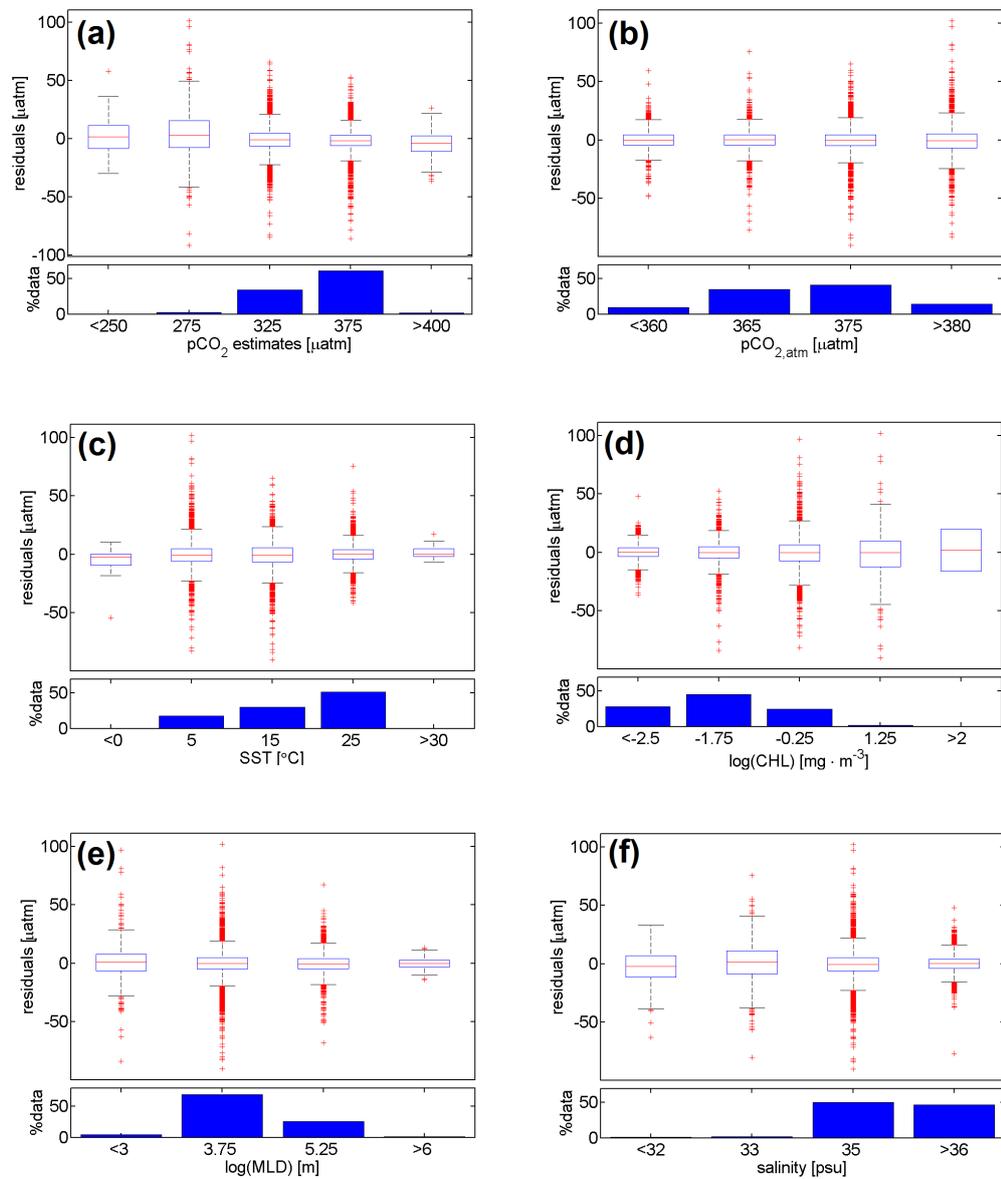


Figure 4.2: Residuals as a function of (a) fitted $p\text{CO}_2$, (b) atmospheric $p\text{CO}_2$, (c) sea-surface temperature (SST), (d) natural log of surface chlorophyll ($\log(\text{CHL})$), (e) natural log of mixed layer depth ($\log(\text{MLD})$), and (f) sea-surface salinity (SSS). The upper plot in each panel depicts the residuals, shown as a box-and-whiskers plot. The red line in the box show the median, the blue-outlined box indicates the 25 and 75 percentiles and red pluses mark residuals outside this interval. The lower plot shows the relative number of observations within each bin.

estimated $p\text{CO}_2$, and also do not show any dependence on the magnitude of the independent variables. Each bin median of the residuals is close to zero, with the strongest spread occurring in the low $p\text{CO}_2$ bins around $275 \mu\text{atm}$ which coincides with low SST at around 5°C and high $\log(\text{CHL})$ concentrations at around -0.25 to $1.25 \text{ mg} \cdot \text{m}^{-3}$. Figure 4.2 further shows that large residuals, most of which stem from regions characterized by

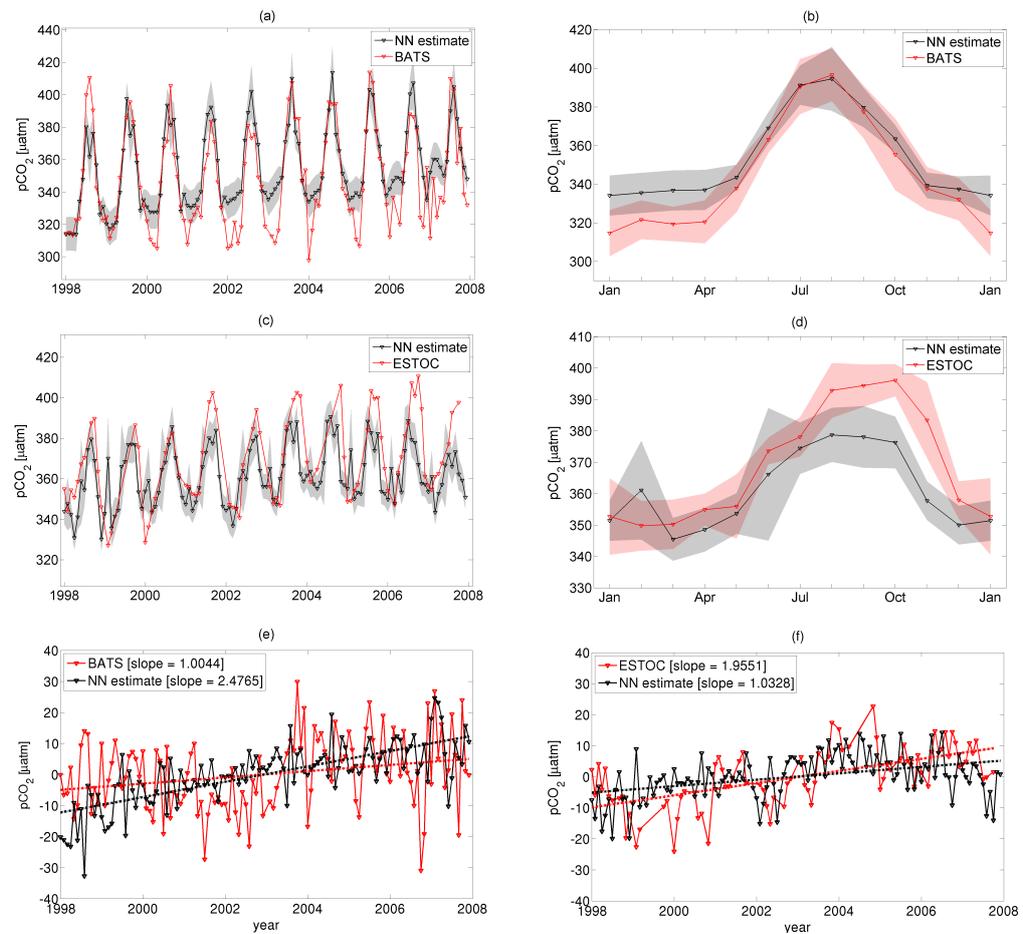


Figure 4.3: Sea surface $p\text{CO}_2$ from neural network (NN) estimates and time-series stations: (a)–(b) from the BATS Hydrostation “S” (Bates, 2007; Gruber *et al.*, 2002) and (c)–(d) from the ESTOC (González-Dávila *et al.*, 2007) timeseries stations. In (a) and (c) the actual timeseries data are compared, while in (b) and (d), the long-term mean seasonal cycle is evaluated. Grey shading shows the uncertainty of the neural network estimate based on its RMSE in (a) and (c), and the average RMSE in (b) and (d). Red shading shows the standard deviation of the mean seasonal cycle for each timeseries station. The deseasonalised timeseries (monthly timeseries minus mean seasonal cycle) at BATS and ESTOC are illustrated in plots (e) and (f) respectively. Linear fits are illustrated by the dashed lines and the slope of each fit is shown in the figure legend).

strong horizontal $p\text{CO}_2$ gradients, are independent of the data density.

The residuals indicate that the combined SOM–FFN method fulfils most tests for a good fit and does not contain any major biases. In particular, there is no indication of a substantial degeneration of the fits as a function of data density. Regions with high spatial or temporal variability are the least well fitted, while the fits for most of the less variable open ocean are very good.

4.2 Validation with independent observations

A second and more robust test of the model's ability to predict basin wide $p\text{CO}_2$ fields was conducted by comparing output values to independent data. To this end the network-based $p\text{CO}_2$ estimates are compared with observations from two timeseries stations in the subtropical North Atlantic, i.e., the combined record from BATS (Bermuda Atlantic Time Series Station) and Hydrostation "S" (Bates, 2007; Gruber *et al.*, 2002) located in the northwestern Sargasso Sea near Bermuda (32.17°N , 64.50°W), and the European Station for Time Series in the Ocean (ESTOC, e.g. González-Dávila *et al.*, 2007) located in the eastern subtropical gyre near the Canary Islands (29.04°N , 15.5°W). Sea surface $p\text{CO}_2$ at BATS has been calculated from the reported DIC and TALK concentrations using CO2SYS as explained in section 3.3.1 of this thesis. These stations provide near monthly coverage over the time period estimated by the neural network method. At ESTOC, $p\text{CO}_2$ is reported, however, it is not directly measured, but calculated from directly observed carbonate system parameter. In months where more than one measurement was taken at each timeseries station, the average value was used to compare the sea surface $p\text{CO}_2$ to the SOM-FFN estimates. Furthermore, SOM-FFN estimates are not centred at the exact geographical position of both timeseries stations, therefore the four closest surrounding $1^\circ \times 1^\circ$ grid-boxes are interpolated to the exact location of the timeseries stations, weighted by their distance.

Figure 4.3 shows the comparison between the neural network estimates with both timeseries for the period between 1998 through 2007 and the mean seasonal cycle within this period. While the phase of the seasonal cycle is captured well, figure 4.3 shows that the neural network estimates in general overestimate the $p\text{CO}_2$ in winter at Bermuda from January to April and underestimate the autumn maxima at ESTOC from August to November. This underestimation of the seasonal amplitude is likely linked to the validation approach which was implemented to prevent the neural network from overtraining (see chapter 2). The neural network estimates further show a decrease in the summer sea surface $p\text{CO}_2$ in the eastern subtropical gyre (figure 4.3 (c)) from 2005 onwards which is not seen in the ESTOC data. The decadal mean difference between BATS data and neural network estimates is $7.56 \mu\text{atm}$ with a root mean squared error (RMSE) of $17.53 \mu\text{atm}$. Similar to BATS, the decadal mean difference between ESTOC data and the estimates in

this study is $-8.06 \mu\text{atm}$ with an RMSE of $16.85 \mu\text{atm}$.

Figure 4.3 (e) and (f) shows the deseasonalised (monthly timeseries minus mean seasonal cycle) for both stations and dashed lines represent the linear fit to the data. It can be clearly seen that there is substantial high frequency variability at both stations. The linear fit suggests a stronger increase in sea surface $p\text{CO}_2$ at BATS and a lower increase at ESTOC. The differences in the trend mainly stem from the first years of the analysis period where the least amount of data exist to train the neural network, however, given the strong variability in both the timeseries data and the neural network estimates, none of the linear trend lines are statistically different from 0.

As a last test, data from the recently updated SOCAT v2 database (Bakker *et al.*, 2014) are used, which provides new independent data points within the study period to validate the results. An additional 3065 gridded observations, spread over the entire Atlantic Ocean have been added for the study region from 1998 through 2007, equivalent to an extra 15% of the total number of data in the Atlantic Ocean used to train the neural network. Figure 4.4 shows the temporal mean and standard deviation of the residuals, similar to figure 4.1. The largest misfit between the network estimates and the SOCAT v2 observations can be identified along the Gulf Stream and North Atlantic Current, confirming that the method has difficulties to fully capture all variability within this region. Overall, the neural network estimates have a RMSE of $22.83 \mu\text{atm}$ and a bias of $4.85 \mu\text{atm}$. When excluding data north of 40°N , where the largest misfits are obtained, the results improve with a RMSE of $16.29 \mu\text{atm}$ and a mean difference of $-1.12 \mu\text{atm}$ similar to the numbers obtained from the independent timeseries stations. This suggests that over most of the ocean, the SOM–FFN method succeeds in predicting the observed $p\text{CO}_2$ at any given time and place to within about $22 \mu\text{atm}$, and a bias of a few μatm .

4.3 Uncertainty of the air–sea flux

The uncertainty of the air–sea CO_2 flux stems from the error in the estimated $\Delta p\text{CO}_2$ and the uncertainty of the gas transfer coefficient (Takahashi *et al.*, 2009). This uncertainty is estimated for the integrated flux over 4 RECCAP/Ocean Inversion regions (see table 4.2 for region borders Gurney *et al.*, 2008) rather than for each $1^\circ \times 1^\circ$ grid–cell.

The $\Delta p\text{CO}_2$ estimate is subject to two main sources of errors, i.e., the error derived

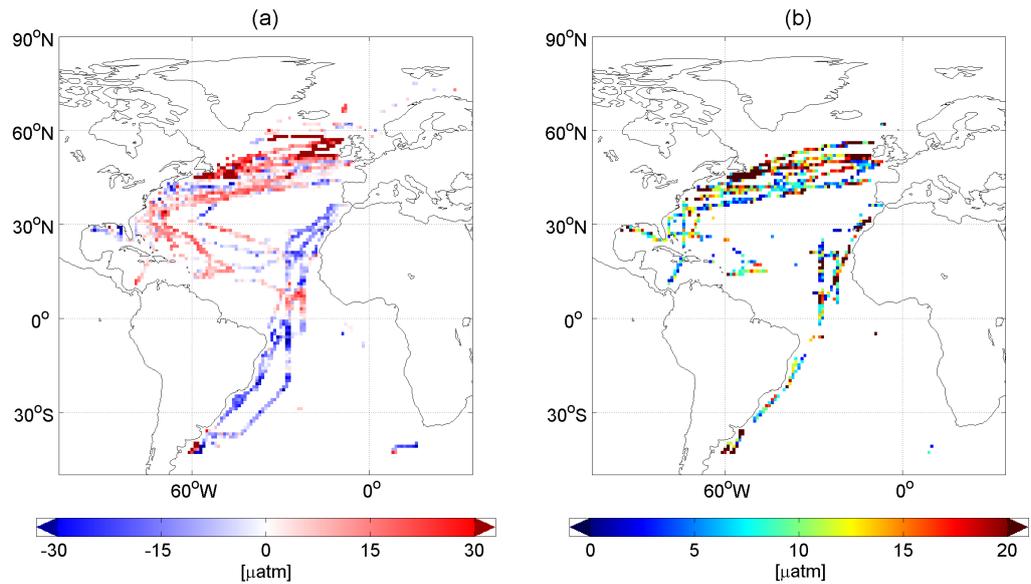


Figure 4.4: (a) Temporal mean residuals and (b) standard deviation of the residuals in μatm between neural network estimates and independent data points obtained from the SOCAT v2 gridded observations (Bakker *et al.*, 2014) for the period from 1998–2007. Pixels that have a value in (a) but not in (b) indicate where only 1 observation in time is available.

from discretizing the original observations into monthly $1^\circ \times 1^\circ$ bins and the error of the neural network method to interpolate the data in time and space. There are sources of uncertainty, such as the $p\text{CO}_2$ measurement error that are not accounted for here, as they are assumed to be small compared to accounted sources of uncertainty.

For the discretizing error associated with the gridding of the observations, a value of $5 \mu\text{atm}$ is used as reported by Sabine *et al.* (2013), while the RMSE value of about $10 \mu\text{atm}$ (see table 4.1) is adopted for the interpolation error. When computing next the error of the mean over a larger scale regions, it is inappropriate to assume that each of the estimates is independent, as these errors are spatially correlated, hence the effective number of degrees of freedom has to be adjusted to those grid boxes, that are not correlated.

To estimate the discretization error associated with gridding for each RECCAP/Ocean Inversion region, the spatial decorrelation length scale of 400 km estimated by Jones *et al.* (2012) is used to compute the effective number of degrees of freedom. The uncertainty of the mean is then estimated by dividing the standard deviation by the square root of this number. The effective number of degrees of freedom is the number of data points that are not correlated, i.e. those that have a distance further than the given decorrelation length. This results in an uncertainty between $1 \mu\text{atm}$ and $2 \mu\text{atm}$ for the individual regions.

To estimate the spatial mean of the neural network error for each RECCAP/Ocean Inversion region, the spatial correlation is estimated by analysing the semi–variogram of the $p\text{CO}_2$ residuals (see e.g. Kalkhan, 2011). For each RECCAP/Ocean Inversion region (see table 4.2 for region borders), the $p\text{CO}_2$ residuals are first divided into 5 randomly chosen mutually–exclusive ensembles, with the exception of the subtropical North Atlantic, where a total of 10 ensembles is used, due to the larger amount of data. These ensembles are random subsamples of the entire residual pool. For each ensemble, the semi–variance of the residuals and their point–to–point Haversine distance matrix is computed, and then an exponential function of the form:

$$y = a + b \cdot \exp\left(\frac{-x}{c}\right) \quad (4.1)$$

is fitted to the semi–variogram in order to estimate the correlation length (parameter c) between the residuals. The variables a , b and c are adjusted to minimize the misfit between the distance variable (x) and the semi–variance (y). The semi–variograms are very sensitive to extreme values of the residuals. Therefore, Chauvenet’s criterion (see e.g. Glover *et al.*, 2011) is used to reject extreme values prior to the computation and the fit. The potential biasing effect of their removal is accounted for by using several different ensembles per region.

Figure 4.5 shows the semi–variograms of all ensembles in the Atlantic Ocean. Correlation lengths of the residuals vary between 9 km, where the ensembles are well below the distance between 2 neighbouring grid boxes, and 532 km. However, in all cases the semi–variogram shows a large lag 0 correlation, (semi–variance at 0 km distance varies between 20–60 μatm^2 within the different ensembles in the different regions) indicating the residuals within one grid cell are correlated with each other, leading to a substantial reduction of the degrees of freedom. Similar to the discretizing error the uncertainty of the mean is then estimated by dividing the RMSE of $10\mu\text{atm}$ by the square root of the effective number of degrees of freedom, i.e. those grid boxes that have a distance further than the calculated decorrelation length. This results in an uncertainty estimate between different regions ranging from 1 μatm to nearly 4 μatm .

Adding the error from the gridding and the neural network together, and assuming a

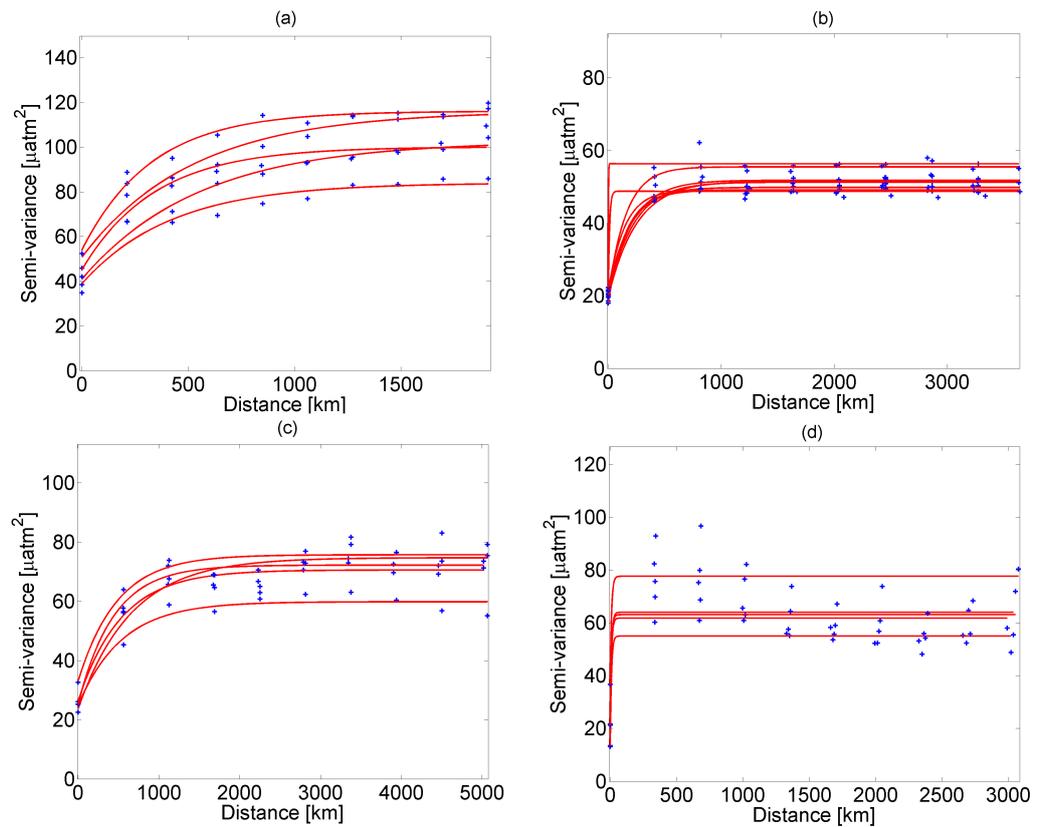


Figure 4.5: Empirical semi-variograms of randomly chosen ensembles of the residuals as a function of distance in (a) the subpolar North Atlantic with a median decorrelation length of 430 km, (b) subtropical North Atlantic with a median decorrelation length of 166 km, (c) the Equatorial Atlantic with a median decorrelation length of 532 km and (d) the South Atlantic with a median decorrelation length of 9 km. Region borders are listed in table 4.2.

mean error of $0.2 \mu\text{atm}$ for the atmospheric $p\text{CO}_2$ (Takahashi *et al.*, 2009), yields a total $\Delta p\text{CO}_2$ error for the 4 regions between 2 and $6 \mu\text{atm}$. With a mean gas transfer rate in the Atlantic Ocean of $0.05 \text{ mol C} \cdot \text{m}^{-2} \cdot \text{yr}^{-1} \cdot \mu\text{atm}^{-1}$ (Takahashi *et al.*, 2009) this results in a flux error between ± 0.03 and $\pm 0.06 \text{ Pg C} \cdot \text{yr}^{-1}$ for the individual RECCAP/Ocean Inversion regions and an overall basin error of $\pm 0.07 \text{ Pg C} \cdot \text{yr}^{-1}$ calculated by standard error propagation.

Furthermore, following Sweeney *et al.* (2007), a random error of 30% in the gas-transfer velocity is assumed. For the long term mean estimate of the Atlantic Ocean ($-0.45 \text{ Pg C} \cdot \text{yr}^{-1}$ from 44°S to 79°N and west of 30°E) the error due to the piston velocity uncertainty is $\pm 0.13 \text{ Pg C} \cdot \text{yr}^{-1}$. This results in a total uncertainty estimate for the Atlantic Ocean of $\pm 0.15 \text{ Pg C} \cdot \text{yr}^{-1}$, or roughly 33%, with the largest contribution stemming from the uncertain gas transfer velocity.

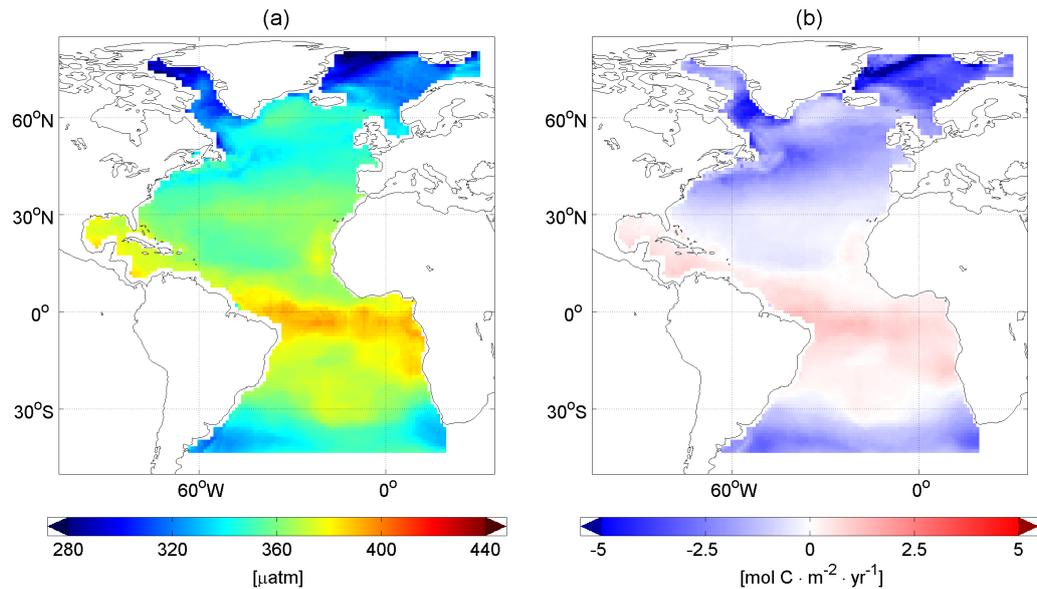


Figure 4.6: Maps of (a) the decadal mean surface ocean $p\text{CO}_2$ and (b) the CO_2 flux density in $\text{mol C} \cdot \text{m}^{-2} \cdot \text{yr}^{-1}$ for the Atlantic Ocean. Negative flux densities indicate CO_2 uptake by the ocean.

4.4 Decadal mean $p\text{CO}_2$ and air–sea CO_2 flux

The decadal mean sea surface $p\text{CO}_2$ (figure 4.6 (a)) shows lowest values in the northern North Atlantic, especially the Labrador Sea, the Greenland Sea and the Norwegian Sea with $p\text{CO}_2$ below $320 \mu\text{atm}$ and in the mid latitudes, along the Gulf Stream and North Atlantic Current (see figure 1.5) and in the South Atlantic south of 30°S . The highest $p\text{CO}_2$ values can be identified in the Equatorial Atlantic, in the North Atlantic along the North Equatorial Current and the tropical and subtropical South Atlantic northwards of 30°S . Further high values are estimated at 60°N around the Irminger Basin and 30°N in the subtropical North Atlantic along the Canaries Current.

The high equatorial $p\text{CO}_2$ area appears to be strongly asymmetric, in agreement with previous studies (Takahashi *et al.*, 2009). Bakker (1998) investigates the relation between the surface ocean circulation, as well as heat exchange, and the sea surface $f\text{CO}_2$ pattern. Bakker (1998) draws the conclusion, that northwards of the Equator, the northwards transport of high $f\text{CO}_2$ waters along the Equatorial Current, and rainfall within the inter-tropical convergence zone contribute to the asymmetric distribution. Southwards of the Equator the subtropical gyre experiences a heat gain, hence the high sea surface $f\text{CO}_2$,

whereas in the northern hemisphere, the subtropical gyre describes a net heat loss region. Similar as described in Bakker (1998), warming of surface waters increases the $p\text{CO}_2$ along the Canaries Current in the northern hemisphere, and the Benguela Current in the southern hemisphere. In contrast, North Atlantic Current waters move northwards and initially high $p\text{CO}_2$ decreases towards the north, linked to the cooling of the surface water.

The decadal mean $p\text{CO}_2$ distribution from the neural network method is generally very similar to that estimated by Takahashi *et al.* (2009), with some important exceptions (figure 4.7). To produce this comparison plot, the estimates are binned to the same resolution ($4^\circ \times 5^\circ$) as the original climatology of Takahashi *et al.* (2009). Furthermore, the estimates are corrected to the year 2000 by subtracting $4.5 \mu\text{atm}$ on the basis of the assumption that the surface ocean follows an atmospheric trend of $1.5 \mu\text{atm}$ per year (Takahashi *et al.*, 2009) and the fact that the estimate is centered around 2003. The strongest differences can be identified in the high latitudes of the North Atlantic within the Labrador Sea, the Greenland Sea and the Norwegian Sea. For the entire Atlantic a mean difference of $0.38 \mu\text{atm}$ and a RMSE of $6.45 \mu\text{atm}$ is calculated. In the South Atlantic, the largest differences occur, where the least amount of observations are available.

Given the overall small bias and the low RMSE between the two different methods to interpolate the data, it appears that the long-term mean surface ocean $p\text{CO}_2$ can be robustly estimated from the available observations.

The CO_2 flux density (figure 4.6 (b)) largely follows the $p\text{CO}_2$ pattern, although with some notable differences. Overall, the North Atlantic is a strong sink for atmospheric CO_2 in the mid and high latitudes, whereas the low latitudes act as a source for atmospheric CO_2 . The strongest CO_2 uptake in the North Atlantic occurs along the Gulf Stream and the North Atlantic Current, as well as in the Labrador, Norwegian, and Greenland Seas, and in the South Atlantic south of 30°S .

The estimates suggest a decadal mean flux of $-0.44 \pm 0.15 \text{ Pg C} \cdot \text{yr}^{-1}$ for the Atlantic Ocean from 44°S to 76°N and 100°W to 19°E ($-0.45 \pm 0.15 \text{ Pg C} \cdot \text{yr}^{-1}$ from 44°S to 79°N and west of 30°E). This is in good agreement with Schuster *et al.* (2013) who provided a "best" estimate of $-0.49 \pm 0.11 \text{ Pg C} \cdot \text{yr}^{-1}$ (derived from the mean fluxes of the $p\text{CO}_2$ climatology and the Ocean Inversion fluxes within the RECCAP project).

