

A novel empirical approach to diagnose patterns of air-sea carbon dioxide fluxes and ocean acidification

**Author:** Sasse, Tristan

Publication Date: 2013

DOI: https://doi.org/10.26190/unsworks/2482

#### License:

https://creativecommons.org/licenses/by-nc-nd/3.0/au/ Link to license to see what you are allowed to do with this resource.

Downloaded from http://hdl.handle.net/1959.4/52867 in https:// unsworks.unsw.edu.au on 2024-04-27

### A Novel Empirical Approach to Diagnose Patterns of Air-Sea CO<sub>2</sub> Fluxes and Ocean Acidification

Tristan P. Sassé

Climate Change Research Centre, Faculty of Science, The University of New South Wales, Australia

PhD Thesis

June 2013

Supervisors: Ben McNeil, Gab Abramowitz, Matthew England

A Thesis submitted in fulfillment of the requirements for the degree of Doctor of Philosophy

#### THE UNIVERSITY OF NEW SOUTH WALES Thesis/Dissertation Sheet

Surname or Family name: Sassé

First name: Tristan

Abbreviation for degree as given in the University calendar: PhD

School: Climate Change Research Centre

Title: A novel empirical approach to diagnose patterns of air-sea CO<sub>2</sub> fluxes and ocean acidification

Understanding the oceans role in mitigating atmospheric CO2 and climate requires a good constraint on spatiotemporal variability in the ocean carbon system. However, large spatiotemporal data limitations hamper our ability to quantify and understand patterns of ocean carbon dynamics. Here, I have developed a novel empirical approach to predict inorganic CO2 concentrations (total inorganic carbon ( $C_T$ ), total alkalinity ( $A_T$ ) and partial pressure of CO<sub>2</sub> (pCO<sub>2</sub>)) in the global ocean mixed-layer using standard hydrographic parameters (SHP; temperature, salinity, dissolved oxygen and nutrients) in order to provide independent constraints and insights on our understanding of ocean carbon dynamics, air-sea gas exchange and ocean acidification.

The novel technique, called SOMLO (Self-Organizing Multiple-Linear Output), couples a neural-network clustering algorithm with a multiple-linear regression to derive empirical relationships using bottle-data. Deploying and testing the SOMLO approach on a newly synthesized global bottle-dataset showed significant improvements over traditional linear approaches; improving global predictive skill by 19% for  $C_{\rm T}$ , with a global capacity to predict  $C_{\rm T}$  to within ±10.9 µmol kg<sup>-1</sup> (±9.2 µmol kg<sup>-1</sup> for  $A_{\rm T}$  and ±22.5 µatm for pCO<sub>2</sub>). In particular, the new non-linear method improved predictive skill in the most complex and dynamically important regions of the ocean (equatorial Pacific and Southern Ocean) by up to 30%.

The SOMLO approach was then applied to monthly SHP climatologies (WOA09) in order to diagnose monthly ocean surface  $C_{T}$ ,  $A_{T}$ and pCO2 patterns for the nominal year of 2000. Based on this analysis, patterns of air-sea CO2 flux were diagnosed and found to be broadly consistent with the global underway  $pCO_2$  database, suggesting a contemporary oceanic CO<sub>2</sub> uptake of 1.10±0.25 PgC yr<sup>-</sup> for the year of 2000. However, significant differences were found in 30% of the ocean, particularly in the equatorial Pacific and Southern Oceans.

For ocean acidification, seasonality in CO2 was found to bring forward the onset of aragonite under-saturation by about 19 years on average, exposing an additional  $24 \times 10^6$  km<sup>2</sup> to at least month-long corrosive aragonite conditions by 2100. Strong seasonal variability was also found to obscure the detection of anthropogenic ocean acidification by ~45 years in subtropical regions and ~60 years in higher latitudes.

Declaration relating to disposition of project thesis/dissertation

I hereby grant to the University of New South Wales or its agents the right to archive and to make available my thesis or dissertation in whole or in part in the University libraries in all forms of media, now or here after known, subject to the provisions of the Copyright Act 1968. I retain all property rights, such as patent rights. I also retain the right to use in future works (such as articles or books) all or part of this thesis or dissertation.

I also authorise University Microfilms to use the 350 word abstract of my thesis in Dissertation Abstracts International (this is applicable to doctoral theses only).

.....

Tristan Sasse Signature

13/09/2013

Witness

Date

The University recognises that there may be exceptional circumstances requiring restrictions on copying or conditions on use. Requests for restriction for a period of up to 2 years must be made in writing. Requests for a longer period of restriction may be considered in exceptional circumstances and require the approval of the Dean of Graduate Research.

FOR OFFICE USE ONLY

Date of completion of requirements for Award:

Other name/s: Peter

Faculty: Science

### **Originality Statement**

'I hereby declare that this submission is my own work and to the best of my knowledge it contains no materials previously published or written by another person, or substantial proportions of material which have been accepted for the award of any other degree or diploma at UNSW or any other educational institution, except where due acknowledgement is made in the thesis. Any contribution made to the research by others, with whom I have worked at UNSW or elsewhere, is explicitly acknowledged in the thesis. I also declare that the intellectual content of this thesis is the product of my own work, except to the extent that assistance from others in the project's design and conception or in style, presentation and linguistic expression is acknowledged.'

Signed Tristan Sasse

Date 13/09/2013

### **Copyright Statement**

'I hereby grant the University of New South Wales or its agents the right to archive and to make available my thesis or dissertation in whole or part in the University libraries in all forms of media, now or here after known, subject to the provisions of the Copyright Act 1968. I retain all proprietary rights, such as patent rights. I also retain the right to use in future works (such as articles or books) all or part of this thesis or dissertation.

I also authorise University Microfilms to use the 350 word abstract of my thesis in Dissertation Abstract International (this is applicable to doctoral theses only).

I have either used no substantial portions of copyright material in my thesis or I have obtained permission to use copyright material; where permission has not been granted I have applied/will apply for a partial restriction of the digital copy of my thesis or dissertation.'

Signed Tristan Sasse

Date 13/09/2013

### **Authenticity Statement**

'I certify that the Library deposit digital copy is a direct equivalent of the final officially approved version of my thesis. No emendation of content has occurred and if there are any minor variations in formatting, they are the result of the conversion to digital format.'

Signed Tristan Sasse Date 13/09/2013

### **Acknowledgments**

I wish to acknowledge all the support I have received over the past 4 years which has allowed me to complete this PhD thesis. My first thanks goes to my supervisor Ben McNeil, his direction, insight and encouragement is the reason I have reached my goals – thank you for believing in me even when I doubted myself. A further special note to my co-supervisor Gab Abramowitz, who provided technical support and an unwavering willingness to assist with any statistical analysis, I thoroughly enjoyed discussing the more technical issues with him. I have also greatly appreciated the encouragement, input and friendship of Emily Shaw, and technical assistance from David Fuchs. I am also thankful to the many other students and researchers at the Climate Change Research Centre for their kind encouragement. For financial assistance, I thank the University of New South Wales.

Finally, I am extremely grateful to my friends and family for their unconditional encouragement over the past 4 years, especially my parents Ruth and Peter Sassé and partner Ellie Meynink, thank you for your loving support throughout all my endeavours and for maintaining my sanity.

### Abstract

Understanding the oceans role in mitigating atmospheric  $CO_2$  and climate requires a good constraint on spatiotemporal variability in the ocean carbon system. However, large spatiotemporal data limitations hamper our ability to quantify and understand patterns of ocean carbon dynamics. Here, I have developed a novel empirical approach to predict inorganic  $CO_2$  concentrations (total inorganic carbon ( $C_T$ ), total alkalinity ( $A_T$ ) and partial pressure of  $CO_2$  ( $pCO_2$ )) in the global ocean mixed-layer using standard hydrographic parameters (SHP; temperature, salinity, dissolved oxygen and nutrients) in order to provide independent constraints and insights on our understanding of ocean carbon dynamics, air-sea gas exchange and ocean acidification.

The novel technique, called SOMLO (Self-Organizing Multiple-Linear Output), couples a neural-network clustering algorithm with a multiple-linear regression to derive empirical relationships using bottle-data. Deploying and testing the SOMLO approach on a newly synthesized global bottle-dataset showed significant improvements over traditional linear approaches; improving global predictive skill by 19% for  $C_{\rm T}$ , with a global capacity to predict  $C_{\rm T}$  to within ±10.9 µmol kg<sup>-1</sup> (±9.2 µmol kg<sup>-1</sup> for  $A_{\rm T}$  and ±22.5 µatm for pCO<sub>2</sub>). In particular, the new non-linear method improved predictive skill in the most complex and dynamically important regions of the ocean (equatorial Pacific and Southern Ocean) by up to 30%.

The SOMLO approach was then applied to monthly SHP climatologies (WOA09) in order to diagnose monthly ocean surface  $C_{\rm T}$ ,  $A_{\rm T}$  and  $p{\rm CO}_2$  patterns for the nominal year of 2000. Based on this analysis, patterns of air-sea CO<sub>2</sub> flux were diagnosed and found to be broadly consistent with the global underway  $p{\rm CO}_2$  database, suggesting a contemporary oceanic CO<sub>2</sub> uptake of 1.10±0.25 PgC yr<sup>-1</sup> for the year of 2000. However, significant differences were found in 30% of the ocean, particularly in the equatorial Pacific and Southern Ocean.

For ocean acidification, seasonality in  $CO_2$  was found to bring forward the onset of aragonite under-saturation by about 19 years on average, exposing an additional  $24 \times 10^6$  km<sup>2</sup> to at least month-long corrosive aragonite conditions by 2100. Strong seasonal variability was also found to obscure the detection of anthropogenic ocean acidification by ~45 years in subtropical regions and ~60 years in higher latitudes.

# Contents

Orig	ginality	Statement	i
Сор	yright	Statement	ii
Autl	henticit	y Statement	ii
Ack	nowled	gments	iii
Abs	tract		iv
List	of Abb	reviations	xii
List	of Tab	les	xiii
List	of Figu	ires	xv
Cha	pter 1.	Introduction	1
<b>Cha</b> 1.1	<b>pter 1</b> . Motivat	Introduction	<b>1</b> 2
<b>Cha</b> 1.1 1.2	<b>pter 1</b> . Motivat Ocean c	Introduction ion arbon dynamics	<b>1</b> 2 4
Cha 1.1 1.2	pter 1. Motivat Ocean c 1.2.1	Introduction ionarbon dynamics Drivers of variability	<b>1</b> 2 4 5
Cha 1.1 1.2 1.3	pter 1. Motivat Ocean c 1.2.1 Global u	Introduction ion arbon dynamics Drivers of variability uptake of anthropogenic CO <sub>2</sub>	<b>1</b> 2 4 5 7
Cha 1.1 1.2 1.3	pter 1. Motivat Ocean c 1.2.1 Global u 1.3.1	Introduction ion	<b>1</b> 2 4 5 7 9
Cha 1.1 1.2 1.3	<b>pter 1.</b> Motivat Ocean c 1.2.1 Global u 1.3.1 1.3.2	Introduction ion	<b>1</b> 2 4 5 7 9 10
Cha 1.1 1.2 1.3	pter 1. Motivat Ocean c 1.2.1 Global c 1.3.1 1.3.2 Implicat	Introduction ion	<b>1</b> 2 4 5 7 9 10 11
Cha 1.1 1.2 1.3 1.4 1.5	pter 1. Motivat Ocean c 1.2.1 Global u 1.3.1 1.3.2 Implicat Spatiote	Introduction ion	<b>1</b> 2 4 5 7 9 10 11 12
Cha 1.1 1.2 1.3 1.4 1.5	pter 1. Motivat Ocean c 1.2.1 Global c 1.3.1 1.3.2 Implicat Spatiote 1.5.1	Introduction ion	<b>1</b> 2 4 5 7 9 10 11 12 14

1.6	Empiri	cal approaches to investigate ocean carbon dynamics	17
1.7	Researc	ch goals and objectives	
Cha	apter 2.	A novel empirical approach to diagnose seasonal ocean surface carbon from bottle data: SOMLO	21
Abs	tract		
2.1	Introdu	iction	23
2.2	Global	bottle measurements and training dataset	
	2.2.1	Defining mixed-layer depths	30
	2.2.2	Identifying coastal samples	
	2.2.3	Biases in the global dataset	
2.3	Norma	lization of <i>C</i> <sub>T</sub> measurements	
	2.3.1	Revelle factor approach	
	2.3.2	Uncertainties in the correction method	
2.4	Testing	g algorithm skill: a Systematic Independent Test (SIT)	
2.5	Traditi	onal MLR approach	40
	2.5.1	Method description	40
	2.5.2	Robust MLR routine	41
	2.5.3	Ad-hoc versus universal MLR	
	2.5.4	Optimal MLR parameter combination	44
	2.5.5	MLR results	46
2.6	Neural	network approach	48
	2.6.1	Overview of neural network approach	48
	2.6.2	Initialization of model constraints	50
	2.6.3	SOM training routine	50
	2.6.4	Supervised SOM training	51

	2.6.5	Completing the input-output mapping	52
	2.6.6	Predicting with the SOMLO/SOMM models	53
2.7	Applica	tion to the global ocean	53
	2.7.1	Defining optimal predictor parameters	53
	2.7.2	Importance of geography	57
	2.7.3	SOMM versus SOMLO: Importance of the local linear optimizer	61
	2.7.4	Using different neural network algorithms	62
2.8	Evaluat	ing the empirical approaches	63
	2.8.1	Measuring the improvement over traditional MLR	63
	2.8.2	SOMLO regional error assessment	64
	2.8.3	Is there a temporal bias in SOMLO predictions?	69
	2.8.4	Significance of anthropogenic C <sub>T</sub> corrections	70
	2.8.5	How effective is a bathymetric approach for identifying coastal data?	72
	2.8.6	Are the neurons capturing the system?	72
	2.8.7	SOMLO model without Arctic data	73
	2.8.8	Stochastic nature of the SOM	75
2.9	Applica stations	tion to the Bermuda Atlantic and Hawaiian Ocean time-series	76
	2.9.1	Predicting the North Atlantic seasonal cycle for inorganic carbon (BATS)	76
	2.9.2	How well does SOMLO capture inter-annual signals?	79
2.10	Compar	ison to previous techniques	81
2.11	Diagnos	sing global $C_{\rm T}$ and $A_{\rm T}$ distributions	82
	2.11.1	WOA09 unit conversions	83
	2.11.2	Spatial variability	84

	2.11.3	Seasonal variability	88
2.12	Conclus	sion	92
Cha	apter 3.	Implications of natural $CO_2$ variability for ocean acidification	95
Abs	tract		96
3.1	Introdu	ction	97
	3.1.1	Onset of aragonite under-saturation	98
	3.1.2	Detecting Ocean acidification trends	99
3.2	Diagno	sing global pH and $\Omega_{Ar}$ distributions	100
	3.2.1	Spatiotemporal variability	101
	3.2.2	Error analysis	106
3.3	Projecti	ing past and future carbonate distributions	108
	3.3.1	Influence of air-sea $p$ CO <sub>2</sub> disequilibrium	109
3.4	Estimat	ing the onset of aragonite under-saturation	112
3.5	Detection	on-times for ocean acidification	118
	3.5.1	Application to the global ocean	120
3.6	Conclus	sion	122
Cha	apter 4.	A new constraint on global air-sea CO <sub>2</sub> fluxes using bottle carbon data	125
Abs	tract		126
4.1	Introdu	ction	126
4.2	Global	training dataset	127
4.3	Neural	network overview	129
4.4	Applica	ation to the global dataset	130
4.5	Testing	the new approach	131

	4.5.1	Probing the model for seasonal bias	. 132
	4.5.2	Influence of anthropogenic correction	. 134
	4.5.3	How well does SOMLO capture temporal variability at BATS?	. 135
4.6	Diagnos	sing monthly <i>p</i> CO <sub>2</sub> climatologies	. 136
	4.6.1	Evaluating SOMLO-predictions in the subtropical North Atlantic	. 139
	4.6.2	Comparison to T-09 ocean surface $pCO_2$ climatology	. 141
	4.6.3	Why not use the $C_{\rm T}$ and $A_{\rm T}$ climatologies to compute $p{\rm CO}_2$ ?	. 143
4.7	Comput	ing air-sea CO <sub>2</sub> exchange	. 144
4.8	Air-sea	CO <sub>2</sub> flux pattern	. 146
	4.8.1	Zonal flux distribution	. 150
4.9	Oceanic	CO <sub>2</sub> uptake estimates and uncertainty	. 150
	4.9.1	Net-contemporary air-sea CO <sub>2</sub> flux	. 152
	4.9.2	Coastal air-sea CO <sub>2</sub> exchange	. 152
4.10	Conclus	ion	. 153

### Chapter 5. Closing remarks

5.1	Summa	ry of findings	. 157
5.2	Signific	cance and way forward	. 160
	5.2.1	Understanding ocean carbon dynamics	. 160
	5.2.2	Constraining biogeochemical provinces	. 161
	5.2.3	Influence of natural seasonal variability	. 161
	5.2.4	A new constraint on air-sea CO <sub>2</sub> flux	. 162
5.3	Acknow	vledgements	. 163

### References

157

Appendix A. T	he ocean carbon system	187
A1.1 Inorganic car	rbon chemistry	. 187
A1.2 Photosynthes	sis	189
A1.3 Calcification	1	190
References		. 190
Appendix B. S	summary of cruise data	192
References		202
Appendix C. N	Aultiple-Linear Regression equations	203
Appendix D. P	Principle Component Regression	205
Appendix E. G S	Geographic boundaries used to partition the global IT predictions	206
Appendix F. N	/lonthly pH and $\Omega_{Ar}$ distributions	207
Appendix G. T	able of regional air-sea CO <sub>2</sub> flux estimates	210
References		210

# List of Abbreviations

Abbreviation/Symbol	Meaning
A <sub>T</sub>	Total Alkalinity
AAIW	Antarctic intermediate cruise
AOU	Apparent oxygen utilization
BATS	Bermuda-Atlantic time-series
$C_{\mathrm{T}}$	Total dissolved inorganic carbon
CARINA	Carbon dioxide in the Atlantic Ocean
Chl-a	Chlorophyll- <i>a</i>
CLIVAR	Climate variability and predictability project
DO or O	Dissolved oxygen
ESTOC	European station for time-series in the ocean
	at the Canary Islands
GLODAP	Global ocean data analysis project
НОТ	Hawaiian ocean time-series
ICP	Ice-coverage percentage
JGOFS	Joint global ocean flux study
LOESS	locally weighted scatterplot smoothing
MCL	Multi-collinearity
MLD	Mixed-laver depth
MLR	Multiple-linear regressions
N	Nitrate
NACP	North Atlantic carbon project
OGCM	Ocean general circulation model
P	Phosphate
PACIFICA	Pacific ocean interior carbon
$p \cap \Omega_2$	Partial pressure of carbon dioxide
PCR	Principle component analysis
PFI	Prince Edward Islands
Ρσ	Peta-grams of carbon
nnm	Parts per million
R	Revelle factor
RSF	Residual standard error
S	Salinity
SHP	Standard hydrographic parameters
Si	Silicate
SIT	Systematic Independent Test
SOM	Self-organizing man
SOMI	Self_organizing multiple_linear output
SNR	Signal_to_noise ratio
S101C	Sea_surface salinity
SSS	Sea surface temperature
т	Temperature
ı VIF	Variance inflation factor
WOCE	World ocean circulation experiment
	A regonite seturation state
S2Ar	Aragointe saturation state