Table 4.2: Comparison of regional and basin-wide decadal mean CO_2 fluxes in $\text{Pg C} \cdot \text{yr}^{-1}$ from the neural network-based method (column VIII) compared to a range of other methods. This includes (I) the $p\text{CO}_2$ climatology of Takahashi *et al.* (2009) and the Tier 1 methodologies described in Schuster *et al.* (2013) which include (II) Ocean Inversions (Gruber *et al.*, 2009), (III) Atmospheric Inversions (Peylin *et al.*, 2013), (IV) Ocean Biogeochemical Models, as well as observation-based results including (V) a SOCAT v1.5 based multi parameter regression (Schuster *et al.*, 2013) and (VI) an estimate based on the $p\text{CO}_2$ database of Takahashi *et al.* (2009) updated by McKinley *et al.* (2011). (VII) lists the best estimate from combining the different RECCAP methodologies as described in Schuster *et al.* (2013)

| Region | (I) $p\text{CO}_2$ climatology [Pg C/yr] | (II) Ocean Inversion [Pg C/yr] | (III) Atmospheric Inversion [Pg C/yr] | (IV) OBGC models [Pg C/yr] | (V) SOCAT MPR [Pg C/yr] | (VI) $p\text{CO}_2$ database [Pg C/yr] | (VII) RECCAP best estimate [Pg C/yr] | (VIII) Neural Network [Pg C/yr] |
|-----------------------------|---|---|--|-------------------------------------|----------------------------------|---|---|--|
| 49°N-76°N (West of 19°E) | -0.23 ±0.12 | -0.19 ±0.06 | -0.28 ±0.03 | -0.17 ±0.02 | -0.07 ±0.04 | -0.30 ±0.13 | -0.21 ±0.06 | -0.20 ±0.07 |
| 18°N-49°N | -0.19 ±0.09 | -0.34 ±0.08 | -0.31 ±0.03 | -0.13 ±0.03 | -0.18 ±0.09 | -0.24 ±0.16 | -0.26 ±0.06 | -0.19 ±0.07 |
| 18°S-18°N | 0.11 ±0.05 | 0.13 ±0.06 | 0.12 ±0.05 | 0.15 ±0.06 | 0.10 ±0.05 | 0.12 ±0.14 | 0.12 ±0.04 | 0.11 ±0.07 |
| 44°S-18°S | -0.10 ±0.05 | -0.17 ±0.05 | -0.13 ±0.02 | -0.17 ±0.01 | -0.25 ±0.12 | -0.21 ±0.23 | -0.14 ±0.04 | -0.16 ±0.06 |
| Atlantic Ocean 44°S-76°N | -0.42 ±0.17 | -0.56 ±0.13 | -0.60 ±0.07 | -0.32 ±0.07 | -0.40 ±0.16 | -0.63 ±0.34 | -0.49 ±0.11 | -0.44 ±0.15 |

Table 4.2 lists the long–term mean fluxes for the Atlantic Ocean as well as for the four individual ocean RECCAP/Ocean Inversion regions considered by Schuster *et al.* (2013). While the basin average flux is well within the uncertainty range of the best estimate from Schuster *et al.* (2013), the subtropical North Atlantic (18°N to 49°N) mean flux is just outside the uncertainty range of the RECCAP best estimate, however their combined uncertainties overlap. In general, the neural network fluxes are close to those of the $p\text{CO}_2$ climatology of Takahashi *et al.* (2009) with the exception of the subtropical South Atlantic (44°S to 18°S) where the long term mean flux is closest to the results of the Ocean Inversion and the Ocean Biogeochemical Models. The main carbon sink region is estimated to be the high latitude North Atlantic with strong uptake throughout the year and a decadal average uptake of $-0.20 \pm 0.07 \text{ Pg C} \cdot \text{yr}^{-1}$, with general agreement between methodologies.

It has to be noted here that there is a mix between methods, and they are strictly not comparable (descriptions as well as strengths and weaknesses of each method individually have been introduced in section 1.2.1 and section 1.6). Observation–based estimates (e.g. Takahashi *et al.*, 2009) and atmospheric inversions provide an estimate of the contemporary air–sea flux, while process based model estimates provide an estimate of the anthropocentric air–sea CO_2 flux including carbon derived from river output. One would therefore expect somewhat stronger model fluxes, however, table 4.2 reveals the weakest carbon uptake by the ocean models. On the one hand this might indicate that the actual discrepancy between model and observation based estimates is actually larger than it appears in table 4.2, but on the other hand also could indicate that the Atlantic Ocean river input of carbon is much smaller than estimated on a global scale.

In chapter 2, four sensitivity runs were introduced, where the input data products were replaced, in order to test the sensitivity of the results to the product chosen. Comparing the decadal mean flux of $-0.45 \pm 0.15 \text{ Pg C} \cdot \text{yr}^{-1}$ (from 44°S to 79°N and west of 30°E) to the results derived from the sensitivity runs SR1–4 reveals that the choice of input data products does not significantly influence the long term mean result (see table 4.3). The decadal mean fluxes from the sensitivity runs range from -0.41 ± 0.14 (SR2) up to $-0.48 \pm 0.16 \text{ Pg C} \cdot \text{yr}^{-1}$ (SR4) and are therefore well within the estimated uncertainty range.

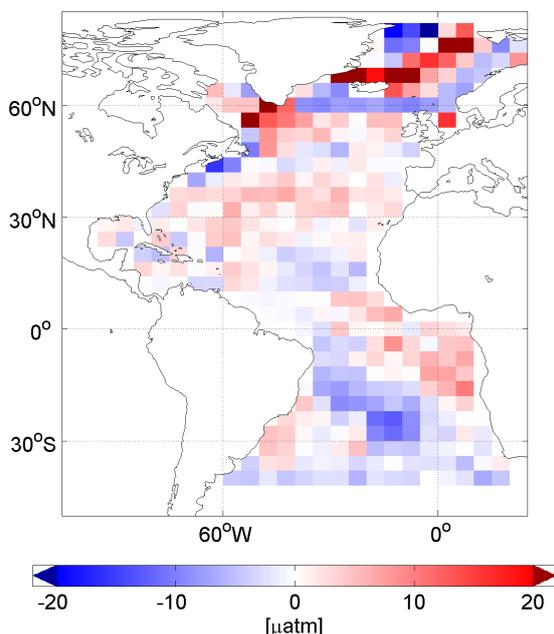


Figure 4.7: Difference in the surface ocean $p\text{CO}_2$ in μatm between the decadal mean neural network estimates (this study), corrected to the year 2000, and the estimates from the climatology of Takahashi *et al.* (2009). Positive differences indicate higher $p\text{CO}_2$ for the neural network estimates. The neural network estimates have been interpolated to the $4 \times 5^\circ$ grid of the Takahashi climatology.

Table 4.3: Summary of the sensitivity runs as described in section 2.1.2 (SR1 – SSS climatology; SR2 – MLD climatology; SR3 – SODA SST and SR4 – no chlorophyll-*a*) in comparison to the SOM-FFN standard run, including temporal mean CO_2 flux, flux trends and the standard deviation of the inter-annual variability ($\text{std}(\text{IAV})$). Results indicate that the sensitivity runs are within the combined uncertainty, i.e., are not statistically different from the standard run, with the exception of the CO_2 flux trend of SR2, indicating the sensitivity of the method to changes in the mixed layer product.

| sensitivity run | temporal mean CO_2 flux [$\text{Pg C} \cdot \text{yr}^{-1}$] | CO_2 flux trend [$\text{Pg C} \cdot \text{yr}^{-1} \cdot \text{decade}^{-1}$] | $\text{std}(\text{IAV})$ [$\text{Pg C} \cdot \text{yr}^{-1}$] |
|-----------------|--|---|--|
| standard run | -0.45 ± 0.15 | -0.15 ± 0.04 | ± 0.036 |
| SR1 | -0.44 ± 0.14 | -0.19 ± 0.03 | ± 0.029 |
| SR2 | -0.41 ± 0.14 | -0.26 ± 0.03 | ± 0.031 |
| SR3 | -0.44 ± 0.15 | -0.14 ± 0.04 | ± 0.036 |
| SR4 | -0.48 ± 0.16 | -0.23 ± 0.04 | ± 0.036 |

4.5 Seasonality

The seasonal cycle of the 10-year mean sea surface $p\text{CO}_2$ exhibits strong latitudinal differences (figure 4.8). The weakest seasonal signals are found north of 60°N , south of 40°S and near the Equator from 10°S to 10°N . The temperate North Atlantic (40° to 60°N) has a distinct seasonal cycle with high $p\text{CO}_2$ from October to April and low values from May

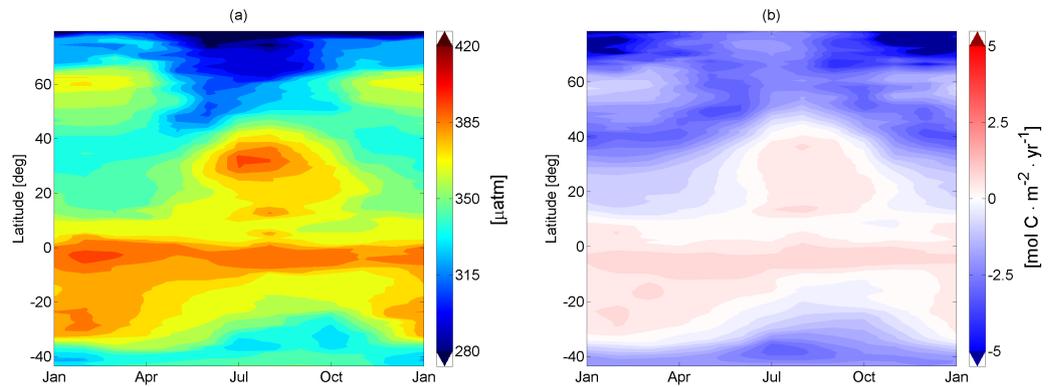


Figure 4.8: Hovmöller plot of the zonally averaged long term mean seasonal cycle of (a) $p\text{CO}_2$ in μatm and (b) the CO_2 flux density in $\text{mol C} \cdot \text{m}^{-2} \cdot \text{yr}^{-1}$. Negative (blue) fluxes indicate an ocean CO_2 sink.

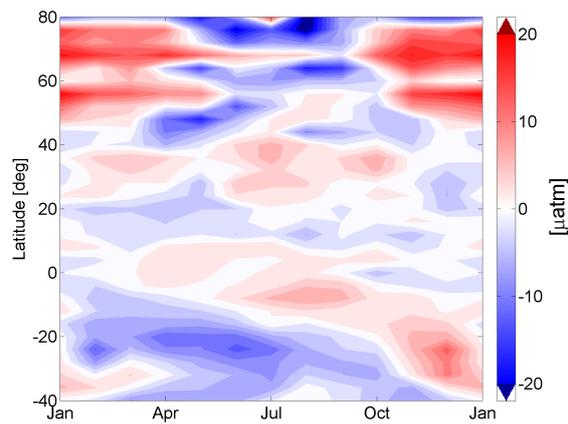


Figure 4.9: Zonally averaged difference in the long term mean seasonal cycle of the surface ocean $p\text{CO}_2$ between the neural network estimates (this study), corrected to the year 2000 and those from the climatology of Takahashi *et al.* (2009). Positive differences indicate higher partial pressures of CO_2 in the neural network based estimates.

to September. By contrast, the subtropical Atlantic between 10°N to 40°N and 10°S to 40°S has low partial pressures in winter and a seasonal maximum in the warmer summer months.

The mean seasonal cycle of the neural network–based estimates of the sea surface $p\text{CO}_2$ agrees relatively well with the seasonal cycle estimated by Takahashi *et al.* (2009), but substantial differences exist at the regional level (figure 4.9). Again, the SOM–FFN estimates are corrected to the year 2000 in order to be comparable. The strongest difference can be identified in the high latitude North Atlantic, where a stronger seasonal cycle

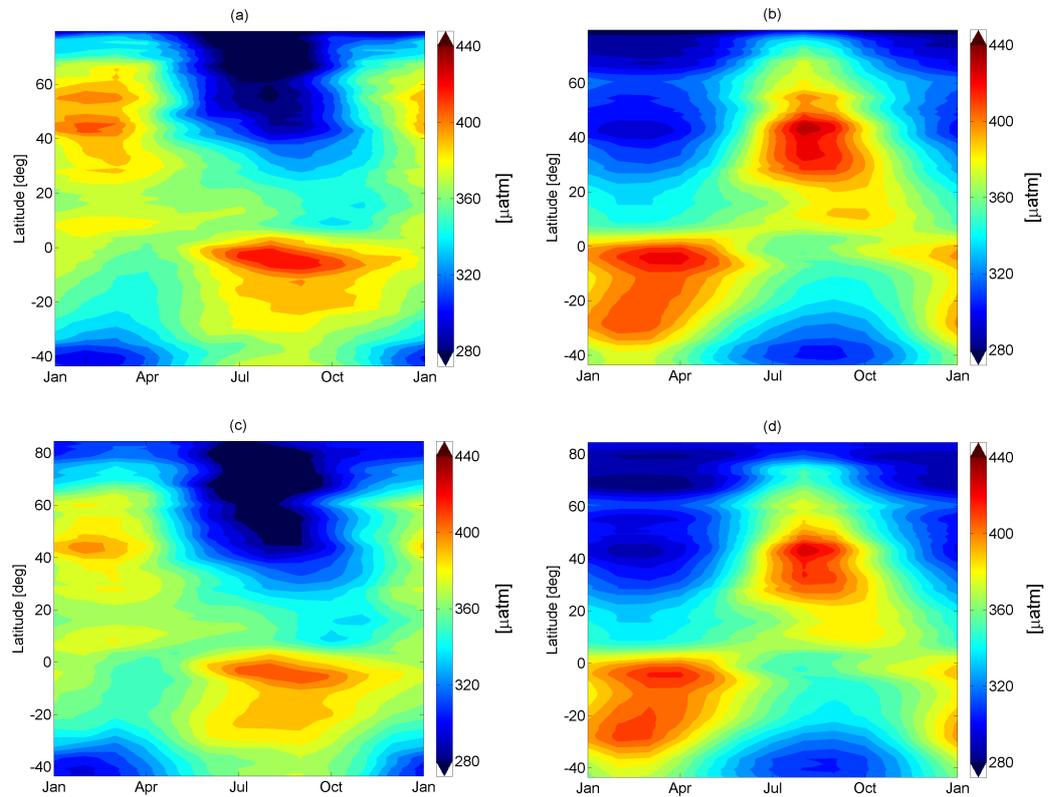


Figure 4.10: Hovmöller plots of the mean seasonal cycle of the zonally averaged sea surface $p\text{CO}_2$ driven by (a) the non-thermal component, such as changes in circulation, mixing, and biology, and (b) the thermal component, i.e., the changes in temperature. Figure (c) and (d) show the non-thermal and thermal component respectively for the Takahashi *et al.* (2009) climatology. The decadal mean $p\text{CO}_2$ has been added to both components.

is estimated by the SOM-FFN compared to the $p\text{CO}_2$ climatology. The SOM-FFN values are higher in winter and lower in summer, and differences exceed $10 \mu\text{atm}$. In comparison, the differences throughout the Atlantic Ocean are mostly within the calculated RMSE of the method.

To determine the drivers behind the seasonal cycles, the long-term mean seasonal cycle of the SOM-FFN results and the Takahashi *et al.* (2009) climatology are split in each grid cell into a thermal and into a non-thermal component (see equations 1 and 2 in Takahashi *et al.*, 2002).

The former is driven by the seasonal changes in temperature and is computed on the basis of the well known temperature sensitivity of $p\text{CO}_2$ (see equation 2 in Takahashi *et al.*, 1993), i.e., assuming a 4% change in $p\text{CO}_2$ per unit change in SST. The same SST product as used for the network training is employed for the computation of the thermal

component. The non-thermal component is computed by the difference between the mean seasonal cycle and the thermal component.

The seasonal cycles of the thermally and non-thermally driven partial pressures tend to cancel each other (figure 4.10), consistent with previous analyses (Takahashi *et al.*, 2002; Sarmiento and Gruber, 2006). In both hemispheres the non-thermally driven $p\text{CO}_2$ decreases in the warmer summer months due to increasing biological production and reduced vertical mixing resulting in increased stratification. The thermally driven seasonal cycle on the other hand follows the increase in sea surface temperature and causes an increase in the sea surface $p\text{CO}_2$ due to a reduced solubility of CO₂. Comparing figure 4.10 with figure 4.8 reveals that the non-thermal seasonal cycle of the sea surface $p\text{CO}_2$ dominates over the thermally driven seasonal cycle polewards of 40°N. In contrast, the seasonal cycle in the subtropical North and South Atlantic is driven by the thermal $p\text{CO}_2$ component. The thermal and non-thermal driven seasonal cycle of the equatorial band and in the South Atlantic south of 35°S compensate each other, resulting in little seasonal variability within each band. This is in good agreement with the climatology of Takahashi *et al.* (2009) illustrated in figure 4.10 (c) and (d).

The seasonal cycle of the CO₂ flux is largely driven by the seasonal cycle of the sea surface $p\text{CO}_2$ with only modest modifications by the seasonal cycles of wind and solubility (figure 4.8 (b)). In large areas of the high latitude North Atlantic (~ 60°N) in winter, the sea surface $p\text{CO}_2$ is supersaturated, whereas in spring and summer the sea surface $p\text{CO}_2$ strongly decreases due to biological activity. This is less pronounced in the air-sea flux, due to the seasonal variability of atmospheric $p\text{CO}_2$ and winds. The most negative flux densities, i.e. the strongest uptake of CO₂ by the ocean, can be identified in most northern latitudes and in the subtropics of the northern and southern hemisphere in winter. The neural network estimates shows a strong seasonal CO₂ outgassing in summer in the northern subtropics, driven by the increasing $p\text{CO}_2$, with a 6-month difference in the southern hemisphere.

4.6 CO₂ trends and inter-annual variability

The main driving variable for trends in the sea surface $p\text{CO}_2$ is the atmospheric CO₂, but within the study period the neural network estimates show that these trends are not

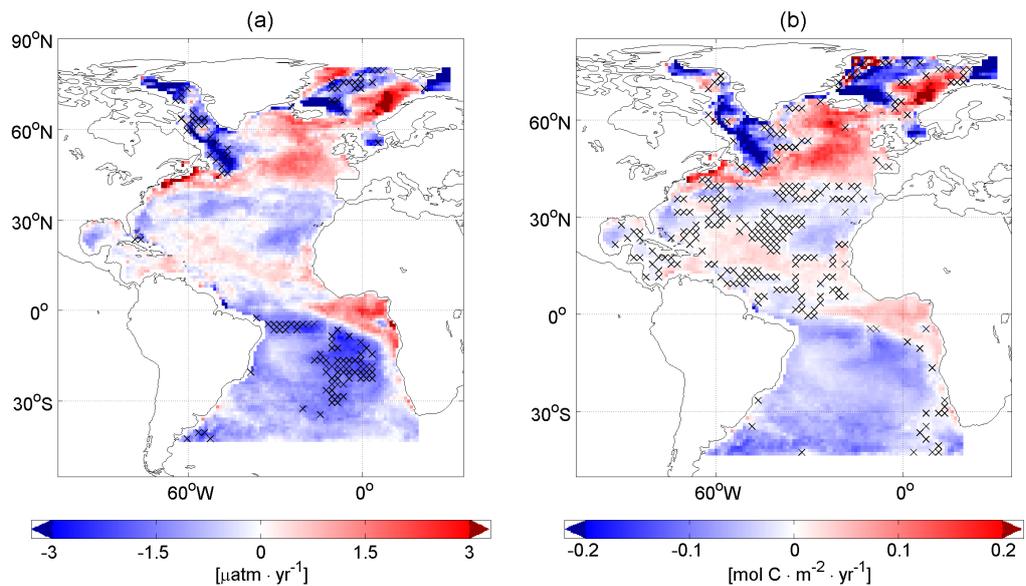


Figure 4.11: Maps of linear trends (a) in sea surface $p\text{CO}_2$ relative to that in the atmosphere and (b) in the CO_2 flux over the period 1998–2007. The relative trend in sea surface $p\text{CO}_2$ was computed by subtracting the atmospheric mean trend. Areas with cross-hatch indicate where the trend is outside the 95% confidence level ($p \geq 0.05$). Trends were derived by first applying a 12 month running mean to each pixel to deseasonalize the data and then calculating the slope of a linear fit to these deseasonalized data.

in parallel. Across large areas of the Atlantic, the 10-year trend of surface ocean $p\text{CO}_2$ is estimated to be lower than that of atmospheric $p\text{CO}_2$ (figure 4.11 (a)), but there are notable exceptions. In this plot, the atmospheric $p\text{CO}_2$ trend has been subtracted from the long-term mean sea surface $p\text{CO}_2$ trend for each $1^\circ \times 1^\circ$ pixel, so that positive values indicate a rate of sea surface $p\text{CO}_2$ increase faster than of the atmosphere and vice versa for negative values. Table 4.1 shows that the bias between estimates and observations is fairly constant at each year individually, suggesting that trends are captured well where observations exist.

The strongest increase in ocean surface $p\text{CO}_2$ relative to that in the atmosphere is found in the North Atlantic poleward of 40°N along the Gulf Stream and North Atlantic Current. Here, the neural network suggests an increase in sea-surface $p\text{CO}_2$ of more than twice the atmospheric increase. Metzl *et al.* (2010) investigated the sink trend over a similar time period (2001 to 2008) in the North subpolar gyre (53°N to 57.5°N , 45°W to 35°W and 57.5°N to 62°N , 40°W to 25°W). These authors found a particularly strong increase in the winter sea surface $f\text{CO}_2$ of 5.8 ± 1.1 and $7.2 \pm 1.3 \mu\text{atm} \cdot \text{yr}^{-1}$. While this is much stronger than suggested here (see figure 4.11), both studies agree on the North

Atlantic subpolar gyre having a trend towards a stronger increase of the sea surface $p\text{CO}_2$ over the ~ 2000 to ~ 2007 period. McKinley *et al.* (2011) found a similar trend toward a weaker undersaturation in their subtropical seasonally-stratified region around 40°N for the period 1993 until 2005, but did not identify a significant trend in the subpolar gyre over the same period. This may reflect differences in the time period, as their analyses with an earlier start, i.e., pre 1990, also suggest a trend toward a weaker undersaturation.

Decadal trends in surface ocean $p\text{CO}_2$ in the Labrador Sea and some parts of the Icelandic Sea were much smaller than in the atmosphere, leading to an overall small trend for the entire region north of 40°N . As the low latitudes of the North Atlantic (0 to 40°N) have close to zero trend relative to the atmosphere, the entire North Atlantic $p\text{CO}_2$ trend of $1.80 \pm 0.77 \mu\text{atm} \cdot \text{yr}^{-1}$ is also very close to that of the atmosphere of $1.90 \pm 0.34 \mu\text{atm} \cdot \text{yr}^{-1}$ from 1998 through 2007.

As expected from the uptake of anthropogenic CO₂ by the surface ocean, the majority of the ocean $p\text{CO}_2$ trend stems from the non-thermal part, i.e., the increase in surface ocean DIC. Splitting the trend into thermal and non-thermal component shows on average a linear trend of $1.46 \pm 1.75 \mu\text{atm} \cdot \text{yr}^{-1}$ for the non-thermal component, while the thermally driven trend is on average $0.37 \pm 1.47 \mu\text{atm} \cdot \text{yr}^{-1}$. However, given their uncertainty, they are not statistically different from zero. The small difference in the increase in surface ocean $p\text{CO}_2$ relative to the atmosphere results in an almost steady strength of the Atlantic carbon sink over time north of the Equator ($-0.01 \pm 0.02 \text{Pg C} \cdot \text{yr}^{-1} \cdot \text{decade}^{-1}$).

Trends for the South Atlantic show a weaker increase in the sea surface $p\text{CO}_2$ relative to the atmosphere with the exception of the eastern South Atlantic and parts along the South American coast. On average, surface ocean $p\text{CO}_2$ increased only by $0.98 \pm 0.97 \mu\text{atm} \cdot \text{yr}^{-1}$ over the 1998 through 2007 period, resulting in a carbon sink increase of $-0.14 \pm 0.02 \text{Pg C} \cdot \text{yr}^{-1} \cdot \text{decade}^{-1}$. Similar to the North Atlantic, the non-thermal component of the $p\text{CO}_2$ with an average trend of $0.76 \pm 1.30 \mu\text{atm} \cdot \text{yr}^{-1}$ appears to be stronger compared to $0.19 \pm 0.79 \mu\text{atm} \cdot \text{yr}^{-1}$ of the thermal component, but given their uncertainty, both trends are again indistinguishable from zero.

Taking the North and South Atlantic together, the trend over the entire study region is one toward a stronger sink over the 10-year period with an overall mean trend of $1.46 \pm 0.76 \mu\text{atm} \cdot \text{yr}^{-1}$ and a flux trend of $-0.15 \pm 0.04 \text{Pg C} \cdot \text{yr}^{-1} \cdot \text{decade}^{-1}$.

The sensitivity runs (see table 4.3) reveal that trend estimates are barely influenced by the choice of the input data product, with the exception of SR2, i.e., where the ECCO2 MLD product was replaced with the de Boyer Montegut (de Boyer Montegut *et al.*, 2004) MLD climatology. While $p\text{CO}_2$ trends are statistically indistinguishable between the neural network estimate ($1.46 \pm 0.76 \mu\text{atm} \cdot \text{yr}^{-1}$) and SR1–4 (1.42 ± 0.59 , 1.25 ± 0.48 , 1.48 ± 0.77 and $1.37 \pm 0.73 \mu\text{atm} \cdot \text{yr}^{-1}$ respectively), this is not always true for the fluxes. Here, SR2 reveals a flux trend ($-0.26 \pm 0.03 \text{Pg C} \cdot \text{yr}^{-1} \cdot \text{decade}^{-1}$) outside the combined uncertainty interval with the neural network estimate ($-0.15 \pm 0.04 \text{Pg C} \cdot \text{yr}^{-1} \cdot \text{decade}^{-1}$).

It is not possible to conclude from the data whether the 10-year trends identified here are part of truly long-term trends (30 years or longer), or whether they are part of decadal time-scale fluctuations (Thomas *et al.*, 2008; Gruber, 2009; McKinley *et al.*, 2011). The most recent studies by McKinley *et al.* (2011) and Fay and McKinley (2013) suggest the latter to be the case. The authors show that short term trends on timescales similar to this study are significantly influenced by the chosen start and end year and strongly reflect climate mode signals such as ENSO and NAO. However, 50-year trends in heat storage (Levitus *et al.*, 2012) and interior ocean oxygen changes in the North Atlantic (Stendardo and Gruber, 2012) indicate that this region has been subject to multi-decadal changes, particularly in the subpolar gyre. It is also tempting to point out that the resulting pattern of a decreasing sink in large areas of the North Atlantic and an increasing sink in the South Atlantic appears to be mirrored in the observation of a faster rate of accumulation in anthropogenic CO₂ in the South compared to the North Atlantic (Wanninkhof *et al.*, 2010). One needs to be careful, though, as the surface ocean trends are for the sum of natural and anthropogenic CO₂, while the ocean interior trends are for anthropogenic CO₂ only.

The largest year-to-year variability in sea surface $p\text{CO}_2$ is found within the North Atlantic north of 40°N and in the eastern Equatorial and South Atlantic. In contrast, the subtropics in both hemispheres show much less year-to-year variability.

Integrating the monthly air-sea CO₂ flux estimates for each year over the Atlantic Ocean reveals the largest year-to-year differences during the second half of the study period (figure 4.12 (a)), where annual mean fluxes range from $-0.39 \pm 0.13 \text{Pg C} \cdot \text{yr}^{-1}$

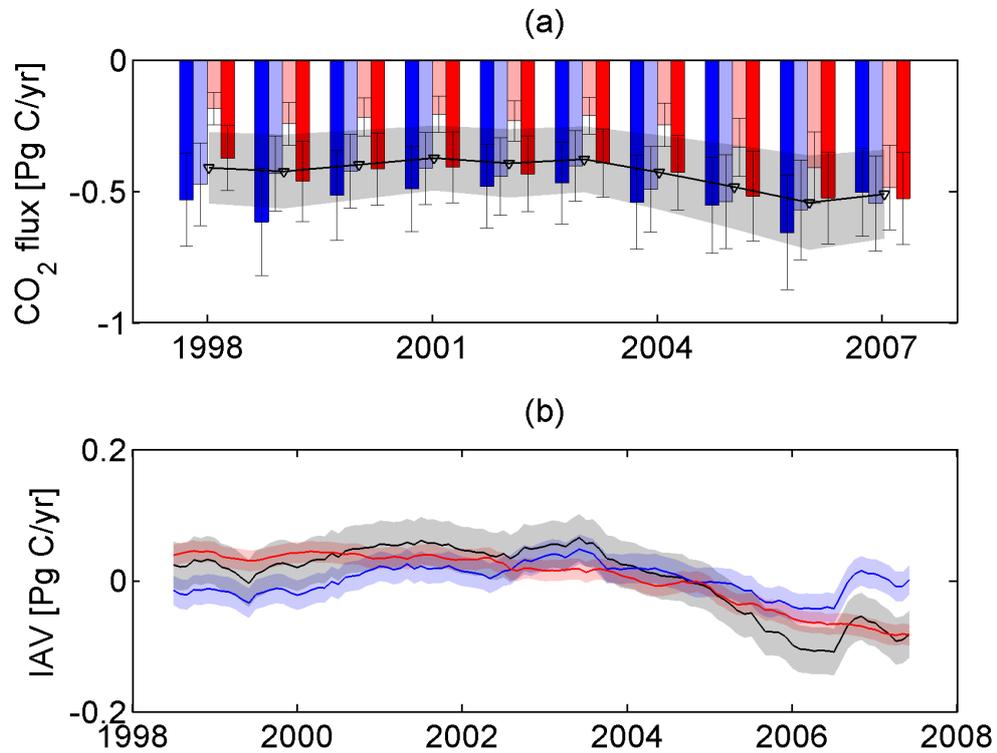


Figure 4.12: (a) Seasonal and annual mean fluxes from 1998–2007 in the Atlantic Ocean (44°S to 79°N and west of 30°E). Dark blue shows the results for the northern hemisphere winter months (DJF), light blue the spring months (MAM), light red the summer months (JJA), dark red the autumn months (SON). The annual mean flux is plotted as a black line on top. (b) inter-annual variability (calculated using a 12 month running mean) for the northern hemisphere (blue line), the southern hemisphere (red line) and the entire Atlantic Ocean (black line). The decadal mean has been removed from the IAV.

in 2001 up to $-0.56 \pm 0.18 \text{ Pg C} \cdot \text{yr}^{-1}$ in 2006. Figure 4.12 (b) illustrates the inter-annual variabilities (IAV) for both hemispheres and the entire Atlantic Ocean. The IAV, calculated as a 12 month running average, is fairly constant from 1998 to 2004 with a weak flux decrease in the northern hemisphere counterbalanced by a weak increase in the southern hemisphere. After 2004 the Atlantic Ocean sink increases mainly due to increases in the southern hemisphere. The standard deviations of the IAV (calculated as a 12 month running average and further detrended) for the Atlantic Ocean north of the Equator, south of the Equator and the entire basin are $\pm 0.02 \text{ Pg C} \cdot \text{yr}^{-1}$, $\pm 0.02 \text{ Pg C} \cdot \text{yr}^{-1}$ and $\pm 0.04 \text{ Pg C} \cdot \text{yr}^{-1}$ respectively, which indicates limited inter-annual variability in the Atlantic Ocean in both hemispheres.