## List of Tables

Table 2.1:	Previous empirical approaches to predict upper ocean $A_{\rm T}$ and $C_{\rm T}$	
	concentrations	27
Table 2.2:	Data sources of the global merged dataset	29
Table 2.3:	Universal and ad-hoc MLR results for (a) $C_{\rm T}$ and (b) $A_{\rm T}$	47
Table 2.4:	Summary of the three independent datasets used to constrain the	
	general configuration of SOMLO	54
Table 2.5:	Skill comparison between the traditional MLR and SOMLO	
	approaches for (a) $C_{\rm T}$ and (b) $A_{\rm T}$	64
Table 2.6:	Regional and global SOMLO skill evaluation	67
Table 2.7:	RSE values of SIT predictions partitioned into 5 regions and by	
	season	70
Table 2.8:	SOMLO skill comparison between near-coast and open-ocean	
	regions	72
Table 2.9:	Global RSE values [µmol kg <sup>-1</sup> ] computed for model predictions	
	under the optimal configuration and with two increases in SOM	
	neuron size	73
Table 2.10:	RSE values of SIT predictions below 70°N, where the optimal	
	SOMLO configurations were trained and tested using a global	
	dataset either with or without Arctic data	75
Table 2.11:	Mean and first standard deviation in RSE values of the three	
	independent subsets (see Sect. 2.6.2) when tested 100 times each	75
Table 2.12:	Statistical summary of independent predictions at the BATS and	
	HOT stations.	79
Table 2.13:	Skill comparison between optimal SOMLO models and previous	
	MLR approaches.	82
Table 2.14:	Summary of units used in the WOA09 products and global bottle	
	dataset	83
Table 3.1:	Summary of observed air-sea $pCO_2$ growth rates	111
Table 4.1:	RSE values of SIT predictions partitioned into 5 regions and by	
	season	134

Table 4.2:	Global skill comparison between optimal SOMLO models trained	
	and tested using data either corrected, or not corrected for	
	anthropogenic CO <sub>2</sub> uptake.	134
Table 4.3:	Comparison between global RSE values computed using SIT	
	predictions of the $pCO_2$ -SOMLO model and computed $pCO_2$ values	
	via the optimal $C_{\rm T}$ and $A_{\rm T}$ SOMLO SIT predictions	144
Table 4.4:	Global anthropogenic (anthro.) or net-contemporary air-sea $CO_2$	
	flux estimates of SOMLO and other independent methods	155
Table B1:	Cruise summary of global combined bottle dataset	192
Table C1.	Ad-hoc and universal A <sub>T</sub> regression co-efficients	203
Table C2.	Ad-hoc and universal C <sub>T</sub> regression co-efficients	204
Table G1:	Regional contemporary air-sea CO <sub>2</sub> flux estimates using SOMLO	
	predicted and Takahashi et al. (2009) ocean surface $pCO_2$	
	distributions	210

# List of Figures

Figure 1.1:	Time-evolving net oceanic CO <sub>2</sub> uptake estimates since 1960 for the	
	expected steady-state (black line) and the mean uptake of five	
	different ocean general circulation models (OGCM) forced with	
	changes in wind, heat and freshwater fluxes (Red dots)	9
Figure 1.2:	Illustration of the different air-sea CO2 flux signals for pre-	
	industrial (~1750) to present times.	13
Figure 2.1:	Global distribution of mean MLD for (a) summer months (b) winter	
	months and (c) seasonal-mean difference (winter minus summer)	31
Figure 2.2:	(a) Spatial distribution of the global training dataset, together with	
	(b) seasonal and (c) yearly histograms partitioned into Southern	
	(light shade) and Northern (dark shade) Hemispheres	33
Figure 2.3:	Anthropogenic correction applied to mixed-layer $C_{\rm T}$ measurements	36
Figure 2.4:	Global distribution of bottle-derived Revelle factors (R).	37
Figure 2.5:	Difference in annual-mean atmospheric CO2 relative to the year	
	2000 ( $\Delta CO_2$ ) observed at Mauna Loa (closed circles) and a global	
	estimate derived from multiple measurement stations (open	
	triangles)	38
Figure 2.6:	Schematic diagram of the robust forward MLR routine	42
Figure 2.7:	Spatiotemporal division of the global training dataset for the ad-hoc	
	MLR approach	44
Figure 2.8:	Global RSE values calculated using the SIT predictions of the	
	universal MLR model	45
Figure 2.9:	Global RSE values calculated using SIT predictions of the universal	
	C <sub>T</sub> MLR model	46
Figure 2.10	Schematic diagram of neural network training and prediction	
	phases	49
Figure 2.11	Distribution of measurements constituting the T1 independent	
	dataset	54
Figure 2.12	:RSE values for the $C_{\rm T}$ (open triangles) and $A_{\rm T}$ (black circles)	
	SOMLO models when applied to the three independent datasets	55

Figure 2.13	: Optimal residual standard error (RSE) values for (a) $C_{\rm T}$ and (b) $A_{\rm T}$	
	SOMLO models	56
Figure 2.14	Global map after longitude values were shifted by 160° West	58
Figure 2.15	Skill of optimal SOMLO models with geographical constraints	;9
Figure 2.16	Distribution of assigned neurons for optimal $C_{\rm T}$ SOM models	
	trained with (a) biogeochemical information only and (b)	
	biogeochemical and geographical information	50
Figure 2.17	Skill comparison between the SOMLO (open triangles) and SOMM	
	(closed triangles) models in capturing $C_{\rm T}$	51
Figure 2.18	Distribution of global SIT residual errors (i.e. predicted - in situ)	
	for optimal (a) $C_{\rm T}$ and (b) $A_{\rm T}$ SOMLO models	55
Figure 2.19	Geographical distribution of the 277 samples with a SIT residual	
	error greater than $\pm 50 \mu\text{mol } \text{kg}^{-1}$ for $C_{\text{T}}$ and/or $A_{\text{T}}$ and are located	
	within 300 km of a major coastline	56
Figure 2.20	(a) and (c) Global SIT predictions versus in situ measurements for	
	optimal $C_{\rm T}$ and $A_{\rm T}$ SOMLO models respectively. (b) and (d)	
	Residual error density distributions for optimal $C_{\rm T}$ and $A_{\rm T}$ models	
	respectively	58
Figure 2.21	:RSE values for optimal SOMLO $C_{\rm T}$ and $A_{\rm T}$ SIT predictions	
	partitioned by season	59
Figure 2.22	: Annual RSE difference between $C_{\rm T}$ models trained and tested using	
	measurements that were either correct or non-corrected for	
	anthropogenic CO <sub>2</sub> uptake	1
Figure 2.23	Distribution of measurements assigned to a neuron containing at	
	least one sample collected within the Arctic region (above 70°N)7	/4
Figure 2.24	BATS in situ and independently predicted seasonal cycles for (a)	
	$C_{\rm T}$ and (b) $A_{\rm T}$	17
Figure 2.25	Distribution of assigned neurons in the North-west Atlantic region	
	for optimal $C_{\rm T}$ SOMLO model (30°N to 70°N and 40°W to 85°W)	/8
Figure 2.26	In situ and independently predicted BATS $C_{\rm T}$ measurements	
	partitioned into years with a loess (locally weighted scatterplot	
	smoothing) line fit	30
	¥3/1	

Figure 2.27: In situ and independently predicted HOT $C_{\rm T}$ measurements with a	
loess (locally weighted scatterplot smoothing) line fit	1
Figure 2.28: Global distributions of (a) bottle $C_{\rm T}$ measurements corrected to the	
year 2000, (b) annual-mean ocean surface SOMLO $C_{\rm T}$ predictions	
for the nominal year of 2000, and (c) GLODAP-v1.1 0m $C_{\rm T}$	
distribution of Key et al. (2004). Values are given in $\mu$ mol kg <sup>-1</sup>	6
Figure 2.29: Global distributions of (a) bottle $A_{\rm T}$ measurements, (b) annual-	
mean ocean surface SOMLO $A_{\rm T}$ predictions, and (c) GLODAP-v1.1	
0m $A_{\rm T}$ distribution of Key et al. (2004). Values are given in µmol	
kg <sup>-1</sup>	7
<b>Figure 2.30:</b> Seasonal amplitudes for SOMLO predicted ( <b>a</b> ) $C_{\rm T}$ and ( <b>b</b> ) $A_{\rm T}$ ocean	
surface distributions	9
<b>Figure 2.31:</b> Monthly SOMLO-predicted ocean surface $C_{\rm T}$ distributions	0
<b>Figure 2.32:</b> Monthly SOMLO-predicted ocean surface $A_{\rm T}$ distributions	1
Figure 3.1: Global distributions of (a) bottle-derived pH measurements	
corrected to the year 2000 and (b) annual-mean SOMLO-derived	
pH predictions for the nominal year of 2000 1	02
<b>Figure 3.2:</b> Global distributions of (a) bottle-derived $\Omega_{Ar}$ measurements	
corrected to the year 2000 and (b) annual-mean SOMLO-derived	
$\Omega_{Ar}$ predictions for the nominal year of 2000 (see Appendix F for	
monthly distribution plots)1	03
Figure 3.3: Seasonal amplitudes for SOMLO-derived (a) pH and (b) $\Omega_{Ar}$	
distribution	05
Figure 3.4: SIT predictions versus bottle-derived measurements for (a) pH and	
(b) aragonite saturation-state ( $\Omega_{Ar}$ ). Panels (c) and (d) present the	
residual error density distributions for pH and $\Omega_{Ar}$ respectively	07
Figure 3.5: Atmospheric $CO_2$ projections between 1990 and 2100 as	
determined by the Representative Concentration Pathways (RCP)	
project (Meinshausen et al., 2011) 1	09
<b>Figure 3.6:</b> Future aragonite under-saturation ( $\Omega_{Ar} < 1$ ) onset at locations in the	
(a) north Atlantic and (b) Southern Ocean 1	13

Figure 3.7:	Estimated aragonite under-saturation onset years for (a) annual-	
	mean, (b) one-month, and (b) the temporal difference between the	
	annual-mean and one-month estimates.	115
Figure 3.8:	(a) Area exposed to at least monthly (blue) and annual-mean (red)	
	aragonite under-saturation conditions within the 21st century. The	
	latitudinal extension (blue region) represents ${\sim}24{\times}10^6~km^2.$ (b)	
	Evolution of ocean surface area exposed to under-saturation	
	conditions for at least monthly (blue line) and annual-mean (red	
	line). At centuries end, the difference between the two estimates is	
	$\sim 24 \times 10^6$ km <sup>2</sup> . The area labeled PEI represents the pteropod study	
	region of Hunt et al. (2008) around the Prince Edward Islands	117
Figure 3.9:	Example of the detection-time approach at a location in the north	
	Atlantic (40.5°N and 30.5°W). Panels (a) and (b) illustrate the	
	seasonal pH and $\Omega_{Ar}$ cycles used to constrain the magnitude of	
	natural noise (red dashed lines), panels (c) and (d) presents the	
	anthropogenic change relative to the year 1990 in 5 year	
	increments, and panels (e) and (f) are the calculated signal-to-noise	
	ratios (SNR, see Eq. 3.2)	119
Figure 3.10	:Global estimates for the first year robust anthropogenic trends can	
	be detected from a 1990 baseline for (a) pH and (b) aragonite	
	saturation state. The BATS and HOT time-series locations are	
	shown in panel (a).	121
Figure 4.1:	Global distribution of mixed-layer bottle-derived $pCO_2$	
	measurements	128
Figure 4.2:	(a) SIT $pCO_2$ predictions versus bottle-derived $pCO_2$ concentrations	
	and (b) residual error density distribution	132
Figure 4.3:	RSE values for $pCO_2$ SIT predictions partitioned by season	133
Figure 4.4:	In situ and independently predicted annual $pCO_2$ cycle at the BATS	
	hydrographic station between 1989 to 2006	135

- Figure 4.5: (a) Annual-mean SOMLO pCO<sub>2</sub> predictions for the nominal year of 2000 and (b) seasonal variability in pCO<sub>2</sub> calculated as the difference between the maximum and minimum predictions in each 1°×1° grid cell.

- Figure 4.9: Comparison between SOMLO-predicted and T-09 pCO<sub>2</sub> climatology values in (a) equatorial Pacific (18°S to 18°N) and (c) Southern Ocean (South of 44°S). Panels (b) and (d) present the residual error distributions for independent SOMLO predictions in the equatorial Pacific and Southern Ocean, respectively.

- **Figure 4.14:** Zonally integrated annual contemporary air-sea CO<sub>2</sub> fluxes using SOMLO-predicted (red line) and T-09 *p*CO<sub>2</sub> estimates (blue line)......150

Figure D1:	Principle Component Regression schematic.	. 205
Figure E1:	Geographical boundaries used to partition the global systematic	
	independent test (SIT) predictions.	. 206
Figure F1:	Monthly SOMLO-derived ocean surface $\Omega_{Ar}$ distributions for the	
	nominal year of 2000.	. 208
Figure F2:	Monthly SOMLO-derived ocean surface pH distributions for the	
	nominal year of 2000.	. 209

# Chapter 1. Introduction

### 1.1 Motivation

For ten thousand years prior to the industrial revolution (~1750 AD), cycling of carbon between the ocean, land and atmosphere maintained a relatively stable atmospheric carbon dioxide (CO<sub>2</sub>) concentration between 260 to 280 ppm (Etheridge et al., 1996; Petit et al., 1999). The ensuing release of ~530 Pg (1 Pg =  $10^{15}$  g) of carbon (PgC) through humanities consumption of geological carbon (~360 PgC; Boden et al., 2012) and land use changes (~170 PgC; Friedlingstein et al., 2010; Houghton, 2003), has resulted in the near exponential increase in atmospheric CO<sub>2</sub> to be 392.6 ppm in 2012 (Dlugokencky and Tans, 2012). As a potent greenhouse gas, the accumulation of  $CO_2$ alters the global radiative energy budget by increasing the atmospheres infrared opacity - commonly known as greenhouse warming (Forster et al., 2007). Already, humanderived  $CO_2$  has contributed ~60% to the positive shift in planetary radiative forcing since 1750 (Forster et al., 2007), and its relative contribution will continue to grow with rising emissions (Hofmann et al., 2006). Humanities demand for fossil-fuel derived energy (Andres et al., 2012) could result in atmospheric  $CO_2$  levels in excess of 730 ppm by centuries end if no emission reduction policies are implemented (Friedlingstein et al., 2006; Meinshausen et al., 2011). The concomitant 2-4°C rise in global-mean surface temperature will significantly change both regional and global climate patterns, including sea-level rise, precipitation trends and extreme weather events (Meehl et al., 2007); all of which present large consequences for society.

Quantifying the redistribution of anthropogenic  $CO_2$  in the global carbon system is crucial for understanding future climate change, since only the fraction that remains airborne contributes to greenhouse warming. Within the global carbon system, only two reservoirs exchange  $CO_2$  fast enough to influence atmospheric concentrations on seasonal and longer time-scales; the oceans and terrestrial biosphere. Comparison between annual atmospheric  $CO_2$  accumulation and anthropogenic  $CO_2$  emissions revealed that on average, 57% of man-made  $CO_2$  has been absorbed by the land and ocean carbon sinks each year since 1959 (Le Quéré et al., 2009). Despite this, there are large year-to-year variations in this airborne-fraction as a result of natural variability in oceanic and terrestrial  $CO_2$  sinks (Sarmiento et al., 2010).

Partitioning between the terrestrial and ocean sinks represents an important challenge for constraining future atmospheric  $CO_2$  levels, since carbon residence time differs distinctly between the two systems. In the terrestrial biosphere, carbon storage ranges from seconds to decades, whereas the ocean can sequester anthropogenic carbon for thousands of years (Archer et al., 2009). Due to extreme heterogeneity in both space and time within the terrestrial carbon system (Raupach, 2011), partitioning between these two sinks has largely relied on better oceanic  $CO_2$  uptake constraints.

Several independent studies estimate the global ocean sequesters ~30% (~2.0 PgC) of anthropogenic CO<sub>2</sub> emissions each year (e.g., Gruber et al., 2009; Manning and Keeling, 2006; Takahashi et al., 2009). Despite the importance of this CO<sub>2</sub> sink from a climate change perspective, when CO<sub>2</sub> dissolves in the ocean it acts as a weak acid lowering seawater pH. The uptake of ~155 Pg of anthropogenic carbon (Khatiwala et al., 2012) has already decreased the average ocean surface pH by ~0.1 units relative to preindustrial times (Feely et al., 2004). The immediate impact of ocean acidification is a decrease in the ability of multiple marine organisms to perform their physiological processes (Fabry et al., 2008; and references within). This includes the ability of calcifying organisms to secrete and preserve their calcium carbonate (CaCO<sub>3</sub>) shells and skeletons; a process that plays a crucial role in the global carbon cycle. Additional to alterations in ocean carbon chemistry, an increasing number of studies are identifying changes in the natural cycling of carbon induced by shifts in the global climate system (e.g., Le Quéré et al., 2007, 2010; Lovenduski et al., 2007; Matear and Lenton, 2008). The influence of these alterations is estimated by models to have

dampened oceanic uptake of anthropogenic  $CO_2$  by ~18% since 1989 (McNeil and Matear, 2012).

Understanding how the ocean will respond to a higher  $CO_2$  world requires accurate knowledge of natural variability in the ocean carbon system on seasonal to inter-annual time-scales. Despite significant efforts over recent decades to establish a global measurement network, large spatial and temporal data limitations hamper efforts to constrain higher-frequency variability and detect secular trends over the majority of the ocean. In order to advance our understanding of how natural variability in the ocean influences atmospheric  $CO_2$  levels, it's important to understand the drivers of ocean carbon variability.

### 1.2 Ocean carbon dynamics

The world's oceans act as a large reservoir of carbon that continuously exchanges  $CO_2$  with the atmosphere. Exchange at the air-sea interface is driven by the air-sea gradient in partial pressure of  $CO_2$  ( $pCO_2$ ), where lower sea-surface concentrations relative to the atmosphere tend to absorb atmospheric  $CO_2$ , while opposite conditions typically outgas  $CO_2$ . Each year, approximately 90 PgC is naturally exchanged at the air-sea interface in both directions (Sarmiento and Gruber, 2002). Despite a near zero net annual flux of natural  $CO_2$ , a recent synthesis of sea-surface  $pCO_2$  measurements suggests the pattern of  $CO_2$  exchange is not homogenous across the global ocean, but exhibits large spatial and temporal variability (Takahashi et al., 2009). For example, in the Southern Ocean, North Atlantic and North Pacific the ocean is estimated to act as a large annual-net sink for atmospheric  $CO_2$ , while the equatorial Pacific and far south of the Southern Ocean are suggested to be large annual-net sources of  $CO_2$  to the atmosphere (Takahashi et al., 2009).

#### 1.2.1 Drivers of variability

The capacity of the oceans to store  $CO_2$  is immense. On multi-centennial time-scales, the oceans will absorb up to 80% of anthropogenic  $CO_2$  (Archer et al., 1997). This is because  $CO_2$  does not simply dissolve in sea-water, but reacts with water molecules to form bicarbonate ( $HCO_3^-$ ) and carbonate ( $CO_3^{2-}$ ) ions (see *Appendix A* for more details). The sum of dissolved  $CO_2$ ,  $HCO_3^-$  and  $CO_3^{2-}$  is known as total dissolved inorganic carbon ( $C_T$ ) and is a conservative parameter with respect to changes in state. For a typical ocean surface pH of 8.2,  $C_T$  is found in the approximate ratio of 1:90:9 for dissolved  $CO_2$ ,  $HCO_3^-$  and  $CO_3^{2-}$ , respectively (Sarmiento and Gruber, 2006). Since oceanic  $pCO_2$  is only influenced by the concentration of dissolved  $CO_2$ , processes that influence the speciation of  $C_T$  provides a first insight into the drivers of oceanic  $pCO_2$ variability.

From the laws of thermodynamics, oceanic temperature and pressure are identified as parameters driving  $C_{\rm T}$  speciation due to their influence on the free energy in the system for chemical reactions; including the solubility of CO<sub>2</sub> in seawater. Several data-based studies have also identified salinity as a minor player in  $C_{\rm T}$  speciation and CO<sub>2</sub> solubility (e.g., Mehrbach et al., 1973; Weiss, 1974). These data-based studies show that cold and saline surface waters exhibit much lower pCO<sub>2</sub> concentrations than warmer and fresher waters. It follows that cold and saline waters typically contain more  $C_{\rm T}$  than the global surface mean due to the oceans enhanced capacity to absorb atmospheric CO<sub>2</sub>. Temperature and salinity also drive ocean circulation pathways through the formation of deep waters in regions where surface waters are denser than the underlying ocean. Consequently, physical-solubility processes act in unison to pump atmospheric CO<sub>2</sub> into the ocean surface decades to centuries later.

Another important parameter for understanding ocean carbon dynamics is total alkalinity ( $A_{\rm T}$ ), which is related to the charge balance of the ocean (see *Appendix A*). Alkalinity helps define the oceans buffering capacity by providing negative ions (HCO<sub>3</sub><sup>-</sup>, CO<sub>3</sub><sup>2-</sup> and others) to which can neutralize H<sup>+</sup> ions. The effect of increasing  $A_{\rm T}$  results in a concomitant decrease in oceanic *p*CO<sub>2</sub>. For a full definition of  $A_{\rm T}$  see Zeebe and Wolf-Gladrow (2001) or Sarmiento and Gruber (2006).

Biological processes reduce ocean surface  $pCO_2$  through the fixation of dissolved CO<sub>2</sub> into organic material by photoplanktonic organisms (see *Appendix A*). Consumption of nutrients and hydrogen ions during photosynthesis slightly increases  $A_T$ , resulting in a further, albeit small, reduction in  $pCO_2$ . Approximately 25% of synthesized organic material is then transported into the ocean interior via settling and advection processes, where it is largely remineralised back into  $C_T$  and nutrients by heterotrophic organisms (Falkowski et al., 1998). Oceanic CO<sub>2</sub> uptake is therefore dependent on the strength of this pump, which is sensitive to physical changes (e.g., temperature and circulation) and the availability of nutrients and light (Langdon and Atkinson, 2005).

In addition to the synthesis of organic material, some planktonic organisms (e.g., coccolothophorids) and zooplankton (e.g., pteropods) consume carbonate ions to form calcium carbonate (CaCO<sub>3</sub>) shells or skeletons. The net result of CaCO<sub>3</sub> precipitation is a reduction in  $A_T$  and  $C_T$  in the ratio of 2:1 respectively (see *Appendix A*). Following from the above discussion, biogenic CaCO<sub>3</sub> formation therefore increases oceanic  $pCO_2$ , while dissolution has the opposite effect. The strength of this process is dependent on ocean pH, with more acidic conditions typically restricting CaCO<sub>3</sub> production (Riebesell et al., 2000).

Global carbon export from the ocean surface to the deep sea through organic material and  $CaCO_3$  is estimated to be between 11-16 PgC yr<sup>-1</sup> (Boyd and Trull, 2007; Falkowski et al., 2000) and 0.4-1.8 PgC yr<sup>-1</sup> (Berelson et al., 2007) respectively. This

indicates that organic matter fixation of  $CO_2$  accounts for approximately 90% of the net biological influence on ocean surface inorganic carbon.

Biological and physical-solubility processes interact on different spatial and temporal time-scales to drive the observed distribution of inorganic carbon species and air-sea  $CO_2$  flux. Model simulations suggest the net effect of these processes regulates atmospheric  $CO_2$  by ~250 ppm on longer time-scales (Cameron et al., 2005).

### 1.3 Global uptake of anthropogenic CO<sub>2</sub>

The recent anthropogenic increase in atmospheric  $CO_2$  is inducing oceanic uptake via air-sea  $pCO_2$  disequilibrium. The characteristic time-scale for the ocean's top 40 meters to establish air-sea equilibrium is about 6 months (Sarmiento and Gruber, 2006), indicating that on decadal time-scales, the rate-limiting step for anthropogenic  $CO_2$ uptake is transport into the ocean interior (Sarmiento et al., 1992). Since biological processes are not limited by  $CO_2$  availability in the upper sunlit ocean, export production is assumed to have changed little since preindustrial times. This implies that physical-solubility processes are the dominate removal pathway for anthropogenic  $C_T$ into the ocean interior. In particular, ocean regions that exhibit strong upwelling, such as the eastern equatorial Pacific and along upwelling margins in the Southern Ocean, will absorb large amounts of anthropogenic  $CO_2$  since interior waters were last exposed to the atmosphere when ambient atmospheric  $CO_2$  concentrations were much lower.

Several independent methods have been proposed to quantify the oceans anthropogenic  $C_{\rm T}$  inventory using in situ hydrographic measurements. Methods based on the "back calculation" approach estimates the small anthropogenic  $C_{\rm T}$  signal by correcting for the large biological and physical-solubility influences on measured values of  $C_{\rm T}$  (e.g., the  $\Delta C^*$  method of Gruber et al. (1996)). Other approaches use tracers such as chlorofluorocarbons (CFCs) to constrain the evolution of a water parcel, which are then used to diagnose anthropogenic  $C_{\rm T}$  distributions (e.g., Hall et al., 2002; Khatiwala et al., 2009; McNeil et al., 2003), while changes in the oceanic  ${}^{13}$ C-to- ${}^{12}$ C ratio (Suess effect) can be used to quantify anthropogenic CO<sub>2</sub> uptake (e.g., McNeil et al., 2001a; Olsen et al., 2006; Quay et al., 1992). These independent approaches estimate a global oceanic  $C_{\rm T}$  increase to be between 122 and 157 PgC, or ~30% of net anthropogenic CO<sub>2</sub> emissions (energy + land-use) over the period 1750 to 2010 (Khatiwala et al., 2012).

Underlying all these data-based estimates are assumptions that biological and physical-solubility pathways for  $C_{\rm T}$  transport into the ocean interior have remained in steady-state throughout the anthropocene. Although this premise was likely adequate for most of the 20<sup>th</sup> century, recent studies have identified shifts in the oceans natural cycling of carbon due to climate and chemistry related alterations. For example, decadal-scale trends in ocean surface temperature and salinity via heat uptake (Levitus et al., 2005; Lyman et al., 2010) and intensification of the global hydrological cycle (Durack and Wijffels, 2010), are influencing both the solubility of CO<sub>2</sub> and ocean circulation pathways. At the same time, ocean warming leads to increased stratification, which dampens the upwelling of nutrient rich subsurface waters crucial for phytoplanktonic growth (Behrenfeld et al., 2006; Boyce et al., 2010). Shifting wind patterns is also influencing circulation and seasonal mixing processes, resulting in either enhanced or diminished upwelling of waters rich in  $C_{\rm T}$  and nutrients (Le Quéré et al., 2007; Lovenduski et al., 2007). Chemistry-related alterations arise from the uptake of ~155 Pg of anthropogenic carbon over the last 250 years (Khatiwala et al., 2012). The resulting decrease in ocean surface pH (global mean of 0.1; Feely et al., 2004) restricts the ability of calcifying organisms to form and preserve their CaCO<sub>3</sub> shells and skeletons (Fabry et al., 2008; and references within). Elevated CO<sub>2</sub> levels in the ocean may also alter nutrient uptake stoichiometric ratios (Boyd and Doney, 2002).

To estimate the extent of alterations in the oceans steady-state carbon cycle for anthropogenic  $CO_2$  uptake, Sarmiento et al. (2010) compared a new suite of coupled ocean-atmosphere models forced with observed trends in wind, heat and freshwater

fluxes, to a data-based steady-state model that assumes constant circulation and biology (Fig. 1.1). Since the late 1980's, the time-varying ocean models absorbed on average 0.35 PgC yr<sup>-1</sup> (or 31%) less CO<sub>2</sub> then expected if increasing atmospheric CO<sub>2</sub> was the only function influencing uptake. This model-based result affirms the emergence of a non-steady-state signal that may have already dampened anthropogenic CO<sub>2</sub> uptake by 18% since 1989 (McNeil and Matear, 2012).



**Figure 1.1:** Time-evolving net oceanic  $CO_2$  uptake estimates since 1960 for the expected steady-state (black line) and the mean uptake of five different ocean general circulation models (OGCM) forced with changes in wind, heat and freshwater fluxes (Red dots). The expected steady-state  $CO_2$  uptake is derived using the Mikaloff-Fletcher et al. (2006) estimate for the nominal year of 1995 and projected between 1960 and 2010 by Sarmiento et al. (2010). The average net  $CO_2$  uptake (red dots) was taken as a combined mean between the five different OGCMs with minimum and maximum bounds shaded in light red (Le Quéré et al., 2007; Lovenduski et al., 2007; Matear and Lenton, 2008; Rodgers et al., 2008; Wetzel et al., 2005). This figure was taken from McNeil and Matear (2012).

#### 1.3.1 Detecting the non-steady-state signal

If the natural cycling of carbon remained in steady-state throughout the last two centuries, the rate of increase in ocean surface  $pCO_2$  would have roughly followed the atmosphere. This provides a basis for detecting the effect of climate- and chemistry-related alterations on air-sea CO<sub>2</sub> exchange. In regions where the observed growth-rate

in ocean surface  $pCO_2$  is slower than atmospheric CO<sub>2</sub>, the oceanic sink can be interpreted as increasing, in contrast, faster oceanic  $pCO_2$  growth rates indicates a decreasing sink for atmospheric CO<sub>2</sub>.

Development of autonomous  $pCO_2$  measuring devices have provided more than 6 million "underway" ocean surface  $pCO_2$  measurements since 1972 (Takahashi et al., 2012). This accumulated dataset has allowed researchers to detect secular growth trends in ocean regions where temporal data coverage is sufficient to account for natural variability (~27% of the global open-ocean (Takahashi et al., 2009)). From a suite of regional studies, a global pattern is emerging that suggests high-latitude CO<sub>2</sub> sinks are decreasing in strength, while low to mid-latitude regions appear to be tracking as expected (Lenton et al., 2012).

#### 1.3.2 Diagnosing the signal

While the "underway"  $pCO_2$  network has provided great insight into the evolving airsea CO<sub>2</sub> flux strength, it cannot be used by itself to constrain a mechanistic understanding of the changing ocean carbon system. To diagnose the non-steady-state signal requires coinciding measurements of  $C_T$ , pH, total alkalinity ( $A_T$ ) and other carbon/climate relevant tracers (temperature, salinity, dissolved oxygen, nutrients, wind speed etc.) in order to partition the effects of biology and mixing. From an observational perspective, only a few time-series stations exist with sufficient temporal data coverage to diagnose the specific drivers of ocean carbon change (e.g. Bermuda Atlantic (BATS; Bates, 2007) and Hawaiian Ocean (HOT; Keeling et al., 2004)). However, due to spatial correlation length scales ranging from tens to hundreds of kilometers for inorganic carbon (Lenton et al., 2006; Li et al., 2005; Sweeney et al., 2002), any observed change can only be assumed over the local ocean region.

It is therefore important to gain a better understanding of natural variability across the global ocean carbon system from which we can contextualize any observed change. Over the past four decades, significant time and resources have been devoted to

establishing a global carbon measurement network. Oceanographic measurement programs like the Geochemical Ocean Sections (GEOSECS) in the 1970s, and World Ocean Circulation Experiment (WOCE)/Joint Global Ocean Flux Study (JGOFS) in the 1990s, have collected thousands of in situ bottle measurements of carbon and carbon relevant parameters (Key et al., 2004, 2010). Currently, the US and International Climate Variability and Predictability program (CLIVAR) is reoccupying key hydrographic surveys from the WOCE/JGOFS programs to monitor decadal changes in the ocean carbon system. Although this accumulated dataset amounts to ~330,000 global measurements, short-term variability in the upper ocean hampers efforts to constrain natural seasonal variability. Numerical models can be used in conjunction with one-time cruise measurements, atmospheric CO<sub>2</sub> inversions, or climate observations to understand drivers of the non-steady-state change (e.g., Le Quéré et al., 2007; Lovenduski et al., 2007; Matear and Lenton, 2008), however, without a sufficient baseline of measurements to validate model output, the diagnostic capability of numerical simulations remains somewhat limited. These challenges, both observationally and within models, limit our ability to detect and diagnose the nonsteady-state changes in ocean carbon.

### 1.4 Implications of oceanic CO<sub>2</sub> uptake

The immediate impact of additional  $CO_2$  in the ocean system is a lower pH and carbonate ion concentration – commonly referred to as ocean acidification (Zeebe et al., 2008). A growing number of studies indicate detrimental impacts for multiple marine ecosystems under ocean acidification conditions (Fabry et al., 2008; and references within).

The effect a reduction in  $CO_3^{2-}$  ions will have for calcifying organisms can be inferred from the saturation state of CaCO<sub>3</sub> minerals ( $\Omega$ )

$$\Omega = [Ca^{2+}][CO_3^{2-}]/K_{sp}^*$$
(1.1)
where  $K_{sp}^{*}$  represents the apparent stoichiometric solubility product for calcite or aragonite, and brackets the total concentrations in solution. When the saturation state falls below 1 (referred to as aragonite or calcite under-saturation) seawater becomes net corrosive for CaCO<sub>3</sub> minerals, resulting in enhanced dissolution and lower calcification rates (Fabry et al., 2008; and references within). Although the ocean surface is currently super-saturated with respect to both calcite and aragonite (i.e.  $\Omega$ >1), a model study by Orr et al. (2005) suggests that the annual-mean saturation state for aragonite (the more soluble CaCO<sub>2</sub> mineral) will fall below 1 in the Southern Ocean and some subarctic regions by centuries end. However, natural carbon variability has been shown to induce earlier under-saturation conditions in the Southern Ocean by 20 years relative to annualmean projections (McNeil and Matear, 2008), which highlights the importance of constraining high-frequency variability. Since current data limitations hamper our capacity to constrain higher-frequency natural variability globally, this represents a significant gap in our ability to understand future ocean acidification implications over the majority of the ocean.

Detecting robust secular trends in pH and  $\Omega$  (referred to as ocean acidification trends) has also been the focus of intense research over recent years due its importance for understanding future implications for marine ecosystems (e.g., Byrne et al., 2010; Feely et al., 2012; Vázquez-Rodríguez et al., 2012; Wootton et al., 2008). Despite a growing network of measurements, large natural variability relative to the small anthropogenic signal hinders efforts to discern robust ocean acidification trends. To determine the best locations for detecting ocean acidification requires an accurate understanding of higher-frequency natural variability.

## 1.5 Spatiotemporal patterns of air-sea CO<sub>2</sub> exchange

Quantifying where atmospheric  $CO_2$  is invading the ocean and at what rate is fundamental for understanding the impacts of oceanic  $CO_2$  uptake. Several independent methods have been proposed to observationally diagnose spatial patterns of air-sea  $CO_2$  exchange based on oceanic/atmospheric measurements. However, it's important to first understand the different flux estimates derived via these independent methods.

Exchange of  $CO_2$  at the air-sea interface was in a long-term steady-state prior to the industrial revolution (~11,000 years), which includes the out-gassing of reverine derived carbon, and is referred to here as the natural (or preindustrial) flux (Fig. 1.2). The ensuing perturbation in atmospheric  $CO_2$  due to the combustion of fossils fuels and land-use change has resulted in a human-induced air-sea  $CO_2$  exchange; which is aptly referred to as the anthropogenic flux. As the natural (or preindustrial) carbon cycle has shifted in response to changes in climate and ocean chemistry, a non-steady-state outgassing signal has emerged. Currently, air-sea  $CO_2$  exchange reflects the superposition of natural, anthropogenic and the non-steady-state flux signals. Here, this integrated flux signal is referred to as the contemporary flux.



**Figure 1.2:** Illustration of the different air-sea  $CO_2$  flux signals for pre-industrial (~1750) to present times.

Some methodologies estimate anthropogenic flux based on the assumption that natural  $CO_2$  exchange has remained in steady-state throughout the anthropocene. In this study, flux estimates derived via these methods are referred to as steady-state anthropogenic, and can be conceptualized in Fig. 1.2 as anthropogenic minus non-steady-state flux signals.

Care needs to be taken when comparing contemporary flux estimates derived via different approaches. When the contemporary signal is computed via summing estimates of the natural and steady-state anthropogenic fluxes, it will be different from other methods that can additionally resolve the emerging non-steady-state signal.

## 1.5.1 Inversion methods

Inversion approaches work on the premise that the distribution of a conservative tracer within a reservoir (e.g.  $CO_2$  in the atmosphere or anthropogenic  $C_T$  in the ocean interior) directly reflects any exchange with external reservoirs and transport within the medium. It therefore stands to reason that if transport can be reversed, regional fluxes of the tracer can be quantified.

Atmospheric inversions capture regional contemporary  $CO_2$  sources and sinks using atmospheric  $CO_2$  measurements and transport models (e.g., Baker et al., 2006; Patra et al., 2005). Continuous  $CO_2$  measurements from a global array of stations have allowed atmospheric inversion studies to diagnose seasonal to inter-annual flux signals (Baker et al., 2006; Gurney et al., 2004). However, these results are not fully independent, as they require prior constraints on spatial air-sea fluxes and model transport signals. Furthermore, it has been shown that regional flux estimates are strongly sensitive to the chosen  $CO_2$  observation network (Law et al., 2003; Patra et al., 2006), and that they have a poor capacity to partition flux signals between the land and marginal seas.

Ocean inversions, on the other hand, capture decadal-scale anthropogenic flux signals for a defined set of oceanic regions using data-based estimates of anthropogenic

 $C_{\rm T}$  concentrations and ocean transport models (Gloor et al., 2003; Mikaloff-Fletcher et al., 2006). Since all transport models used in the most recent study were not forced with climate-related alterations (Mikaloff-Fletcher et al., 2006), this approach captures the steady-state anthropogenic flux signal.

Ocean inversions have also been used to capture the natural air-sea flux signal via data-based estimates of the gas-exchange component of  $C_{\rm T}$  (Mikaloff-Fletcher et al., 2007). Although regional flux estimates range from -0.33 to 0.36 PgC yr<sup>-1</sup>, when integrated globally these regional differences largely cancel out, leaving a residual of  $0.03\pm0.08$  PgC yr<sup>-1</sup> (Mikaloff-Fletcher et al., 2007). However, ocean inversions cannot capture the natural out-gassing of riverine-derived carbon. To account for this, an estimated out-gassing of  $0.45\pm0.2$  PgC yr<sup>-1</sup> (Jacobson et al., 2007a) was added to their global estimate to arrive at a final natural air-sea CO<sub>2</sub> flux (Gruber et al., 2009).

Uncertain transport models, empirically derived tracers from measured  $C_{\rm T}$  and other hydrographic properties (i.e.  $\Delta C^*$  method of Gruber et al.(1996)), and uncertain riverine-derived CO<sub>2</sub> out-gassing all contribute to uncertainties in ocean inversion estimates.

A joint ocean/atmosphere inversion has also been developed to capture both preindustrial and steady-state anthropogenic  $CO_2$  fluxes (Jacobson et al., 2007a, b).

#### 1.5.2 Direct approach

Constraints on the air-sea  $pCO_2$  gradient and  $CO_2$  gas transfer can be used to calculate contemporary air-sea  $CO_2$  fluxes (Wanninkhof, 1992). Takahashi et al. (1997, 2002, 2009) has interpolated large ocean surface  $pCO_2$  measurement networks to diagnose monthly  $CO_2$  flux patterns for the nominal years of 1990, 1995 and 2000, respectively. Despite a current global network of almost 6.4 million "underway"  $pCO_2$  measurements (Takahashi et al., 2012), when coastal samples are removed and measurements are averaged over a 1°×1° grid-scale for each month and year, the dataset reduces to ~115,000 independent samples. From a spatial perspective, the distribution is heavily biased to the Northern Hemisphere. For example, ~38% of measurements were collected in the temperate North Atlantic/Pacific (18°N to 44°N), while only 14% were collected in the Southern Ocean (>44°S), yet the Southern Ocean covers 30% more open-ocean area. Spatiotemporal bias, together with strong natural variability in some ocean regions (e.g. equatorial Pacific), limits the Takahashi et al. (2009) approach to a course  $4^{\circ} \times 5^{\circ}$  resolution without the faculty to resolve inter-annual variability.

Some researchers have used the "underway"  $pCO_2$  network in conjunction with self-organizing maps (e.g., Friedrich and Oschlies, 2009b; Telszewski et al., 2009; Watson et al., 2009) or multi-linear regressions (e.g., Arrigo et al., 2010; Chierici et al., 2009; Cosca et al., 2003) to derive empirical relationships using sea-surface temperature and salinity. Since "underway" pCO<sub>2</sub> measurements generally have no coinciding biogeochemical information to help constrain the system (i.e. dissolved oxygen or nutrients), some researchers have additionally incorporated satellite-based estimates of Chlorophyll a and/or mixed-layer depth to help constrain the influence of phytoplanktonic growth and mixing (e.g., Arrigo et al., 2010; Telszewski et al., 2009). These empirical models have then been applied to basin wide predictor variable maps to diagnosed inter-annual flux fields. While this approach has only been applied on a basin-wide scale, a global application is currently being developed (Peter Landschützer; personal communication). Other researchers have related Takahashi et al. (2009)  $pCO_2$ climatology values to sea-surface temperature (SST), which are then used to project around the year 2000 using observed trends in SST to capture inter-annual flux variability (Park et al., 2010).