This Atlantic Ocean low variability is further confirmed by the sensitivity runs (see

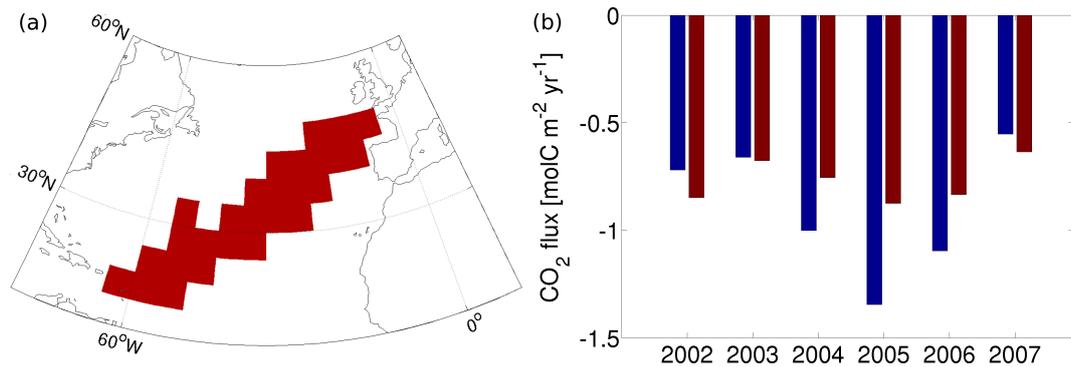


Figure 4.13: Annual mean fluxes within (a) $5^\circ \times 5^\circ$ boxes along the UK-Caribbean line as used by Watson *et al.* (2009). Blue bars in (b) indicate the annual average CO₂ flux density co-located to observations and red bars indicate the average flux density for the entire area in (a). Both red and blue bars are calculated from the SOM-FFN estimates presented in this chapter.

table 4.3), ranging from $\pm 0.03 \text{ Pg C} \cdot \text{yr}^{-1}$ (SR1, SR2) to $\pm 0.04 \text{ Pg C} \cdot \text{yr}^{-1}$ (SR3, SR4), indicating that the result is not sensitive with regards to the data choice. This shows that the main findings are statistically indistinguishable from those derived without chlorophyll-*a* (SR4), indicating the possibility to expand the analysis period back in time in future studies.

Somewhat in contrast to the basin-wide findings here, Watson *et al.* (2009) found strong variability along the UK-Caribbean line based on in-situ observations between 2002 through 2007. Figure 4.13 (b) illustrates the findings of this study for the same region (illustrated in figure 4.13 (a)) as investigated in Watson *et al.* (2009). Blue bars indicate the average flux density for only those grid boxes where co-located observations exist, to be comparable to Watson *et al.* (2009), whereas red bars show the average flux density for the entire area, both estimated by the SOM-FFN outputs. Blue bars indicate roughly $0.2\text{-}0.3 \text{ mol C} \cdot \text{m}^{-2} \cdot \text{yr}^{-1}$ larger uptake fluxes than identified in Watson *et al.* (2009), which can be explained by the difference in wind products (CCMP in this study compared to NCEP/NCAR in Watson *et al.* (2009)) and gas transfer formulations (Sweeney *et al.* (2007) in this study compared to Nightingale *et al.* (2000) in Watson *et al.* (2009)). Year-to-year variabilities within the blue bars, however, compare well with the findings in Watson *et al.* (2009), indicating strong variability in the observations that are not reflected in the basin-wide estimates. This is further confirmed by the red bars. When observations are extrapolated to the entire area in figure 4.13 (a), a large amount of variability is lost, resulting in a damping of the uptake minimum minus uptake maximum from $\sim 0.8 \text{ mol}$

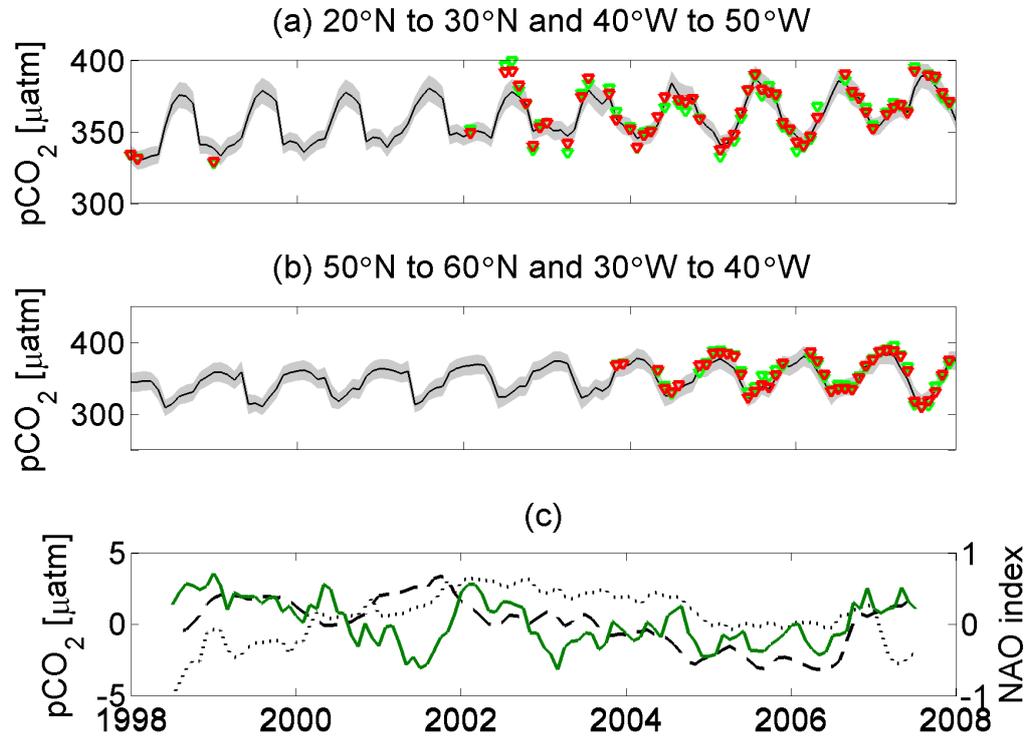


Figure 4.14: (a) Temporal evolution of the $p\text{CO}_2$ (in μatm) in the subtropical (20°N to 30°N ; 40°W to 50°W) and (b) in the subpolar box (50°N to 60°N ; 30°W to 40°W). The black line shows the spatial average $p\text{CO}_2$ within each $10^\circ \times 10^\circ$ box. Red triangles illustrate the average sea surface $p\text{CO}_2$ from the gridded SOCAT v1.5 database, within each box where observations are available and the green triangles represent the average of the neural network $p\text{CO}_2$ of those $1^\circ \times 1^\circ$ pixels which have co-located $p\text{CO}_2$ observations in SOCAT v1.5. (c) $p\text{CO}_2$ anomalies (left axis in μatm – detrended and smoothed using a 12 month running average filter) of the SOM-FFN estimates in (a) and (b) compared to the NAO index (right axis – smoothed using a 12 month running average filter). The dashed line shows the anomaly for the subtropical box, the dotted line for the subpolar box and the green line illustrates the NAO index.

$\text{C} \cdot \text{m}^{-2} \cdot \text{yr}^{-1}$ (blue bars) to $\sim 0.2 \text{ mol C} \cdot \text{m}^{-2} \cdot \text{yr}^{-1}$ (red bars). On the one hand, this might indicate, that results over larger areas are "too smooth" and do not reflect the observed variability, however, on the other hand this also indicates that temporal and spatial heterogeneity of the observations plays an important role and needs to be accounted for, when estimating the year-to-year variability of the integrated CO₂ flux.

Inter-annual variability of the sea-surface $p\text{CO}_2$ in the North Atlantic has previously been linked to variations in the North Atlantic Oscillation (NAO) (e.g. Gruber *et al.*, 2002; Schuster and Watson, 2007; Thomas *et al.*, 2008). Here, the effect of the NAO is investigated by focusing on two $10^\circ \times 10^\circ$ boxes, one located in the subtropical North Atlantic (20°N to 30°N and 40°W to 50°W) and the other in the subpolar North Atlantic (50°N

to 60°N and 30°W to 40°W). Figure 4.14 illustrates the $p\text{CO}_2$ estimate and their anomalies for each box together with the NAO index. Results show a weak, but significant ($p \leq 0.05$) positive correlation in the subtropics ($R=0.32$) and a negative correlation in the subpolar box ($R=-0.31$). This pattern is consistent with that identified by Thomas *et al.* (2008) on the basis of a modeling study (see also summary by Gruber (2009), the recent multi-model analysis by Keller *et al.* (2012), and time-series analyses (e.g. Gruber *et al.*, 2002; Bates, 2007) for the BATS site). The correlation patterns are derived from the neural network estimates, hence the NAO signal stems from the signal of the input data. Clearly, an important driver are the NAO-associated SST anomalies, but these are strongly modified by the various physical and biogeochemical changes that are driven by the NAO-induced changes in heat fluxes and windstress (see e.g. Keller *et al.*, 2012).

4.7 Summary and conclusion

A novel 2-step neural network approach is used to create monthly ocean surface $p\text{CO}_2$ fields from 1998 through 2007 in the Atlantic Ocean using the SOCAT v1.5 $p\text{CO}_2$ measurements. Independent testing indicates that the estimates are accurate to within 22.8 μatm for the entire Atlantic Ocean, with an improved fit of 16.2 μatm for data south of 40°N. The results of this study suggest a decadal mean CO_2 flux from 1998 through 2007 of $-0.45 \pm 0.15 \text{ Pg C} \cdot \text{yr}^{-1}$ for the Atlantic Ocean from 44°S to 79°N and west of 30°E. This result is in good accordance with the recent assessment from the RECCAP project (Schuster *et al.*, 2013). The strongest seasonal variability in the predicted sea surface $p\text{CO}_2$ and air-sea fluxes was identified within the subtropics in the northern and southern hemisphere, i.e., the zones where the temperature effect dominates the seasonal cycle of sea surface $p\text{CO}_2$.

Trends in sea surface $p\text{CO}_2$ suggest that in large areas poleward of 40°N the rate of increase in oceanic $p\text{CO}_2$ was faster than the atmosphere, leading to a regional weakening of the carbon sink strength. However, this is counterbalanced on the basin scale by weaker surface ocean $p\text{CO}_2$ trends elsewhere in the North Atlantic. The South Atlantic, in contrast to the North Atlantic, shows an increasing carbon sink strength throughout the study period. In total, the Atlantic Ocean carbon sink increased by $-0.15 \pm 0.04 \text{ Pg C} \cdot \text{yr}^{-1} \cdot \text{decade}^{-1}$ during 1998–2007. The standard deviation of the inter-annual variability of the

spatially integrated CO₂ flux within the study period in the Atlantic Ocean was $\pm 0.04 \text{ Pg C} \cdot \text{yr}^{-1}$ with both low inter-annual variability in the southern hemisphere ($\pm 0.02 \text{ Pg C} \cdot \text{yr}^{-1}$) and in the northern hemisphere ($\pm 0.02 \text{ Pg C} \cdot \text{yr}^{-1}$).

It would be beneficial to extend the study period to further investigate responses to climate modes such as the NAO and to investigate multi-decadal variabilities. The absence of chlorophyll-*a* permits to prolong the analysis period and this option will be explored in subsequent work. However, chlorophyll-*a* is a simple, but important proxy representing the relation between biology and $p\text{CO}_2$ and our results provide no evidence that chlorophyll can be neglected when considering longer timescales. Chlorophyll-*a* is available from models before the launch of satellite observations, but these products to-date have not achieved sufficient reliability.

The $p\text{CO}_2$ product shows that the data collection and synthesis effort of the marine carbon community makes it possible to investigate the seasonal to inter-annual variability of the ocean carbon sink on a basin-scale based on observations. Future measurements are expected to increase the accuracy of such observation-based estimates, and further improvements of the methods used to model the observations will result in providing better historical estimates and more accurate products for these important fluxes.

Chapter 5

Variability of the global ocean carbon sink

”The brightest flashes in the world of thought are
incomplete until they have been proven to have
their counterparts in the world of fact.”

(John Tyndall)

In this chapter the 2-step SOM-FFN $p\text{CO}_2$ and CO_2 flux results, derived from the method presented in chapter 2, are presented for the global ocean from 1998 through 2011. Firstly, the quality of the $p\text{CO}_2$ product is determined by examination of the $p\text{CO}_2$ residuals, between the SOM-FFN estimates and gridded observations of the SOCAT v2 database (Bakker *et al.*, 2014). Secondly, the SOM-FFN $p\text{CO}_2$ estimates are validated with independent timeseries products. It will be shown that the global ocean contemporary uptake flux and the anthropogenic CO_2 uptake flux, calculated from the ocean net flux plus the riverine-derived carbon flux, are in good agreement with recent studies. In this chapter, the sea surface $p\text{CO}_2$ trends and the CO_2 flux variability will be examined for the global ocean, as well as the four major ocean sub-basins, namely the Atlantic Ocean, the Pacific Ocean, the Indian Ocean and the Southern Ocean, and the main driver for the global ocean sea surface $p\text{CO}_2$ variability within the study period will be investigated. Finally, the air-sea CO_2 flux estimates will be used to estimate the anthropogenic carbon budget from 1998 through 2011.

5.1 Residuals and validation

The neural network $p\text{CO}_2$ estimates for the global ocean obtain a good fit when compared to the SOCAT v2 data. Table 5.1 reveals a mean r^2 of 0.88, a root mean squared error of $12.05 \mu\text{atm}$ and a small overall bias of $-0.12 \mu\text{atm}$ over the entire time period from 1998–2011 for a total of 105196 gridded observations. This good fit further holds for each year individually with an r^2 ranging from 0.84 to 0.90, a RMSE ranging from $9.89 \mu\text{atm}$ to $15.79 \mu\text{atm}$ and an annual bias within $\pm 1 \mu\text{atm}$ (table 5.1).

The temporal mean $p\text{CO}_2$ residuals for each pixel, shown in figure 5.1, reveal a non-uniform distribution in space. The largest model to observation discrepancies occur in the high latitudes of both hemispheres, in regions with strong horizontal gradients such as on the Patagonian shelf, in the eastern Equatorial Pacific and the eastern Equatorial Atlantic, as well as in the Gulf Stream and North Atlantic Current region. Open ocean regions and gyres show a better fit between the SOM-FFN $p\text{CO}_2$ estimates and the SOCAT v2 gridded data product. The spatial patterns in the standard error of the residuals (figure 5.1 (b)) suggest, that these discrepancies have likely been caused by variabilities within the $p\text{CO}_2$ data that are uncorrelated to the variabilities within the chosen set of input data and

Table 5.1: Statistical measures of the comparison of the neural network–based estimates of $p\text{CO}_2$ with the SOCAT v2 gridded observations (Bakker *et al.*, 2014)

| Period | r^2 | RMSE [μatm] | bias [μatm] | # data |
|-----------|-------|--------------------------|--------------------------|--------|
| 1998-2011 | 0.88 | 12.05 | -0.12 | 105196 |
| 1998 | 0.89 | 10.85 | 0.36 | 5923 |
| 1999 | 0.87 | 13.63 | -0.07 | 4039 |
| 2000 | 0.85 | 15.79 | -0.05 | 4827 |
| 2001 | 0.87 | 15.61 | -0.90 | 4624 |
| 2002 | 0.88 | 11.41 | -0.12 | 6582 |
| 2003 | 0.85 | 11.53 | 0.08 | 6993 |
| 2004 | 0.88 | 11.20 | -0.18 | 8085 |
| 2005 | 0.89 | 10.91 | -0.32 | 9243 |
| 2006 | 0.88 | 11.88 | 0.10 | 11153 |
| 2007 | 0.87 | 11.99 | -0.37 | 11111 |
| 2008 | 0.90 | 11.51 | -0.11 | 9504 |
| 2009 | 0.84 | 12.86 | 0.29 | 9650 |
| 2010 | 0.88 | 9.98 | -0.39 | 9234 |
| 2011 | 0.88 | 12.42 | -0.11 | 4228 |

can therefore not be fully reconstructed.

5.2 Validation with independent observations

To check the goodness of fit of the global SOM–FFN $p\text{CO}_2$ estimates, they are compared with available observations from moorings and timeseries stations (available from <http://cdiac.ornl.gov/oceans/Moorings/>), which are not included in SOCAT v2 and hence have not been used to train the feed–forward neural network. The geographical location and the statistical comparison of all stations can be found in table 5.2. The gridded SOM–FFN estimates are generally not centred at the exact geographical position of timeseries stations. Therefore the values at the four closest surrounding grid–boxes are interpolated to the exact location of the timeseries station, weighted by their distance, to compare them to each station. The sea surface $p\text{CO}_2$ for stations BATS (Bates, 2007; Gruber *et al.*, 2002) and K2 (Wakita *et al.*, 2010) has been calculated from the reported DIC and TALK concentrations using CO2SYS as explained in section 3.3.1 of this thesis. All other timeseries stations (listed in table 5.2) report either calculated or directly measured sea surface $p\text{CO}_2$ or $f\text{CO}_2$. If $f\text{CO}_2$ was reported it has been converted to $p\text{CO}_2$ using equation 2.1, using reported pressures and temperatures.

In general, root mean squared errors (RMSE) at the different stations are similar to

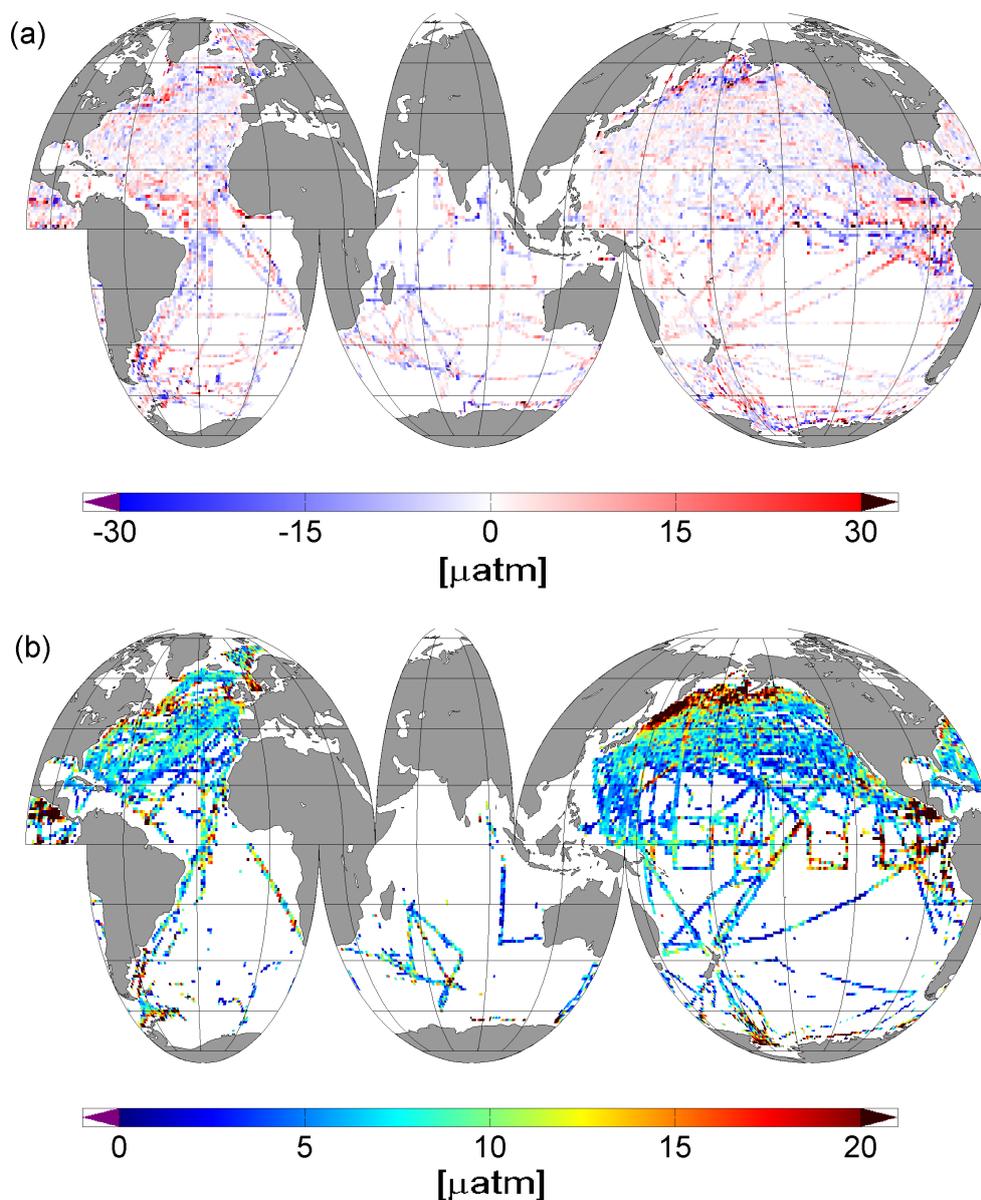


Figure 5.1: Temporal mean residuals (a) and standard deviation of the residuals (b) in μatm between neural network estimates and SOCAT v2 gridded observations (Bakker *et al.*, 2014) for the period from 1998–2011.

those calculated from the SOCAT v2 dataset (see table 5.1) ranging from 11 to 15 μatm , with the exception of the Irminger Sea (Olafsson, 2007) and K2 (Wakita *et al.*, 2010) stations, where the residuals show a larger spread of 25 to 27 μatm . There is a mainly negative mean offset between the SOM–FFN $p\text{CO}_2$ and all independent stations ranging from $-1 \mu\text{atm}$ up to $-7 \mu\text{atm}$, with the exception of BATS (Bates, 2007; Gruber *et al.*, 2002) and K2 (Wakita *et al.*, 2010) where the offset is positive. At stations where the largest RMSE’s were identified (K2, Irminger Sea) the bias is small. The input parameters

Table 5.2: Statistical comparison between the neural network estimates, and data from timeseries and mooring stations. Stations include the Bermuda Atlantic Timeseries Station (BATS) (Bates, 2007; Gruber *et al.*, 2002), the Hawaiian Ocean Timeseries station (HOT) (Dore *et al.*, 2009), the European Station for Timeseries in the Ocean (ESTOC) (González-Dávila *et al.*, 2007), the Irminger Sea station (Olafsson, 2007), the PAPA mooring (Sutton *et al.*, 2013) and the K2 time-series station (Wakita *et al.*, 2010)

| Station | location | time period | RMSE [μatm] | bias [μatm] |
|----------|-------------------|-------------------|--------------------------|--------------------------|
| BATS | 32.17°N, 64.50°W | 1998 through 2009 | 14.57 | 4.79 |
| HOT | 22.75°N, 158.00°W | 1998 through 2010 | 11.44 | -4.30 |
| ESTOC | 29.04°N, 15.50°W | 1998 through 2009 | 14.50 | -7.15 |
| Irminger | 64.30°N, 28.00°W | 1998 through 2006 | 24.97 | -0.99 |
| PAPA | 50.12°N, 144.83°W | 2007 through 2010 | 11.70 | -6.36 |
| K2 | 47.00°N, 160.00°E | 2007 through 2009 | 26.80 | 0.63 |

at these two stations (located furthest north of all stations) are not able to fully reconstruct the sea surface $p\text{CO}_2$ variability, hence the neural network estimate is close to the SOCAT v2 mean of this region (which resembles the mean at both stations) in order to reduce the mean squared error between estimates and SOCAT v2 $p\text{CO}_2$ observations (see chapter 2).

Figure 5.2 shows the visual comparison between two of the longest running timeseries stations and the SOM–FFN estimates, i.e., the combined record from BATS (Bermuda Atlantic Time Series Station) and Hydrostation "S" (Bates, 2007; Gruber *et al.*, 2002) and the Hawaiian Ocean Timeseries station (HOT, Dore *et al.*, 2009). Both timeseries stations provide near monthly coverage over the time period. While the seasonal cycle is fairly well captured at both stations, the winter minimum at BATS appears to be underestimated in the SOM–FFN estimates, as identified in the Atlantic Ocean study (chapter 4). It is remarkable that although the network was trained with data from the SOCAT v2 database (data marked with blue triangles from the MOSEAN and WHOTS stations), the comparison shows a better agreement with the independent HOT data (figure 5.2 (d)), likely due to the large amount of data within the same province, hence the MOSEAN and WHOTS data have less weight. The MOSEAN and WHOTS station data in SOCAT v2 have been further flagged as bad data in a version 2 release note.

5.3 Uncertainty of the air–sea CO₂ flux

Similar to the Atlantic study in chapter 4, the uncertainty of the air–sea $p\text{CO}_2$ flux stems from the error in the estimated $\Delta p\text{CO}_2$ and the uncertainty of the gas transfer coefficient

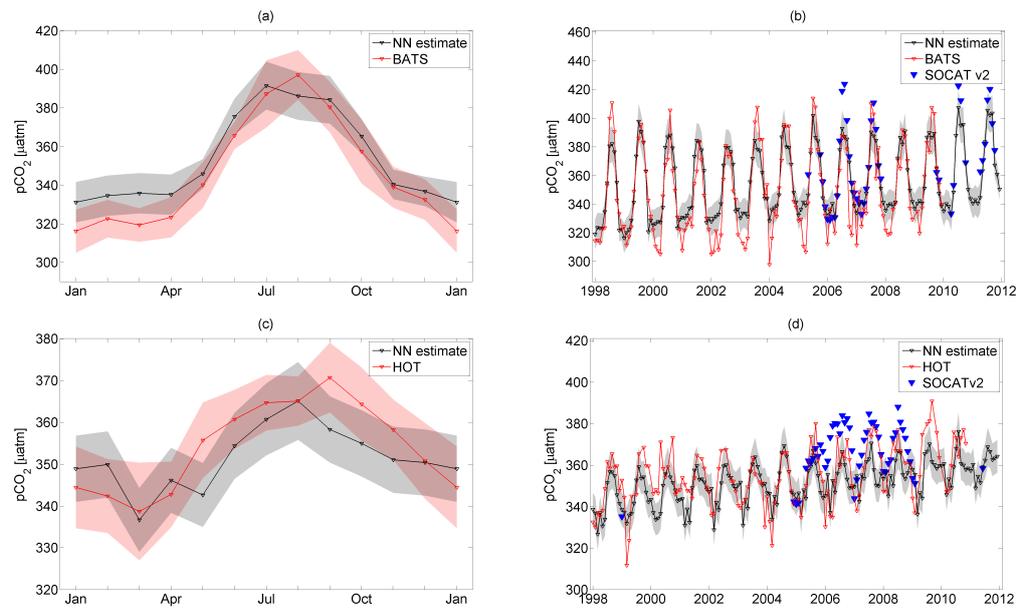


Figure 5.2: Long term seasonal cycle and mean seasonal cycle of the neural network estimates compared to BATS Hydrostation "S" (Bates, 2007; Gruber *et al.*, 2002) (a)–(b) and HOT (Dore *et al.*, 2009) (c)–(d) timeseries stations. Grey shading shows the uncertainty based on the RMSE of the SOM–FFN estimate. Pink shading shows the standard deviation of the the mean seasonal cycle for each timeseries station. Blue triangles show co–located observations available in SOCAT v2. Blue triangles close to the HOT Station are from the combined timeseries of the MOSEAN and WHOTS stations, whereas data close to the BATS station are from individual cruises.

(Takahashi *et al.*, 2009). Therefore, the global flux uncertainties are based on 3 main sources, namely the uncertainties derived from the gridding of the underway observations into $1^\circ \times 1^\circ$ bins, The uncertainty derived from the spread between gridded observations and SOM–FFN $p\text{CO}_2$ estimates and finally, the uncertainty of the gas transfer formulation, conducted for the 11 global RECCAP/Ocean Inversion regions. These regions are illustrates in figure 5.3 based on Gurney *et al.* (2008).

The discretization or gridding error is estimated using the global mean decorrelation length scale of 400 km by Jones *et al.* (2012). Dividing the standard deviation of $5 \mu\text{atm}$ as reported by Sabine *et al.* (2013) by the total number of effective degrees of freedom, where the effective number of degrees of freedom is the number of data points that are not correlated, i.e. those that have a distance further than the given decorrelation length. This results in an uncertainty ranging from about $1 \mu\text{atm}$ to $2 \mu\text{atm}$ for the individual regions.

In order to estimate the error derived from the neural network approach for each RECCAP/Ocean Inversion region, the spatial correlation is determined by analysing the semi–variograms of the residuals (see figure 5.4) within each region. For each RECCAP/Ocean Inversion region, the residuals are first divided into 5 randomly chosen mutually–exclusive

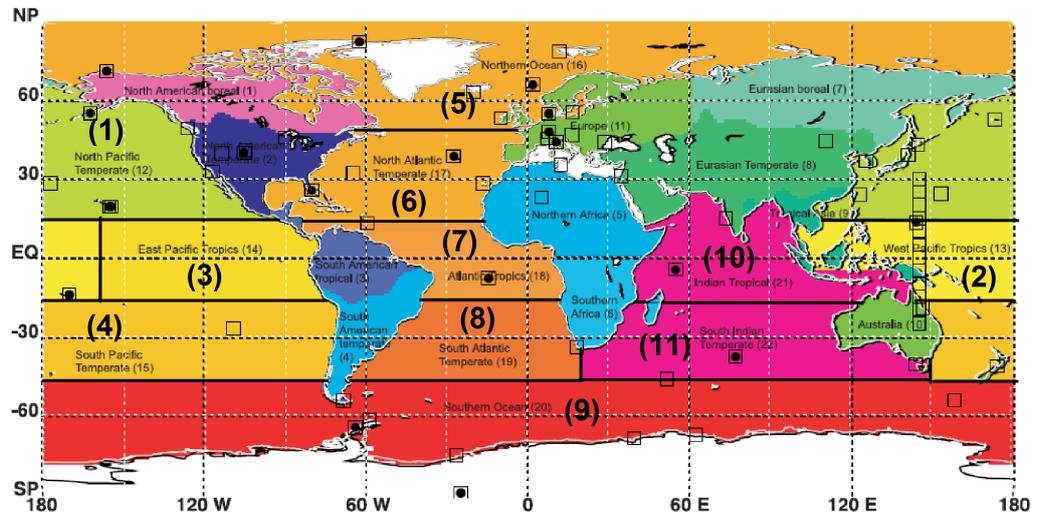


Figure 5.3: Map of the Ocean Inversion regions as illustrated in Gurney *et al.* (2008). Ocean region numbers, as they are used in this thesis, are enlarged for better visualization.

ensembles, with the exception of the subtropical/temperate North Atlantic and the subtropical/temperate North Pacific (regions (1) and (6) in figure 5.3), where a total of 15 ensembles is used, due to the larger amount of data, the subpolar Atlantic Ocean (region (5) in figure 5.3), where 10 ensembles are used and the northern and southern Indian Ocean (region (10) and region (11) in figure 5.3), where only 3 ensembles are used, due to the data sparsity within this region from 1998 through 2011. An exponential function is then fit to the semi-variogram, according to equation 4.1, resulting in autocorrelation lengths between 10 km and 957 km for the individual regions. In all cases the semi-variogram shows a large lag 0 correlation, reducing the effective number of degrees of freedom. Overall this results in a $p\text{CO}_2$ uncertainty for each region individually ranging from 1 μatm to 7 μatm , based on dividing the RMSE of the global residuals (12.05 μatm , see table 5.1) by the square root of the number of decorrelated data points.

Adding the error from the data gridding and the SOM–FFN mapping together, and assuming a mean error of 0.2 μatm for the atmospheric $p\text{CO}_2$ (Takahashi *et al.*, 2009), yields a total $\Delta p\text{CO}_2$ uncertainty for the 11 regions between 2 and 9 μatm . With a global mean gas transfer rate of $0.05 \text{ mol C} \cdot \text{m}^{-2} \cdot \text{yr}^{-1} \cdot \mu\text{atm}^{-1}$ (not accounting for regional variability) this results in a flux uncertainty between ± 0.02 and $\pm 0.11 \text{ Pg C} \cdot \text{yr}^{-1}$ for the individual regions and a global mean uncertainty of $\pm 0.19 \text{ Pg C} \cdot \text{yr}^{-1}$ calculated by standard error propagation. This is roughly 12% of the estimated SOM–FFN based long term mean integrated carbon flux ($-1.54 \text{ Pg C} \cdot \text{yr}^{-1}$), which will be introduced in the

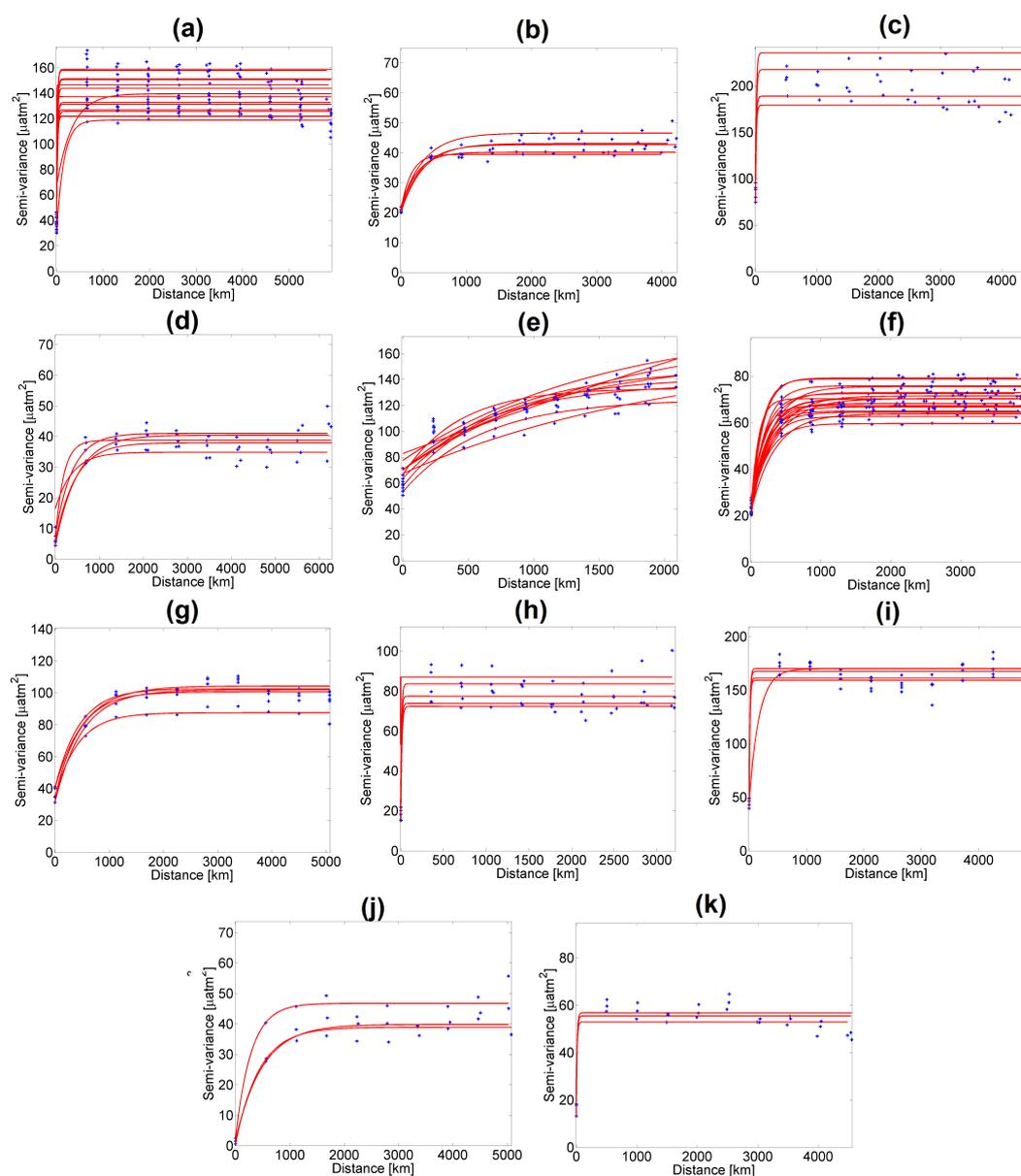


Figure 5.4: Empirical semi-variograms of randomly chosen ensembles of the residuals as a function of distance for the 11 RECCAP/Ocean Inversion regions (a)–(k). Correlation lengths range from 10 km in the South Atlantic (region (8) in figure 5.3) to 957 km in the subpolar North Atlantic (region (5) in figure 5.3)

subsequent section.

To estimate the uncertainty of the gas transfer velocity, four of the most commonly used gas transfer velocity estimates from Wanninkhof (1992) (equation 5.1), Wanninkhof and McGillis (1999) (equation 5.2), Nightingale *et al.* (2000) (equation 5.3) and Sweeney *et al.* (2007) (equation 5.4) are used. These formulations are:

$$k_w = 0.31 \cdot u^2 \cdot (\text{Sc}/660)^{-0.5} \quad (5.1)$$

$$k_w = 0.0283 \cdot u^3 \cdot (\text{Sc}/660)^{-0.5} \quad (5.2)$$

$$k_w = (0.333 \cdot u + 0.222 \cdot u^2) \cdot (\text{Sc}/600)^{-0.5} \quad (5.3)$$

$$k_w = 0.27 \cdot (\text{Sc}/660)^{-0.5} \cdot u^2 \quad (5.4)$$

The uncertainty is then computed from the standard deviation of the four resulting flux densities, leading to a global mean uncertainty of $\pm 0.62 \text{ Pg C} \cdot \text{yr}^{-1}$ which is roughly 40% of the long term mean integrated carbon flux ($-1.54 \text{ Pg C} \cdot \text{yr}^{-1}$, excluding the Arctic Ocean).

Combining the above uncertainties using standard error propagation yields $\pm 0.65 \text{ Pg C} \cdot \text{yr}^{-1}$ for the long term mean flux estimate, or roughly 42%. This is slightly less than the suggested 50% by Takahashi *et al.* (2009).

5.4 Long term mean $p\text{CO}_2$ and air–sea flux in the global ocean

The highest $p\text{CO}_2$ values can be identified in the tropical zones of the global ocean, particularly in the eastern Equatorial Pacific upwelling area, the northern Indian Ocean, as well as along the Californian Current and in the high latitude North Pacific (figure 5.5). Considering the long term mean $p\text{CO}_2$ from 1998 through 2011, these high partial pressure areas are supersaturated resulting in a positive, or outgassing flux, of carbon dioxide from the ocean into the atmosphere (see figure 5.5). The lowest sea surface $p\text{CO}_2$ values are found in the high latitude North Atlantic, along the Gulf Stream, along the Kuroshio

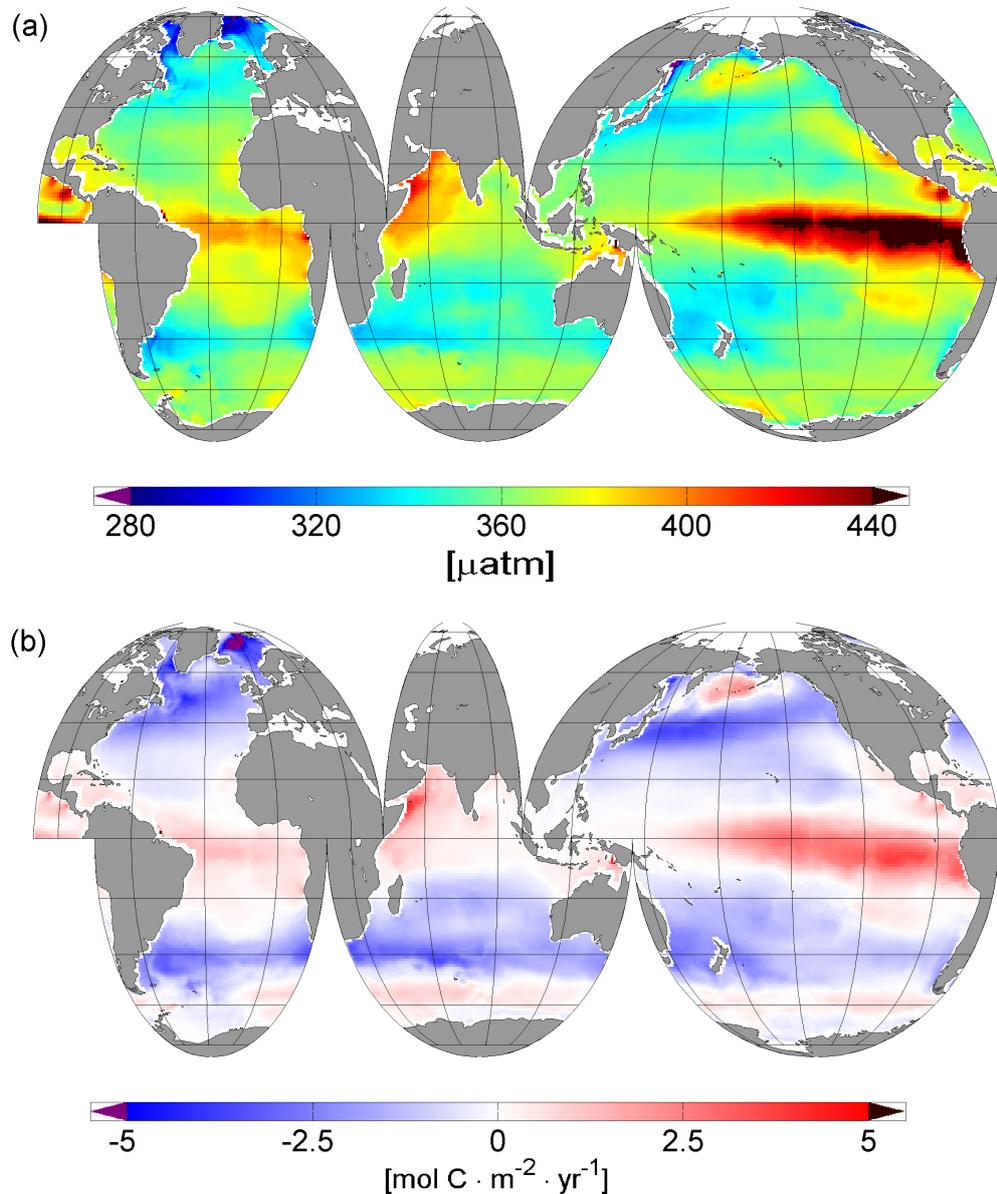


Figure 5.5: Map of (a) temporal mean surface ocean $p\text{CO}_2$ and (b) CO_2 flux density in $\text{mol C} \cdot \text{m}^{-2} \cdot \text{yr}^{-1}$ for the global ocean (excluding the Arctic Ocean). Negative flux densities indicate CO_2 uptake by the ocean.

Current and North Pacific Current and in the subtropical bands of the southern hemisphere. These regions correspond to the global ocean's major sink regions according to the 14-year averaged SOM–FFN $p\text{CO}_2$ based air–sea flux estimates.

The long term mean $p\text{CO}_2$ patterns largely follow the climatology patterns of Takahashi *et al.* (2009). The comparison of the global SOM–FFN $p\text{CO}_2$ estimates with the Takahashi *et al.* (2009) climatology yields a RMSE of $6.13 \mu\text{atm}$ and a mean difference of $0.98 \mu\text{atm}$. These numbers are very similar to those obtained in the Atlantic Ocean

study in chapter 4. Figure 5.6 shows the difference between the SOM–FFN $p\text{CO}_2$ estimates averaged onto the $4^\circ \times 5^\circ$ grid and the Takahashi *et al.* (2009) climatology. The long term mean $p\text{CO}_2$ estimates are centred between 2004 and 2005 and therefore need to be corrected to the year 2000. Unlike the Atlantic study in chapter 4, this is done by removing 4.5 times the actual atmospheric trend for each grid box. The atmospheric $p\text{CO}_2$ trend is computed from the GLOBALVIEW-CO2 (2011) data, which were converted to $p\text{CO}_2$. The comparison shows that the largest differences between the year 2000 corrected SOM–FFN $p\text{CO}_2$ and the Takahashi *et al.* (2009) climatology occur in the high latitudes of both hemispheres and the Equatorial Pacific. The differences in the Equatorial Pacific are likely linked to the circumstance, that Takahashi *et al.* (2009) removed observations from ENSO years from their analysis, while the SOM–FFN estimates include observations in ENSO years.

The 14–year averaged integrated CO_2 flux, derived from the SOM–FFN method, is estimated to be $-1.54 \pm 0.65 \text{ Pg C} \cdot \text{yr}^{-1}$ (excluding the Arctic Ocean), which is close to the estimate of Takahashi *et al.* (2009) corrected for under sampling of $-1.6 \pm 0.9 \text{ Pg C} \cdot \text{yr}^{-1}$. However, these global flux estimates refer to different reference years, as the SOM–FFN based flux estimate is centred in 2004, whereas Takahashi *et al.* (2009) uses a reference year 2000. The SOM–FFN based estimate refers to the contemporary air–sea flux, as there is no direct way to distinguish between the natural and the anthropogenic component when estimates are derived from observations directly. However, considering the natural outgassing flux of riverine carbon of $0.45 \pm 0.18 \text{ Pg C} \cdot \text{yr}^{-1}$ (Jacobson *et al.*, 2007) and additionally adding an air–sea flux of $-0.12 \pm 0.06 \text{ Pg C} \cdot \text{yr}^{-1}$ for the Arctic Ocean (Schuster *et al.*, 2013), an estimate of the anthropogenic carbon flux of $-2.11 \pm 0.68 \text{ Pg C} \cdot \text{yr}^{-1}$ is derived for the period 1998 through 2011, which is close to recent estimates (e.g. Wanninkhof *et al.*, 2013b).

Temporal mean air–sea CO_2 flux estimates for the global ocean and the individual ocean basins are summarized in table 5.3 and will be discussed in the following subsections.

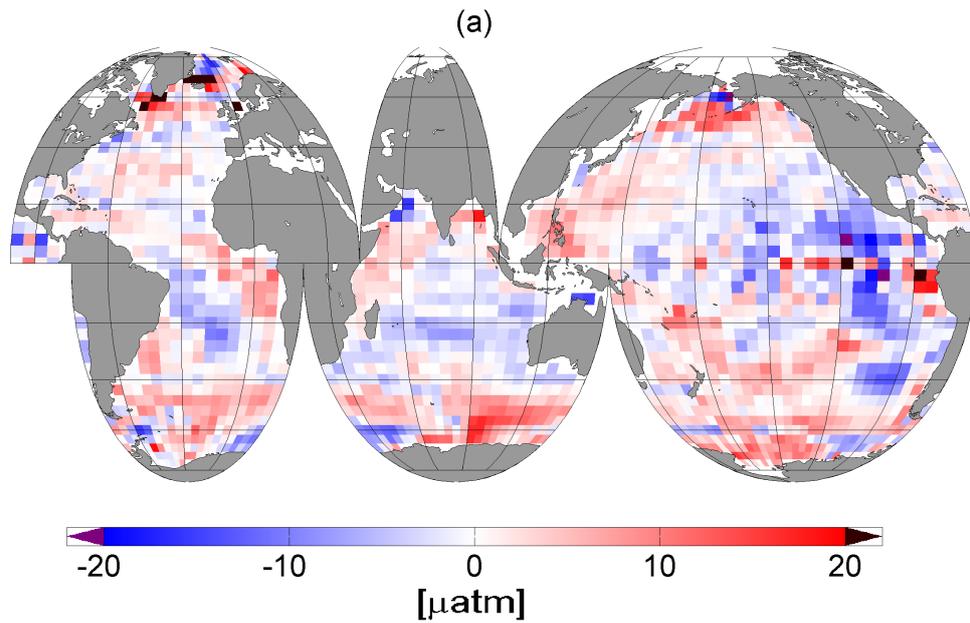


Figure 5.6: Difference in the surface ocean $p\text{CO}_2$ in μatm between the decadal mean neural network estimates (this study), corrected to the year 2000, and the estimates from the climatology of Takahashi *et al.* (2009). Positive differences indicate higher $p\text{CO}_2$ for the neural network estimates.

Table 5.3: Temporal mean CO_2 fluxes, CO_2 flux trends and the standard deviation of the inter-annual variability ($\text{std}(\text{IAV})$) for the major ocean basins and the global ocean, summarized from the text. Basin-wide uncertainties were calculated using standard error propagation of the individual uncertainty estimates derived for the 11 RECCAP/Ocean Inversion regions as presented in section 5.3. Trends were estimated from the slope of a linear fit, applied to the 12-month running averaged timeseries of the basin-wide integrated data. Trend uncertainties are estimated from the standard deviation of the residuals of the linear fit. The standard deviation of the IAV ($\text{std}(\text{IAV})$) was estimated from the standard deviation of the 12-month running averaged and detrended timeseries.

| basin | temporal mean CO_2 flux [$\text{Pg C} \cdot \text{yr}^{-1}$] | CO_2 flux trend [$\text{Pg C} \cdot \text{yr}^{-1} \cdot \text{decade}^{-1}$] | $\text{std}(\text{IAV})$ [$\text{Pg C} \cdot \text{yr}^{-1}$] |
|----------------|--|---|--|
| global ocean | -1.54 ± 0.65 | -1.09 ± 0.13 | ± 0.13 |
| Pacific Ocean | -0.41 ± 0.31 | -0.29 ± 0.13 | ± 0.12 |
| Atlantic Ocean | -0.54 ± 0.16 | -0.36 ± 0.03 | ± 0.04 |
| Southern Ocean | -0.23 ± 0.09 | -0.36 ± 0.07 | ± 0.06 |
| Indian Ocean | -0.35 ± 0.10 | -0.13 ± 0.02 | ± 0.02 |

5.5 Seasonality

Figure 5.7 (a) shows the seasonal cycle of the zonally averaged CO_2 flux density of the entire Pacific Ocean north of 44°S . The Equatorial Pacific is a strong outgassing region throughout the entire year, linked to the Ekman upwelling of carbon-rich waters from deeper layers. There is a distinct temperature-driven seasonal cycle in the lower latitudes (equatorwards of 40°N and 40°S) of both hemispheres resulting in wintertime uptake of

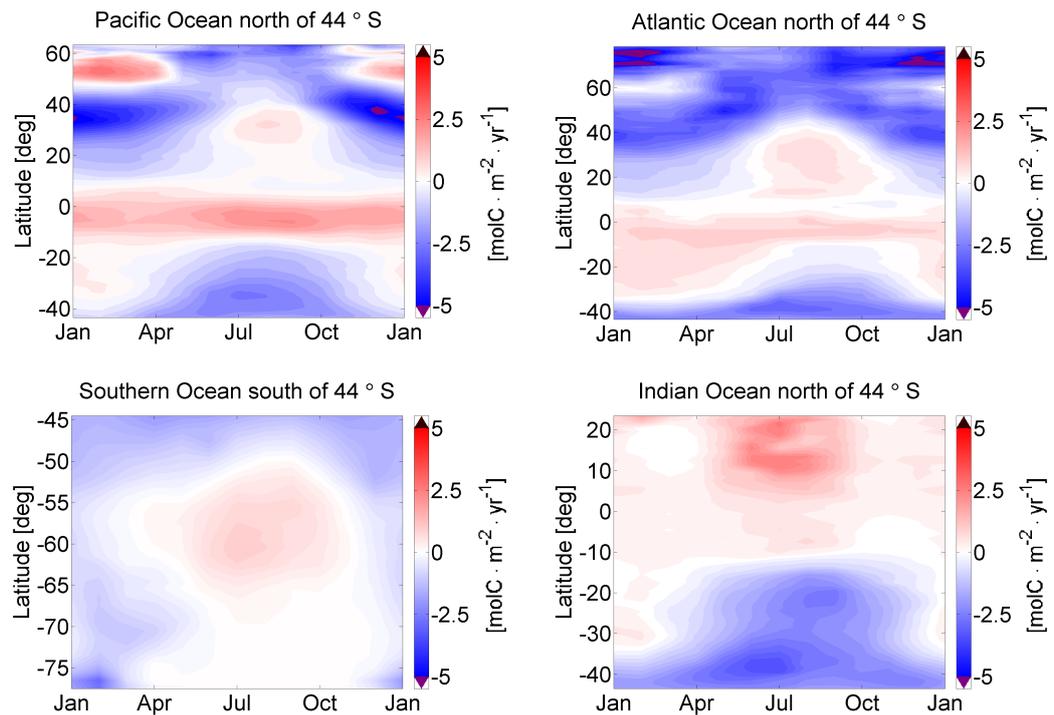


Figure 5.7: Seasonal cycle of the air–sea flux density for the 4 major ocean basins, namely (a) the Pacific Ocean, (b) the Atlantic Ocean, (c) the Southern Ocean and (d) the Indian Ocean. Negative or blue values indicate uptake of CO₂ by the ocean.

CO₂ by the ocean and summertime outgassing. The high latitude North Pacific (north of 40°N) shows an anti–phased cycle compared to the subtropical cycle. Here, the $p\text{CO}_2$ is drawn down by biological production from late spring to summer.

The seasonal cycle of the Atlantic Ocean (44°S–79°N and west of 30°E) flux density has been discussed in detail in Landschützer *et al.* (2013) and chapter 4 of this thesis. Figure 5.7 (b) shows the seasonal cycle of the Atlantic Ocean flux density, considering the additional timeperiod from 2008 through 2011. The seasonal cycle is very similar to that of the Pacific Ocean, as both basins are stretching along similar latitudes. However, there are significant differences. The equatorial band of the Atlantic Ocean shows lower partial pressures throughout the entire year than that of the Pacific Ocean. Furthermore, the band from 40°N to 60°N has a much stronger seasonal amplitude in the Pacific Ocean than in the Atlantic Ocean.

The SOM–FFN based CO₂ flux estimates show that the flux density (figure 5.7 (c)) in the Southern Ocean (south of 44°S) follows a clear seasonal pattern, with CO₂ uptake in the southern hemisphere summer (December to March). In contrast there is outgassing of CO₂ from 50°S to 65°S in the southern hemisphere winter (July to October). Only a

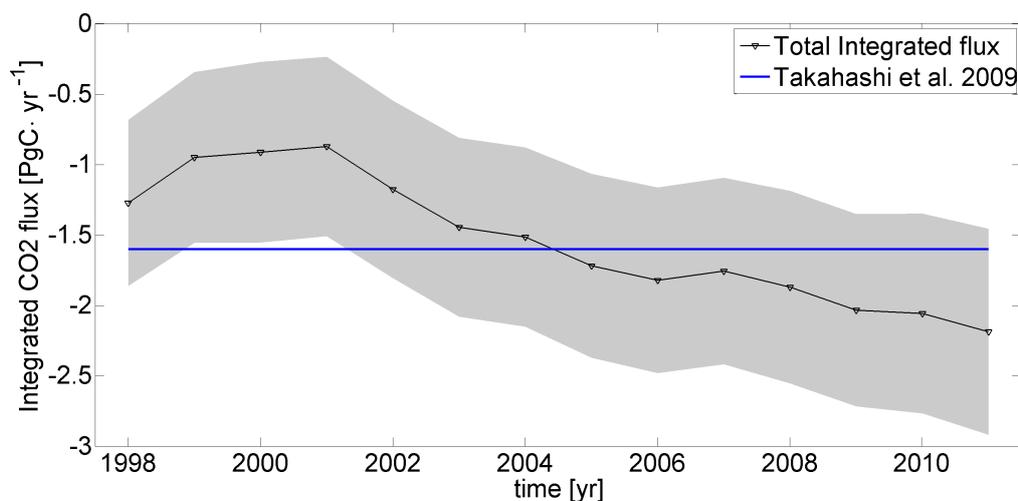


Figure 5.8: Timeseries of the annual mean Integrated air-sea CO₂ flux in Pg C · yr⁻¹ (black line). The gray shading indicates the uncertainty interval estimated in section 5.3. The blue line represents the estimate from Takahashi *et al.* (2009), corrected for undersampling. The integrated flux shows a flux reduction from 1998 through 2001 and a strong flux increase from 2001 onwards. The processes behind this global variability and trend are discussed for each basin separately in sections 5.6.1 to 5.6.4.

small band in the northern part of this basin (from 44°S to 50°S) remains an uptake region all year round.

The seasonal cycle of the Indian Ocean (north of 40°S) flux density (figure 5.7 (d)) shows a distinct equatorial high $p\text{CO}_2$ band from 10°N to 10°S barely changing throughout the entire year. Northwards, the neural network estimates show similar high CO₂ partial pressures from October to May compared to the Equator, but with an additional monsoon driven (see e.g. Sarma *et al.* (2013)) partial pressure increase from May to October. South of the Equator the Indian Ocean shows a very similar seasonal pattern compared to the southern hemisphere cycles of the Atlantic and the Pacific Ocean.

5.6 Inter-annual variability and trends

The global air-sea flux undergoes strong year-to-year variabilities within the time period from 1998 through 2011 with a minimum carbon uptake of -0.87 ± 0.64 Pg C · yr⁻¹ in 2001 and a maximum uptake of -2.19 ± 0.73 Pg C · yr⁻¹ in 2011. Figure 5.8 shows the timeseries of the contemporary CO₂ flux estimated by the SOM-FFN method. The linear trend evolving within this 14-year time period (linear fit to the estimates, smoothed using a 12 month running mean filter) is estimated to be -1.09 ± 0.13 Pg C · yr⁻¹ · decade⁻¹.

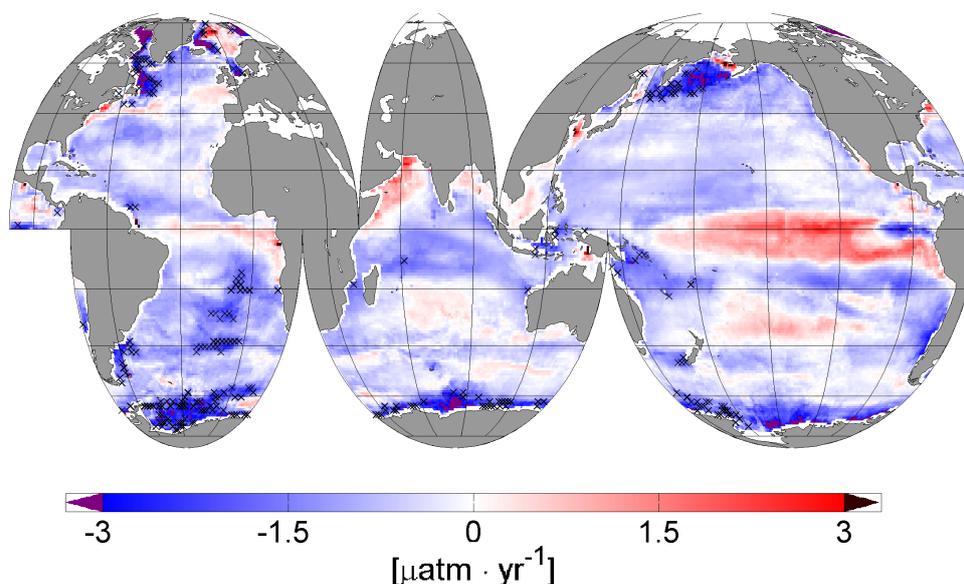


Figure 5.9: Linear trends in sea surface $p\text{CO}_2$ relative to that in the atmosphere over the period 1998–2011. The relative trend in sea surface $p\text{CO}_2$ was computed by subtracting the atmospheric mean trend. Areas with cross-hatch indicate where the trend is outside the 95% confidence level ($p \geq 0.05$). Trends are derived by applying a 12 month running mean to each pixel and are calculated as the slope of a linear fit.

Uncertainties of the trend are estimated as the standard deviation of the trend residuals (the smoothed estimates minus the data along the fitted linear trend line). However when only the start and end year annual mean CO_2 flux is taken into account, the trend reduces to $-0.70 \text{ Pg C} \cdot \text{yr}^{-1} \cdot \text{decade}^{-1}$ or roughly decreases by 30%, indicating that the global mean trend over such a short time period is considerably influenced by the inter-annual variability. This issue has been previously addressed in McKinley *et al.* (2011) and Fay and McKinley (2013).