Although the direct approach does not capture the anthropogenic flux signal, it is common in the literature to subtract estimates of the natural  $CO_2$  exchange and quote this value as the anthropogenic flux signal (e.g., Takahashi et al., 2009; Wanninkhof et al., 2013). However, this approach does not account for the non-steady-state out-gassing signal, and therefor under-estimates the anthropogenic flux. For clarity within this work, flux estimates derived via this approach are referred to as net-contemporary (see Fig. 1.2).

## 1.6 Empirical approaches to investigate ocean carbon dynamics

The introduction of autonomous measuring devices, such as gliders, moorings, ships of opportunity and ARGO floats (Roemmich et al., 2004), has amounted to an order of magnitude more standard hydrographic measurements (SHP; temperature, salinity, dissolved oxygen and nutrients) than bottle carbon data (Boyer et al., 2009). In particular, remote locations with harsh environments (e.g. the Southern Ocean) are where improved data coverage is most striking. Since bottle carbon data are nearly always complimented with a suit of SHP measurements, which are known to modulate inorganic carbon variability (see *Sect. 1.2.1*), it should be implicit that we can derive empirical relationships between carbon parameters and SHP. If a robust model is established, it can be applied to the much larger SHP datasets to dramatically improve on current carbon data limitations, with immediate implications for advancing our understanding of ocean carbon dynamics.

Divergent biological and mixing regimes throughout the ocean make it difficult to apply linear empirical techniques on a global scale. Researchers have therefore traditionally partitioned the global bottle dataset geographically, hydrographically and temporally in an attempt to better constrain the system and provide insights on anthropogenic  $CO_2$  accumulation, air-sea  $CO_2$  flux patterns and biological matter production (Bates et al., 2006; Lee et al., 2000b, 2006; McNeil et al., 2007). Employing ad-hoc methods however leads to issues with these empirical methods, such as boundary discontinuities, influence of outliers that are not consistent with the bulk biogeochemical dynamics of a region, or statistical issues when applying linear regressions to a small sample group; all of which contribute to uncertainties in model predictions. Furthermore, since the most recent global empirical studies for  $C_{\rm T}$  (Lee et al., 2000b) and  $A_{\rm T}$  (Lee et al., 2006), the number of mixed-layer measurements has increased by ~45%, thereby providing a much larger, and more accurate data pool to constrain the global system.

## 1.7 Research goals and objectives

Several questions motivated the research presented in this thesis:

1) Can we develop and use non-linear empirical approaches to quantify and understand mixed-layer carbon dynamics ( $C_{\rm T}$  and  $A_{\rm T}$ ) on seasonal to inter-annual time-scales?

Previous empirical approaches relating bottle-based inorganic carbon to SHP have employed subjective data partitioning methods in order to better constrain the system. In *Chapter 2*, I explore the ability of a non-linear empirical technique to capture  $C_{\rm T}$  and  $A_{\rm T}$  concentrations in the global mixed-layer using a newly synthesized bottle data network without the need for any traditional data partitioning that may contribute to uncertainties in previous model predictions.

2) How important is natural CO<sub>2</sub> seasonality in influencing the onset and detection of future ocean acidification?

The implications of natural  $CO_2$  variability for future ocean acidification remain largely unknown due to current spatiotemporal data limitations. In *Chapter 3*, I firstly diagnose monthly ocean surface climatologies for properties related to ocean acidification via the approach developed in *Chapter 2*. These monthly climatologies are then used to investigate the magnitude and influence of seasonal  $CO_2$  variability for the onset of aragonite under-saturation and detecting future ocean acidification.

3) Can the global bottle carbon network be used to provide a new and independent constraint on the patterns of air-sea  $CO_2$  exchange across the global ocean?

Despite significant improvements over recent years, our understanding of spatial and temporal air-sea  $CO_2$  flux variability remains limited, since methods either require

the use of uncertain transport models or suffer from interpolation biases from regionally sparse "underway" oceanic  $pCO_2$  measurements (see *Sect. 1.5*). In *Chapter 4*, I deploy the new non-linear approach, as developed in *Chapter 2*, on ~22,000 bottle-derived  $pCO_2$  measurements to diagnose global monthly ocean surface  $pCO_2$  climatologies. This provides a new and independent constraint on the patterns of ocean surface  $pCO_2$ and contemporary air-sea  $CO_2$  exchange. Chapter 2.

# A novel empirical approach to diagnose seasonal ocean surface carbon from bottle data: SOMLO

This chapter is an edited version of a paper published in Biogeosciences (BG).

Sasse, T. P., McNeil, B. I., and Abramowitz, G. (2013), A novel method for diagnosing seasonal to inter-annual surface ocean carbon dynamics from bottle data using neural networks, Biogeosciences, 10(6), 4319-4340, DOI: 10.5194/bg-10-4319-2013

## Abstract

The ocean's role in modulating the observed 1-7 PgC yr<sup>-1</sup> inter-annual variability in atmospheric CO<sub>2</sub> growth rate is an important, but poorly constrained process due to current spatiotemporal limitations in ocean carbon measurements. Here, a non-linear empirical approach is developed to predict inorganic  $CO_2$  concentrations (total carbon dioxide  $(C_{\rm T})$  and total alkalinity  $(A_{\rm T})$  in the global ocean mixed-layer from hydrographic properties (temperature, salinity, dissolved oxygen and nutrients). The benefit of this approach is that once the empirical relationship is established, it can be applied to hydrographic datasets that have better spatiotemporal coverage, and therefore provide an additional constraint to diagnose ocean carbon dynamics globally. Previous empirical approaches have employed multiple-linear regressions (MLR) and relied on subjective geographic and temporal partitioning of carbon data to constrain complex global carbon dynamics in the mixed-layer. Synthesizing a new global  $C_T/A_T$  carbon bottle dataset consisting of ~33,000 measurements in the open-ocean mixed-layer, a neural network based approach is developed to better constrain the non-linear carbon system. The approach classifies features in the global biogeochemical dataset based on their similarity and homogeneity in a self-organizing map (SOM). After the initial SOM analysis, which includes geographic constraints, a local linear optimizer is applied to the neural network, which considerably enhances the predictive skill of the new approach. This new approach is called SOMLO, or self-organizing multiple-linear output. Using independent bottle carbon data, the traditional MLR analysis is compared to the novel SOMLO approach to capture spatial  $C_{\rm T}$  and  $A_{\rm T}$  distributions. Globally, the SOMLO approach improves predictive skill in  $C_{\rm T}$  by 19% relative to the traditional MLR, with a global capacity to predict  $C_{\rm T}$  to within 10.9 µmol kg<sup>-1</sup> (9.2 µmol kg<sup>-1</sup> for  $A_{\rm T}$ ). The nonlinear SOMLO approach is particularly powerful in complex but important regions like the Southern Ocean, North Atlantic and equatorial Pacific, where residual standard errors were reduced by 25-30% over traditional linear methods. The SOMLO technique is further tested using the Bermuda Atlantic time-series (BATS) and Hawaiian ocean time-series (HOT) datasets, where hydrographic data was capable of explaining 90% of the seasonal cycle and inter-annual variability at those multi-decadal time-series stations. Finally, the model is applied to global monthly climatologies of the standard hydrographic parameters to diagnose monthly  $C_{\rm T}$  and  $A_{\rm T}$  distributions in the global open-ocean.

## 2.1 Introduction

The ocean's role in modulating rising atmospheric carbon dioxide (CO<sub>2</sub>) levels has been found to be very important over a number of decades (Quay et al., 1992; Gruber et al. 1996; Sabine et al., 2004; Khatiwala et al., 2012). A variety of data-based estimates suggest net oceanic uptake for CO<sub>2</sub> to be  $2.1\pm1.0$  PgC yr<sup>-1</sup> (1 Pg =  $10^{15}$  g) since the year 2000, or about 25-30% of total anthropogenic CO<sub>2</sub> emissions (fossil-fuel + land use) over that period (Jacobson et al., 2007a; Khatiwala et al., 2009; Manning and Keeling, 2006; McNeil et al., 2003; Mikaloff-Fletcher et al., 2006; Takahashi et al., 2009). Between 1990 and 2009, atmospheric CO<sub>2</sub> accumulation rates vary between 1-7 PgC yr<sup>-1</sup>, indicating large inter-annual variability from both the terrestrial and oceanic reservoirs (Sarmiento et al., 2010). Although our long-term, decadal-scale understanding of oceanic CO<sub>2</sub> uptake has advanced, our shorter-term understanding (seasonal to inter-annual) of ocean carbon dynamics remains poorly constrained due to current data limitations.

Atmospheric CO<sub>2</sub> observations, inversion techniques and ocean models suggest a large range for inter-annual variability in oceanic CO<sub>2</sub> uptake (0.1-1.5 PgC yr<sup>-1</sup>) (Bender et al., 2005; Le Quéré et al., 2003; Patra et al., 2006; Rayner et al., 2008). However, from an oceanic perspective, our understanding of natural variability in the ocean carbon system has come about sporadically, dominated by regional time-series measurement programs (e.g., Bermuda Atlantic time-series (BATS; Bates, 2007) and Hawaiian Ocean time-series (HOT; Keeling et al., 2004)). Without a better understanding of shorter-scale natural variability, the ability to constrain and understand the time-evolving capacity for the ocean to absorb atmospheric  $CO_2$  in a high- $CO_2$ world will be limited, particularly since some evidence suggests the ability for the ocean to absorb  $CO_2$  has slowed since the late 1980s as a consequence of decadal-scale trends in winds and oceanic circulation (Le Quéré et al., 2010; McNeil and Matear, 2012; Sarmiento et al., 2010).

Standard hydrographic measurements in the ocean (temperature, salinity, dissolved oxygen and nutrients) are sampled and analyzed much more frequently than inorganic carbon. With the deployment of satellites, gliders and ARGO floats providing an immense capacity for capturing short-term seasonal to inter-annual variability in the oceans, the question is, can this new information be used to help infer and diagnose short-term carbon dynamics in the ocean?

The oceans inorganic carbon system can be fully constrained by knowing any two measurements within its inorganic carbon constituents; partial pressure of CO<sub>2</sub>  $(pCO_2)$ , total dissolved carbon dioxide  $(C_T)$ , total alkalinity  $(A_T)$  or pH. Significant time and resources have been devoted on national and international levels to survey the global oceanic  $C_T$  and  $A_T$  distribution. However, even with approximately 330,000 bottle measurements taken sporadically over the past 30 years, our ability to globally understand natural seasonal  $C_T$  and  $A_T$  dynamics has been hindered due to the large spatiotemporal limitations in this current accumulated dataset (Key et al., 2004).

Autonomous  $pCO_2$  measuring devices, mounted mainly onto commercial shipping vessels, has resulted in a global network of approximately 6.4 million ocean surface  $pCO_2$  measurements (Takahashi et al., 2012). This  $pCO_2$  dataset has given us the best idea of seasonal (Takahashi et al., 2009; herein after referred to as T-09) to inter-annual (McKinley et al., 2011; Park et al., 2010; Telszewski et al., 2009) CO<sub>2</sub>

variability within the ocean. However, the global  $pCO_2$  dataset cannot inform us on some very important processes and biogeochemical dynamics that modulate atmospheric CO<sub>2</sub>. The oceans biological carbon export flux has been estimated to be between 11-16 PgC yr<sup>-1</sup> from satellite-based chlorophyll a measurements (Falkowski et al., 2000), some 5-8 times the net oceanic CO<sub>2</sub> absorption from the atmosphere. Small changes in the biological carbon flux will therefore have large and important implications for atmospheric  $CO_2$ . However, this large signal is yet to be constrained from inorganic carbon data itself, since it requires constraints on mixed-layer carbon dynamics rather than just sea-surface constraints like the  $pCO_2$  climatology. Secondly, without equivalent  $A_T$  or  $C_T$  measurements,  $pCO_2$  by itself cannot provide insights into partitioning the biological carbon pump into both organic and calcification components, particularly important with regard to future ocean acidification impacts. Previous estimates on this "rain ratio" (organic/calcifier export flux) have needed to assume a constant Redfield ratio on nutrient changes in the oceans mixed-layer (Sarmiento et al., 2002). Finally, spatiotemporal deficiencies in the  $pCO_2$  measurement network, particularly in remote locations like the Southern Ocean, introduce uncertainties in the direct evaluation of short-term variability. To understand seasonal to inter-annual variability in these regions requires methods that have better spatiotemporal coverage than is constrained by historical  $pCO_2$  sampling. Here, I seek to diagnose seasonal to inter-annual  $C_{\rm T}$  and  $A_{\rm T}$  concentrations in the mixed-layer, which will provide independent, but important additional constraints to the global sea-surface  $pCO_2$ climatology.

To varying degrees, concentrations of  $C_{\rm T}$  and  $A_{\rm T}$  are influenced by the solubility of CO<sub>2</sub>, biological processes, vertical and lateral water transport and direct CO<sub>2</sub> exchange with the atmosphere (Sarmiento and Gruber, 2006). Ocean mixing is largely controlled by density dynamics via temperature (T) and salinity (S) variations in the ocean, which also regulate the solubility of CO<sub>2</sub> (Weiss, 1974). Information on nitrate (N), silicate (Si), phosphate (P) and dissolved oxygen (DO or O) variations provide insight into the biological influences on oceanic inorganic carbon (Anderson and Sarmiento, 1994). From this, it should be implicit that empirical relationships can be derived between these standard hydrographical parameters and the carbon constituents. If a robust empirical relationship is established, it could be applied to the order of magnitude more in situ measurements of these standard hydrographic parameters (Boyer et al., 2009), or the objectively analyzed  $1^{\circ} \times 1^{\circ}$  climatologies (e.g. Locarnini et al. (2010)), to give new constraints on seasonal to inter-annual carbon dynamics in the mixed-layer.

The use of the global sea-surface  $pCO_2$  dataset would be ideal to develop such empirical algorithms. However, these continuous  $pCO_2$  measurements generally have no coinciding biogeochemical information (i.e. DO or nutrients) that could help establish an empirical relationship. Some have used satellite-based Chlorophyll *a* measurements to help constrain ocean surface  $pCO_2$  with varying degrees of success (Chen et al., 2011; Chierici et al., 2009; Telszewski et al., 2009). The benefits of using ship-based bottle measurements of  $C_T$  and  $A_T$  is that they are almost always complemented by a suite of hydrographic and biogeochemical parameters (T, S, DO and nutrients) that can be used to help derive empirical relationships.

Wallace (1995) verified a multiple-linear regression (MLR) concept by successfully capturing  $C_{\rm T}$  using T, S, Si and apparent oxygen utilization (AOU) in the North Atlantic. Several studies have since investigated this MLR approach in capturing the surface distribution of  $C_{\rm T}$  and  $A_{\rm T}$  (Table 2.1).

**Table 2.1:** Previous empirical approaches to predict upper ocean  $A_T$  and  $C_T$  concentrations. T = temperature, S = salinity, DO = dissolved oxygen, AOU = Apparent Oxygen Utilization, N = Nitrate (NO<sub>3</sub><sup>-</sup>), Si = silicate (SiO<sub>4</sub>), P = phosphate (PO<sub>4</sub><sup>3-</sup>), Chl-*a* = Chlorophyll *a*, Lat = Latitude, Long = Longitude.

Ocean	Response	Predictor variables	$N^{\mathrm{a}}$	RSE <sup>b</sup>	Authors	
Region	variable			[µmol kg <sup>-1</sup> ]		
Global	$NA_T^c$	Т	1740	5	(Millero et al., 1998)	
Global	$A_{\mathrm{T}}$	T, $T^2$ , S, S <sup>2</sup> , Long	5692	8.1	(Lee et al., 2006)	
Indian	$A_{\mathrm{T}}$	T, S, N, AOU, Depth, Lat, P	2363	4.5 - 6.4 <sup>d</sup>	(Bates et al., 2006)	
Southern	$A_{\mathrm{T}}$	S, N, Si	1200	8.1	(McNeil et al., 2007)	
Arctic	$A_{\mathrm{T}}$	T, S	853	26.9, 75	(Arrigo et al., 2010)	
Global	$NC_T^{c}$	$T, T^2, N$	~4900	7	(Lee et al., 2000b)	
Indian	$C_{\mathrm{T}}$	T, S, N, AOU, Depth, Lat, P	2395	4.4 - 6.0 <sup>d</sup>	(Bates et al., 2006)	
Southern	$C_{\mathrm{T}}$	T, S, DO, N, Si	1032	8	(McNeil et al., 2007)	
Arctic	$C_{\mathrm{T}}$	Chl-a, T, S	853	33.4, 61.6, 17.3	(Arrigo et al., 2010)	

<sup>a</sup> Number of measurements used in the study

<sup>b</sup> Residual standard error as quoted by the authors

<sup>c</sup> Salinity normalized concentrations of  $C_{\rm T}$  and  $A_{\rm T}$  (×35/S)

<sup>d</sup> Range of RSE values presented for the four monsoonal/inter-monsoonal seasons

Divergent biological and mixing regimes throughout the ocean have made it difficult to use linear empirical techniques on a global scale. Researchers have traditionally partitioned the global bottle dataset based on subjective geographical, hydrographical and temporal boundaries in an attempt to improve the ability of linear approaches to capture the non-linear relationship between inorganic carbon and the standard hydrographic parameters. Here, a non-linear empirical modelling approach is used to avoid this subjective partitioning and is shown to deliver considerable improvements in predictability. The non-linear model uses a self-organizing map (SOM; Kohonen, 1988) to classify (or cluster) measurements of hydrographic parameters and  $C_T/A_T$  separately for each group. The SOM model has already been found well suited in extracting features of the ocean surface  $pCO_2$  dataset in the North Atlantic using a combination of modelled and remotely sensed parameters to constrain the system, (Friedrich and Oschlies, 2009b, a; Lefèvre et al., 2005; Telszewski et al., 2009).

To contextualize this work, a traditional MLR approach is first used to diagnose global seasonal carbon dynamics in the ocean. To do this, the MLR approach is deployed on a newly synthesized  $C_T/A_T$  bottle dataset of ~33,000 mixed-layer samples. Next, the SOM-based approach is presented and used to diagnose seasonal carbon dynamics on a global scale, which better accounts for non-linearities that would limit the ability of the MLR. To compare the MLR and new SOM approach, a systematic independent test is developed to assess the model's skill. The BATS and HOT in situ time-series are then used as an explicit test for the new approach. Finally, the capacity of the model is shown to capture coherent spatial and temporal carbon fields over the global open-ocean surface.

## 2.2 Global bottle measurements and training dataset

The extraordinary effort to collate and synthesis bottle hydrographic and biogeochemical data has been conducted by several groups; including GLODAP (Global Ocean Data Analysis Project; Key et al., 2004), CARINA (CARbon dioxide IN the Atlantic ocean; CARINA Group, 2009b, a, 2010) and PACIFICA (PACIFic ocean Interior CArbon; Suzuki et al., 2013).

Precision in bottle  $C_{\rm T}$  and  $A_{\rm T}$  measurements has consistently improved over the past 30 years as a result of advances in techniques and apparatus (e.g., Bradshaw et al., 1981; Johnson et al., 1987). In particular, the introduction of standard operating procedures and certified reference materials (Department of Energy, 1994; Dickson et al., 2003; Dickson et al., 2007) has achieved quality consistency within independent laboratory measurements, which is currently estimated to be  $\pm 2 \mu \text{mol kg}^{-1}$  (Dickson et al., 2007). Despite this, systematic measurement biases exist between independent laboratories. To account for this issue when combining data, a secondary Quality Control (QC) method was incorporated by the project groups to identify and smooth out any systematic offsets, as outlined in Tanhua et al. (2010). The internal consistency of the CARINA  $C_{\rm T}/A_{\rm T}$  dataset is estimated to be  $\pm 2.5 \mu \text{mol kg}^{-1}$  (Tanhua et al., 2010).

More recent additional measurements included in the global dataset underwent a 1<sup>st</sup> QC check to remove measurements flagged as bad or questionable under the World Ocean Circulation Experiment (WOCE) convention (Joyce and Corry, 1994). No 2<sup>nd</sup> QC was implemented since accuracy in recent measurements is estimated at  $\pm 2 \mu mol$  $kg^{-1}$ .

For this work, 470 cruises from GLODAP, PACIFICA, CARINA, CLIVAR and miscellaneous sources were merged with the BATS and HOT measurements to form the global carbon training dataset, as shown in Table 2.2 (see Appendix B for full cruise summary). Some cruise datasets were incorporated into two or more of the synthesis projects under different names or with slight variations in quoted time and position. To identify and remove duplicated data from the combined global dataset, the entire dataset was checked for measurements with exact and near exact position and time values. The global dataset was then refined to be within the mixed-layer, non-coastal and data post 1980 due to large uncertainties in early measuring techniques. The final number of usable  $C_{\rm T}/A_{\rm T}$  discrete measurements in the global mixed-layer was ~33,000.

Source	Number of Measurements
CARINA	12599
PACIFICA	9690
GLODAP	6674
<b>CLIVAR</b> <sup>a</sup>	1689
AAIW <sup>b</sup>	755
BATS <sup>c</sup>	705
$HOT^d$	540
NACP <sup>e</sup>	291
Miscellaneous	192
Total	33135
<sup>a</sup> Climate	Variability and Predictability
<sup>b</sup> Antarcti	c Intermediate Cruise
c Bermud	a Atlantic time-series
<sup>d</sup> Hawaiia	n Ocean time-series

 
 Table 2.2: Data sources of the global merged dataset. The second column indicates the
number of samples with a least one carbon parameter measurement (see Appendix B for full cruise summary).

<sup>e</sup> North Atlantic Carbon Program

#### 2.2.1 Defining mixed-layer depths

Derived by wind stress and air-sea heat exchange, the mixed-layer depth (MLD) describes the maximum penetration depth of the quasi-homogeneous region of surface water (Kara et al., 2003). Typically ranging from 20m in summer months, to 500m during the winter season in some parts of the ocean (de Boyer et al., 2004; see Fig. 2.1), including MLD measurements is an important additional constraint on carbon dynamics that is added from bottle measurements.

Discriminating mixed-layer measurements for each cast was conducted via a bivariant linear interpolation from a regular  $2^{\circ} \times 2^{\circ}$  gridded MLD climatology developed by de Boyer et al. (2004) (Fig.2.1). Their methodology was based on a change in potential density from a 10m reference measurement of 0.03 kg m<sup>-3</sup>. Approximately 900,000 CTD profiles including Argo data up to September 2008 were used to constrain the MLD climatology.



**Figure 2.1:** Global distribution of mean MLD for (**a**) summer months (**b**) winter months and (**c**) seasonal-mean difference (winter minus summer). Summer and winter months are defined as December to February and June to August for the Southern Hemisphere, respectively. Boreal summer/winter months are opposite. These seasonal MLD distributions were derived using the MLD climatologies of de Boyer et al. (2004).

#### 2.2.2 Identifying coastal samples

Carbon biogeochemical dynamics in coastal zones have been shown to be divorced from the open-ocean system due to terrigenous influences. For example, sediment upwelling, anthropogenic influences on coastal ecosystems, and nutrient/carbon delivery from rivers have all been identified as processes perturbing coastal biogeochemical dynamics from the open-ocean (Cotrim da Cunha et al., 2007; Gibbs et al., 2006; Jickells, 1998; Seitzinger et al., 2005). To mitigate these biases from the oceanic dataset, all casts with a seafloor bathymetry of 500m or less were removed from the global training dataset. The bathymetric depth for each cast was linearly interpolated from NOAA's 1 arcminute global relief product re-gridded to 10 arcseconds (Amante and Eakins, 2009). Eliminating coastal influences reduces the global dataset by ~9%, but is important when applying the neural network approach.

## 2.2.3 Biases in the global dataset

While the distribution of measurements in the refined dataset exhibits no spatial bias over the major ocean basins (Fig. 2.2a), there are approximately 45% less wintertime measurements than were collected during summertime (Fig. 2.2b). This seasonal distribution is examined here as a potential cause for bias when applying the empirical approach.



**Figure 2.2:** (a) Spatial distribution of the global training dataset, together with (b) seasonal and (c) yearly histograms partitioned into Southern (light shade) and Northern (dark shade) Hemispheres. Southern Hemisphere seasons are defined as summer (December to February), autumn (March to May), winter (June to August) and spring (September to November), Northern Hemisphere seasons are opposite.

## 2.3 Normalization of $C_{T}$ measurements

Global atmospheric CO<sub>2</sub> concentrations during the 1980's, 1990's and 2000's increased at 1.60±0.56, 1.47±0.66 and 1.90±0.38 ppm yr<sup>-1</sup> respectively (Conway, T. and Tans, P., NOAA/ESRL, www.esrl.noaa.gov/gmd/ccgg/trends). Since the air-sea pCO<sub>2</sub> gradient represents the thermochemical driving potential for CO<sub>2</sub> exchange, growth rates in ocean surface pCO<sub>2</sub> have closely tracked atmospheric CO<sub>2</sub> (e.g., Bates, 2007; Feely et al., 2006; McKinley et al., 2011; Takahashi et al., 2009). To account for anthropogenic CO<sub>2</sub> uptake, all mixed-layer measurements were corrected to the reference year 2000 by calculating the change in mixed-layer  $C_{\rm T}$  in equilibrium with the atmospheric CO<sub>2</sub> increase using observed Revelle factors. This approach is somewhat equivalent to that of T-09, whereby all pCO<sub>2</sub> measurements values were corrected to the year 2000 using a rate of 1.5 µatm yr<sup>-1</sup>.

#### 2.3.1 Revelle factor approach

The Revelle factor (*R*) quantifies the relationship between the fractional change in oceanic  $pCO_2$  and  $C_T$  in an otherwise static system (Eq. 2.1). For small changes, the partial derivatives in *R* can be expressed as a delta, which then provides a well suited empirical means to account for anthropogenic biases in  $C_T$  measurements.

$$R = \frac{\partial \ln p \text{CO}_2}{\partial \ln C_{\text{T}}} = \frac{C_{\text{T}}}{p \text{CO}_2} \frac{\Delta p \text{CO}_2}{\Delta C_{\text{T}}}$$
(2.1)

Rearranging Eq. 2.1 illustrates how the anthropogenic  $C_{\rm T}$  component ( $\Delta C_{\rm T}$ ) can be constrained if  $C_{\rm T}$ ,  $p\rm CO_2$  and R are known together with the anthropogenic change in oceanic  $p\rm CO_2$  ( $\Delta p\rm CO_2$ ) via

$$\Delta C_{\rm T} = \frac{C_{\rm T}}{R} \frac{\Delta p \rm CO_2}{p \rm CO_2} \tag{2.2}$$

Revelle factors and  $pCO_2$  concentrations were calculated using bottle measurements of  $C_T$ ,  $A_T$ , temperature and salinity (phosphate and silicate concentrations

were also used where available) via the CO2SYS program developed by Pierrot et al. (2006). Selection of the Mehrbach et al. (1973) carbonic acid constants, as refitted by Dickson and Millero (1987), was based on comparison studies by Lee et al. (2000a), McNeil et al. (2007), Millero et al. (2002), and Wanninkhof et al. (1999), and maintained consistency with the GLODAP and CARINA products (Key et al., 2004; Pierrot et al., 2010). Here, the anthropogenic rate of increase in mixed-layer  $pCO_2$  is assumed to be equilibrium with the observed annual atmospheric CO<sub>2</sub> growth rate at the Mauna Loa measurement site (Tans, P., NOAA/ESRL, www.esrl.noaa.gov/gmd/ccgg/trends and Keeling, R. Scripps Institute of Oceanography, scrippsco2.ucsd.edu/). The final expression to correct  $C_{\rm T}$  measurements to the reference year 2000 is given by

$$C_{\mathrm{T}(2000)}^{\mathrm{sea}} = C_{\mathrm{T}(\mathrm{in\ situ\ year})}^{\mathrm{sea}} + \left(\frac{\mathrm{CO}_{2(2000)}^{\mathrm{atm}} - \mathrm{CO}_{2(\mathrm{in\ situ\ year})}^{\mathrm{atm}}}{p\mathrm{CO}_{2(\mathrm{in\ situ\ year})}^{\mathrm{sea}}}\right) \frac{C_{\mathrm{T}(\mathrm{in\ situ\ year})}^{\mathrm{sea}}}{R}$$
(2.3)

where superscripts sea and atm represent seawater and atmosphere respectively.

Of the 31,438 global mixed-layer  $C_{\rm T}$  measurements, 28% (8,711) were missing at least one crucial parameter to constrain the anthropogenic correction via the proposed technique (i.e.  $A_{\rm T}$ , temperature and/or salinity). Rather than discarding these samples, the 22,727 corrected  $C_{\rm T}$  measurements were used to constrain the anthropogenic correction via a 4-D linear interpolation in latitude, longitude, in situ pressure and calculated annual anthropogenic rate of  $C_{\rm T}$  increase. The skill of this approach was evaluated by dividing the 22,727 measurements into 10 equal subsets, which were then each used as an independent test. This independent testing showed that interpolated values were accurate to within  $\pm 0.08 \ \mu {\rm mol} \ {\rm kg}^{-1} \ {\rm yr}^{-1}$  (or 8% of the mean annual rate of  $C_{\rm T}$  increase).

The global-mean rate of increase in mixed-layer  $C_{\rm T}$  concentration was found to be 0.996 µmol kg<sup>-1</sup> yr<sup>-1</sup> (Fig. 2.3), which is consistent with the 1 µmol kg<sup>-1</sup> yr<sup>-1</sup> anthropogenic  $C_{\rm T}$  correction rate used by Lee et al. (2000b) for measurements between 30°N and 30°S, and is consistent with reported rates of increase observed at the HOT (Winn et al., 1998) and BATS (Bates, 2007) time-series stations.



**Figure 2.3:** Anthropogenic correction applied to mixed-layer  $C_{\rm T}$  measurements (defined by  $C_{\rm T}$  correction =  $C_{\rm T}$  (in situ) –  $C_{\rm T}$  (correct to the year 2000)). Deploying a linear regression between Year and  $C_{\rm T}$  correction (red line) indicates the mean rate of increase in global mixed-layer  $C_{\rm T}$  measurements is 0.996 µmol kg<sup>-1</sup> yr<sup>-1</sup> (r-squared value of 0.96).

Although there are regions of the ocean where upwelling and sea-ice inhibit airsea gas exchange, resulting in considerable CO<sub>2</sub> disequilibrium (e.g. Southern Ocean and equatorial Pacific), the anthropogenic CO<sub>2</sub> correction technique used here, like those for T-09 and Lee et al. (2000b), will be biased in these regions. However, by performing a test using no anthropogenic CO<sub>2</sub> correction (see *Sect. 2.8.4*), it was found that this anthropogenic correction had a very low impact on the models ability to predict global  $C_{\rm T}$ . This is in part due to the large natural fingerprint of  $C_{\rm T}$  (±50 µmol kg<sup>-1</sup>) relative to the small changes (~1 µmol kg<sup>-1</sup> yr<sup>-1</sup>) resulting from anthropogenic CO<sub>2</sub> uptake.

#### 2.3.2 Uncertainties in the correction method

Two assumptions that could introduce uncertainties in the applied anthropogenic  $CO_2$  corrections include a) a static Revelle factor over the correction period, and b) the atmospheric  $CO_2$  record observed at Mauna Loa is representative on a global scale.

The direct result of additional  $CO_2$  in the oceans is an increase in the Revelle factor (Egleston et al., 2010). Between 1800 and 1994, the absorption of ~118 Pg of anthropogenic carbon is estimated to have increased *R* by 1 unit (Sabine et al., 2004). Since bottle-derived *R* values range from ~8 in (sub)tropical regions to 15 in higher latitudes (Fig. 2.4), this relatively small increase in *R* over the past ~200 years suggests that any change over the average correction period applied here of 5.7 years will be insignificant.





To ensure the observed annual-mean change in atmospheric CO<sub>2</sub> ( $\Delta$ CO<sub>2</sub>) at Mauna Loa is applicable on a global scale, the  $\Delta$ CO<sub>2</sub> term at Mauna Loa ( $\Delta$ CO<sub>2</sub> = CO<sub>2(2000)</sub> – CO<sub>2(year)</sub>) is compared to a global-mean  $\Delta$ CO<sub>2</sub> estimate derived from multiple measurement stations (Conway, T. and Tans, P., NOAA/ESRL, www.esrl.noaa.gov/gmd/ccgg/trends) (Fig. 2.5). The high degree of similarity between the two temporal records indicates that no bias is introduced though applying the observed Mauna Loa  $\Delta CO_2$  on a global scale. This high correlation comes as no surprise, since the time for inter-hemisphere mixing is approximately 1 year for atmospheric CO<sub>2</sub> (Bowman and Cohen, 1997).



**Figure 2.5:** Difference in annual-mean atmospheric  $CO_2$  relative to the year 2000 ( $\Delta CO_2$ ) observed at Mauna Loa (closed circles) and a global estimate derived from multiple measurement stations (open triangles).

## 2.4 Testing algorithm skill: a Systematic Independent Test (SIT)

Most empirical studies report statistical errors calculated as the residual standard error (RSE) from linear regressions. For example,  $C_{\rm T}$  in the Indian Ocean was reported to be predicted to within ±5 µmol kg<sup>-1</sup> using a suite of hydrographic parameters (Bates et al., 2006), ±8 µmol kg<sup>-1</sup> for the Southern Ocean (McNeil et al., 2007) and ±7 µmol kg<sup>-1</sup> for a global dataset (Lee et al., 2000b). However, an independent dataset not used in the

regressions is needed to accurately report true statistical uncertainty for any empirical approach.

Here, a 'Systematic Independent Test' (SIT) approach was developed in order to compare the MLR and NN empirical approaches consistently. The SIT method evaluates the algorithm's skill through an independent test of each cruise or time-series without using it in the training or regression dataset. This implies that for a training data pool consisting of *n* cruises and *i* time series, n + i unique algorithms with identical model configurations are used to predict the excluded cruise or time series measurements. Calculating the residual standard error (RSE; Eq. 2.4) using all (or a subset) of the independent cruises and time-series predictions then provides a better and accurate estimate of the algorithms global (or regional) skill. In Eq. 2.4, the independent predictions and in situ measurements are represented by  $y_{indp-pred}$  and  $y_{in-situ}$  respectively, and *N* represents the number of discrete samples.

$$RSE = \sqrt{\frac{\sum (y_{indp-pred} - y_{in-situ})^2}{N-2}}$$
(2.4)

The reason for independently testing each cruise dataset individually, rather than a randomly selected subset of data, is due to similar concentrations of carbon and auxiliary measurements within local casts of the same cruise. As there are typically two to three measurements within each cast of the training dataset, the independent prediction of one of these measurements will give a misleading representation of the model's true skill, as the remaining two measurements with a very similar "biogeochemical fingerprint" will be used to train the algorithm. The independent prediction of an entire cruise will therefore provide a more robust measure of the algorithm's skill.

## 2.5 Traditional MLR approach

#### 2.5.1 Method description

Multiple-linear regression is a numerical estimation of the linear relationship between a set of predictor variables,  $\mathbf{x} = (x_1 \dots x_n \dots x_N)$ , and response variable, y:

$$y = \beta_0 + \sum_{n=1}^{N} \beta_n x_n$$
 (2.5)

where  $\beta_0$  and  $\beta_n$  represent the intercept and empirically derived coefficients respectively.

Multi-collinearity (MCL) between predictor variables and non-normality of the residual errors are both issues that may affect the predictive and diagnostic ability of a model. MCL refers to the scenario when two or more predictor variables are linearly correlated to a high extent, or one predictor variable is a near linear combination of others. When MCL is present, it can result in misleading model results and hypothesis test conclusions (Faraway, 2004). A quantitative test for the presence and extent of MCL is the Variance Inflation Factor (VIF), which indicates the inflation to the coefficients standard errors due to MCL. Interaction terms between collinear variables can be incorporated into the model to reduce the effects of MCL.

When the distribution of residual errors is deviated from normal, the estimated coefficients may not be optimal, resulting in inexact hypothesis tests and confidence intervals. Visual inspection of Q-Q plots is a typical approach to identify skewness within the residual errors. Although more quantitative methods are available, such as the Shapiro-Wilk, they should only be used in conjunction with subjective visualization of Q-Q plots. Depending on the degree of departure from normality, actions can range from transforming the response variable in cases of severe Cauchy distribution, to simply ignoring the issue when only short tailed skewness is observed. The simplest pre-emptive measure to reduce the effects of non-normality is through a robust MLR regression, which reduces the influence of outliers (Faraway, 2004).

In order to minimize the effects of MCL and non-normality, a forward stepwise robust MLR routine was developed here to constrain the empirical relationships between  $C_{\text{T}}/A_{\text{T}}$  and the standard hydrographic parameters.

#### 2.5.2 Robust MLR routine

Following the schematic in Fig. 2.6, the routine initiates by ranking predictor variables  $p_1, ..., p_n, ..., p_N$  according to their degree of linear correlation to the response variable, y; where  $p_{n,1}$  represents the parameter with the highest correlation. The primary model  $(M_1)$  is then established by applying a least-squares MLR between the top ranked predictor variable  $(p_{n,1})$  and y to constrain the regression coefficients  $\beta_0$  and  $\beta_{n,1}$ . The routine then expands on  $M_1$  in step 3 by regressing the top two ranked predictor variables (m = 2); where m represents the modelled predictor variable with the lowest correlation to y.

To determine if MCL exists in the expanded model  $(M_m)$ , VIF values are calculated for each modelled variable in  $M_m$  and compare to VIF values for the same variables modelled in  $M_{m-1}$ . The existence of MCL is identified if the VIF value for any predictor variable  $p_{n,i}$  (where i < m) increased by 5. For the scenario when MCL is detected, the model is updated with interaction terms between the newly added predictor variable  $(p_{n,m})$  and any modelled variable with a VIF increase greater than 5. An analysis of variance (ANOVA) between the previous model  $(M_{m-1})$  and expanded model  $(M_m^*)$  is then applied to evaluate the significance of the newly added predictor variable and interaction terms. If the expanded model is found to statistically constrain the system with a higher degree of skill with a 95% confidence interval, the updates are accepted and the routine returns to step 3 to incorporate the next lowest ranked predictor variable (i.e. m = m + 1).

If MCL is not detected, a null-hypothesis test based on the t-statistic is applied to determine if the coefficient of the new predictor variable  $(\beta_{n,m})$  is significantly different from 0 (i.e. the new predictor is important in constraining the system). If it

does not differ from 0 within a 95% confidence interval, the new predictor variable is defined as insignificant and is subsequently rejected from the model. The routine then returns to step 3 to again expand  $M_m$  with the next lowest ranked predictor variable.

Once each predictor variable has had an opportunity to update the model (i.e. m = I), any desired higher order terms are incorporated into the model on the provision that the first order term was found to be statistically significant. The routine then prunes the model through an iterative process to remove insignificant terms based on the t-test. Once all terms are found to be statistically significant, the final stage of the routine applies a robust MLR to the set of significant terms to reduce any influence from outliers.

- Step 1) Rank predictor variables according to Pearson correlation:  $(p_{n,1}, \dots, p_{n,i}, \dots, p_{n,l})$
- Step 2) Primary least-squares MLR model (m = 1):  $M_1 = \beta_0 + \beta_{n,1} p_{n,1}$
- Step 3) Expand on model (m = m + 1):





Step 8) Apply robust MLR to optimal parameter combination

Figure 2.6: Schematic diagram of the robust forward MLR routine.