Besides temporal variabilities, regional trend patterns emerge depending on the time period. Figure 5.9 shows the linear $p\text{CO}_2$ trends for each $1^\circ \times 1^\circ$ pixel from 1998 through 2011 with the linear atmospheric trend subtracted from the oceanic trend. The atmospheric $p\text{CO}_2$ trend is computed from the GLOBALVIEW-CO2 (2011) data, which were converted to $p\text{CO}_2$. Most areas of the global ocean show a weaker accumulation of carbon dioxide in the ocean than in the atmosphere, in particular the South Atlantic Ocean and high latitude areas in both hemispheres. However, regionally, the surface ocean exhibits the atmospheric partial pressure, e.g. in the high $p\text{CO}_2$ areas in the tropics, along the Gulf Stream and in parts of the South Indian Ocean and South Pacific Ocean.

Basin-scale variabilities and trends for the major ocean basins individually are discussed in the next sections.

5.6.1 The Pacific Ocean

Throughout the time period from 1998 to 2011 the Pacific Ocean north of 44°S undergoes considerable year-to-year variabilities with an annual integrated uptake flux minimum of $-0.12 \pm 0.32 \text{ Pg C} \cdot \text{yr}^{-1}$ in the year 2000 and an uptake maximum of $-0.64 \pm 0.31 \text{ Pg C} \cdot \text{yr}^{-1}$ in 2006. Figure 5.10 (a) shows a reduction in the carbon sink of about $0.5 \text{ Pg C} \cdot \text{yr}^{-1}$ from 1999 through 2001 caused by the La Niña response after the major ENSO event in 1997–1998, in line with a regional study of Feely *et al.* (2006) conducted in the Equatorial Pacific. In general, ENSO patterns appear to be the dominant mode of the global variability of the SOM–FFN based flux estimate, which will be discussed below in sections 5.6.5.

The linear trend, derived from the slope of the linear fit to the CO₂ flux estimate, smoothed using a 12 month running mean filter, shows an increase of the carbon sink of $-0.29 \pm 0.13 \text{ Pg C} \cdot \text{yr}^{-1} \cdot \text{decade}^{-1}$. Splitting the basin into northern and southern hemisphere reveals, that this trend stems almost entirely from the northern hemisphere ($-0.26 \pm 0.06 \text{ Pg C} \cdot \text{yr}^{-1} \cdot \text{decade}^{-1}$, compared to $-0.03 \pm 0.07 \text{ Pg C} \cdot \text{yr}^{-1} \cdot \text{decade}^{-1}$ south of the Equator). Overall, this linear trend estimate is strongly influenced by the La Niña event in the early years of the period (see figure 5.10). When shifting the start year of the linear trend calculation to 2003 the Pacific Ocean shows a positive trend, i.e. a trend towards a decreasing carbon sink of $0.17 \pm 0.06 \text{ Pg C} \cdot \text{yr}^{-1} \cdot \text{decade}^{-1}$. The inter-annual variability (1σ of the deseasonalized and detrended data) is estimated to be $\pm 0.12 \text{ Pg C} \cdot \text{yr}^{-1}$. The Pacific Ocean numbers are summarized in table 5.3.

5.6.2 The Atlantic Ocean

The Atlantic Ocean variability and trends have been discussed Landschützer *et al.* (2013) for the time period 1998 through 2007 and in chapter 4 of this thesis. Landschützer *et al.* (2013) identified a steady carbon sink from 1998 up to the year 2004 and further an increasing carbon sink from 2004 onwards, mainly driven by a strong increase in the southern hemisphere. This study provides an update of this estimate, by extending the

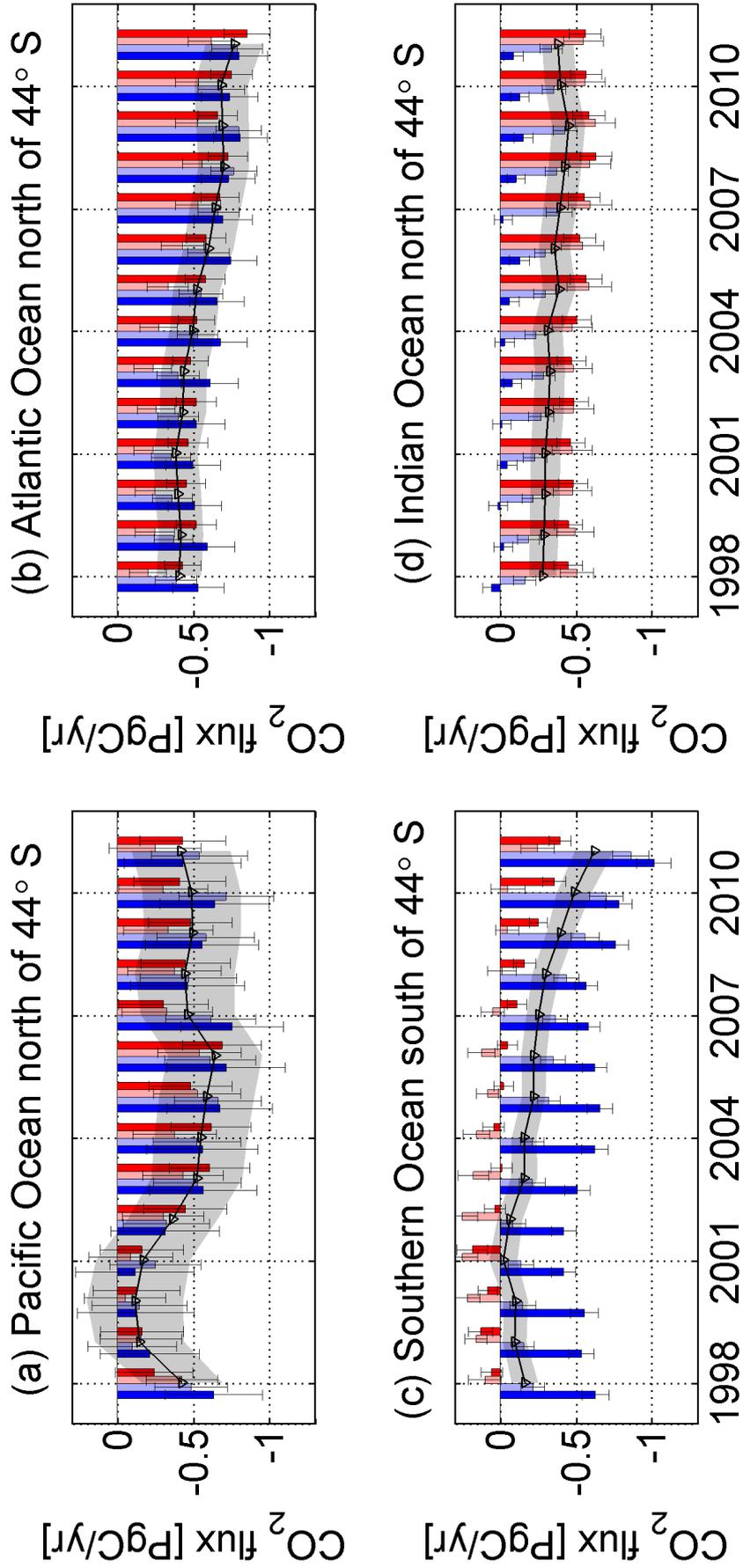


Figure 5.10: Seasonal and annual mean fluxes from 1998–2011 for the four major ocean basins, (a) the Pacific Ocean, (b) the Atlantic Ocean, (c) the Southern Ocean and (d) the Indian Ocean. Dark blue shows the results for the northern hemisphere winter months (DJF), light blue the spring months (MAM), light red the summer months (JJA), dark red the autumn months (SON). The annual mean flux is plotted as a black line on top. Error bars indicate the uncertainty range of the air–sea flux estimate in section 5.3.

time period of the study to December 2011 and by additional observational data in the SOCAT v2 dataset, in particular for the years 2005 to 2007 (Bakker *et al.*, 2014). Figure 5.10 (b) shows a close to steady carbon flux from 1998 through 2003 and further an increase in the carbon sink, i.e. one year earlier as identified in Landschützer *et al.* (2013). The Atlantic Ocean carbon uptake peaks in the year 2011 with an annual mean uptake of $-0.77 \pm 0.19 \text{ Pg C} \cdot \text{yr}^{-1}$. The linear trend suggests an increase in the Atlantic Ocean carbon sink of $-0.36 \pm 0.03 \text{ Pg C} \cdot \text{yr}^{-1} \cdot \text{decade}^{-1}$, which is more than twice the value obtained in Landschützer *et al.* (2013) for the period from 1998 through 2007. While the trend in Landschützer *et al.* (2013) almost entirely originated in the southern hemisphere, the extended timeseries shows a similar flux trend in both hemispheres ($-0.18 \pm 0.01 \text{ Pg C} \cdot \text{yr}^{-1} \cdot \text{decade}^{-1}$ north of the Equator, compared to $-0.19 \pm 0.02 \text{ Pg C} \cdot \text{yr}^{-1} \cdot \text{decade}^{-1}$ south of the Equator). However, the standard deviation of the inter-annual variability (calculated as the standard deviation of the detrended 12 month running mean monthly integrated carbon flux in the Atlantic Ocean) remains unchanged compared to the estimate from Landschützer *et al.* (2013) of $\pm 0.04 \text{ Pg C} \cdot \text{yr}^{-1}$.

5.6.3 The Southern Ocean

The Southern Ocean is the largest of the four basins and remains strongly under-sampled with respect to its large surface area (see figure 5.1). The SOM-FFN $p\text{CO}_2$ based flux estimates for the Southern Ocean show the largest range between flux minimum ($-0.02 \pm 0.08 \text{ Pg C} \cdot \text{yr}^{-1}$ in 2001) and maximum ($-0.62 \pm 0.13 \text{ Pg C} \cdot \text{yr}^{-1}$ in 2011) for all basins (figure 5.10).

The Southern Ocean carbon sink decreases from 1998 to 2001 and from there on strongly increases until 2011. From 1998 through 2011 the Southern Ocean shows a linear flux trend of $-0.36 \pm 0.07 \text{ Pg C} \cdot \text{yr}^{-1} \cdot \text{decade}^{-1}$ with moderate inter-annual variability of $\pm 0.06 \text{ Pg C} \cdot \text{yr}^{-1}$. Furthermore, the Southern Ocean shows large differences in seasonal trends. Summer (December, January and February) trends are the least pronounced ($-0.26 \pm 0.11 \text{ Pg C} \cdot \text{yr}^{-1} \cdot \text{decade}^{-1}$) compared to autumn (March, April, May) trends ($-0.48 \pm 0.11 \text{ Pg C} \cdot \text{yr}^{-1} \cdot \text{decade}^{-1}$), which illustrates the season with the strongest CO_2 flux increase. The Southern Ocean mean CO_2 flux, trend and variability are summarized in table 5.3.

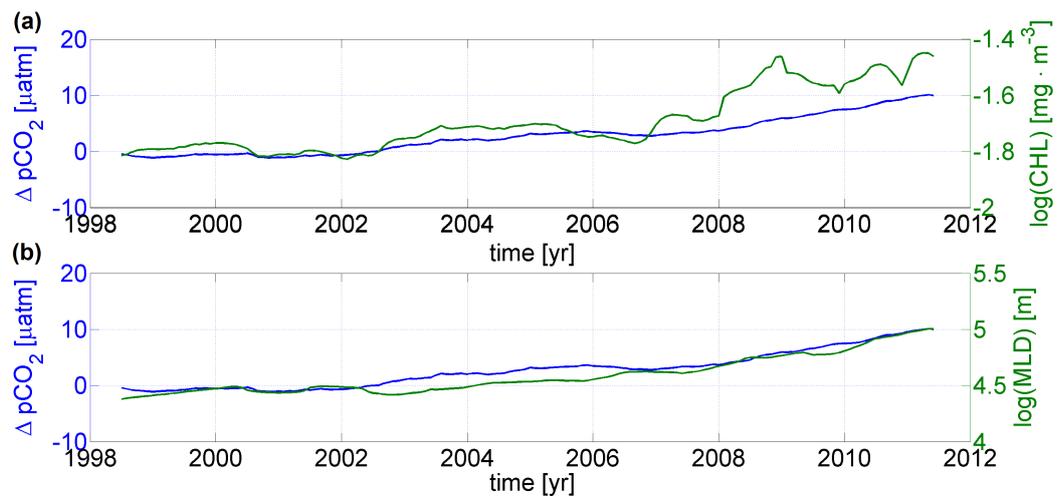


Figure 5.11: Comparison between the deseasonalized air-sea partial pressure difference of CO₂ ($\Delta p\text{CO}_2$ – increasing numbers reflect a stronger $p\text{CO}_2$ increase in the atmosphere compared to the ocean and an increase in the CO₂ uptake flux of the ocean) with (a) the natural logarithm of the deseasonalized chlorophyll-a ($\log(\text{CHL})$) and (b) the natural logarithm of the deseasonalized mixed layer depth ($\log(\text{MLD})$) in the Southern Ocean between 44°S and 58°S

Lenton *et al.* (2013) found that most of the air-sea CO₂ flux between 1990 and 2010 occurs between 44°S – 58°S, while the Southern Ocean south of 58°S has only a small annual net flux. Overall, the Southern Ocean mean flux ($-0.23 \pm 0.09 \text{ Pg C} \cdot \text{yr}^{-1}$) compares well with the Takahashi *et al.* (2009) limatology-based estimate from Lenton *et al.* (2013) ($-0.27 \pm 0.13 \text{ Pg C} \cdot \text{yr}^{-1}$). Figure 5.9 reveals a clear increase in the air-sea $\Delta p\text{CO}_2$ south of 58°S, whereas north of this latitude trends are less pronounced, relative to the atmospheric mean trend. Therefore the Southern Ocean is further divided into two sub basins to investigate the CO₂ flux trend.

The CO₂ trend south of 58°S is estimated as $-0.16 \pm 0.02 \text{ Pg C} \cdot \text{yr}^{-1} \cdot \text{decade}^{-1}$ from 1998 through 2011. North of 58°S the ocean uptake of CO₂ increases by $-0.20 \pm 0.06 \text{ Pg C} \cdot \text{yr}^{-1} \cdot \text{decade}^{-1}$, which is stronger compared to south of 58°S due to the larger surface area. There is only limited inter-annual variability ($\pm 0.02 \text{ Pg C} \cdot \text{yr}^{-1}$) south of 58°S. In contrast the area north of 58°S shows stronger inter-annual variability in the air-sea CO₂ flux ($\pm 0.05 \text{ Pg C} \cdot \text{yr}^{-1}$).

A possible explanation for the trends and the variability of the ocean band between 44°S and 58°S can be provided by the index polarity of the SAM (see figure 1.9 (c)). Figure 5.11 shows the partial pressure difference between the ocean and the atmosphere in comparison with the chlorophyll-a concentration and the mixed layer depth. High index polarity of the SAM leads to an increase of the westerly winds resulting in deeper mixing.

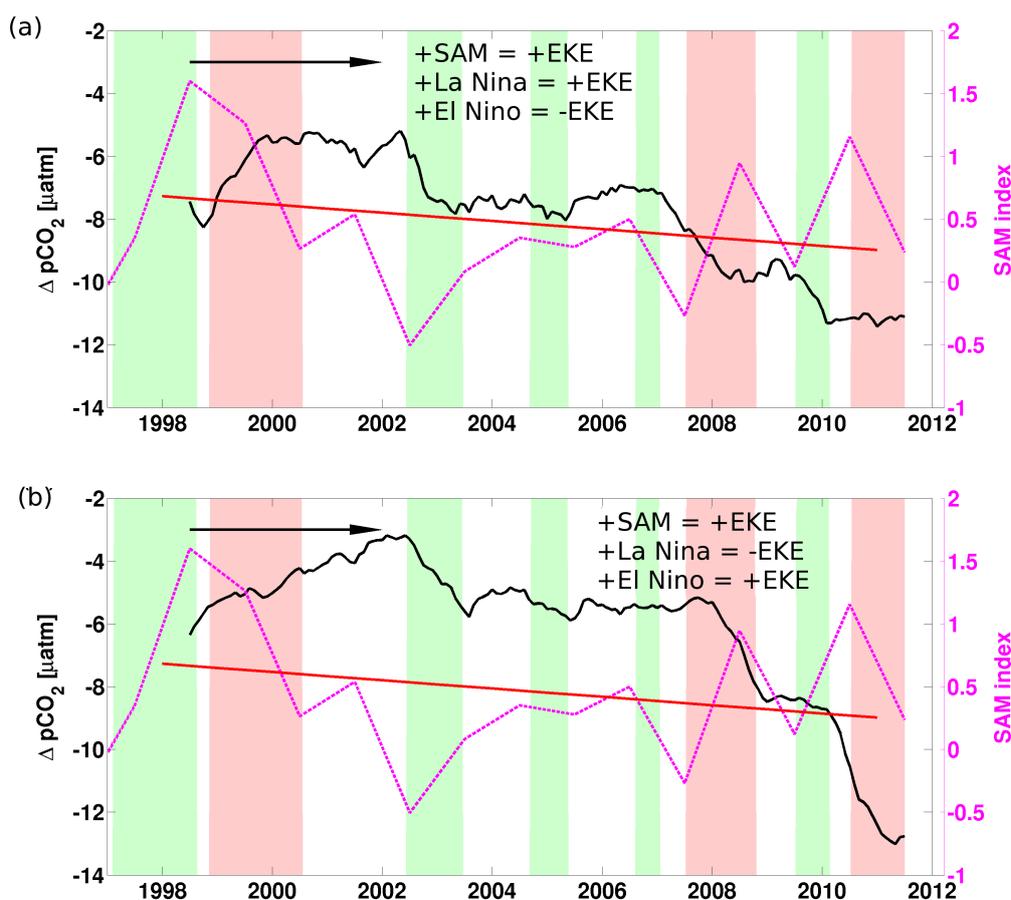


Figure 5.12: Comparison between the deseasonalized sea–air (for better comparison) partial pressure difference of CO_2 ($\Delta p\text{CO}_2$ – decreasing numbers reflect a stronger $p\text{CO}_2$ increase in the atmosphere compared to the ocean and an increase in the CO_2 uptake flux of the ocean) with the annual mean Southern Annular Mode (SAM) index (Marshall, 2003) for (a) the Pacific sector of the Southern Ocean (160°E to 60°W and 58°S to 44°S) and (b) the Atlantic and Indian sector of the Southern Ocean (60°W to 160°E and 58°S to 44°S). Green background indicates El Niño phases, where the Multivariate ENSO Index (MEI, Wolter and Timlin (2011)) is higher than 0.5, red background indicates La Niña phases where the MEI is smaller than -0.5 and white background indicates neutral ENSO phases. The black arrow illustrates the potential timelag of ~ 3 years between the 1998 high SAM index and the $\Delta p\text{CO}_2$ minimum as suggested by the eddy kinetic energy (EKE) response timelag (Morrow *et al.*, 2010). The red linear regression line illustrates the reduction in the SAM index trend from 1998 onwards

The $\Delta p\text{CO}_2$ increase and the resulting strengthening of the Southern Ocean carbon sink is consistent with recent findings from Fay and McKinley (2013). These authors argue for a strengthening of the Southern Ocean carbon sink linked to the weakening of the SAM index since the early 2000s. However, the increase in CHL in figure 5.11 (a) and the increase in MLD in figure 5.11 (b) do not support the argument that this increase stems from a reduction in vertical mixing. However, MLD might not be a good proxy for vertical mixing (at least not when considering averages over large regions) and trends in chlorophyll-*a* might not be driving the $p\text{CO}_2$, but rather be another emerging property.

Therefore, figure 5.12 (a) and (b) illustrates the $\Delta p\text{CO}_2$ in comparison to the annual averaged SAM index directly as well as positive (El Niño - green background) and negative (La Niña - red background) ENSO phases for (a) the Pacific sector of the Southern Ocean (160°E to 60°W) and (b) the Atlantic and Indian sector (60°W to 160°E). From 1965 onwards, the SAM index has continued to increase (see figure 1.9) leading to a peak in 1998. From 1998 until the mid 2000s, however, the SAM index shows a negative trend, which was also reported by Fay and McKinley (2013). Fay and McKinley (2013) argue that the negative index trend in the SAM results in a reduction in the vertical mixing and in a reinvigoration of the Southern Ocean carbon sink, which is supported by this study.

The $\Delta p\text{CO}_2$ timeseries in figure 5.12 suggests however, that the $p\text{CO}_2$ response to the SAM is not linear and not consistent throughout all sectors of the Southern Ocean (see figure 5.12 (a) and (b)). Meredith and Hogg (2006) found anomalously high eddy kinetic energy (EKE) in the Antarctic Circumpolar Current (ACC) during the 2000–2002 period, which these authors interpret as a 2–3 year lagged response to a significant peak in the circumpolar eastward wind stress quantified by the SAM. Morrow *et al.* (2010) support this conclusion, arguing that the EKE response after the 1998 SAM peak is further enhanced in the Pacific sector of the Southern Ocean by the La Niña event of 1999. In general Morrow *et al.* (2010) found that El Niño events counteract the effect of positive SAM events, whereas La Niña events enhance positive SAM phases in the Pacific sector and vice versa in the Atlantic and Indian sector.

It seems plausible that the enhanced EKE in response to the SAM/La Niña peak in 1998 led to enhanced mixing, resulting in an increase in the sea-surface $p\text{CO}_2$, peaking in \sim 2001/2002. This can be identified in both the Pacific, as well as the Atlantic and Indian sector, hence arguably, the strong SAM signal dominates over the La Niña response in the Atlantic and Indian sector. The period between 2002 and 2008 illustrates a fairly constant $\Delta p\text{CO}_2$, hence the atmospheric $p\text{CO}_2$ increase is leveled by the ocean, potentially due to reduced mixing (compared to the 2000–2002 period) linked to the reduction of the SAM index after the 1998 event. The last years of the analysis period from 2008 onwards illustrate a strong trend towards enhanced strengthening of the ocean CO_2 sink, particularly in the Atlantic and Indian sector. Throughout the last years of the analysis (\sim 2007 onwards), figure 5.12 (a) and (b) suggest a stronger link between ENSO phases and

the $\Delta p\text{CO}_2$ variability, i.e., a strong $\Delta p\text{CO}_2$ decrease during La Niña phases in the Atlantic and Indian sector, compared to decreasing $\Delta p\text{CO}_2$ during the 2009–2010 El Niño phase in the Pacific sector. This is however not completely in line with the EKE response to ENSO as suggested by Morrow *et al.* (2010), as these authors argue for an ENSO response–timelag of 1–2 years, whereas here the best match is found without any timelag.

The previous paragraphs emphasize the potential link between the SAM and ENSO climate modes to the variability and trends estimated in the Southern Ocean carbon sink. So far only large scale regions were considered, hence in order to investigate all the processes in the Southern Ocean that lead to the reinvigoration of the carbon sink it would be beneficial to extend the analysis and look at regional variabilities and the response of all input variables to the SAM and ENSO. Moreover, it would be beneficial to quantify regional effects that lead to a reduction/increase of the Southern Ocean carbon sink, e.g. driven by surface water cooling/warming, as well as increase/decrease in chlorophyll linked to regional enhanced/reduced wind or eddy induced mixing. Regional differences could further shed new light on the counter-intuitive finding, that basin wide both mixed layer depth and $\Delta p\text{CO}_2$ are increasing. This is suggested for future work.

5.6.4 The Indian Ocean

Compared to the other ocean basins, the Indian Ocean shows the lowest year-to-year variability, with a maximum CO_2 uptake in 2009 of $-0.45 \pm 0.10 \text{ Pg C} \cdot \text{yr}^{-1}$ and a minimum uptake in 1998 of $-0.28 \pm 0.10 \text{ Pg C} \cdot \text{yr}^{-1}$. This basin shows the weakest increase in the carbon sink compared to the other basins from 1998 through 2011 with a linear trend of $-0.13 \pm 0.02 \text{ Pg C} \cdot \text{yr}^{-1} \cdot \text{decade}^{-1}$. The inter-annual variability is estimated to be $\pm 0.02 \text{ Pg C} \cdot \text{yr}^{-1}$ and is substantially smaller than e.g. in the Pacific Ocean. Indian Ocean mean CO_2 flux, trend and variability are summarized in table 5.3.

5.6.5 The global ocean

The results of the individual ocean basins shed new light on the variability and trends identified in figure 5.8. Particularly the strong trend of $\sim 1 \text{ Pg C} \cdot \text{yr}^{-1}$ is likely a reflection of intra-decadal processes and not an indication of an inter-decadal trend signal. It is the result of the 1999–2001 La Niña event, which led to a reduction of the Pacific Ocean (and

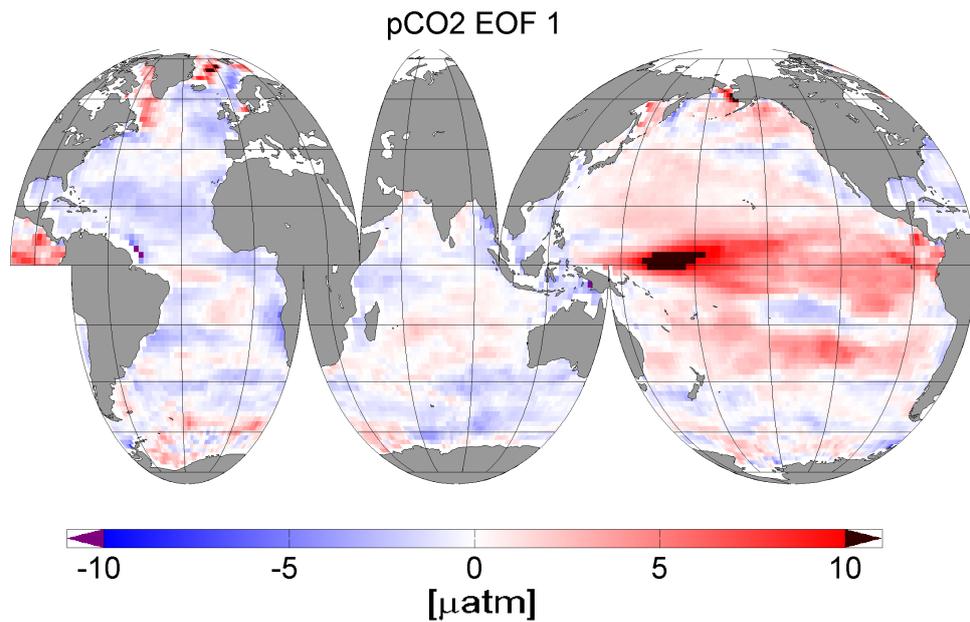


Figure 5.13: Spatial distribution of the amplitude of the first (leading) EOF of the sea surface $p\text{CO}_2$. Data are re-scaled to 2° longitude \times 2° latitude

hence the global ocean) carbon sink of $\sim 0.5 \text{ Pg C} \cdot \text{yr}^{-1}$. Furthermore, the peak in the SAM index in 1998 possibly illustrates the enhanced vertical mixing in the ACC region with a time lag of $\sim 2\text{--}3$ years, resulting in an almost saturated Southern Ocean in 2001. The following reduction of the SAM index potentially suggests a reduction in vertical mixing and the reinvigoration of the Southern Ocean carbon sink, additionally modified by ENSO, resulting in a basin-wide trend signal of $-0.36 \pm 0.07 \text{ Pg C} \cdot \text{yr}^{-1} \cdot \text{decade}^{-1}$ for the 1998 through 2011 period. The trend contribution of the Atlantic Ocean is equally strong compared to the Southern Ocean (-0.36 ± 0.03), where particularly trends in the South Atlantic appear to be reflected in the faster accumulation of anthropogenic CO_2 in the ocean interior (see e.g. Wanninkhof *et al.*, 2010). Finally, Khatiwala *et al.* (2013) suggest a trend contribution from atmospheric CO_2 , attributed to the rise in anthropogenic CO_2 , of $0.35 \text{ Pg C} \cdot \text{yr}^{-1} \cdot \text{decade}^{-1}$ based on ocean inventory changes.

To investigate the dominant mode of variability within the global ocean, an empirical orthogonal function (EOF) analysis was conducted. Due to the size of the covariance matrix the SOM-FFN $p\text{CO}_2$ estimates are re-scaled to a coarser spatial resolution of 2° longitude \times 2° latitude. Furthermore, $p\text{CO}_2$ fields are deseasonalized, using a 12 month running average, and detrended by removing the linear trend for each grid box from 1998 through 2011. Figure 5.13 shows the spatial structure of the first or leading EOF. It clearly

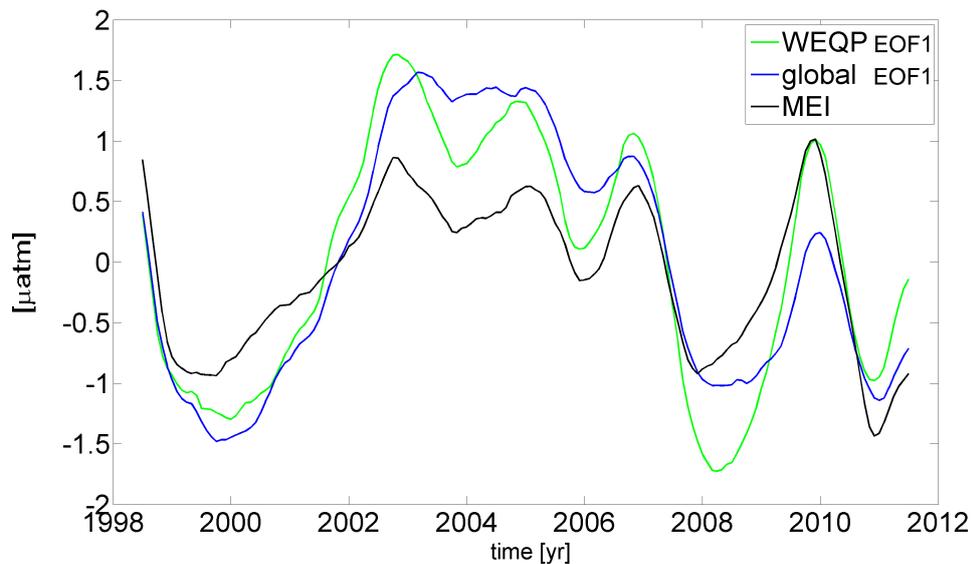


Figure 5.14: Time series of the leading EOF's of the western Equatorial Pacific (WEQP, 15°S to 15°N and west of 160°W) and the global ocean, as well as the Multivariate ENSO Index (MEI)

shows that the largest amount of variability stems from the western Equatorial Pacific, which is the region strongly influenced by the ENSO signal. Figure 5.14 compares the timeseries of the leading EOF from the western Equatorial Pacific and the global ocean to the MEI (<http://www.esrl.noaa.gov/psd/enso/mei/>). The comparison shows, that ENSO is globally the leading mode of variability with a correlation coefficient of 0.85 when correlating the timeseries of the leading global EOF with the Multivariate ENSO Index. However, the EOF analysis shows that the leading EOF, which strongly correlates with the MEI, explains 72% of the total CO₂ flux variance in the western Equatorial Pacific (15°S to 15°N and west of 160°W), but only about 28% of the global variance.