This MLR routine is well suited for optimizing the model and dampening the influence of outliers that cannot be reasonably identified as bad measurements. This aspect is particularly important when the global dataset is subject to ad-hoc geographical and/or temporal separation methods, where measurements not consistent with the bulk biogeochemical dynamics within a region have the potential to influence the model.

Although the primary interest here is to optimize the models ability to predict mixed-layer  $C_{\rm T}$  and  $A_{\rm T}$  concentrations, the regression fits can provide important information on the relationship between individual predictor parameters and the response variable. The regression routine was therefore developed to minimize any violation of the underlying MLR assumptions in order to provide future opportunities for diagnostic investigation.

#### 2.5.3 Ad-hoc versus universal MLR

To investigate the application of the traditional MLR method, the skill of using one single regression globally (universal MLR) was compared to an ad-hoc approach that partitions the dataset into regions (ad-hoc MLR). The ad-hoc approach was based on dividing the global carbon dataset on the geographical and temporal guidelines outlined by Lee et al. (2000b, 2006) and Bates et al. (2006). In this way, the global dataset was partitioned into 5 geographic regions to constrain the  $A_{\rm T}$  system, and 11 geographic regions, 8 of which were subjected to further partitioning into summer and winter months to constrain  $C_{\rm T}$  (see Fig. 2.7). The universal method simply uses the entire global dataset without division.



**Figure 2.7:** Spatiotemporal division of the global training dataset for the ad-hoc MLR approach. Black boundaries are common for both  $C_{\rm T}$  and  $A_{\rm T}$  models, while red boundaries are for  $C_{\rm T}$  only, and blue for  $A_{\rm T}$  only. A red asterix indicates that MLR's were developed for both summer months (November to April for austral hemisphere) and winter months (May to October for austral hemisphere) to constrain  $C_{\rm T}$ . Boreal summer/winter seasons are opposite.

## 2.5.4 Optimal MLR parameter combination

Defining the optimal predictor combination for the universal model was conducted via a two stage process. In the first stage, the model was initially trained and tested (SIT) using all available parameters, and then each individual parameter was excluded to evaluate its importance in constraining the system. This identified salinity, followed by temperature as the two most important parameters for capturing mixed-layer  $C_{\rm T}$ , while salinity was the only parameter to strongly influence the predictive skill of the  $A_{\rm T}$  model (Fig. 2.8).



**Figure 2.8:** Global RSE values calculated using the SIT predictions of the universal MLR model. Excluded parameter represent the variable not used in the MLR training and testing; including phosphate (P), nitrate (N), silicate (Si), dissolved oxygen (DO), salinity (S) and temperature (T). The dashed lines represent the RSE values for models trained using all parameters.

To then converge on the optimal parameter combination for  $C_{\rm T}$ , the model was first trained and tested using only temperature and salinity information. Dissolved oxygen and nutrient parameters were then individually included to ascertain their importance (Fig. 2.9a). This revealed that phosphate and nitrate both provide important additional information beyond temperature and salinity, while dissolved oxygen did not improve the models predictive skill. Finally, testing the universal model using the predictor combinations TSPN, TSPNSi and TSPNO revealed only a slight improvement beyond TSP and TSN (Fig. 2.9b). Despite these improvements, the initial model using all available parameters captured global  $C_{\rm T}$  with the lowest RSE, and was therefore defined as the optimal predictor combination. The inclusion of second order terms for temperature and salinity did not further improve the models global skill for  $C_{\rm T}$ .



**Figure 2.9:** Global RSE values calculated using SIT predictions of the universal  $C_T$  MLR model. Parameter Set represents the combination of parameters used to train and test the MLR; including phosphate (P), nitrate (N), silicate (Si), dissolved oxygen (O), salinity (S) and temperature (T).

Applying the same approach using parameters beyond salinity to capture  $A_T$  identified the optimal predictor combination to be T, S, S<sup>2</sup>, DO, P and Si.

The optimal parameter combinations identified for the universal models were additionally used when applying the ad-hoc MLR approach. It is important to note that the MLR routine (*Sect. 2.5.2*) has the ability to define an optimal sub-set of predictors to capture the local system, and can include interaction terms when necessary.

## 2.5.5 MLR results

When universally applying the traditional MLR on the ~33,000 global mixed-layer  $C_{\rm T}$  measurements, the statistical regression RSE is 15.1 µmol kg<sup>-1</sup> when using T, S, DO, P,

N and Si as predictors (Table 2.3). If applying the ad-hoc geographical and temporal separations, the statistical regression RSE reduces to 13.2  $\mu$ mol kg<sup>-1</sup>. However, when the independent test (SIT) is used to evaluate the regressions, errors increase to be 16  $\mu$ mol kg<sup>-1</sup> for the ad-hoc approach and 15.6  $\mu$ mol kg<sup>-1</sup> for the global regression. For  $A_T$ , optimal predictors were found to be T, S, S<sup>2</sup>, DO, P and Si, while a global MLR algorithm captured the signal to within 11  $\mu$ mol kg<sup>-1</sup> using the SIT approach. All empirical relationships for the global and ad-hoc MLR models can be found in *Appendix C*.

The MLR approach and results provides a framework to attempt to develop a better method that captures any potential non-linear biases that are contributing to errors of  $\pm 16 \ \mu\text{mol} \ \text{kg}^{-1}$  in  $C_{\text{T}}$  predictions, and  $\pm 11 \ \mu\text{mol} \ \text{kg}^{-1}$  for  $A_{\text{T}}$  on a global scale.

				RSE [µmol kg <sup>-1</sup> ]							
				Regression		Independent test (SIT)					
Region	Zone <sup>a</sup>	$N^{\mathrm{b}}$	N cruises <sup>c</sup>	Ad-hoc	Universal	Ad-hoc	Universal				
(a) <i>C</i> <sub>T</sub>											
Subtropical	1	5388	109	11.9	17.1	15.2	17.3				
Eq. Pacific	2	752	14	11.3	16.8	18.9	17.7				
North Atlantic	3	4626	69	13.2	15.5	15.5	16.2				
North Pacific	4	2344	112	17.7	17.2	16.8	17.5				
Southern Ocean	5	7856	75	12.5	12.4	16.4	12.8				
Global		20966	289	13.2	15.1	16.0	15.6				
(b) $A_{\mathrm{T}}$											
Subtropical	1	4917	94	10.2	10.2	11.0	10.4				
Eq. Pacific	2	513	7	6.9	12.4	9.4	13.0				
North Atlantic	3	3181	53	7.7	10.0	7.9	10.1				
North Pacific	4	1956	88	14.3	16.4	14.8	16.6				
Southern Ocean	5	6084	58	8.0	9.1	9.4	9.8				
Global		16651	224	9.5	10.8	10.4	11.1				

**Table 2.3:** Universal and ad-hoc MLR results for (a)  $C_{\rm T}$  and (b)  $A_{\rm T}$ .

<sup>a</sup> Corresponding geographical region in Fig. 2.7

<sup>b</sup> Number of measurements in the corresponding region

<sup>c</sup> Number of unique cruises/time series in the region
## 2.6 Neural network approach

The human brain is the most complex organ in the human body, acting as a non-linear information processing system. Its ability to perform complex tasks, such as pattern recognition, is attributed to the brains capacity to organize its processing units, known as neurons, in a highly structured manner through a learning process of its sounding environment (Haykin, 1999). From the inception of this recognition, numerous algorithms have been developed to mimic this learning process in solving highly non-linear problems. For geosciences, technological advances enabling scientists to collate databases of greater size and complexity are increasingly seeking neural networks to aid in the solution of complex and non-linear systems.

## 2.6.1 Overview of neural network approach

The self-organizing map (SOM) is a neural network algorithm which uses an iterative approach to classify multi-dimensional datasets into discrete groups, or neurons, usually arranged in a 2-dimensional grid (Kohonen, 1988). Using an algorithm that employs discrete clustering is appealing, as it removes the need for the type of ad-hoc data partitioning discussed in *Sect. 2.5.3*. This has led to application of SOMs in a wide range of disciplines (e.g., Abramowitz, 2005; Hales et al., 2012; Hsu et al., 2002; Pöllä et al., 2009).

Figure 2.10 illustrates the routine of SOM training and prediction. For a training dataset of *P* samples consisting of predictor variables *x* and response variable *y*, the SOM clustering process allocates each sample to one of *J* neurons (sometimes also called clusters, nodes or groups). The neurons are typically arranged in a 2 dimensional  $A \times B$  matrix so that a node is represented as  $j_{a,b}$ . The clustering algorithm aims to ensure that nodes that are nearby in this matrix contain samples that have similar values of the predictor variables *x*. The y = f(x) input-output mapping is then completed by performing a linear regression between *x* and *y* separately for each neuron.

These SOM and regression parameters can then be used to make predictions of y for an independent set of Q predictor samples  $(x_1, ..., x_q, ..., x_Q)$ . First, each  $x_q$  is allocated to a SOM neuron based on its similarity to the SOM weights from the training dataset. This is the "winning neuron" for a particular sample  $j(x_q)$ . Then the regression parameters for  $j(x_q)$  are used to predict  $y_q$ .

#### Training Phase





Here, two variants to this approach are explored. The first, as described above, uses a multiple-linear regression at each neuron, which is described here as selforganizing multiple-linear output (SOMLO). The second takes the mean of all response values belonging to a node, which is referred to here as the self-organizing map mean (SOMM). Both approaches are now described in more detail.

#### 2.6.2 Initialization of model constraints

For this implementation of the SOM algorithm, the input-output pairs  $(x_p, y_p)$ ,  $1 \le p \le P$  in the training dataset are some subset of x = (T, S, DO, N, Si, P), and  $y = C_T$  or  $A_T$ . To ensure each predictor variable has an equal opportunity to define the features of the SOM during the training routine, the variables were normalized (zero-mean) and scaled by their standard deviation so that their distribution and range are similar. For nitrate, phosphate and silicate, due to the exponential decay in their distribution of measurements from low to high concentrations, these variables were  $\log_{10}$  scaled.

The *J*-neuron SOM used here is structured in a hexagonal topology (Fig. 2.10). Careful consideration needs to be exercised when defining the size of the SOM, as too few neurons will not capture all important features, and too many will over-fit the training dataset. Each neuron  $(j_{a,b})$  is then assigned an initial weighting vector ( $\boldsymbol{\omega}$ ) of length equal to the number of input variables (i.e. the SHP), and whose values are randomly selected from the input variable range.

#### 2.6.3 SOM training routine

Once all the neuron weights have been initialized, training is an iterative process designed to cluster the *P* samples into *J* neurons. For each iteration step of the model  $(\tau)$ , the input data samples are individually presented to the SOM in a random order and the neuron whose weights are closest to the current input sample is declared the "winning neuron" for that sample, using

distance 
$$(\boldsymbol{x}_{p,}\boldsymbol{\omega}_{j}) = \left[\sum_{n=1}^{N} (x_{p,n} - \omega_{j,n})^{2}\right]^{0.5}$$
 (2.6)

That is, the "winning neuron",  $j(\mathbf{x}_p)$ , for sample  $\mathbf{x}_p$ , is simply the neuron that minimizes this distance. Once the winning neuron is established, the weights of the winning neuron, as well as the neurons in its topological neighbourhood, are adjusted towards the value of the current sample value  $(\mathbf{x}_p)$  via

$$\boldsymbol{\omega}_{j}(\tau+1) = \boldsymbol{\omega}_{j}(\tau) + h_{j,j(\boldsymbol{x}_{p})}(\boldsymbol{x}_{p} - \boldsymbol{\omega}_{j}(\tau))$$
(2.7)

In this expression,  $h_{j,j(x_p)}$  determines the extent to which a node's weight is brought closer to the current sample value (termed a 'learning rate',  $h \le 1$ ). It also determines the size of the neighbourhood around the winning node that receives a significant adjustment. Here, the learning rate is calculate via

$$h_{j,j(\boldsymbol{x}_p)} = \eta(\tau) \exp\left(-\frac{d_{j,j(\boldsymbol{x}_p)}}{2\sigma^2(\tau)}\right)$$
(2.8)

where,  $d_{j,j(\mathbf{x}_p)}$  represents the discrete distance in the SOM topology between the winning neuron  $j(\mathbf{x}_p)$  and an arbitrary neuron j, and  $\sigma^2(\tau)$  and  $\eta(\tau)$  are the neighbourhood width and learning rate respectively. As the model progresses through iterations,  $\sigma^2(\tau)$  ensures that the neighbourhood width shrinks from a value that significantly adjusts most of the neurons, to finish with only adjusting the winning neuron. Similarly, the learning rate  $\eta(\tau)$  decreases with iterations, so that regional features of the SOM gradually develop as iterations continue.

## 2.6.4 Supervised SOM training

The form of the model used here is known as a supervised SOM, whereby distributional information of the response parameter ( $C_T$  or  $A_T$ ) is used as an additional constraint beyond the hydrographic information (T, S, DO, etc.) in clustering the global dataset into the set of *J* neurons.

A supervised form of the SOM was first suggested by Kohonen (2001) and later developed by Melssen et al. (2006). In this approach, each neuron is assigned an initial weighting vector ( $\omega$ ) of length equal to number of predictor variables, as described above, together with a weighting number ( $\gamma$ ) that is randomly selected from within the response variable range. The distance measure then incorporates both the weighting vector and weighting number in establishing the winning neuron. For every iteration step ( $\tau$ ), each data sample ( $x_p, y_p$ ) is presented to the SOM model twice. In the first pass, the winning neuron is identified using

$$j(\boldsymbol{x}_{p}, y_{p}) = \min_{j} \left( (1 - \alpha(\tau)) \left[ \sum_{n=1}^{N} (x_{p,n} - \omega_{j,n})^{2} \right]^{0.5} + \alpha(\tau) |y_{p} - \gamma_{j}| \right)$$
(2.9)

where  $0 < \alpha(\tau) < 1$  is responsible for regulating the relative weight of the similarity measures. By initially setting  $\alpha(\tau)$  to 0.75, more weight is given to the response variable  $(y_p)$  in identifying the winning neuron. As  $\alpha(\tau)$  reduces linearly with iteration to 0.5, both similarity measures are given equal weighting in identifying the winning neuron. Once the winning neuron is established, the weighting vectors ( $\omega$ ) are updated using the same approach as presented in *Sect. 2.6.3*.

In the second pass, the winning neuron is determined using

$$j(\boldsymbol{x}_{p}, \boldsymbol{y}_{p}) = \min_{j} \left( \alpha \left( \tau \left[ \sum_{n=1}^{N} \left( x_{p,n} - \omega_{j,n} \right)^{2} \right]^{0.5} + \left( 1 - \alpha(\tau) \right) |\boldsymbol{y}_{p} - \boldsymbol{\gamma}_{j}| \right)$$
(2.10)

where more weight is now given to the predictor variables  $(x_p)$  in identifying the winning neuron. Once the winning neuron has been identified, weighting numbers  $(\gamma)$  are adjusting following the same routine presented in *Sect. 2.6.3*. By alternating the adjustment between the weighting vectors and weighting numbers the model converges on the optimal distribution of the global dataset into the set of *J* neurons.

### 2.6.5 Completing the input-output mapping

The y = f(x) is completed here in one of two ways. First, the mean of all output values  $(y_p)$  belonging to a node is used – the SOMM. Alternatively, MLRs are derived using the training data assigned to the winning neuron to establish this relationship (see Fig. 2.10). Here, MLRs are derived after the SOM training through the application of either a principal component regression (PCR; see *Appendix D* for details), or the forward stepwise robust MLR routine (see *Sect. 2.5.2*).

To ensure confidence in regression coefficients, a minimum threshold value of 10 times the number of predictor parameters was implemented. If the number of data points assigned to the winning neuron is below this threshold value, data from the second most similar neuron is merged with the winner, and then third, until the data pool reaches the threshold limit.

#### 2.6.6 Predicting with the SOMLO/SOMM models

For any independent input data vector  $(\mathbf{x}_q)$ , the output value  $(y_q)$  can then be predicted using the SOM trained above via a two-step process. First, determine which neuron each new data sample is closest to using the distance measure in *Sect. 2.6.3* (Eq. 2.6). Then the output value (of  $C_T$  or  $A_T$ ) is determined using either the mean value of the winning neuron's training output values (using the SOMM), or the regression parameters established with the training data.

## 2.7 Application to the global ocean

A two phase process was employed to converge on the optimal SOMLO configuration for the ocean carbon mixed-layer dataset. Firstly, three unique subsets of ocean carbon data were extracted to ascertain which hydrographic parameter combination worked best. In the second phase, the SIT approach was applied to make an out-of-sample assessment of the global skill of the model.

### 2.7.1 Defining optimal predictor parameters

Correlations between hydrographic parameters may lead to redundancy in the information predictor variables provide. To investigate the importance of each variable in informing the SOM or constraining the MLR, independent tests were performed that exclude the variables one at a time. These test the models ability to capture three unique independent datasets that each represents about 10% of the global carbon dataset

(Table 2.4). As an example, Fig. 2.11 presents the spatial distribution of the T1 independent dataset, which represents 11.4% of the global training dataset.

**Table 2.4:** Summary of the three independent datasets used to constrain the general configuration of SOMLO.

Indonandant datasat	Number of	Percentage of	
muependent dataset	measurements	global dataset	
T1	3769	11.4	
T2	2919	8.8	
Т3	3391	10.2	
Total	10079	30.4	



Figure 2.11: Distribution of measurements constituting the T1 independent dataset.

To explore the optimal SOM configuration, 800 iteration steps were used to train the SOM, using neuron map sizes ranging from 9 to 529 for every different input variable combination, with the ultimate aim to converge on the model with the lowest RSE. It's important to note that the same datasets were used to train and validate the model for each test. This is because some bottle samples do not contain the entire suite of dissolved oxygen and nutrient measurements, so by using the same datasets to train and test each model combination eliminates any potential bias in the independent testing of SOMLO. Salinity was found to be the most important parameter for capturing the mixedlayer carbon signal, followed by temperature, then nutrients (Fig. 2.12).



**Figure 2.12:** RSE values for the  $C_T$  (open triangles) and  $A_T$  (black circles) SOMLO models when applied to the three independent datasets. Numbers under the dotted lines are the optimal number of neurons to constrain the system. Excluded parameter represents the variable not used in the SOMLO training and testing; including phosphate (P), nitrate (N), silicate (Si), dissolved oxygen (DO), salinity (S) and temperature (T). The dashed lines represent the RSE values for the optimal models trained using all parameters.

The final optimal parameter set and SOM neuron size using the three independent tests were (SOPSi, 25) and (TSPO, 56) for the global  $A_T$  and  $C_T$  models respectively (Fig. 2.13). The SOMLO model using the PCR constrained the  $C_T$  system with a higher skill than the robust MLR, while  $A_T$  was better constrained using the robust MLR model.



**Figure 2.13:** Optimal residual standard error (RSE) values for (**a**)  $C_{\rm T}$  and (**b**)  $A_{\rm T}$  SOMLO models. Numbers above the line represent the optimal number of neurons, while Parameter Set represents the combination of parameters used to train and test the SOMLO model; including phosphate (P), nitrate (N), silicate (Si), dissolved oxygen (O), salinity (S) and temperature (T). The dashed line represents the RSE value for the optimal model trained using all parameters.

The addition of phosphate beyond temperature, salinity and dissolved oxygen improved the prediction of  $C_{\rm T}$  by ~27% or 5.1 µmol kg<sup>-1</sup> (Fig. 2.13). Without air-sea gas

exchange modulating its behavior, phosphate likely provides clearer constraints on organic matter production and respiration than dissolved oxygen alone. The redundancy of nitrate for both  $C_{\rm T}$  and  $A_{\rm T}$  (Fig. 2.13) is likely due to the near constant stoichiometric uptake rate of phosphate and nitrate by photosynthesizing organisms (Revelle and Suess, 1957). The preference of phosphate over nitrate may be the result of continual production of organic matter by nitrogen-fixers after the nitrate pool is completely depleted (Gruber and Sarmiento, 1997), or from the re-naming of samples where only 'nitrate + nitrite' was listed to nitrate in the GLODAP and CARINA products (Key et al., 2004).