5.7 A global carbon budget (1998 through 2011)

Fossil fuel burning, land use change and cement production are the major contributors to the increase of CO₂ in the atmosphere. Le Quéré *et al.* (2009) report that between 1959 and 2008 only 43% of the emitted carbon has remained in the atmosphere. This raises the important questions "Where does the remainder of the emitted carbon go?" and "How variable are these sinks?". In this chapter, the global ocean anthropogenic uptake of CO₂ was estimated to be $-2.11 \pm 0.68 \text{ Pg C} \cdot \text{yr}^{-1}$ (including the correction for riverine-derived carbon). The global carbon budget can be closed using the global

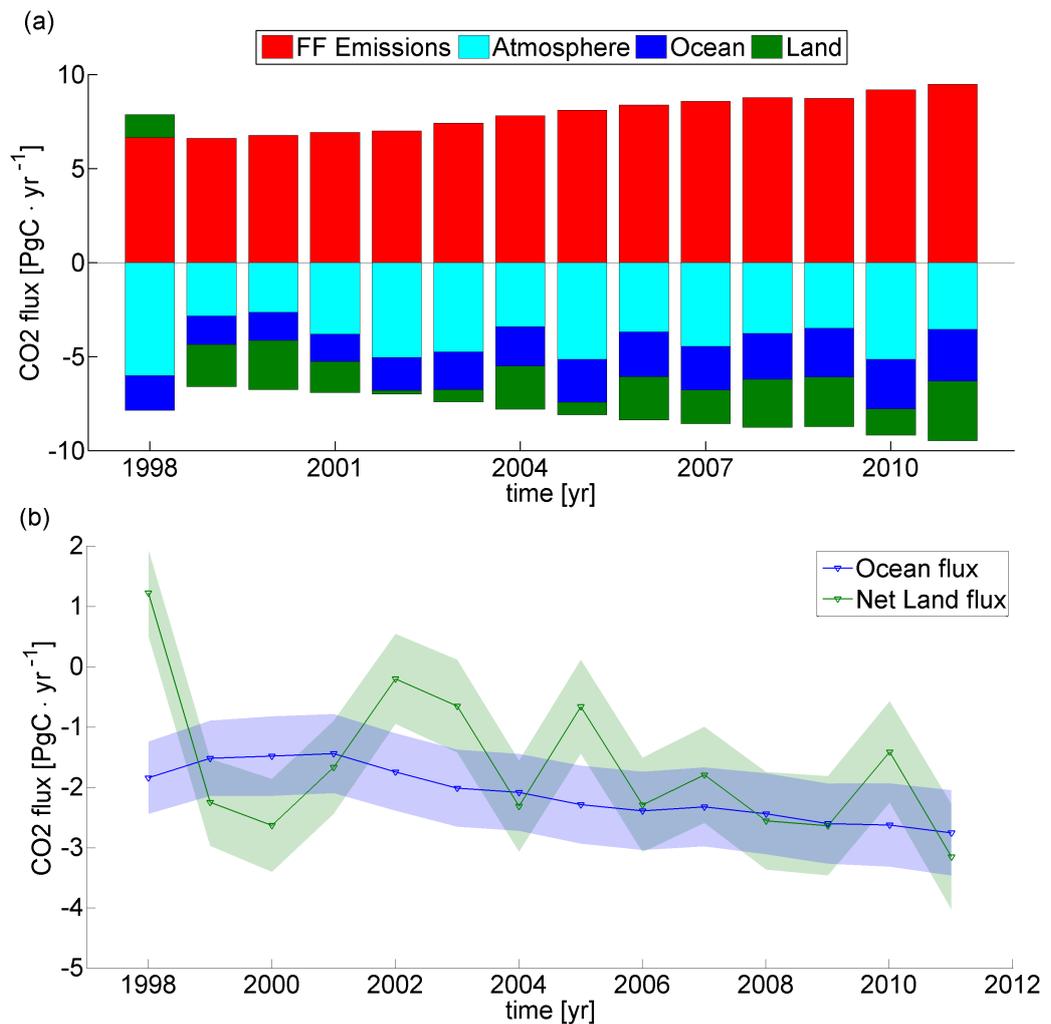


Figure 5.15: Global Carbon Budget (a) as the sum of sources (positive) and sinks (negative) for each year from 1998 through 2011 and (b) for the global ocean CO₂ sink and the resulting land flux individually.

ocean carbon sink numbers, provided that the actual global carbon emissions from fossil fuel burning (Marland *et al.* (2005); http://cdiac.ornl.gov/trends/emis/meth_reg.html), land use change (Houghton (2003); http://cdiac.ornl.gov/trends/emis/meth_reg.html) and the atmospheric accumulation rate (Ballantyne *et al.* (2012); Ed Dlugokencky and Pieter Tans, NOAA/ESRL, www.esrl.noaa.gov/gmd/ccgg/trends/) are known.

Figure 5.15 (a) shows the annual mean anthropogenic carbon budget derived from this study, while figure 5.15 (b) compares the ocean sink (net flux including riverine-derived carbon) with the net land sink (land flux plus land use change) for each year directly. The net land flux is derived from the emitted fossil fuel, the atmospheric accumulation rate and the estimated neural network ocean CO₂ flux following equation 1.4:

$$F'_{\text{land}} = E_{\text{mff}} - F_{\text{ocean}} - \frac{dG_{\text{atm}}}{dt} \quad (5.5)$$

the net land flux can further be used to calculate the residual land CO₂ uptake flux when subtracting land use change. This is calculated via:

$$F_{\text{land}} = F'_{\text{land}} - E_{\text{mluc}} \quad (5.6)$$

In general the net land sink (F'_{land}) is subject to much stronger variability compared to the ocean carbon sink, ranging from a net outgassing flux in post El Niño 1998 of $1.2 \pm 0.7 \text{ Pg C} \cdot \text{yr}^{-1}$ to a strong sink of $-3.1 \pm 0.8 \cdot \text{yr}^{-1}$ in 2011. Throughout the period 1998 through 2011, land use change describes a fairly constant source of carbon, leading to a residual land flux (F_{land}) ranging from $-0.36 \pm 0.87 \text{ Pg C} \cdot \text{yr}^{-1}$ in 1998 to $-4.06 \pm 1.01 \text{ Pg C} \cdot \text{yr}^{-1}$ in 2011.

Average fossil fuel emissions and land use change combined from 1998 through 2011 are $8.94 \pm 0.64 \text{ Pg C} \cdot \text{yr}^{-1}$, hence the ocean was taking up on average roughly 24% of the global emitted CO₂ from 1998 through 2011. On average, the atmospheric accumulation rate was $-4.13 \pm 0.17 \text{ Pg C} \cdot \text{yr}^{-1}$, and the resulting average residual land flux was $-2.7 \pm 0.9 \text{ Pg C} \cdot \text{yr}^{-1}$, hence roughly 46% of the emitted carbon accumulated in the atmosphere, while about 30% of the emitted carbon was taken up by land from 1998 through 2011.

This is in good agreement with the recent carbon budget of Le Quéré *et al.* (2013) from 2001 through 2011. These authors report emissions from fossil fuels of $8.3 \pm 0.4 \text{ Pg C} \cdot \text{yr}^{-1}$ and from land use change of $1.0 \pm 0.5 \text{ Pg C} \cdot \text{yr}^{-1}$ and an atmospheric accumulation rate of $-4.3 \pm 0.1 \text{ Pg C} \cdot \text{yr}^{-1}$. The largest difference stems from the ocean flux component, which in Le Quéré *et al.* (2013) is estimated to be $-2.5 \pm 0.5 \text{ Pg C} \cdot \text{yr}^{-1}$. Closing the budget, Le Quéré *et al.* (2013) estimate a residual land flux of $-2.6 \pm 0.8 \text{ Pg C} \cdot \text{yr}^{-1}$ within the 2001 through 2011 time period.

5.8 Summary and conclusion

In this chapter, the SOM–FFN method is used to estimate the sea surface $p\text{CO}_2$ on a global scale from 1998 through 2011. The validation of the results show that they are in good agreement with observations from the SOCAT v2 database, with a small bias of $-0.12 \mu\text{atm}$ and a RMSE of $12.05 \mu\text{atm}$, but also compare reasonably well with independent timeseries stations and mooring data with a bias ranging from close to zero to $-7.15 \mu\text{atm}$ and a RMSE ranging between $11.44 \mu\text{atm}$ and $26.80 \mu\text{atm}$.

The mean results compare well with results from existing studies. The net air–sea flux of this study is estimated to be $-1.54 \pm 0.65 \text{ Pg C} \cdot \text{yr}^{-1}$ for the global ocean south of 79°N , which adds up to $-1.66 \pm 0.65 \text{ Pg C} \cdot \text{yr}^{-1}$ when the recent Arctic estimate of Schuster *et al.* (2013) is included. This number is similar to the undersampling corrected estimate of $-1.6 \pm 0.9 \text{ Pg C} \cdot \text{yr}^{-1}$ obtained from Takahashi *et al.* (2009). Including an estimate for riverine–derived carbon (Jacobson *et al.*, 2007) results in an anthropogenic air–sea CO_2 flux of $-2.11 \pm 0.68 \text{ Pg C} \cdot \text{yr}^{-1}$, which compares well with the recent estimates of $2.0 \text{ Pg C} \cdot \text{yr}^{-1}$ obtained from Wanninkhof *et al.* (2013b).

Within the study period from 1998 through 2011, the global integrated flux of CO_2 undergoes considerable year–to–year changes. From 1998 through 2011 an increasing carbon sink of $-1.09 \pm 0.13 \text{ Pg C} \cdot \text{yr}^{-1} \cdot \text{decade}^{-1}$ is estimated, which is however strongly influenced by intra–decadal signals, such as ENSO and SAM, and is therefore unlikely part of a long–term trend signal. The strongest flux increase is found in the Atlantic Ocean and the Southern Ocean with $-0.36 \pm 0.03 \text{ Pg C} \cdot \text{yr}^{-1} \cdot \text{decade}^{-1}$ and $-0.36 \pm 0.07 \text{ Pg C} \cdot \text{yr}^{-1} \cdot \text{decade}^{-1}$, respectively. While in chapter 4 the Atlantic Ocean sink increase was almost entirely driven by the southern hemisphere, the prolonged timeseries in chapter 5 reveals an equally strong increase in both hemispheres of the Atlantic Ocean. The strong increase in the Southern Ocean carbon uptake is potentially linked to the 1998 SAM peak followed by a weakening of the index polarity of the SAM index.

Throughout the study period, the Pacific Ocean shows the largest standard deviation of the IAV of $\pm 0.12 \text{ Pg C} \cdot \text{yr}^{-1}$. An EOF analysis shows, that this signal is driven by the ENSO mode, which appears to be the dominant mode of the global sea surface $p\text{CO}_2$ variability. The standard deviation of the IAV for the Southern Ocean, the Atlantic Ocean and the Indian Ocean is estimated to be $\pm 0.06 \text{ Pg C} \cdot \text{yr}^{-1}$, $\pm 0.04 \text{ Pg C} \cdot \text{yr}^{-1}$ and ± 0.02

$\text{Pg C} \cdot \text{yr}^{-1}$ respectively.

Over the 1998 through 2011 period, roughly 24% of the emitted CO_2 was taken up by the ocean. This estimate implies a global average residual land flux of $-2.70 \pm 0.93 \text{ Pg C} \cdot \text{yr}^{-1}$ over the same period, which accounts for 30% of the emitted carbon. From 1998 through 2011 the land sink has been more variable than the global ocean CO_2 sink.

While variabilities within the considered periods of this thesis are mainly driven by natural climate modes, like e.g. ENSO, which explains roughly 1/4 of the global variability alone, the increasing trend in sea surface $p\text{CO}_2$ is mainly driven by the anthropogenic impact on the climate system. However, trends over periods of 10–14 years as presented in this chapter are not necessarily representative for a longer term trend of several decades, but are more influenced by naturally occurring intra-decadal variabilities, hence it is questionable that a linear trend assumption is suitable for trend estimates on short timescales. The strong reduction of the CO_2 uptake in the post 1998 La Niña years strongly shapes the CO_2 flux trends identified. It is therefore essential to distinguish between anthropogenic trends and natural variabilities in order to predict future increase.

It is essential to understand the variability of the ocean carbon sink in order to accurately quantify the different pathways of emitted carbon, hence the results in this thesis illustrate that ocean measurements are important not only for the purpose of inferring the ocean carbon sink, but also for the quantification of the global land sink and the closure of the global carbon budget. Furthermore, the basin-wide and global sea surface $p\text{CO}_2$ maps form an accurate basis to evaluate global biogeochemical models and are suitable to improve future climate projections.

Chapter 6

General discussions, conclusions and future research

"Only a Sith deals in absolutes."

(Obi-Wan Kenobi)

6.1 Summary of the main findings

A new 2-step neural network approach has been developed to overcome most of the limitations of currently used data interpolation methods. The SOM-FFN method is capable of capturing a large amount of variability due to the non-linear predictor-observation relationship on a fine $1^\circ \times 1^\circ$ spatial grid, which makes it suitable to investigate the inter-annual variability of the air-sea CO_2 flux. The method determines the non-linear relationships between the surface ocean $p\text{CO}_2$ observations and a set of input data to produce basin-wide sea surface maps of $p\text{CO}_2$ on a monthly basis. The network gathers information from similar ocean biogeochemical provinces and provides regional $p\text{CO}_2$ estimates, which are then used to investigate the changing distribution of the sea surface $p\text{CO}_2$ in the Atlantic Ocean and the global ocean. The method, however, relies on the assumption that the ocean carbon sink and its variability can be estimated as a function of proxy data, which are subjectively chosen. Furthermore, the method relies on ocean carbon measurements in order to establish a correct relationship between input data and target $p\text{CO}_2$.

Observations form the basis of the methods and the results presented in this thesis. Chapter 2 provides an insight on how sea surface $p\text{CO}_2$ measurements are obtained, analysed and quality controlled. Bottle data, collected along the UK-Caribbean section, reveal three distinct water masses. The first, west of 55°W , shows an autumn maximum of sea surface $p\text{CO}_2$. The second, from 24°W to 55°W , shows the temperature driven gyre seasonality with the lowest partial pressures in the cooler winter and spring months and $p\text{CO}_2$ maximum in summer. The third region, east of 24°W (north of $\sim 40^\circ\text{N}$), shows the reversed seasonal cycle influenced by biological production in agreement with Takahashi *et al.* (2002). Along the line, DIC and TALK changes counteract the temperature driven $p\text{CO}_2$ increase with contributions to the $p\text{CO}_2$ change of all drivers, namely $> 200 \mu\text{atm}$ from SST and $< -200 \mu\text{atm}$ from DIC and TALK. Changes in freshwater does not have a strong affect on the sea surface $p\text{CO}_2$

The validation of the $p\text{CO}_2$ estimates obtained from the method presented in chapter 2, conducted for the Atlantic Ocean from 1998 through 2007 in chapter 4 and the global ocean from 1998 through 2011 illustrate the great potential of the method to reproduce available observations from the SOCAT database (Pfeil *et al.*, 2013; Sabine *et al.*, 2013; Bakker *et al.*, 2014) within a small error margin. Furthermore, no temporal biases exist

and the error is independent of the available data density. Independent testing shows that the model–data error increases, however, observations from timeseries stations and independent data products are reconstructed within a reasonable small error.

The results in chapter 4 and 5 illustrate that the Atlantic Ocean from 1998 through 2007 and the global Ocean from 1998 through 2011 have been major sinks for CO₂ on average ($-0.45 \pm 0.15 \text{ Pg C} \cdot \text{yr}^{-1}$ and $-1.54 \pm 0.65 \text{ Pg C} \cdot \text{yr}^{-1}$, respectively). The sink strength strongly varies in time and from ocean basin to basin. From 1998 through 2011 the Pacific Ocean shows the strongest variability of all basins ($\pm 0.12 \text{ Pg C} \cdot \text{yr}^{-1}$, 1σ) linked to the ENSO climate mode, while the Southern Ocean and the Atlantic Ocean show the strongest increase in the CO₂ flux ($-0.36 \pm 0.07 \text{ Pg C} \cdot \text{yr}^{-1} \cdot \text{decade}^{-1}$ and $-0.36 \pm 0.03 \text{ Pg C} \cdot \text{yr}^{-1} \cdot \text{decade}^{-1}$, respectively). Trends are sensitive to the chosen start and end year (Fay and McKinley, 2013), which is reflected in the difference between trend estimates for the Atlantic Ocean from 1998 through 2007 ($-0.15 \pm 0.04 \text{ Pg C} \cdot \text{yr}^{-1} \cdot \text{decade}^{-1}$) and from 1998 through 2011 ($-0.36 \pm 0.03 \text{ Pg C} \cdot \text{yr}^{-1} \cdot \text{decade}^{-1}$). The global ocean flux estimates further have the potential to close the global carbon budget and to provide an estimate of the atmosphere–land flux of CO₂ based on ocean CO₂ measurements.

6.2 General discussion and conclusions

Until recently, observation–based $p\text{CO}_2$ estimates were limited to climatologies (Takahashi *et al.*, 2009) or ocean sub–basins (Telszewski *et al.*, 2009; Nakaoka *et al.*, 2013) due to methodological limitations and data sparsity. Recently, with increasing numbers of available observations, new methods were developed to estimate seasonal (Rödenbeck *et al.*, 2013) as well as the inter–annual variability of the air–sea CO₂ flux (Park *et al.*, 2010). The combined 2–step SOM–FFN method presented in this thesis represents a novel method within a new generation of statistical approaches to interpolate available observations in time and space. Similar to Sasse *et al.* (2013), the global ocean is in a first step clustered into biogeochemical provinces using a SOM technique, however, the second step (the FFN) does not assume linearity, hence it is capable to reconstruct large amounts of variability due to the non–linear input–output relationships. A complete assessment of the existing methods to create global sea surface $p\text{CO}_2$ maps is currently in progress

(SOCOM – Surface Ocean $p\text{CO}_2$ Mapping inter-comparison), which includes the inter-comparison e.g. of neural network based methods (Sasse *et al.*, 2013; Landschützer *et al.*, 2013; Nakaoka *et al.*, 2013), estimates derived from empirical relationships (Park *et al.*, 2010) and from an ocean mixed layer scheme (Rödenbeck *et al.*, 2013).

One important question arising from observational studies is: "Where do we need to set the focus for future measurements, to improve current estimates?". This question can be answered by looking at figure 1.10, illustrating two major issues:

- there are several ocean areas which remain largely unobserved, particularly in the southern hemisphere
- the majority of the global ocean has been observed in less than four calendar months

Firstly, without ocean sided CO_2 observations in certain areas current data interpolation methods and models can not be validated, hence large uncertainties remain regarding the strength of the ocean carbon sink in unobserved areas. Secondly, without a good sea surface $p\text{CO}_2$ observation coverage in time, the quantification of seasonal as well inter-annual variabilities remains a challenge. Figure 5.1 reveals, that the high latitude North Pacific, as well as the Equatorial Pacific and the high latitude North Atlantic Ocean are amongst the regions with the largest bias and standard deviation of the $p\text{CO}_2$ residuals, although they are well observed compared to the remainder of the global ocean. This indicates that these regions show $p\text{CO}_2$ variability in time, which can not fully be reconstructed with the chosen input data in chapter 4 and 5. In conclusion, the recommendation towards the sea surface CO_2 measurement community, drawn from this thesis is to make observations in data-poor areas and to continue the measurement effort in high $p\text{CO}_2$ variability areas, such as the high latitude North Atlantic, high latitude North Pacific, and the Equatorial Pacific.

The quality of surface ocean CO_2 observations has been highlighted in chapter 3. Wanninkhof *et al.* (2013a) reports uncertainties related to measurements that received a SOCAT flag A or B of $\leq 2 \mu\text{atm}$ and $\leq 5 \mu\text{atm}$ for flags C or D. Currently, however, the model-observation error exceeds these values (e.g. RMSE in chapter 5 of $12.05 \mu\text{atm}$), hence, in conclusion, it is more favourable to reduce the misfit between $p\text{CO}_2$ estimates and observations, rather than the measurement uncertainty, as the first contributes more to

Table 6.1: Comparison of the contemporary (observation–based estimates without river input) and the anthropogenic flux (corrected by the input of riverine carbon) between the SOM–FFN results with recent studies. Annotations: ^(a) the flux estimate was derived from the best estimate of the available RECCAP methodologies; ^(b) Global Carbon Budget v1.3 numbers were used to calculate the sink from 1998 through 2011; ^(c) SOM–FFN estimate including the Arctic Ocean best estimate of Schuster *et al.* (2013) of $-0.12 \text{ Pg C} \cdot \text{yr}^{-1}$ for both the anthropogenic and the contemporary flux

| | Time period or reference year | Anthropogenic flux [Pg C · yr ⁻¹] | Contemporary flux [Pg C · yr ⁻¹] |
|---|-------------------------------------|---|--|
| Wanninkhof <i>et al.</i> (2013b) ^(a) | 1990-2009 | -2.0 | - |
| Sasse <i>et al.</i> (2013) | 2000 | - | -1.55±0.32 |
| Takahashi <i>et al.</i> (2009) | 2000 | -2.0±1.0 | -1.6±0.9 |
| Gruber <i>et al.</i> (2009) | 1995-2000 | -2.2±0.3 | -1.7±0.4 |
| Le Quéré <i>et al.</i> (2013) ^(b) | 1998-2011 | -2.4±0.5 | - |
| SOM-FFN ^(c) | 1998-2011 | -2.11±0.68 | -1.66±0.66 |

the overall uncertainty of the air–sea CO₂ flux. This may be achieved e.g. by collecting more observations in high $p\text{CO}_2$ variability areas.

In general the largest source of uncertainty estimated for the air–sea CO₂ flux in chapter 4 and chapter 5 stems from the formulation of the gas transfer velocity, reflecting the disadvantage of the $\Delta p\text{CO}_2$ method. Hence, in order to reduce the uncertainty of CO₂ flux estimates, derived from the $\Delta p\text{CO}_2$ method, attention needs to be drawn towards the formulation of the kinetic gas transfer term. It would be favourable for the community to provide an uncertainty metric regarding different formulation from the literature with regards to the different wind products.

In general, the uncertainty estimate of the SOM–FFN $p\text{CO}_2$ fields provides room for improvement. So far, regional deviations in the residuals derived from the difference between SOM–FFN $p\text{CO}_2$ estimates and the SOCAT gridded observations have been ignored in the studies conducted within this thesis. Furthermore, decorrelation length estimates of the residuals can be potentially improved by adopting a different approach (see e.g. Jones *et al.*, 2012).

The global ocean long term mean CO₂ flux results (chapter 5) compare well with recent studies. Table 6.1 lists the global ocean flux estimate derived from the SOM–FFN method together with recent global estimates of the contemporary and anthropogenic CO₂ flux, as reported in the literature. Although time periods vary for the individual flux estimates, derived from a variety of methods, the temporal mean carbon flux estimates

show strong agreement, hence it can be concluded that the global mean SOM–FFN based CO₂ flux provides a robust estimate.

Table 6.2 lists the long term mean CO₂ flux, as well as variability and trend estimates from the studies conducted in this thesis and recently derived estimates for the major global ocean sub–basins from the RECCAP project for comparison. The SOM–FFN based CO₂ flux results show good agreement with the long term mean estimates derived from the different studies, with the exception of the Southern Ocean.

Although significant differences between methodologies occur, the standard deviation of the IAV's reported in the Atlantic Ocean (Schuster *et al.*, 2013) are consistently smaller compared to the Pacific Ocean (Ishii *et al.*, 2013), which is reflected in the SOM–FFN results. Results from the Southern Ocean (Lenton *et al.*, 2013) and Indian Ocean (Sarma *et al.*, 2013) do not include the standard deviation of the IAV, but the maximum value of the IAV and the IAV amplitude (temporal maximum of the IAV minus temporal minimum), respectively. The SOM–FFN derived maximum of the IAV compares well with models and inversion studies in the Southern Ocean and the IAV amplitude is on the lower end of the range estimated within the Indian Ocean RECCAP study.

Trend estimates show the largest source of disagreement between the SOM–FFN results and the RECCAP studies (see table 6.2). This is likely linked to different time periods investigated. While CO₂ flux trends in RECCAP are calculated from 1990 through 2009, SOM–FFN CO₂ trends are for the period 1998 through 2011. This difference in the investigated time period is particularly pronounced in the Southern Ocean. In conclusion, short term trends (10–14 years as discussed in this thesis) have to be viewed with caution, as they are likely influenced by climate modes, such as ENSO, NAO and SAM.

The influence of climate modes, in particular ENSO, NAO and SAM has been discussed within this thesis in chapters 4 and 5. The phasing of the ENSO event is the main driver of the CO₂ flux variability, explaining roughly 28% of the global variability alone. This is particularly reflected in the Pacific Ocean IAV estimate ($\pm 0.12 \text{ Pg C} \cdot \text{yr}^{-1}$, 1σ) of the air–sea CO₂ flux. Increasing air–sea CO₂ flux trends in the Southern Ocean from 1998 through 2011 illustrated in chapter 5 are possibly linked to the recent weakening of the SAM phase as suggested by Fay and McKinley (2013). A weak but significant air–sea CO₂ flux to NAO correlation was further established within the subtropical and subpolar

Table 6.2: Comparison between the SOM–FFN estimates from 1998 through 2011 and the range of estimates derived from the RECCAP project (see section 1.4 for the full range of available methodologies) from 1990 through 2009 for the major ocean basins. Annotations: (*) refers to the best estimate; (+) refers to the median of all methodologies; (a) based on ocean biogeochemical (OBGC) models and inversions; (b) based on OBGC; (c) from atmospheric inversions; (d) numbers in brackets stem from the Atlantic Ocean study (chapter 4), whereas regular values stem from the global ocean study (chapter 5); (e) SOM–FFN estimate including the Arctic Ocean best estimate of Schuster *et al.* (2013) of $-0.12 \text{ Pg C} \cdot \text{yr}^{-1}$; (f) excluding the Arctic Ocean; (g) IAV calculation method corresponds to those reported in the RECCAP studies, i.e. $\text{std}(\text{IAV})$, $\text{max}(\text{IAV})$, IAV amplitude

| | Atlantic + Arctic Ocean 44°S-90°N | Pacific Ocean 44°S-60°N | Southern Ocean 75°S-44°S | Indian Ocean 44°S-30°N |
|--|---|----------------------------|---|------------------------------|
| Schuster <i>et al.</i> (2013): | | | | |
| temporal mean [$\text{Pg C} \cdot \text{yr}^{-1}$] | $-0.61 \pm 0.06^{(*)}$ | - | - | - |
| std(IAV) [$\text{Pg C} \cdot \text{yr}^{-1}$] | ± 0.015 to ± 0.055 | - | - | - |
| trend [$\text{Pg C} \cdot \text{yr}^{-1} \cdot \text{decade}^{-1}$] | 0.017 ± 0.026 to -0.290 ± 0.007 | - | - | - |
| Ishii <i>et al.</i> (2013): | | | | |
| temporal mean [$\text{Pg C} \cdot \text{yr}^{-1}$] | - | $-0.40 \pm 0.21^{(*)}$ | - | - |
| std(IAV) [$\text{Pg C} \cdot \text{yr}^{-1}$] | - | ± 0.08 to ± 0.29 | - | - |
| trend [$\text{Pg C} \cdot \text{yr}^{-1} \cdot \text{decade}^{-1}$] | - | - | - | - |
| Lenton <i>et al.</i> (2013): | | | | |
| temporal mean [$\text{Pg C} \cdot \text{yr}^{-1}$] | - | - | $-0.42 \pm 0.07^{(+)}$ | - |
| max(IAV) [$\text{Pg C} \cdot \text{yr}^{-1}$] | - | - | $0.10^{(b)}$ to $0.11^{(c)}$ | - |
| trend ^(a) [$\text{Pg C} \cdot \text{yr}^{-1} \cdot \text{decade}^{-1}$] | - | - | $-0.09 \pm 0.04^{(b)}$ to $0.00 \pm 0.03^{(c)}$ | - |
| Sarma <i>et al.</i> (2013): | | | | |
| temporal mean [$\text{Pg C} \cdot \text{yr}^{-1}$] | - | - | - | $-0.37 \pm 0.08^{(+)}$ |
| IAV amplitude [$\text{Pg C} \cdot \text{yr}^{-1}$] | - | - | - | $0.05^{(b)}$ to $0.24^{(c)}$ |
| trend ^(b) [$\text{Pg C} \cdot \text{yr}^{-1} \cdot \text{decade}^{-1}$] | - | - | - | -0.01 |
| SOM-FFN ^(d) : | | | | |
| temporal mean ^(e) [$\text{Pg C} \cdot \text{yr}^{-1}$] | -0.66 ± 0.17 (-0.56 ± 0.16) | -0.41 ± 0.31 | -0.23 ± 0.09 | -0.35 ± 0.10 |
| IAV ^(g) [$\text{Pg C} \cdot \text{yr}^{-1}$] | $\pm 0.04^{(f)}$ | ± 0.12 | 0.11 | 0.08 |
| trend [$\text{Pg C} \cdot \text{yr}^{-1} \cdot \text{decade}^{-1}$] | $-0.36 \pm 0.03^{(f)}$ ($-0.15 \pm 0.04^{(f)}$) | -0.29 ± 0.13 | -0.36 ± 0.07 | -0.13 ± 0.02 |

North Atlantic in chapter 4, but this signal is weak compared to e.g. ENSO. It can be concluded, that these links provide a step forward in our understanding of the variability of the global ocean carbon sink and how future climate variability may effect the global ocean sink strength.

At last, the SOM–FFN based global air–sea CO₂ flux estimates have great potential to estimate the global land CO₂ flux, based on sea surface *p*CO₂ measurements, and to close the global carbon budget. Within a comparable time–frame, the land flux estimates in chapter 5 show good agreement with recent estimates from Le Quéré *et al.* (2013).

6.3 Future research

Several conclusions/outcomes drawn from this thesis deserve further investigation:

- Extension of the SOM–FFN study period would help to establish the response of the air–sea CO₂ flux to climate modes such as ENSO, SAM and NAO and corresponding multi–decadal variabilities. Excluding satellite chlorophyll–a from the SOM–FFN input data sets provides an opportunity to prolong the analysis period and this option will be explored in subsequent work. Results from the sensitivity runs in chapter 4 indicate that estimates without chlorophyll–a as an input parameter do not significantly change the major findings, however, chlorophyll–a is a simple, but unique proxy representing the relation between biology and *p*CO₂ and further tests are needed to establish if chlorophyll can be neglected when considering longer timescales.
- In order to investigate the recent reinvigoration of the Southern Ocean carbon sink, it is vital to extend the analysis and look at regional variabilities of both the *p*CO₂ as well as all drivers of the carbonate system in response to the SAM and ENSO climate modes.
- The largest uncertainty with regards to the integrated air–sea CO₂ flux from surface ocean *p*CO₂ observations stems from the gas transfer formulation. In order to decrease the uncertainty of the ocean carbon uptake it is essential to more accurately quantify the gas exchange parametrization.

- It would be beneficial to integrate the open ocean estimates of the air–sea CO₂ fluxes, presented in this thesis, with high resolution coastal estimates. Recently, Chen *et al.* (2013) evaluated data from 165 estuaries and 87 continental shelves and concluded that the former release roughly 0.1 Pg C · yr⁻¹ whereas the latter take up -0.4 Pg C · yr⁻¹, highlighting the importance of coastal regions regarding the global ocean CO₂ uptake.
- The 2–step neural network approach (chapter 2) has great potential to be used for high resolution regional surface ocean pCO₂ estimates, e.g. in shelf seas, under consideration of relevant input data, such as the extend of seasonal stratification, as derived from the difference between mixed layer depth and the bottom depth.
- The SOM–FFN sea surface pCO₂ and air–sea CO₂ flux product shows that the data collection and synthesis effort of the marine carbon community makes it possible to investigate the seasonal to inter–annual variability of the ocean carbon sink based on observations. There are, however, gaps in the global surface ocean CO₂ observation network, particularly in the southern hemisphere, that make it impossible to validate pCO₂ estimates in these regions. Hence, the recommendation towards the sea surface CO₂ measurement community is, to obtain measurements in undersampled areas and to continue the measurement effort in high pCO₂ variability areas.

Appendix A

List of acronyms

Table A.1: List of acronyms

| Abbreviations | Description |
|----------------------|---|
| AABW | Antarctic Bottom Water |
| ACC | Antarctic Circumpolar Current |
| BATS | Bermuda Atlantic Timeseries Station |
| BMU | Best Matching Unit |
| CARBOCHANGE | Changes in Carbon Uptake and Emissions by Oceans in a Changing Climate |
| CARBOOCEAN | Marine Carbon Sources and Sinks Assessment |
| CARINA | Carbon in the Atlantic Ocean |
| CAVASSOO | Carbon Variability Studies by Ships of Opportunity |
| CCMP | Cross Calibrated Multi Platform |
| CDIAC | Carbon Dioxide Information Analysis Centre |
| CHL | Chlorophyll-a |
| CO2SYS | Program Developed for CO ₂ System Calculations |
| CRM | Certified Reference Material |
| DIC | Dissolved Inorganic Carbon |
| DJF | December, January, February |
| ECCO2 | Estimating the Circulation and Climate of the Ocean Phase 2 |

Continued on next page

Table A.1 – continued from previous page

| Abbreviations | Description |
|----------------------|---|
| EKE | Eddy kinetic energy |
| ENSO | El Niño Southern Oscillation |
| EOF | Empirical Orthogonal Function |
| ESRL | Earth System Research Laboratory |
| ESTOC | European Station for Timeseries in the Ocean |
| FFN | Feed-forward Network |
| FINP | Feed-forward Input Dataset |
| FINP2 | Feed-forward Input Dataset 2 |
| FITR | Feed-forward Input Training subset |
| FIVAL | Feed-forward Input Validation subset |
| FRESH | Freshwater |
| GEOSECS | Geochemical Ocean Sections program |
| GLODAP | Global Data Analysis Project |
| HadCRUT | Combined temperature record of the Hadley Centre and the Climatic Research Unit |
| HOT | Hawaiian Ocean Timeseries station |
| IAV | Inter-annual Variability |
| IPCC | Intergovernmental Panel on Climate Change |
| JGOFS | Joint Global Ocean Flux Study |
| JJA | June, July, August |
| LDEO | Lamont-Doherty Earth Observatory |
| MAM | March, April, May |
| MEI | Multivariate ENSO Index |
| MLD | Mixed Layer Depth |
| MPR | Multi-Parameter Regression |
| NADW | North Atlantic Deep Water |
| NAO | North Atlantic Oscillation |
| NCAR | National Centre for Atmospheric Research |

Continued on next page

Table A.1 – continued from previous page

| Abbreviations | Description |
|----------------------|--|
| NCEP | National Centres for Environmental Prediction |
| NOAA | National Oceanic and Atmospheric Administration |
| OACES | Ocean–Atmosphere Exchange Study |
| OBGC | Ocean Biogeochemical |
| PC | Principal component |
| RECCAP | Regional Carbon Cycle Assessment and Processes |
| RMSE | Root Mean Squared Error |
| SAM | Southern Annular Mode |
| sDIC | salinity normalized Dissolved inorganic Carbon |
| SeaWIFS | Sea–viewing Wide Field–of–view Sensor |
| SINP | Self–organizing Map Input dataset |
| SOCAT | Surface Ocean Carbon Atlas |
| SOCOM | Surface Ocean $p\text{CO}_2$ Mapping inter–comparison |
| SODA | Simple Ocean Data Assimilation |
| SOM | Self–Organizing Map |
| SOMMA | Single–Operator Multi–Metabolic Analyser |
| SON | September, October, November |
| SOP | Standard Operating Procedure |
| SR1–4 | Sensitivity runs 1–4 |
| SSS | Sea Surface Salinity |
| SST | Sea Surface Temperature |
| sTALK | salinity normalized Total Alkalinity |
| TALK | Total Alkalinity |
| VINDTA | Versatile Instrument for the Determination of Total Inorganic Carbon and Titration Alkalinity |
| VOS | Voluntary Observing Ships |
| WOCE | World Ocean Circulation Experiment |

Appendix B

List of symbols

Table B.1: Variables, symbols and description

| Variables and symbols | Description | Standard Units |
|------------------------|--------------------------------|--|
| Chapter 1 | | |
| subscript <i>abs</i> | Absorbed radiation | - |
| superscript <i>em</i> | Emitted radiation | - |
| E | Energy | $\text{W} \cdot \text{m}^{-2}$ |
| subscript <i>S</i> | Sun | - |
| subscript <i>E</i> | Earth | - |
| R | Radius | m |
| T | absolute Temperature | K |
| D | Distance between Sun and Earth | m |
| σ | Stefan–Boltzmann constant | $\text{W} \cdot \text{m}^{-2} \cdot \text{K}^{-4}$ |
| α | Earth albedo | - |
| Em | Carbon emissions | $\text{Pg C} \cdot \text{yr}^{-1}$ |
| subscript <i>ff</i> | fossil fuel | - |
| subscript <i>luc</i> | land use change | - |
| subscript <i>atm</i> | atmosphere | - |
| subscript <i>land</i> | land surface | - |
| subscript <i>ocean</i> | ocean surface | - |

Continued on next page

Table B.1 – continued from previous page

| Variables and symbols | Description | Standard Units |
|---------------------------------|--------------------------------------|---|
| F | Carbon flux | Pg C · yr ⁻¹ |
| $\frac{dG}{dt}$ | carbon accumulation rate | Pg C · yr ⁻¹ |
| subscript eq | equilibrium | - |
| subscript w | water | - |
| K ₀ | Solubility of CO ₂ | mol C · l ⁻¹ · atm ⁻¹ |
| <i>p</i> A | partial pressure of molecule A | μatm |
| [A] | concentration of molecule A | mol · l ⁻¹ |
| A ₁₋₃ | constants for solubility calculation | - |
| B ₁₋₃ | constants for solubility calculation | - |
| s | salinity | - |
| z | film layer thickness | m |
| ε | molecular diffusivity | m ² · s ⁻¹ |
| k | gas transfer velocity | cm · hr ⁻¹ |
| κ | transfer resistance factor | - |
| Sc | Schmidt Number | - |
| A,B,C,D | Constants for Sc calculation | - |
| n | Schmidt Number exponent | - |
| ν | kinematic viscosity of water | m ² · s ⁻¹ |
| <i>u</i> * | friction velocity | m · s ⁻¹ |
| $\tau_{Rayolds}$ | Raynolds stress | Pa |
| $\bar{\rho}$ | average density | kg · m ⁻³ |
| subscript gas | in gas phase | - |
| K ₁ , K ₂ | equilibrium constants | mol C · l ⁻¹ |
| Chapter 2 | | |
| <i>f</i> A | fugacity of a non-ideal gas A | μatm |
| superscript <i>surf</i> | surface | - |
| R | universal gas constant | J · K ⁻¹ · mol ⁻¹ |
| P | barometric pressure | Pa |

Continued on next page

Table B.1 – continued from previous page

| Variables and symbols | Description | Standard Units |
|------------------------------|---|--------------------------------|
| B, δ | viral coefficients | - |
| subscript ds | deseasonalised | - |
| subscript $clim$ | climatology | - |
| \mathbf{p} | neural network input vector | - |
| d | Euclidean distance | - |
| \mathbf{W} | weight matrix of a neuron | - |
| subscript i | neuron index | - |
| superscript j | index of input vector | - |
| subscript n | number of input vector elements | - |
| subscript m | number of neurons | - |
| \mathbf{S} | step function | - |
| b | network bias | - |
| a | network layer output element | - |
| subscript est | estimated | - |
| subscript opt | optimal | - |
| r | ratio | % |
| h | modifiable network parameters | - |
| \mathbf{x} | vector containing network weights and biases | - |
| \mathbf{J} | Jacobian matrix | - |
| \mathbf{I} | Identity matrix | - |
| \mathbf{e} | vector containing network errors | - |
| μ | network learning rate | - |
| f_{ice} | percentage of ice over | % |
| u | wind speed at 10m high | $\text{m} \cdot \text{s}^{-1}$ |
| x_A | mole fraction of a gas A | ppm |
| Chapter 3 | | |
| N | Coulometer reading | counts |

Continued on next page

Table B.1 – continued from previous page

| Variables and symbols | Description | Standard Units |
|------------------------------|----------------------------------|--------------------------------------|
| b | background reading | counts · min ⁻¹ |
| a | acid blank reading | counts |
| ρ | density | kg · m ⁻³ |
| subscript s | seawater sample | - |
| a-e | polynomial constants | - |
| γ | buffer factor | $\mu\text{mol} \cdot \text{kg}^{-1}$ |
| Chapter 4 | | |
| a,b,c | coefficients for exponential fit | - |
| Chapter 5 | | |
| F' | net Carbon flux | Pg C · yr ⁻¹ |

References

- Amari, S., N. Murata, K.-R. Muller, M. Finke, and H. H. Yang (1997), Asymptotic Statistical Theory of Overtraining and Cross-Validation, *IEEE TRANSACTIONS ON NEURAL NETWORKS*, 8, 985–996.
- Antonov, J. I., D. Seidov, T. P. Boyer, R. A. Locarnini, A. V. Mishonov, and H. E. Garcia (2010), *World Ocean Atlas 2009, Volume 2: Salinity*, NOAA Atlas NESDIS 69, U.S. Government Printing Office, Washington, D.C.
- Atlas, R., R. N. Hoffman, J. Ardizzone, S. M. Leidner, J. C. Jusem, D. K. Smith, and D. Gombos (2011), A cross-calibrated multiplatform ocean surface wind velocity product for meteorological and oceanographic applications, *Bull. Amer. Meteor. Soc.*, 92, 157–174.
- Bakker, D. C. E. (1998), *Process studies of the air-sea exchange of carbon dioxide in the Atlantic Ocean*, Ph.D. thesis, Rijksuniversiteit Groningen.
- Bakker, D. C. E., B. Pfeil, K. Smith, S. Hankin, A. Olsen, S. R. Alin, C. Cosca, S. Harasawa, A. Kozyr, Y. Nojiri, K. M. O'Brien, U. Schuster, M. Telszewski, B. Tilbrook, C. Wada, J. Akl, L. Barbero, N. R. Bates, J. Boutin, Y. Bozec, W.-J. Cai, R. D. Castle, F. P. Chavez, L. Chen, M. Chierici, K. Currie, H. J. W. de Baar, W. Evans, R. A. Feely, A. Fransson, Z. Gao, B. Hales, N. J. Hardman-Mountford, M. Hoppema, W.-J. Huang, C. W. Hunt, B. Huss, T. Ichikawa, T. Johannessen, E. M. Jones, S. D. Jones, S. Jutterström, V. Kitidis, A. Körtzinger, P. Landschützer, S. K. Lauvset, N. Lefèvre, A. B. Manke, J. T. Mathis, L. Merlivat, N. Metzl, A. Murata, T. Newberger, A. M. Omar, T. Ono, G.-H. Park, K. Paterson, D. Pierrot, A. F. Ríos, C. L. Sabine, S. Saito, J. Salisbury, V. V. S. S. Sarma, R. Schlitzer, R. Sieger, I. Skjelvan, T. Steinhoff, K. F. Sullivan, H. Sun, A. J. Sutton, T. Suzuki, C. Sweeney, T. Takahashi, J. Tjiputra, N. Tsurushima, S. M. A. C. van Heuven, D. Vandemark, P. Vlahos, D. W. R. Wallace, R. Wanninkhof, and A. J. Watson (2014), An update to the Surface Ocean CO₂ Atlas (SOCAT version 2), *Earth System Science Data*, 6(1), 69–90.
- Ballantyne, A. P., C. B. Alden, J. B. Miller, P. P. Tans, and J. W. C. White (2012), Increase in observed net carbon dioxide uptake by land and oceans during the last 50 years, *Nature*, 488, 70–72.
- Bates, N. R. (2007), Interannual variability of the oceanic CO₂ sink in the subtropical gyre of the North Atlantic Ocean over the last 2 decades, *Journal of Geophysical Research*, 112, 26.
- Bennington, V., G. A. McKinley, S. Dutkiewicz, and D. Ullman (2009), What does chlorophyll variability tell us about export and air-sea CO₂ flux variability in the North Atlantic?, *Global Biogeochemical Cycles*, 23, GB3002.

- Broecker, W. S., and T.-H. Peng (1974), Gas exchange rates between air and sea, *Tellus*, 26, 21–35.
- Broecker, W. S., and T.-H. Peng (1982), Tracers in the sea, in: *ELDIGO Press*, p. 690, Lamont Doherty Geological Observatory, Pelisades, NY.
- CARINA-Group (2009), *Carbon in the Atlantic Ocean Region - the CARINA project: Results and Data, Version 1.0.*, Tech. rep., Carbon Dioxide Information Analysis Center, Oak Ridge National Laboratory, U.S. Department of Energy, Oak Ridge, Tennessee.
- Carton, J. A., and B. S. Giese (2008), A Reanalysis of Ocean Climate Using Simple Ocean Data Assimilation (SODA), *Monthly weather review*, 136, 2999–3017.
- Chen, C.-T. A., T.-H. Huang, Y.-C. Chen, Y. Bai, X. He, and Y. Kang (2013), Air-sea exchanges of CO₂ in the world's coastal seas, *Biogeosciences*, 10, 6509–6544.
- Chierici, M., A. Olsen, T. Johannessen, J. Trinañes, and R. Wanninkhof (2009), Algorithms to estimate the carbon dioxide uptake in the northern North Atlantic using ship board observations, satellite and ocean analysis data, *Deep-Sea Research II*, 56, 630–639.
- Cooper, D. J., A. J. Watson, and R. D. Ling (1998), Variation of P_{CO2} along a North Atlantic shipping route (U.K. to the Caribbean): A year of automated observations, *Marine Chemistry*, 60, 147–164.
- Corbière, A., N. Metzl, G. Reverdin, C. Brunet, and T. Takahashi (2007), Interannual and decadal variability of the oceanic carbon sink in the North Atlantic subpolar gyre, *Tellus*, 59B, 168–178.
- de Boyer Montegut, C., G. Madec, A. S. Fischer, A. Lazar, and D. Iudicone (2004), Mixed layer depth over the global ocean: An examination of profile data and a profile-based climatology, *Journal of Geophysical Research*, 109, C12,003.
- Demuth, H., M. Beale, and M. Hagan (2008), *Neural Network Toolbox 6 Users Guide*, The MathWorks, Inc., 3 Apple Hill Drive, Natick, MA.
- Dickson, A. G. (1981), An exact definition of total alkalinity and a procedure for the estimation of alkalinity and total inorganic carbon from titration data, *Deep-Sea Research*, 28A, 609–623.
- Dickson, A. G., and F. J. Millero (1987), A comparison of the equilibrium constants for the dissociation of carbonic acid in seawater media, *Deep-Sea Research*, 34, 1733–1743.
- Dickson, A. G., C. L. Sabine, and J. R. Christian (Eds.) (2007), *Guide to Best Practices for Ocean CO₂ Measurements*, PICES special Publication, IOCCP Report No. 8.
- Doney, S., V. Fabry, R. A. Feely, and J. Kleypas (2009), Ocean Acidification: The Other CO₂ Problem, *Annual Reviews Marine Science*, 1, 169–192.
- Dore, J. E., R. Lukas, D. W. Sadler, and D. M. Church, M. J. and Karl (2009), Physical and biogeochemical modulation of ocean acidification in the central North Pacific, *Proc Natl Acad Sci USA*, 106, 12,235–12,240.
- Etheridge, D. M., L. P. Steele, R. L. Langenfelds, R. J. Francey, J.-M. Barnola, and V. I. Morgan (1996), Natural and anthropogenic changes in atmospheric CO₂ over the last 1000 years from air in Antarctic ice and firn., *Journal of Geophysical Research*, 101, 4115–4128.

- Fay, A. R., and G. A. McKinley (2013), Global trends in surface ocean pCO₂ from in situ data, *Global Biogeochemical Cycles*, 27, 1–17.
- Feely, R. A., T. Takahashi, R. Wanninkhof, M. J. McPhaden, C. E. Cosca, S. C. Sutherland, and M.-E. Carr (2006), Decadal variability of the air-sea CO₂ fluxes in the equatorial Pacific Ocean, *Journal of Geophysical Research*, 111, C08S90.
- Friedrich, T., and A. Oschlies (2009), Neural network-based estimates of North Atlantic surface pCO₂ from satellite data: A methodological study, *Journal of Geophysical Research*, 114, 1–12.
- GLOBALVIEW-CO2 (2011), Cooperative Atmospheric Data Integration Project - Carbon Dioxide, CD-ROM, NOAA ESRL, Boulder, Colorado [Also available on Internet via anonymous FTP to ftp.cmdl.noaa.gov, Path: ccg/co2/GLOBALVIEW].
- Gloor, M., N. Gruber, J. Sarmiento, C. L. Sabine, R. A. Feely, and C. Rödenbeck (2003), A first estimate of present and preindustrial air-sea CO₂ flux patterns based on ocean interior carbon measurements and models, *Geophysical Research Letters*, 30, 1010.
- Glover, D. M., W. J. Jenkins, and S. C. Doney (2011), *Modelling Methods for Marine Science*, Cambridge University Press.
- Gnanadesikan, A., and R. Hallber (2002), *Encyclopedia of Physical Science and Technology*, chap. Physical oceanography, thermal structure and general circulation, Academic Press, San Diego.
- González-Dávila, M., J. M. Santana-Casiano, and E. F. González-Dávila (2007), Interannual variability of the upper ocean carbon cycle in the Northeast Atlantic Ocean, *Geophysical Research Letters*, 34, L07,608.
- Graven, H. D., T. Guilderson, and R. Keeling (2012), Observations of radiocarbon in CO₂ at seven global sampling sites in the Scripps flask network: Analysis of spatial gradients and seasonal cycles, *Journal of Geophysical Research*, 117, D02,303.
- Gruber, N. (2009), Fickle trends in the ocean, *Nature*, 458, 155–156.
- Gruber, N., M. Gloor, S. E. Mikaloff Fletcher, S. C. Doney, S. Dutkiewicz, M. J. Follows, M. Gerber, A. R. Jacobson, F. Joos, K. Lindsay, D. Menemenlis, A. Mouchet, S. A. Müller, J. L. Sarmiento, and T. Takahashi (2009), Oceanic sources, sinks, and transport of atmospheric CO₂, *Global Biogeochemical Cycles*, 23, GB1005.
- Gruber, N., C. D. Keeling, and N. R. Bates (2002), Interannual Variability in the North Atlantic Ocean Carbon Sink, *Science*, 298, 2374.
- Gruber, N., J. L. Sarmiento, and T. F. Stocker (1996), An improved method for detecting anthropogenic CO₂ in the oceans, *Global Biogeochemical Cycles*, 10, 809–837.
- Gurney, K., D. Baker, P. Rayner, and S. Denning (2008), Interannual variations in continental-scale net carbon exchange and sensitivity to observing networks estimated from atmospheric CO₂ inversions for the period 1980 to 2005, *Global Biogeochemical Cycles*, 22, GB3025.
- Hagan, M., H. B. Demuth, and M. H. Beale (1996), *Neural Network Design*, PWS Publishing, Boston, MA.
- Hagan, M. T., and M. Menhaj (1994), Training feed-forward networks with the Marquardt algorithm, *IEEE Transactions on Neural Networks*, Vol. 5, 989–993.

- Haverd, V., M. R. Raupach, P. Briggs, J. G. Canadell, S. J. Davis, R. M. Law, C. P. Meyer, G. P. Peters, C. Pickett-Heaps, and B. Sherman (2013), The Australian terrestrial carbon budget, *Biogeosciences*, *10*, 851–869.
- Houghton, R. A. (2003), Revised estimates of the annual net flux of carbon to the atmosphere from changes in land use and land management 1850-2000, *Tellus Series B - Chemical and Physical Meteorology*, *55*, 378–390.
- Hurrell, J. W. (1995), Decadal trends in the North-Atlantic Oscillation: regional temperatures and precipitation, *Science*, *269*, 676–679.
- Ishii, M., R. A. Feely, K. B. Rodgers, G.-H. Park, R. Wanninkhof, D. Sasano, H. Sugimoto, C. E. Cosca, S. Nakaoka, M. Telszewski, Y. Nojiri, S. E. Mikaloff Fletcher, Y. Niwa, P. K. Patra, V. Valsala, h. Nakano, I. Lima, S. C. Doney, E. T. Buitenhuis, O. Aumont, j. P. Dunne, A. Lenton, and T. Takahashi (2013), Air-sea CO₂ flux in the Pacific Ocean for the period 1990-2009, *Biogeosciences Discussions*, *10*, 12,155–12,216.
- Jacobson, A. R., S. E. Mikaloff Fletcher, N. Gruber, J. Sarmiento, and M. Gloor (2007), A joint atmosphere-ocean inversion for surface fluxes of carbon dioxide: 2. Regional results, *Global Biogeochemical Cycles*, *21*, GB1020.
- Jähne, B., K. O. Münnich, R. Bössinger, A. Dutzi, W. Huber, and P. Libner (1987), On parameters influencing air-water gas exchange, *Journal of Geophysical Research*, *92*, 1937–1949.
- Johnson, K., K. Wills, D. Butler, W. Johnson, and C. Wong (1993), Coulometric total carbon dioxide analysis for marine studies: maximizing the performance of an automated gas extraction system and coulometric detector, *Marine Chemistry*, *44*, 167–187.
- Jones, E. P., and S. D. Smith (1977), A First Measurement of Air-Sea CO₂ Flux by Eddy Correlation, *Journal of Geophysical Research*, *82*, 5990–5992.
- Jones, S. D., C. Le Quéré, and C. Rödenbeck (2012), Autocorrelation characteristics of surface ocean pCO₂ and air-sea CO₂ fluxes, *Global Biogeochemical Cycles*, *26*, GB2042.
- Kalkhan, M. A. (2011), *Spatial Statistics: GeoSpatial Information Modelling and Thematic Mapping*, Taylor & Francis Group.
- Kalnay, E., M. Kanamitsu, R. Kistler, W. Collins, D. Deaven, L. Gandin, M. Iredell, S. Saha, G. White, J. Woollen, Y. Zhu, A. Leetmaa, R. Reynolds, M. Chelliah, W. Ebisuzaki, W. Higgins, J. Janowiak, K. C. Mo, C. Ropelewski, J. Wang, R. Jenne, and D. Joseph (1996), The NCEP/NCAR 40-year reanalysis project, *Bulletin of the American Meteorological Society*, *77*, 437–470.
- Keeling, C. D. (1968), Carbon Dioxide in Surface Ocean Waters, *Journal of Geophysical Research*, *73*, 4543–4553.
- Keeling, R. F., B. B. Stephens, R. G. Najjar, S. C. Doney, D. Archer, and N. Heimann (1998), Seasonal variations in the atmospheric O₂/N₂ ratio in relation to the kinetics of air-sea gas exchange., *Global Biogeochemical Cycles*, *12*, 141–164.
- Keller, K., F. Joos, C. C. Raible, V. Cocco, T. Frölicher, J. P. Dunne, M. Gehlen, L. Bopp, J. C. Orr, J. Tjiputra, C. Heinze, J. Segsneider, T. Roy, and N. Metzl (2012), Variability

- of the ocean carbon cycle in response to the North Atlantic Oscillation, *Tellus*, *64*, 18,738.
- 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 Biogeochemical Cycles*, *18*, GB4031.
- Khatiwala, S., T. Tanhua, S. Mikaloff Fletcher, M. Gerber, S. C. Doney, H. D. Graven, N. Gruber, G. A. McKinley, A. Murata, A. F. Rios, and C. L. Sabine (2013), Global ocean storage of anthropogenic carbon, *Biogeosciences*, *10*, 2169–2191.
- Kiehl, J., and K. Trenberth (1997), Earths annual global mean energy budget, *Bull. Am. Meteorol. Soc.*, *78*, 197–206.
- Knutti, R., T. F. Stocker, F. Joos, and G.-K. Plattner (2003), Probabilistic climate change projections using neural networks, *Climate Dynamics*, *21*, 257–272.
- Kohonen, T. (1987), *Self-Organization and Associative Memory*, 2nd edn., Springer-Verlag, Berlin.
- Kohonen, T. (2001), *Self-Organizing Maps*, 3rd edn., Springer-Verlag, Berlin Heidelberg New York.
- Kolehmainen, M., H. Martikainen, and J. Ruuskanen (2001), Neural networks and periodic components used in air quality forecasting, *Atmospheric Environment*, *35*, 815–825.
- Körtzinger, A. (1999), *Methods of Seawater Analysis*, chap. Determination of carbon dioxide partial pressure (p(CO₂)), pp. 149–158, Verlag Chemie.
- Kottek, M., J. Grieser, C. Beck, B. Rudolf, and F. Rubel (2006), World Map of the Kppen-Geiger climate classification updated, *Meteorologische Zeitschrift*, *15*, 259–263.
- Landschützer, P., N. Gruber, D. C. E. Bakker, U. Schuster, S. Nakaoka, M. R. Payne, T. Sasse, and J. Zeng (2013), A neural network-based estimate of the seasonal to inter-annual variability of the Atlantic Ocean carbon sink, *Biogeosciences*, *10*, 7793–7815.
- Le Quéré, C., R. J. Andres, T. Boden, T. Conway, R. A. Houghton, J. I. House, G. Marland, G. P. Peters, G. R. van der Werf, A. Ahlström, R. M. Andrew, L. Bopp, J. G. Canadell, P. Ciais, S. C. Doney, C. Enright, P. Friedlingstein, C. Huntingford, A. K. Jain, C. Jourdain, E. Kato, R. F. Keeling, K. Klein Goldewijk, S. Levis, P. Levy, M. Lomas, B. Poulter, M. R. Raupach, J. Schwinger, S. Sitch, B. D. Stocker, N. Viovy, S. Zaehle, and N. Zeng (2013), The global carbon budget 1959-2011, *Earth System Science Data*, *5*, 165–185.
- Le Quéré, C., M. R. Raupach, J. G. Canadell, G. Marland, C. P. Bopp, L., T. J. Conway, S. C. Doney, R. A. Feely, P. Foster, P. Friedlingstein, K. Gurney, R. A. Houghton, J. I. House, C. Huntingford, P. E. Levy, M. R. Lomas, J. Majkut, N. Metzl, J. P. Ometto, G. P. Peters, I. C. Prentice, J. T. Randerson, S. W. Running, J. L. Sarmiento, U. Schuster, S. Sitch, T. Takahashi, R. v. G. Viovy, N., and F. I. Woodward (2009), Trends in the sources and sinks of carbon dioxide, *Nature Geoscience*, *2*, 831–836.
- Le Quéré, C., C. Rödenbeck, E. T. Buitenhuis, T. J. Conway, R. Langenfelds, A. Gomez, C. Labuschagne, M. Ramonet, T. Nakazawa, N. Metzl, N. Gillett, and M. Heimann

- (2007), Saturation of the Southern Ocean CO₂ Sink Due to Recent Climate Change, *Science*, 316, 1735–1738.
- Le Quéré, C., T. Takahashi, E. T. Buitenhuis, C. Rödenbeck, and S. C. Sutherland (2010), Impact of climate change and variability on the global oceanic sink of CO₂, *Global Biogeochemical Cycles*, 24, GB4007.
- Le Treut, H., R. Somerville, U. Cubasch, Y. Ding, C. Mauritzen, A. Mokssit, P. T., and M. Prather (2007), *Climate Change 2007: The Physical Science Basis. Contribution of Working Group I to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change*, chap. Historical Overview of Climate Change, pp. 93–127, Cambridge University Press, United Kingdom and New York, NY, USA.
- Lee, K., L. Tong, F. J. Millero, C. L. Sabine, A. G. Dickson, C. Goyet, P. G.-H. and Wanninkhof R., R. A. Feely, and R. M. Key (2006), Global relationships of total alkalinity with salinity and temperature in surface waters of the world's oceans, *Geophysical Research Letters*, 33, L19,605.
- Lefèvre, N., A. J. Watson, A. Olsen, A. F. Rios, F. F. Pérez, and T. Johannessen (2004), A decrease in the sink for atmospheric CO₂ in the North Atlantic, *Geophysical Research Letters*, 31, L07,306.
- Lefèvre, N., A. J. Watson, and A. R. Watson (2005), A comparison of multiple regression and neural network techniques for mapping in situ pCO₂ data, *Tellus*, 57B, 375–384.
- Lenton, A., B. Tilbrook, R. M. Law, D. Bakker, S. C. Doney, N. Gruber, M. Ishii, M. Hoppema, N. S. Lovenduski, R. J. Matear, B. I. McNeil, N. Metzl, S. E. Mikaloff Fletcher, P. M. S. Monteiro, C. Rödenbeck, C. Sweeney, and T. Takahashi (2013), Sea-air CO₂ fluxes in the Southern Ocean for the period 1990-2009, *Biogeosciences*, 10, 4037–4054.
- Levitus, S., J. I. Antonov, T. P. Boyer, O. K. Baranova, H. E. Garcia, R. A. Locarnini, A. V. Mishonov, J. R. Reagan, D. Seidov, E. S. Yarosh, and M. M. Zweng (2012), World ocean heat content and thermosteric sea level change (0-2000 m), 1955-2010, *Geophysical Research Letters*, 39, L10,603.
- Lewis, E., and D. W. R. Wallace (1998), *Program Developed for CO₂ System Calculations, ORNL/CDIAC-105*, Carbon Dioxide Information Analysis Center, Oak Ridge National Laboratory, U.S. Department of Energy, Oak Ridge, Tennessee.
- Liss, P. S., and L. Merlivat (1986), *The Role of Air-Sea Exchange in Geochemical Cycling*, chap. Air-sea gas exchange rates: Introduction and Synthesis, pp. 113–127, D. Reidel, Dordrecht, Netherlands.
- Liss, P. S., and P. G. Slater (1974), Flux of gases across the air-sea interface, *Nature*, 247, 181–184.
- Longhurst, A., S. Sathyendranath, T. Platt, and C. Caverhill (1995), An estimate of global primary production in the ocean from satellite radiometer data, *Journal of Plankton Research*, 17, 1245–1271.
- Lovenduski, N. S., N. Gruber, S. C. Doney, and D. Lima, I. (2007), Enhanced CO₂ outgassing in the Southern Ocean from a positive phase of the Southern Annular Mode, *Global Biogeochemical Cycles*, 21, GB2026.
- Lozier, S. (2010), Deconstructing the Conveyor Belt, *Science*, 328, 1507–1511.