Precipitation and dissolution of calcium carbonates (CaCO<sub>3</sub>) affects the concentration of  $A_{\rm T}$  twice as much as  $C_{\rm T}$  (Sarmiento and Gruber, 2006). As waters high in silicate tend to relate to high biological respiration by diatoms (a non-calcifying organism), and waters of low silicate foster a more conducive environment for calcifying organisms (such as coccolithophores) (Kirchman, 2012), silicate helps constrain the spatial patterns of CaCO<sub>3</sub> cycling which influences  $A_{\rm T}$ .

Salinity's significant importance in constraining  $A_{\rm T}$  is likely related to its influence on the charge difference between anions and cations in seawater (Zeebe and Wolf-Gladrow, 2001), leading to the known high correlation between these two parameters (Millero et al., 1998). In contrast, the addition of temperature to the parameter set was found to be redundant, as pointed out by some earlier studies (e.g. McNeil et al. (2007)).

## 2.7.2 Importance of geography

Carbon data from geographically diverse ocean regions will be clustered into the same neuron when input-output concentrations are similar. For example, a cluster of similar biogeochemical data in the North Atlantic Ocean can be equally represented by those in some parts of the North Pacific Ocean, despite there being little ocean interconnectedness between these two carbon datasets on shorter time-scales. Spatial lengthscales of variability are known to be within ocean basins, not between them, especially those constrained by land. Without applying geographical boundary conditions, non-linearities may be introduced into the final MLR which would limit the models predictive skill. To test this hypothesis, optimal model configurations were trained with the inclusion of geographical input parameters during the training of the SOM, but were excluded as predictor parameters in the linear regressions.

Global position representation through latitude and longitude can be problematic due to the mid-Pacific discontinuity of longitude at  $\pm 180^{\circ}$  and shortening of geographical distance between degrees of longitude towards the poles. As a measure to reduce the influence of longitudinal discontinuity at  $\pm 180^{\circ}$  in the mid-Pacific, all longitude values were shifted by 160° West (or 20° East), thereby setting the 180° discontinuity at a position that bisects continental Africa and Europe (Fig. 2.14).



Figure 2.14: Global map after longitude values were shifted by 160° West.

A normal vector to the Earth ellipsoid (n-vector) was also tested that eliminates both issues mentioned above. The n-vector transforms the 2-D latitude/longitude position system into a 3-D vector whilst maintaining unique vectors for every geographical position. Latitude and longitude values were transformed here using a version of the n-vector presented by Gade (2010)

$$\mathbf{n} = \begin{bmatrix} \sin(\operatorname{latitude}) \\ \sin(\operatorname{longitude})\cos(\operatorname{latitude}) \\ \cos(\operatorname{longitude})\cos(\operatorname{latitude}) \end{bmatrix}$$
(2.11)

The introduction of geographical information was found to be a powerful addition in improving the global skill for  $C_{\rm T}$  by 16% or 2.2 µmol kg<sup>-1</sup>, however there was little improvement for  $A_{\rm T}$  (Fig. 2.15). The optimal SOMLO configuration additionally incorporates longitude and n-vector geographical inputs in constraining  $A_{\rm T}$  and  $C_{\rm T}$  respectively, and increased the optimal number of neurons for  $C_{\rm T}$  to 64.



**Figure 2.15:** Skill of optimal SOMLO models with geographical constraints. Lat and Lon represent latitude and longitude respectively, while n-vector represents the normal vector to the Earths ellipsoid as defined by Gade (2010). Numbers below dashed line represent the optimal number of neurons.

To better understand and visualize why geography is important, the spatial distribution of neurons is compared for  $C_{\rm T}$  models trained with only biogeochemical information, and both biogeochemical and geographical information (Fig. 2.16). To

illustrate the spatial distribution of the assigned neurons for the global carbon dataset, neurons are represented as different colours. Here, each colour represents a neuron, while shades of colour indicate close similarity in the weighting vectors. The broad regions of similarity that are captured when the SOM is constrained by only biogeochemical properties include the Southern Ocean, subtropical gyres, North Pacific and North Atlantic (Fig. 2.16a). However, these ocean "fingerprints" extend beyond the known spatial length scales, for example linking features in the Southern Ocean to that of the North Atlantic, while zonal bands stretch across ocean basins (Fig. 2.16a). When biogeochemical and geographical information are incorporated into the SOM training routine, the resulting distribution preserves the neuron boundaries at known frontal zones, such as the subtropical convergence zone, but is able to constrain the classification of data to be within each ocean basin (Fig. 2.16b). Using geography is an important additional constraint that implicitly shortens the length scales of variability which dominate seasonal mixed-layer dynamics in the ocean. It is important to note that the addition of geography did not alter the optimal parameter set for the technique.



**Figure 2.16:** Distribution of assigned neurons for optimal  $C_T$  SOM models trained with (a) biogeochemical information only and (b) biogeochemical and geographical information. Measurements of the same colour were assigned to the same neuron, while shades of colours represent close similarity of neuron weighting vectors.

#### 2.7.3 SOMM versus SOMLO: Importance of the local linear optimizer

To explore the predictive ability of the SOMM model, optimal model configurations were tested using the three independent datasets but with neuron sizes extending up to 2500 (Fig. 2.17). Using all data, the SOMM model converged on an RSE value of 16  $\mu$ mol kg<sup>-1</sup> in constraining  $C_{\rm T}$ . Although the SOMM is powerful in constraining complex non-linear systems, its ability to accurately predict independent data is dependent on the information coverage provided by the training dataset. If the information provided to SOM by the training dataset does not cover a large portion of the systems variability and drivers, then the predictive skill of SOMM will suffer. Since the global carbon system is highly variable in space and time, the current spatiotemporal deficiencies in the global bottle dataset limits the ability of SOMM to accurately predict  $C_{\rm T}$  on a global scale. Using a local multiple-linear optimizer in addition to the global SOM (i.e. the MLR), is shown here to significantly improve the model's ability to constrain global  $C_{\rm T}$  by ~27%, or 4.4  $\mu$ mol kg<sup>-1</sup>. Similar findings were found for the  $A_{\rm T}$  model.



Figure 2.17: Skill comparison between the SOMLO (open triangles) and SOMM (closed triangles) models in capturing  $C_{\rm T}$ . The dashed lines represent the lowest RSE values for to two models.

Previous empirical approaches to capture ocean surface  $pCO_2$  using the "underway"  $pCO_2$  measurement network have either employed a SOM or MLR model (e.g., Arrigo et al., 2010; Chierici et al., 2009; Telszewski et al., 2009). Specifically, SOM based approaches first cluster a combination of re-analyzed fields of sea surface temperature (SST), salinity (SSS), and satellite-based estimates of Chl-a and MLD. "Underway" pCO<sub>2</sub> measurements are then co-located with Chl a, SST, SSS and MLD values and assigned to a neuron using the same Euclidean distance measure as presented in Sect. 2.6.3. Finally, ocean surface  $pCO_2$  distributions are constrained by taking the mean of all "underway" measurements assigned to each neuron (e.g., Friedrich and Oschlies, 2009a; Telszewski et al., 2009). Conversely, studies based on a MLR approach first derive linear relationships between "underway" pCO<sub>2</sub> and a combination of co-located Chl a, SST, SSS and MLD parameters, and then predict basin-wide  $pCO_2$ maps (e.g., Chen et al., 2011; Chierici et al., 2009; Lefèvre and Taylor, 2002). Combining both the SOM and MLR technique, as done here, is shown to significantly improve the predictive skill of the global bottle-based model. Applying the SOMLO technique to "underway"  $pCO_2$  measurements could therefore provide a more accurate approach to predict ocean surface  $pCO_2$ .

## 2.7.4 Using different neural network algorithms

Although Kohonens SOM algorithm has been found well suited for geophysical problems, there exists a suite of other neural network algorithms, such as k-means, back-propagation and radial-basis function (Haykin, 1999), that may capture the oceans inorganic carbon system with a higher degree of accuracy.

The *k*-means approach clusters any  $M \times N$  dataset into *k* groups, such that the sum of squares from all data samples assigned to each group is minimized (Hartigan and Wong, 1979). During early stages of this PhD work, a comparison test between the supervised-SOM and *k*-means algorithms found that accuracy in the SOM-based approach is slightly better than *k*-means estimates when applied to the bottle-carbon 62

dataset. Based on these early results, the SOM algorithm was further developed to incorporate the MLR routines.

While it's important to pursue novel empirical methods to better constrain the system, the presence systematic and random sources of error in historical bottle measurements will cap empirical accuracy. Precision in historical  $A_{\rm T}$  and  $C_{\rm T}$  measurements is estimated to be between 5-10 µmol kg<sup>-1</sup> prior to the early 1990's, 4-5 µmol kg<sup>-1</sup> after the introduction of certified reference materials by Dickson (1990c), and is currently less than 2 µmol kg<sup>-1</sup> (Key et al., 2010). Since 68% of samples in the global training dataset were collected during the period 1990 to 2003, with only 10% prior to 1990, the benchmark for empirical carbon predictions is estimated here to be ~4-5 µmol kg<sup>-1</sup>. However, this estimate is based on perfect SHP measurements, whereas accuracy in nutrient measurements has been typically no greater than 2% (Key et al., 2010).

## 2.8 Evaluating the empirical approaches

## 2.8.1 Measuring the improvement over traditional MLR

To evaluate the skill of the two independent approaches used here (MLR versus SOMLO), the global SIT predictions for each technique were divided into 5 geographical regions and evaluated globally (Table 2.5). The SOMLO approach was found to improve the predictive skill of  $C_{\rm T}$  by between 11-30% in all 5 regions (Table 2.5). In particular, known complex dynamical regions with global CO<sub>2</sub> importance like the equatorial Pacific, Southern Ocean and North Atlantic are where the non-linear SOMLO approach excelled, improving  $C_{\rm T}$  prediction by between 23-30% (or 4-6 µmol kg<sup>-1</sup>). From a global point of view, SOMLO improves the predictive skill of  $C_{\rm T}$  in the mixed-layer by 19%.

For  $A_{\rm T}$ , the benefits of using SOMLO are much weaker, with only a marginal global improvement by 6.7% (or 0.7 µmol kg<sup>-1</sup>) and even deterioration of detection in the equatorial Pacific and North Atlantic. This is likely related to the known high linear

relationship between alkalinity and salinity (e.g. Millero et al. (1998)), which limits the benefits of SOMLO, since it better constrains more complex non-linear systems.

		RSE <sup>a</sup> [µmol kg <sup>-1</sup> ]				
Region	Zone <sup>b</sup>	$N^{ m c}$	Ad-hoc MLR	SOMLO	% Improvement	
(a) $C_{\mathrm{T}}$						
Subtropical	1	5388	15.2	13.5	11.2	
Eq. Pacific	2	752	18.9	13.3	29.7	
North Atlantic	3	4626	15.5	11.7	24.5	
North Pacific	4	2344	16.8	14.3	14.9	
Southern Ocean	5	7856	16.4	12.7	22.6	
Global		20966	$16.0(15.6)^{d}$	12.9	19.4 (17.4) <sup>d</sup>	
$(\mathbf{b}) \mathbf{A}_{\mathbf{T}}$						
Subtropical	1	4917	11.0	9.2	16.4	
Eq. Pacific	2	513	9.4	9.6	-2.1	
North Atlantic	3	3181	8.0	8.5	-6.3	
North Pacific	4	1956	14.8	14.4	2.7	
Southern Ocean	5	6084	9.4	8.8	6.4	
Global		16651	$10.4(11.1)^{d}$	9.7	$6.7(12.6)^{d}$	

**Table 2.5:** Skill comparison between the traditional MLR and SOMLO approaches for (a)  $C_{\rm T}$  and (b)  $A_{\rm T}$ .

<sup>a</sup> Calculated using the SIT predictions

<sup>b</sup> Corresponding geographical region in Fig. 2.7

<sup>c</sup> Number of measurements

<sup>d</sup> Universal MLR

## 2.8.2 SOMLO regional error assessment

Plotting the distribution of global SIT residual errors permits investigation into the spatial skill of the SOMLO model (Fig. 2.18). Although the Arctic Ocean, Bay of Bengal and Sea of Okhotsk are regions not well constrained by the novel technique, the majority of the ocean maintains a relatively homogenous residual error range (Fig. 2.18). These unconstrained regions are either coastal or marginal seas with known locally complex biogeochemical regimes, so it is understandable that a trained global open-ocean technique will poorly constrain these local regions.



**Figure 2.18:** Distribution of global SIT residual errors (i.e. predicted – in situ) for optimal (a)  $C_{\rm T}$  and (b)  $A_{\rm T}$  SOMLO models, given in µmol kg<sup>-1</sup>. Error values were interpolated around in situ locations using VG gridding software of Ocean Data Viewer (Schlitzer, R.: Ocean Data View, http://odv.awi.de, 2011).

Further scrutinizing the 395 samples with a SIT residual error greater than  $\pm 50$  µmol kg<sup>-1</sup> for  $C_{\rm T}$  and/or  $A_{\rm T}$  revealed that 70% (277) are located within 300 km of a major coastline (Fig. 2.19). Since a study by Gibbs et al. (2006) identified terrestrial influences extend up to 345 km from land, and well beyond the bathymetric defined coastal ocean limit of 500m used here, these anomalous independently predictions are likely the result of land-ocean interactions affecting the carbon and SHP concentrations.



**Figure 2.19:** Geographical distribution of the 277 samples with a SIT residual error greater than  $\pm 50 \text{ }\mu\text{mol kg}^{-1}$  for  $C_{\text{T}}$  and/or  $A_{\text{T}}$  and are located within 300 km of a major coastline.

Separating the SIT predictions into 14 geographical regions (see *Appendix E*) and excluding these anomalous coastal samples then provides the most accurate estimate of SOMLOs regional open-ocean skill (Table 2.6). This reconfirms that the Arctic Ocean and Bay of Bengal are the two regions were the model's skill is poorest. By finally excluding all Arctic Ocean measurements (North of 70°N), the estimate for the global open-ocean accuracy for SOMLO  $C_{\rm T}$  and  $A_{\rm T}$  predictions is 10.9 and 9.2 µmol kg<sup>-1</sup> respectively.

		RS	5E <sup>b</sup>	N <sup>c</sup>	
Region	Zone <sup>a</sup>	$C_{\mathrm{T}}$	$A_{\mathrm{T}}$	$C_{\mathrm{T}}$	$A_{\mathrm{T}}$
Arctic Ocean	1	26.6	22.1	782	795
Sup-Polar North Atlantic	2	11.6	9.0	4425	2641
Subtropical North Atlantic	3	9.1	6.6	1481	1254
Equatorial Atlantic	4	13.7	13.0	654	582
Subtropical South Atlantic	5	10.6	8.7	659	551
Sub-polar North Pacific	6	11.2	147	2053	1615
Subtropical North Pacific	0 7	11.2	8 2	2055	1/1/6
Equatorial Desifie	0	11.1	0.2 9.2	1524	1440 00 <b>2</b>
	0	11.2	8.5	1324	802 1404
Subtropical South Pacific	9	12.3	1.1	1824	1404
Subtropical North Indian	10	22.1	13.4	143	168
(Exc. Bay of Bengal)		(13.9)	(7.5)	(111)	(136)
Equatorial Indian	11	11.8	7.7	512	500
Subtropical South Indian	12	11.5	5.6	1411	1388
Southorn Occor	12	07	00	2050	2000
Southern Ocean	15	8.7	8.8	3950	3088
Sub-Antarctic waters	14	9.5	8.5	2250	14/4
Global		11.8	10.2	24035	17708
Global (below 70°N)		10.9	9.2	23253	16913

Table 2.6: Regional and global SOMLO skill evaluation (see Appendix E for map of spatial division).

<sup>a</sup> For a plot of corresponding geographical regions see *Appendix E* <sup>b</sup> Residual Standard Error [μmol kg<sup>-1</sup>] <sup>c</sup> Number of measurements in the region.

To investigate skewness and bias in SOMLO predictions, global SIT predictions where plotted against their corresponding in situ measurement and the distribution of residual errors examined (Fig. 2.20). For both  $C_{\rm T}$  and  $A_{\rm T}$ , skewness is limited  $(r^2 \ge 0.97)$ , while SOMLO predictions are globally biased by 0.02 and 0.15 µmol kg<sup>-1</sup> for  $C_{\rm T}$  and  $A_{\rm T}$ , respectively. These results provide additional confidence in SOMLO's ability to accurately predict  $C_{\rm T}$  and  $A_{\rm T}$  concentrations for any given set of temperature, salinity, dissolved oxygen, silicate, and phosphate measurements in the global openocean mixed-layer.



**Figure 2.20:** (a) and (c) Global SIT predictions versus in situ measurements for optimal  $C_{\rm T}$  and  $A_{\rm T}$  SOMLO models respectively. (b) and (d) Residual error density distributions for optimal  $C_{\rm T}$  and  $A_{\rm T}$  models respectively.  $r^2$  = r-squared correlation and N = number of samples.

#### 2.8.3 Is there a temporal bias in SOMLO predictions?

The influence of seasonal data bias is investigated via partitioning the global SIT predictions into seasons and calculating the RSE values (Fig. 2.21; Table 2.7). No strong seasonal bias was found on a global scale (Fig. 2.21). On a regional scale, wintertime samples in the Southern Ocean (south of 44°S) represents only ~10% of measurements within this region, yet SOMLO is able to predict wintertime samples with an accuracy of  $\pm 7.7 \mu$ mol kg<sup>-1</sup> for both  $C_{\rm T}$  and  $A_{\rm T}$  (Table 2.7).



**Figure 2.21:** RSE values for optimal SOMLO  $C_T$  and  $A_T$  SIT predictions partitioned by season, where seasons are defined as Summer (December to February), Autumn (March to May), Winter (June to August) and Spring (September to November) for Southern Hemisphere data. Northern Hemisphere seasons differ by 6 months.

**Table 2.7:** RSE values of SIT predictions partitioned into 5 regions and by season. Seasons are defined as summer (December to February), autumn (March to May), winter (June to August) and spring (September to November) for Southern Hemisphere data. Northern Hemisphere seasons differ by 6 months.

		RSE $[\mu mol kg^{-1}]$ (N <sup>a</sup> )				
Region	Latitudinal	Summer	Autumn	Winter	Spring	Net N
	band					
			(a) <i>C</i> <sub>T</sub>			
High North	$44^{\circ}N - 70^{\circ}N$	12.70 (1323)	9.57 (1322)	12.45 (704)	10.54 (971)	4320
Temperate	$18^{\circ}N - 44^{\circ}N$	13.49 (1434)	10.40 (1335)	9.16 (1448)	11.39 (1201)	5418
North						
Tropical	$18^{\circ}\mathrm{S} - 18^{\circ}\mathrm{N}$	10.39 (1224)	14.28 (928)	11.18 (947)	11.86 (1324)	4423
Temperate	$18^{\circ}S - 44^{\circ}S$	11.90 (1037)	11.42 (464)	13.54 (774)	9.73 (1369)	3644
South						
Southern Ocean	South of 44°S	9.78 (1851)	9.58 (1475)	7.74 (585)	7.97 (1537)	5448
Global	Below 70°N	11.63 (6869)	10.84 (5524)	10.85 (4458)	10.27 (6402)	23253
			(b) <i>A</i> <sub>T</sub>			
High North	$44^{\circ}N - 70^{\circ}N$	11.26 (1026)	13.67 (768)	17.69 (304)	11.09 (524)	2622
Temperate	18°N – 44°N	8.64 (1216)	8.91 (970)	7.16 (922)	7.62 (904)	4012
North						
Tropical	18°S – 18°N	8.19 (771)	9.55 (769)	9.20 (579)	9.70 (1155)	3274
Temperate	$18^{\circ}S - 44^{\circ}S$	8.42 (896)	8.17 (387)	6.94 (542)	5.85 (1185)	3014
South						
Southern Ocean	South of 44°S	8.40 (1383)	8.47 (1356)	7.72 (387)	10.12 (869)	3995
Global	Below 70°N	9.04 (5292)	9.86 (4250)	9.36 (2734)	8.75 (4637)	16913
2.2.1.0						

<sup>a</sup> Number of measurements

## 2.8.4 Significance of anthropogenic C<sub>T</sub> corrections

To test the significance of the applied anthropogenic  $C_{\rm T}$  corrections, the optimal  $C_{\rm T}$  model was trained and tested using measurements that were not corrected for anthropogenic CO<sub>2</sub> uptake. Applying the SIT approach globally, revealed a global predictive skill of 13.2 µmol kg<sup>-1</sup>, which is ~26% higher than the optimal model trained with corrected  $C_{\rm T}$  measurements (10.9 µmol kg<sup>-1</sup>). This difference of 2.3 µmol kg<sup>-1</sup> between the two approaches indicates that anthropogenic corrections have a low impact on the models ability to predict global  $C_{\rm T}$ . Despite this, the importance of accounting for anthropogenic CO<sub>2</sub> uptake is illustrated by partitioning the SIT predictions by year and calculating the difference in RSE values between the models trained and tested using corrected (RSE<sub>(yr; corrected</sub>)) and non-corrected  $C_{\rm T}$  measurements (RSE<sub>(yr; not corrected</sub>)):

$$\Delta RSE_{(yr)} = RSE_{(yr;non-corrected)} - RSE_{(yr;corrected)}$$
(2.12)

where yr represents year and spans the global dataset range (i.e. 1981 to 2010). The positive and increasing  $\Delta RSE_{(yr)}$  as year diverges from the normalized year of 2000 indicates that anthropogenic adjustments directly enhances the models global skill (Fig. 2.22). Although this result does not advocate that the applied corrections were globally accurate, it does verify the importance of accounting for anthropogenic CO<sub>2</sub> uptake when deriving empirically relationships for  $C_{\rm T}$  using measurements collected over several decades.