- Lüger, H., R. Wanninkhof, D. W. R. Wallace, and A. Körtzinger (2006), CO₂ fluxes in the subtropical and subarctic North Atlantic based on measurements from a volunteer observing ship, *Journal of Geophysical Research*, *111*, C06,024.
- Lundstedt, H. (1992), Neural Networks and Predictions of Solar-Terrestrial Effects, *Planet and Space Science*, *40*, 457–464.
- Lüthi, D., M. Le Floch, B. Bereiter, T. Blunier, J.-M. Barnola, U. Siegenthaler, D. Raynaud, J. Jouzel, H. Fischer, K. Kawamura, and T. F. Stocker (2008), High-resolution carbon dioxide concentration record 650,000-800,000 years before present, *Nature*, *453*, 379–382.
- Maier, H. R., and G. C. Dandy (2000), Neural networks for the prediction and forecasting of water resources variables: a review of modelling issues and applications, *Environmental Modelling & Software*, *15*, 101–124.
- Marland, G., T. A. Boden, and R. J. Andres (2005), *Regional, and National fossil fuel CO₂ emissions. in Trends: A Compendium of data on global change*, Tech. rep., Carbon Dioxide Information Analysis Center, Oak Ridge National Laboratory, U.S. Department of Energy, Oak Ridge, Tenn., U.S.A.
- Marquardt, D. W. (1963), An Algorithm for Least-Squares Estimation of Nonlinear Parameters, *Journal of the Society for Industrial and Applied Mathematics*, *Vol. II*, 431–441.
- Marshall, G. J. (2003), Trends in the Southern Annular Mode from observations and re-analyses, *Journal of Climate*, *16*, 4134–4143.
- Marshall, J., Y. Kushnir, D. Battisti, P. Chang, A. Czaja, R. Dickson, J. Hurrell, M. McCartney, R. Saravanan, and M. Visbeck (2001), North Atlantic Climate Variability: Phenomena, Impacts and Mechanisms, *International Journal of Climatology*, *21*, 1863–1898.
- Marzban, C., and G. J. Stumpf (1996), A Neural Network for Tornado Prediction Based on Doppler Radar-Derived Attributes, *Journal of Applied Meteorology*, *35*, 617–626.
- McGillis, W. R., J. B. Edson, J. E. Hare, and C. W. Fairall (2001), Direct covariance air-sea CO₂ fluxes, *Journal of Geophysical Research*, *106*, 16,729–16,745.
- McKinley, G. A., A. R. Fay, T. Takahashi, and N. Metzl (2011), Convergence of atmospheric and North Atlantic carbon dioxide trends on multidecadal timescales, *Nature Geoscience*, *4*, 606–610.
- McPhaden, M. J., A. J. Busalacchi, R. Cheney, J.-R. Donguy, K. S. Gage, D. Halpern, M. Ji, P. Julian, G. Meyers, G. T. Mitchum, P. P. Niiler, J. Picaut, R. W. Reynolds, N. Smith, and K. Takeuchi (1998), The Tropical Ocean-Global Atmosphere observing system: A decade of progress, *Journal of Geophysical Research: Oceans*, *103*(C7), 14,169–14,240.
- Mehrbach, C., C. H. Culbertson, and R. M. Hawley, J. E. and Pytkowicz (1973), Measurement of the apparent dissociation constants of carbonic acid in seawater at atmospheric pressure, *Limnology and Oceanography*, *18*, 897–907.
- Menemenlis, D., J. Campin, P. Heimbach, C. Hill, T. Lee, A. Nguyen, M. Schodlok, and H. Zhang (2008), ECCO2: High resolution global ocean and sea ice data synthesis, *Mercator Ocean Quarterly Newsletter*, *31*, 13–21.

- Meredith, M. P., and A. M. Hogg (2006), Circumpolar response of Southern Ocean eddy activity to a change in the Southern Annular Mode, *Geophysical Research Letters*, *33*, L16,608.
- Metzl, N., A. Corbière, G. Reverdin, A. Lenton, T. Takahashi, A. Olsen, T. Johannessen, D. Pierrot, R. Wanninkhof, S. R. Ólafsdóttir, J. Olafsson, and M. Ramonet (2010), Recent acceleration of the sea surface fCO₂ growth rate in the North Atlantic subpolar gyre (1993-2008) revealed by winter observations, *Global Biogeochemical Cycles*, *24*, GB4004.
- Mikaloff Fletcher, S. E., N. Gruber, A. R. Jacobson, S. C. Doney, S. Dutkiewicz, M. Gerber, M. Follows, F. Joos, K. Lindsay, D. Menemenlis, A. Mouchet, S. A. Müller, and J. L. Sarmiento (2006), Inverse estimates of anthropogenic CO₂ uptake, transport, and storage by the ocean, *Global Biogeochemical Cycles*, *20*, GB2002.
- Mikaloff Fletcher, S. E., N. Gruber, A. R. Jacobson, M. Gloor, S. C. Doney, S. Dutkiewicz, M. Gerber, M. Follows, F. Joos, K. Lindsay, D. Menemenlis, A. Mouchet, S. A. Müller, and J. Sarmiento (2007), Inverse estimates of the oceanic sources and sinks of natural CO₂ and the implied oceanic carbon transport, *Global Biogeochemical Cycles*, *21*, GB1010.
- Milankovitch, M. (1920), *Theorie Mathematique des Phenomenes Thermiques produits par la Radiation Solaire*, Gauthier-Villars Paris.
- Milankovitch, M. (1930), *Mathematische Klimalehre und Astronomische Theorie der Klimaschwankungen, Handbuch der Klimologie Band 1 Teil A*, Borntraeger Berlin.
- Millero, F. J., A. G. Dickson, G. Eiseid, C. Goyet, P. Guenther, K. M. Johnson, R. M. Key, K. Lee, D. Purkerson, C. L. Sabine, R. G. Schottle, D. W. R. Wallace, E. Lewis, and W. C. D. (1998), Assessment of the quality of the shipboard measurements of total alkalinity on the WOCE Hydrographic Program Indian Ocean CO₂ survey cruises 1994-1996, *Marine Chemistry*, *63*, 9-20.
- Mintrop, L. (2011), *VINDTA - Versatile INSTRument for the Determination of Titration Alkalinity, MANUAL for VERSIONS 3S and 3C*, Marianda, Tulpenweg 28, D- 24145 Kiel, Germany, v2 edn.
- Mintrop, L., F. Pérez, M. González-Dávila, J. Santana-Casiano, and A. Körtzinger (2000), Alkalinity Determination by Potentiometry: Intercalibration using three Different Methods, *Ciencias Marinas*, *26*, 23-27.
- Morice, C. P., J. J. Kennedy, N. A. Rayner, and P. D. Jones (2012), Quantifying uncertainties in global and regional temperature change using an ensemble of observational estimates: the HadCRUT4 dataset, *Journal of Geophysical Research*, *117*, D08,101.
- Morrow, R., M. L. Ward, A. M. Hogg, and S. Pasquet (2010), Eddy response to Southern Ocean climate modes, *Journal of Geophysical Research*, *115*, C10,030.
- Myhre, G., D. Shindell, F.-M. Breon, W. Collins, J. Fuglestedt, J. Huang, D. Koch, J.-F. Lamarque, D. Lee, B. Mendoza, T. Nakajima, A. Robock, G. Stephens, T. Takemura, and H. Zhang (2013), *Climate Change 2013: The Physical Science Basis. Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change*, chap. Anthropogenic and Natural Radiative Forcing, Cambridge University Press.

- Nakaoka, S., M. Telszewski, Y. Nojiri, S. Yasunaka, C. Miyazaki, H. Mukai, and N. Usui (2013), Estimating temporal and spatial variation of ocean surface pCO₂ in the North Pacific using a self-organizing map neural network technique, *Biogeosciences*, *10*, 6093–6106.
- Nightingale, P., G. Malin, C. Law, A. Watson, P. Liss, M. Liddicoat, J. Boutin, and R. Upstill-Goddard (2000), In situ evaluation of air-sea gas exchange parameterizations using novel conservative and volatile tracers, *Global Biogeochemical Cycles*, *14*(1), 373–387.
- Olafsson, J. (2007), *Irminger Sea cruise data from the 1991-2006 cruises*, CARINA Data Set, Tech. rep., Carbon Dioxide Information Analysis Center, Oak Ridge National Laboratory, US Department of Energy, Oak Ridge, Tennessee.
- Olsen, A., A. M. Omar, R. G. J. Bellerby, T. Johannessen, U. S. Ninnemann, K. R. Brown, K. A. Olsson, J. Olafsson, G. Nondal, C. Kivimäe, S. Kringstad, C. Neill, and S. Olafsdottir (2006), Magnitude and origin of the anthropogenic CO₂ increase and ¹³C Suess effect in the Nordic seas since 1981, *Global Biogeochemical Cycles*, *20*, GB3027.
- Osborn, T. J. (2011), Winter 2009/2010 temperatures and a record-breaking North Atlantic Oscillation index, *Weather*, *66*, 19–21.
- Park, G.-H. and Wanninkhof, R., S. C. Doney, T. Takahashi, K. Lee, R. Feely, C. Sabine, J. Trinañes, and I. Lima (2010), Variability of global net sea-air CO₂ fluxes over the last three decades using empirical relationships, *Tellus*, *62B*, 352–368.
- Peng, T.-H., T. Takahashi, W. S. Broecker, and J. Olafsson (1987), Seasonal variability of carbon dioxide, nutrients and oxygen in the northern North Atlantic surface water: Observations and model, *Tellus*, *39B*, 439–458.
- Peng, T.-H., and R. Wanninkhof (2010), Increase in anthropogenic CO₂ in the Atlantic Ocean in the last two decades, *Deep-Sea Research I*, *57*, 755–770.
- Peylin, P., R. M. Law, K. R. Gurney, F. Chevallier, A. R. Jacobson, T. Maki, Y. Niwa, P. K. Patra, W. Peters, P. J. Rayner, C. Rödenbeck, and X. Zhang (2013), Global atmospheric carbon budget: results from an ensemble of atmospheric CO₂ inversions, *Biogeosciences Discussions*, *10*, 5301–5360.
- Pfeil, B., A. Olsen, D. C. E. Bakker, S. Hankin, H. Koyuk, A. Kozyr, J. Malczyk, A. Manke, N. Metzl, C. L. Sabine, J. Akl, S. R. Alin, N. Bates, R. G. J. Bellerby, A. Borges, J. Boutin, P. J. Brown, W.-J. Cai, F. P. Chavez, A. Chen, C. Cosca, A. J. Fassbender, R. A. Feely, M. González-Dávila, C. Goyet, B. Hales, N. Hardman-Mountford, C. Heinze, M. Hood, M. Hoppema, C. W. Hunt, D. Hydes, M. Ishii, T. Johannessen, S. D. Jones, R. M. Key, A. Körtzinger, P. Landschützer, S. K. Lauvset, N. Lefèvre, A. Lenton, A. Lourantou, L. Merlivat, T. Midorikawa, L. Mintrop, C. Miyazaki, A. Murata, A. Nakadate, Y. Nakano, S. Nakaoka, Y. Nojiri, A. M. Omar, X. A. Padin, G.-H. Park, K. Paterson, F. F. Perez, D. Pierrot, A. Poisson, A. F. Rios, J. M. Santana-Casiano, J. Salisbury, V. V. S. S. Sarma, R. Schlitzer, B. Schneider, U. Schuster, R. Sieger, I. Skjelvan, T. Steinhoff, T. Suzuki, T. Takahashi, K. Tedesco, M. Telszewski, H. Thomas, B. Tilbrook, J. Tjiputra, D. Vandemark, T. Veness, R. Wanninkhof, A. J. Watson, R. Weiss, C. S. Wong, and H. Yoshikawa-Inoue (2013), A uniform, quality controlled Surface Ocean CO₂ Atlas (SOCAT), *Earth System Science Data*, *5*, 125–143.

- Rayner, N. A., D. E. Parker, E. B. Horton, C. K. Folland, L. V. Alexander, D. P. Rowell, E. C. Kent, and A. Kaplan (2003), Global analyses of sea surface temperature, sea ice, and night marine air temperature since the late nineteenth century, *Journal of Geophysical Research*, 108(D14), 4407.
- Redfield, A. C., B. H. Ketchum, and F. A. Richards (1963), *The Sea*, chap. The influence of organisms on the composition of sea-water, pp. 26–77, Wiley, New York.
- Reynolds, R. W., N. A. Rayner, T. M. Smith, D. C. Stokes, and W. Wang (2002), An improved in situ and satellite SST analysis for climate, *Journal of Climate*, 15, 1609–1625.
- Robinson, A., and H. Stommel (1959), The oceanic thermocline and the associated thermohaline circulation, *Tellus*, 3, 295–308.
- Robock, A. (2000), Volcanic eruptions and climate, *Reviews of Geophysics*, 38, 191–219.
- Rödenbeck, C., R. Keeling, D. Bakker, N. Metzl, A. Olsen, C. Sabine, and M. Heimann (2013), Global surface-ocean pCO₂ and sea-air CO₂ flux variability from an observation-driven ocean mixed-layer scheme., *Ocean Science*, 9, 93–216.
- Sabine, C. L., R. A. Feely, N. Gruber, R. M. Key, K. Lee, J. L. Bullister, R. Wanninkhof, C. S. Wong, D. W. R. Wallace, B. Tilbrook, F. J. Millero, T.-H. Peng, A. Kozyr, T. Ono, and A. F. Rios (2004), The oceanic sink for anthropogenic CO₂., *Science (New York, N.Y.)*, 305(5682), 367–71.
- Sabine, C. L., S. Hankin, H. Koyuk, D. C. Bakker, B. Pfeil, A. Olsen, N. Metzl, A. Kozyr, A. Fassbender, A. Manke, J. Malczyk, J. Akl, S. R. Alin, R. G. J. Bellerby, A. Borges, J. Boutin, P. J. Brown, W.-J. Cai, F. P. Chavez, A. Chen, C. Cosca, R. A. Feely, A. González-Dávila, C. Goyet, N. Hardman-Mountford, C. Heinze, M. Hoppema, C. W. Hunt, D. Hydes, M. Ishii, T. Johannessen, R. M. Key, A. Körtzinger, P. Landschützer, S. K. Lauvset, N. Lefèvre, A. Lenton, A. Lourantou, L. Merlivat, T. Midorikawa, L. Mintrop, C. Miyazaki, A. Murata, A. Nakadate, Y. Nakano, S. Nakaoka, Y. Nojiri, A. M. Omar, X. A. Padin, G.-H. Park, K. Paterson, F. F. Perez, D. Pierrot, A. Poisson, A. F. Rios, J. Salisbury, J. M. Santana-Casiano, V. V. S. S. Sarma, R. Schlitzer, B. Schneider, U. Schuster, R. Sieger, I. Skjelvan, T. Steinhoff, T. Suzuki, T. Takahashi, K. Tedesco, M. Telszewski, H. Thomas, B. Tilbrook, D. Vandemark, T. Veness, A. J. Watson, R. Weiss, W. C. S., and H. Yoshikawa-Inoue (2013), Surface Ocean CO₂ Atlas (SOCAT) gridded data products, *Earth System Science Data*, 5, 145–153.
- Sarma, V. V. S. S., A. Lenton, R. M. Law, N. Metzl, P. K. Patra, S. Doney, I. D. Lima, E. Dlugokencky, M. Ramonet, and V. Valsala (2013), Sea-air fluxes in the Indian Ocean between 1990 and 2009, *Biogeosciences*, 10, 7035–7052.
- Sarmiento, J., and N. Gruber (2006), *Ocean Biogeochemical Dynamics*, Princeton University Press.
- Sarmiento, J. M., M. Gloor, N. Gruber, C. Beaulieu, A. R. Jacobson, S. E. Mikaloff Fletcher, S. Pacala, and K. Rodgers (2010), Trends and regional distributions of land and ocean carbon sinks, *Biogeosciences*, 7, 2351–2367.
- Sasse, T. P., B. I. McNeil, and G. Abramowitz (2013), A new constraint on global air-sea CO₂ fluxes using bottle carbon data, *Geophysical Research Letters*, 40, 1594–1599.

- Schuster, U., A. Hannides, L. Mintrop, and A. Kortzinger (2009a), Sensors and instruments for oceanic dissolved carbon measurements, *Ocean Science.*, *5*, 547–558.
- Schuster, U., G. A. McKinley, N. Bates, F. Chevallier, S. C. Doney, A. R. Fay, M. González-Dávila, N. Gruber, S. Jones, J. Krijnen, P. Landschützer, N. Lefèvre, M. Manizza, J. Mathis, N. Metzl, A. Olsen, A. F. Rios, C. Rödenbeck, J. M. Santana-Casiano, T. Takahashi, R. Wanninkhof, and A. J. Watson (2013), Atlantic and Arctic sea-air CO₂ fluxes, 1990–2009, *Biogeosciences*, *10*, 607–627.
- Schuster, U., and A. J. Watson (2007), A variable and decreasing sink for atmospheric CO₂ in the North Atlantic., *Journal of Geophysical Research*, *112*, C11,006.
- Schuster, U., A. J. Watson, N. R. Bates, A. Corbière, M. Gonzalez-Davila, N. Metzl, D. Pierrot, and M. Santana-Casiano (2009b), Trends in North Atlantic sea-surface fCO₂ from 1990 to 2006., *Deep-Sea Research II*, *56*, 620–629.
- Stendardo, I., and N. Gruber (2012), Oxygen trends over five decades in the North Atlantic, *Journal of Geophysical Research*, *117*, C11,004.
- Stocker, T. F., D. Qin, G.-K. Plattner, M. Tignor, S. K. Allen, J. Boschung, A. Nauels, Y. Xia, V. Bex, and P. M. Midgley (Eds.) (2013), *Climate Change 2013: The Physical Science Basis. Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change*, chap. Summary for Policymakers, Cambridge University Press.
- Sutton, A., J. Mathis, C. Sabine, S. Maenner, and S. Musielewicz (2013), *High-resolution ocean and atmosphere pCO₂ time-series measurements from mooring Papa 145W 50N*, Tech. rep., Carbon Dioxide Information Analysis Center, Oak Ridge National Laboratory, US Department of Energy, Oak Ridge, Tennessee.
- Sweeney, C., E. Gloor, A. R. Jacobson, R. M. Key, G. McKinley, J. L. Sarmiento, and R. Wanninkhof (2007), Constraining global air-sea gas exchange for CO₂ with recent bomb ¹⁴C measurements, *Global Biogeochemical Cycles*, *21*, GB2015.
- Takahashi, T., R. A. Feely, R. F. Weiss, R. Wanninkhof, D. W. Chipman, S. C. Sutherland, and T. T. Takahashi (1997), Global air-sea flux of CO₂: An estimate based on measurements of sea-air pCO₂ difference, *Proceedings of the National Academy of Sciences*, *94*, 8292–8299.
- Takahashi, T., J. Olafson, J. Goddard, D. Chipman, and S. Sutherland (1993), Seasonal Variations of CO₂ and Nutrients in the High-Latitude Surface Oceans: A Comparative Study, *Global Biogeochemical Cycles*, *7*(4), 843–878.
- Takahashi, T., and Sutherland (2013), *Climatological mean distribution of pH and carbonate ion concentration in global ocean surface waters in the unified pH scale and mean rate of changes in selected areas*, Tech. rep., Lamont-Doherty Earth Observatory of Columbia University.
- Takahashi, T., S. Sutherland, R. Wanninkhof, C. Sweeney, R. Feely, D. Chipman, B. Hales, G. Friederich, F. Chavez, C. Sabine, A. Watson, D. Bakker, U. Schuster, N. Metzl, H. Yoshikawa-Inoue, M. Ishii, T. Midorikawa, Y. Nojiri, A. Körtzinger, T. Steinhoff, M. Hoppema, J. Olafson, T. Arnarson, B. Tilbrook, T. Johannessen, A. Olsen, R. Bellerby, C. Wong, B. Delille, N. Bates, and H. de Baar (2009), Climatological mean and decadal change in surface ocean pCO₂, and net sea-air CO₂ flux over the global oceans, *Deep-Sea Research II*, *56*, 554–577.

- Takahashi, T., S. C. Sutherland, and A. Kozyr (2008), *Global Ocean Surface Water Partial Pressure of CO₂ Database: Measurements Performed during 1968-2006 (Version 1.0)*, Tech. rep., Carbon Dioxide Information Analysis Center, Oak Ridge National Laboratory, U. S. Department of Energy, Oak Ridge, TN 37831, pp.20.
- Takahashi, T., S. C. Sutherland, and A. Kozyr (2013), *Global Ocean Surface Water Partial Pressure of CO₂ Database: Measurements Performed During 1957-2012 (Version 2012). ORNL/CDIAC-160, NDP-088(V2012)*, Tech. rep., Carbon Dioxide Information Analysis Center, Oak Ridge National Laboratory, U.S. Department of Energy, Oak Ridge, Tennessee.
- Takahashi, T., S. C. Sutherland, C. Sweeney, A. Poisson, N. Metzl, B. Tilbrook, N. Bates, R. Wanninkhof, R. F. Feely, C. Sabine, J. Olafsson, and Y. Nojiri (2002), Global sea-air CO₂ flux based on climatological surface ocean pCO₂, and seasonal biological and temperature effects, *Deep-Sea Research II*, 49, 1601–1622.
- Takahashi, T., R. Wanninkhof, R. Feely, R. Weiss, D. Chipman, N. Bates, J. Olafsson, C. Sabine, and S. Sutherland (1999), Net sea-air CO₂ flux over the global oceans: An improved estimate based on the sea-air pCO₂ difference, in: *Proceedings of the Second International Symposium, CO₂ in the Oceans* (Nojiri, Y., ed.), pp. 9–14, Center for Global Environmental Research, National Institute for Environmental Studies, Tsukuba, Japan.
- Takahashi, T., R. Weiss, R. Wanninkhof, D. Chipman, and R. Feely (2003), Global air-sea flux of CO₂: an estimate based on measurements of sea-air pCO₂ difference, *Proceedings of the National Academy of Science*, 94, 8292–8299.
- Takahashi, T., R. T. Williams, and D. L. Bos (1982), *GEOSECS Pacific Expedition, Volume 3, Hydrographic Data 1973-1974*, chap. Carbonate chemistry, pp. 77–83, National Science Foundation, Washington, D.C.
- Tangang, F. T., W. W. Hsieh, and B. Tang (1997), Forecasting the equatorial Pacific sea surface temperatures by neural network models, *Climate Dynamics*, 13, 135–147.
- Telszewski, M., A. Chazottes, U. Schuster, A. J. Watson, C. Moulin, D. C. E. Bakker, M. Gonzalez-Dávila, T. Johannessen, A. Körtzinger, H. Lüger, A. Olsen, A. Omar, X. A. Padin, A. F. Rios, T. Steinhoff, M. Santana-Casiano, D. W. R. Wallace, and R. Wanninkhof (2009), Estimating the monthly pCO₂ distribution in the North Atlantic using a self-organizing neural network, *Biogeosciences*, 6, 1405–1421.
- Thomas, H., F. A. E. Prowe, I. D. Lima, S. C. Doney, R. Wanninkhof, R. J. Greatbatch, U. Schuster, and A. Corbière (2008), Changes in the North Atlantic Oscillation influence CO₂ uptake in the North Atlantic over the past 2 decades, *Global Biogeochemical Cycles*, 22, GB4027.
- Thorntwaite, C. W. (1948), An Approach toward a Rational Classification of Climate, *Geographical Review*, 38, 55–94.
- Trenberth, K., P. Jones, P. Ambenje, R. Bojariu, D. Easterling, A. Klein Tank, D. Parker, F. Rahimzadeh, J. Renwick, M. Rusticucci, B. Soden, and P. Zhai (2007), *Climate Change 2007: The Physical Science Basis. Contribution of Working Group I to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change*, chap. Observations: Surface and Atmospheric Climate Change, Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA.

- Tsai, C.-T., C. Lin, and J.-N. Shen (2002), Neural network for wave forecasting among multi-stations, *Ocean Engineering*, 29, 1683–1695.
- Tyndall, J. (1861), On the absorption and radiation of heat by gases and vapours, and on the physical connection, *Philos. Mag.*, 22, 277–302.
- Ullman, D. J., G. A. McKinley, V. Bennington, and S. Dutkiewicz (2009), Trends in the North Atlantic carbon sink: 1992-2006, *Global Biochemical Cycles*, 23, GB4011.
- Vesanto, J., J. Himberg, E. Alhoniemi, and J. Parhankangas (2000), *SOM Toolbox for Matlab 5*, Tech. rep., Helsinki University of Technology.
- Wagstaff, K., C. Cardie, S. Rogers, and S. Schrödl (2001), *Constrained k-means clustering with background knowledge*, *Proceedings of the Eighteenth International Conference on Machine Learning*, Tech. rep.
- Wakita, M., S. Watanabe, A. Murata, and M. Honda (2010), *Hydrographic and CO₂ Data Report at Station K2 during the 1999-2008 cruises*, Tech. rep., Carbon Dioxide Information Analysis Center, Oak Ridge National Laboratory, US Department of Energy, Oak Ridge, Tennessee.
- Wanninkhof, R. (1992), Relation between wind speed and gas exchange over the ocean, *Journal of Geophysical Research*, 97, 7373–7383.
- Wanninkhof, R., D. C. E. Bakker, N. Bates, A. Olsen, T. Steinhoff, and A. J. Sutton (2013a), *Incorporation of Alternative Sensors in the SOCAT Database and Adjustments to Dataset Quality Control Flags.*, Tech. rep., Carbon Dioxide Information Analysis Center, Oak Ridge National Laboratory, US Department of Energy, Oak Ridge, Tennessee.
- Wanninkhof, R., S. Doney, J. L. Bullister, N. M. Levine, M. Warner, and N. Gruber (2010), Detecting anthropogenic CO₂ changes in the interior Atlantic Ocean between 1989 and 2005, *Journal of Geophysical Research*, 115, C11 028.
- Wanninkhof, R., and W. McGillis (1999), A cubic relationship between air-sea CO₂ exchange and wind speed, *Geophysical Research Letters*, 26(13), 1889–1892.
- Wanninkhof, R., G. H. Park, T. Takahashi, C. Sweeney, R. Feely, Y. Nojiri, N. Gruber, S. C. Doney, G. A. McKinley, A. Lenton, C. Le Quéré, C. Heinze, J. Schwinger, H. Graven, and S. Khatiwala (2013b), Global ocean carbon uptake: magnitude, variability and trends, *Biogeosciences*, 10, 1983–2000.
- Watson, A. J., U. Schuster, D. C. E. Bakker, N. R. Bates, A. Corbière, M. González-Dávila, T. Friedrich, J. Hauck, C. Heinze, T. Johannessen, A. Körtzinger, N. Metzl, J. Olafsson, A. Olsen, A. Oschlies, X. A. Padin, B. Pfeil, J. M. Santana-Casiano, T. Steinhoff, M. Telszewski, A. F. Rios, D. W. R. Wallace, and R. Wanninkhof (2009), Tracking the variable North Atlantic sink for atmospheric CO₂, *Science (New York, N.Y.)*, 326(5958), 1391–3.
- Wattenberg, H. (1933), *Wissenschaftliche Ergebnisse der Deutschen Atlantischen Expedition, 1925-1927*, chap. Kalziumkarbonat und Kohlensäuregehalt des Meerwassers, pp. 233–331, vol. 8 edn., Walter de Gruyter, Berlin.
- Wegener, A. (1966), *The origin of continents and oceans.*, Courier Dover.

- Weiss, R. F. (1974), Carbon Dioxide in Water and Seawater: The Solubility of a Non-Ideal Gas, *Marine Chemistry*, 2, 203–215.
- Wolter, K. (1987), The Southern Oscillation in surface circulation and climate over the tropical Atlantic, Eastern Pacific, and Indian Oceans as captured by cluster analysis, *J. Climate Appl. Meteor.*, 26, 540–558.
- Wolter, K., and M. S. Timlin (2011), El Niño/Southern Oscillation behaviour since 1871 as diagnosed in an extended multivariate ENSO index (MEI.ext)., *Intl. J. Climatology*, 31, 14pp.
- Woolf, D. K. (1997), *The Sea Surface and Global Change*, chap. Bubbles and their role in gas exchange, pp. 173–205, Cambridge University Press.