**Figure 2.22:** Annual RSE difference between  $C_{\rm T}$  models trained and tested using measurements that were either correct or non-corrected for anthropogenic CO<sub>2</sub> uptake (defined by:  $\Delta RSE_{(yr)} = RSE_{(yr; non-corrected)} - RSE_{(yr; corrected)}$ ). The positive and increasing  $\Delta RSE_{(yr)}$  as year diverges from the normalized year of 2000 indicates that anthropogenic corrections improve the predictive skill of SOMLO.

#### 2.8.5 How effective is a bathymetric approach for identifying coastal data?

Identifying and removing samples influenced by terrestrial processes is an important aspect when deriving empirical relationships for open-ocean regions. To evaluate the appropriateness of identifying coastal data under a bathymetric depth approach, RSE values were calculated for near-coast (within 300 km of a major coastline) and open-ocean measurements using the global SIT predictions, but excluding the 298 measurements already suspected as terrestrially influenced and data North of 70°N.

Accuracy in SOMLOs capacity to predict open-ocean samples is 14.4% better than near-coast samples for  $A_{\rm T}$ , and 10.9% for  $C_{\rm T}$  (Table 2.8). This result indicates that identifying coastal samples under a bathymetric depth approach may not be effective in ocean regions where coastal biogeochemical processes and terrestrial influences are not coupled to a shelf break, but are rather dependent on biotic distributions, as shown by Gibbs et al. (2006). Future attempts to identify coastal measurements should therefore not solely rely on bathymetric depth.

 Table 2.8: SOMLO skill comparison between near-coast and open-ocean regions.

Model	Near-coast	Open-ocean	% difference			
$C_{\mathrm{T}}$	11.9 (4338)	10.6 (18875)	10.9			
$A_{\mathrm{T}}$	10.4 (2856)	8.9 (14014)	14.4			
Residual Standard Error [µmol kg <sup>-1</sup> ]						

<sup>b</sup> Number of in situ measurements

#### 2.8.6 Are the neurons capturing the system?

Optimal model configurations established here may be biased to the three independent subsets that were used to constrain the optimal parameter combination and SOM neuron size (see *Sect. 2.7.1*). To ensure the optimal SOM model captures all important features within the global carbon system, and therefore minimizes any potential influence due to grouping biases, the SIT approach was applied globally using the optimal parameter combinations but with an increases in the optimal SOM neuron size (Table 2.9). This

essentially evaluates the models global skill when the data is partitioned into more neurons than was optimal when tested on the three independent subsets.

**Table 2.9:** Global RSE values  $[\mu mol kg^{-1}]$  computed using model predictions under the optimal configuration and with two increases in SOM neuron size. The small differences indicate that all important features are captured under the optimal model configuration.

	$C_{\rm T}$ model		$A_{\rm T}$ model		
	Number of neurons	RSE	Number of neurons	RSE	
Optimal	64	12.45	25	9.78	
Step 1	72	12.59	30	10.16	
Step 2	81	12.82	36	10.28	

The increase in global RSE values by 0.1 to 0.4  $\mu$ mol kg<sup>-1</sup> for each step in SOM neuron size (Table 2.9) indicates that all important features were captured when using the three independent datasets, and more importantly, that the optimal SOM configuration defined by the three independent test-sets remains valid on a global scale.

## 2.8.7 SOMLO model without Arctic data

Uniqueness in parameter concentrations within the Arctic region (North of 70°N), in particular that of salinity due to intense freshening of the water body, results in the classification of Arctic measurements into features that are near exclusive to the region (Figure 2.23). This suggests Arctic measurements have little influence in constraining the remaining global samples. This facet is particularly important since residual errors in the Arctic are ~100% higher than all other regions (see Table 2.6).



**Figure 2.23**: Distribution of measurements assigned to a neuron containing at least one sample collected within the Arctic region (above 70°N). Numbers represent the neuron each measurement was assigned to (maximum of 64).

To evaluate the influence of Arctic samples on the models global predictive skill, the optimal SOMLO model was trained and tested (SIT) using a global dataset that excluded Arctic Ocean data (samples North of 70°N). Comparison between the global RSE values of this model, and the same samples predicted under the optimal 'all data' model, revealed that SOMLOs ability to predict samples below 70°N differed by 0.1% and 2% between the two  $C_{\rm T}$  and  $A_{\rm T}$  models respectively (Table 2.10). This very small difference confirms that Arctic samples present an insignificant influence on the models capacity to predict carbon samples below 70°N, and that no bias exists in comparing the skill of the global SOMLO model to the universal MLR, which excluded Arctic data when fitting and testing the regression.

	RSE [µ		
	Model with Arctic data	Model without Arctic data	% difference
$C_{\mathrm{T}}$	12.45	12.44	0.1%
$A_{\mathrm{T}}$	9.71	9.9	2%

**Table 2.10:** RSE values of SIT predictions below 70°N, where the optimal SOMLO configurations were trained and tested using a global dataset either with or without Arctic data.

## 2.8.8 Stochastic nature of the SOM

Initialization of the SOM neuron weights is a stochastic process (See *Sect. 2.6.2*), and can therefore lead to predictions that are not reproducible. In this study, this facet is dampened by the small neuron to training dataset ratio (1:375 for  $C_T$  and 1:700 for  $A_T$ ), and the use of 800 training iteration steps, which should converge on a similar grouping of measurements when the model is trained under static conditions (i.e. same parameter combination, number of neurons and training dataset).

As a test to explore the influence of stochastic initialisation, the optimal SOMLO model was trained and tested 100 times using the three independent subsets, and then the RSE values examined for reproducibility (Table 2.11). The small 1<sup>st</sup> standard deviation of 0.2  $\mu$ mol kg<sup>-1</sup> (or 1.6%) around the mean RSE for  $C_{\rm T}$  demonstrates reproducibility in SOMLO predictions, and therefore a negligible influence due to the stochastic SOM initialization.

**Table 2.11:** Mean and first standard deviation in RSE values of the three independent subsets (see *Sect. 2.6.2*) when tested 100 times each. The small 1<sup>st</sup> deviation indicates negligible influence due to the stochastic SOM initialization.

Model	Mean RSE [µmol kg <sup>-1</sup> ]	1 <sup>st</sup> Standard Deviation [μmol kg <sup>-1</sup> ]	% of mean
$C_{\mathrm{T}}$	12.2	0.2	1.6%
$A_{\mathrm{T}}$	8.2	0.1	1.2%

# 2.9 Application to the Bermuda Atlantic and Hawaiian Ocean time-series stations

The SOMLO model was trained using a global  $C_{\rm T}$  and  $A_{\rm T}$  dataset that consisted mostly of sporadic one-time cruises in time. To test how well seasonal to inter-annual variability is captured using the new technique, the carbon time-series data from the BATS and HOT stations were used as an independent test-bed.

### 2.9.1 Predicting the North Atlantic seasonal cycle for inorganic carbon (BATS)

The BATS hydrographic site is a high frequency measurement program of carbon and auxiliary parameters that has been ongoing since 1989 (Bates, 2007). Located in the Sargasso Sea at 31°43'N and 64°10'W (see Fig. 2.25), a strong seasonal cycle exists at BATS that is characterized by warm surface temperatures in summertime increasing stratification, while cooler temperatures during the wintertime enhances the mixed-layer depth, and therefore delivery of nutrient and  $C_{\rm T}$  enriched deep waters to the surface (Bates et al., 1996). To test SOMLOs capacity to reconstruct the BATS seasonal cycle, the global algorithm was first re-trained without using the BATS 1989-2007 carbon time-series dataset, and then the measured monthly hydrographic properties between 1987-2007 were used to independently predict  $C_{\rm T}$  and  $A_{\rm T}$  concentrations at the BATS site. Finally, the independent carbon predictions were compared to the in situ measurements to investigate the temporal skill of the technique. The BATS  $C_{\rm T}/A_{\rm T}$  values were also independently predicted using the traditional MLR approach as a further test.

Figure 2.24a-b shows the measured versus predicted  $C_{\rm T}$  and  $A_{\rm T}$  annual cycles at BATS. Within the uncertainty of the SOMLO prediction, both the magnitude and structure of the seasonal  $C_{\rm T}$  cycle at BATS is well constrained, capturing 90% of the signal (Fig. 2.24a, Table 2.12). For a global MLR approach, the seasonal cycle is overestimated significantly by ~50%, while the ad-hoc approach under-predicts the

seasonal cycle by ~20  $\mu$ mol kg<sup>-1</sup>. For  $A_{\rm T}$ , the small seasonality is captured by both techniques (Fig. 2.24b).



**Figure 2.24:** BATS in situ and independently predicted seasonal cycles for (a)  $C_{\rm T}$  and (b)  $A_{\rm T}$ . Black dots represent in situ measurements and blue shaded region represents the uncertainty in SOMLO predictions.

To gain a better insight into how the SOMLO substantially improves the prediction of the BATS seasonal cycle from the traditional MLR analysis, the neuron distribution for  $C_{\rm T}$  measurements in the North West Atlantic is examined (Fig. 2.25). Applying a traditional ad-hoc MLR analysis requires defining somewhat subjective longitude and latitude boundaries for the data to be used in the linear regressions. Here, as an illustration, the spatial boundaries of 30°N to 70°N and 40°W to 85°W that were also used by Lee et al. (2000b) in their MLR approach are presented. The traditional MLR explicitly uses all carbon data within the prescribed region, whilst the SOMLO approach partitions the data into neurons without any prior geographic constraints. The benefit in this approach is that when SOMLO is applied to a new dataset (in this case BATS) the SOM only uses neurons (data) most consistent with its "biogeochemical fingerprint", and therefore reduces the potential bias that would be introduced from including all data in the regression.



**Figure 2.25:** Distribution of assigned neurons in the North-west Atlantic region for optimal  $C_{\rm T}$  SOMLO model (30°N to 70°N and 40°W to 85°W). Numbers represent the neuron each measurement was assigned to (maximum of 64).

#### 2.9.2 How well does SOMLO capture inter-annual signals?

Inter-annual variability of  $C_{\rm T}$  at BATS is captured to within the uncertainty of the SOMLO technique over the 18 year period (Fig. 2.26). This illustrates a new and powerful way to diagnose year-to-year carbon variability in the ocean by using the many more long-term hydrographic time-series that are available in the ocean (McNeil, 2010). To further test the SOMLO approach in capturing inter-annual variability, the  $C_{\rm T}$  signal at the HOT time-series as reported by Brix et al. (2004) was independently predicted. The SOMLO predictions capture the smoothed inter-annual trend-line at the HOT site to within 85% (Fig. 2.27, Table 2.12).

The BATS and HOT comparisons provide additional confidence that the SOMLO approach provides good constraints on both seasonal and inter-annual variability for  $C_{\rm T}$ . This illustrates the potential of the new technique to be used on a wider scale to help understand the oceans role in modulating atmospheric CO<sub>2</sub>.

**Table 2.12:** Statistical summary of independent predictions at the BATS and HOT stations. Including; RSE of independent predictions, correlation between predictions and in situ measurements, and variance in measured values.

Site	RSE	Correlation	Variance in measured
	[µmol kg <sup>-1</sup> ]	(Pearson)	data $[(\mu mol kg^{-1})^2]$
		$\mathcal{C}_{T}$	
BATS	7.4	0.90	232
HOT	10.0	0.72	180
		A <sub>T</sub>	
BATS	6.9	0.77	108
HOT	7.2	0.90	195







**Figure 2.27:** In situ and independently predicted HOT  $C_{\rm T}$  measurements with a loess (locally weighted scatterplot smoothing) line fit.

# 2.10 Comparison to previous techniques

It's important to emphasize that reported error estimates of previous empirical studies are calculated from the regressions residual error rather than independent tests as done here (see Table 2.1). This means direct comparison between previous studies and results presented here is not valid. The systematic independent test approach (see *Sect. 2.4*) was therefore used to accurately report the differences between SOMLO results and previous traditional MLR approaches.

Two sets of calculations were conducted as shown in Table 2.13. The first set of calculations (MLR<sub>old</sub>) involved taking the regressions from a suite of prior work (Bates et al., 2006; Lee et al., 2000b, 2006; McNeil et al., 2007) and applying them to the new larger dataset within each region. The second set of calculations (MLR<sub>new</sub>) involved developing a new set of regressions using the same geographical and temporal boundaries and predictor variables as the previous authors within the much larger dataset. Using the SIT predictions, the skill of the models were calculated (RSE) and could then be directly compared to SOMLO results (see Table 2.13).

The SOMLO, as shown at BATS and HOT, improved the predictive skill of  $C_{\rm T}$  and  $A_{\rm T}$  in most regions by between 10-40% relative to the new set of regressions, and by 10-48% compared to previous regressions (Table 2.13). Globally, the SOMLO reduces the error in predicted  $C_{\rm T}$  by 28% beyond the MLR method used to conduct the only global analysis (Lee et al., 2000b). For  $A_{\rm T}$ , the new technique reduces the global error by 9% relative the MLR regressions presented by the only other global empirical study (Lee et al., 2006).

**Table 2.13:** Skill comparison between optimal SOMLO models and previous MLR approaches.  $MLR_{old}$  involves applying a suit of regressions from previous work to the new larger dataset,  $MLR_{new}$  involved developing a new set of regressions using the same ad-hoc partitioning approaches as previous authors, finally, SOMLO is the skill of the optimal SOMLO models. *N* represent the number of measurements used to calculate the RSE values, and % the percentage improvement between the MLR<sub>old/new</sub> and SOMLO approaches.

Study			R	RSE [µmol k	g <sup>-1</sup> ]	% diff	erence <sup>d</sup>	
Region	Model	Ν	MLR <sub>old</sub>	MLR <sub>new</sub>	SOMLO	MLR <sub>old</sub>	MLR <sub>new</sub>	Author
Global <sup>a</sup>	$C_{\mathrm{T}}$	13881	22.0	17.8	12.8	42	28	(Lee et al., 2000b)
Indian Ocean <sup>b</sup>	$C_{\mathrm{T}}$	2052	15.2	21.4	13.0	14	39	(Bates et al., 2006)
Southern Ocean	$C_{\mathrm{T}}$	4196	17.3	8.8	9.0	48	-2	(McNeil et al., 2007)
Global (exc. North	$A_{\mathrm{T}}$	10360	11.7	10.9	10.7	9	2	(Lee et al.,
Pacific) <sup>c</sup>		(8995)	(10.3)	(10.4)	(9.9)			2006)
Indian Ocean <sup>b</sup>	$A_{\mathrm{T}}$	2042	9.4	11.8	7.1	24	40	(Bates et
Southern Ocean	$A_{\mathrm{T}}$	4196	10.3	10.3	9.3	10	10	al., 2006) (McNeil et al., 2007)

<sup>a</sup> Using only surface data (above 30m)

<sup>b</sup> Only measurements from within the mixed-layer defined here were used to constrain new and test previous regressions.

<sup>c</sup> The North Pacific empirical regression of Lee et al. (2006) included an interaction term between temperature and longitude. Here, longitude values were taken to range from  $0^{\circ}$ -360°.

<sup>d</sup> Calculated by ((MLR<sub>old (or new)</sub> – SOMLO) / MLR<sub>old (or new)</sub>)×100

# 2.11 Diagnosing global $C_{T}$ and $A_{T}$ distributions

Large historical and recent datasets up until 2008 of temperature, salinity, dissolved oxygen and nutrients has allowed researchers to objectively interpolate global monthly  $1^{\circ} \times 1^{\circ}$  resolution climatologies – the World Ocean Atlas 2009 project (WOA09;

Antonov et al., 2010; Garcia et al., 2010a; Garcia et al., 2010b; Locarnini et al., 2010). These WOA09 regularly gridded maps have provided important boundary and/or initial conditions for numerical models and aided in evaluating model performance. However, irregular sampling density in the SHP network, particularly nutrients (Boyer et al., 2009), limits the WOA09 monthly distributions to be only representative of large-scale features, without the ability to resolve smaller-scale features such as boundary currents and eddy fields (Boyer et al., 2009). Here, the WOA09 ocean surface (0m) climatologies were used as an independent dataset (see Fig. 2.10) to diagnose monthly  $C_{\rm T}$  and  $A_{\rm T}$  distributions for the nominal year of 2000 via the global bottle-trained SOMLO model.

## 2.11.1 WOA09 unit conversions

Parameter units used in the WOA09 products and global bottle dataset are listed in Table 2.14. To assimilate WOA09 nutrient units to the bottle dataset, nutrient concentrations were multiplied by seawater density computed via the equation of state approximation of Brydon et al. (1999). For dissolved oxygen, WOA09 concentrations were first converted from ml to  $\mu$ mol using a conversion factor of 44.6596 (derived from the molar volume of oxygen gas at standard temperature and pressure; 22.3916 L mol<sup>-1</sup>), and then multiplied by seawater density.

	Dataset				
Variable	Global bottle dataset	WOA09			
Temperature	°C	°C			
Salinity	psu	psu			
Dissolved oxygen	µmol kg <sup>-1</sup>	ml $L^{-1}$			
Phosphate	µmol kg <sup>-1</sup>	µmol L <sup>-1</sup>			
Silicate	µmol kg <sup>-1</sup>	µmol L <sup>-1</sup>			

 Table 2.14: Summary of units used in the WOA09 products and global bottle dataset.
#### 2.11.2 Spatial variability

Large-scale features in SOMLO-predicted annual-mean  $C_{\rm T}$  and  $A_{\rm T}$  distributions are in good agreement with the bottle measurements, and follow our broader understanding of spatial carbon variability (Fig. 2.28 and 2.29). In the Southern Ocean, carbon distributions are found in longitudinally homogenous bands driven by the Antarctic Circumpolar Current (ACC), and higher  $C_{\rm T}$  concentrations relative to the global-mean resulting from strong upwelling of CO<sub>2</sub> enriched subsurface waters and cooler surface temperatures enhancing CO<sub>2</sub> solubility (McNeil et al., 2007; Metzl et al., 2006). In equatorial upwelling regions, cold waters enriched with remineralized organic material are brought to surface resulting in elevated  $C_{\rm T}$  and  $A_{\rm T}$  concentrations (Feely et al., 2002). As the surface waters are then transported laterally from the site of upwelling, biological processes and loss of  $CO_2$  to the atmosphere reduces  $C_T$  to some of the lowest concentrations observed globally. For A<sub>T</sub>, maxima concentrations are found in the central subtropical gyres (~25°), where stronger evaporation relative to precipitation elevates salinity concentrations resulting in higher A<sub>T</sub> values (Lee et al., 2006; Millero et al., 1998). Conversely, freshwater input from rivers and seasonal ice melt lowers  $A_{\rm T}$  in regions like the Bay of Bengal (George et al., 1994) and Arctic marginal waters (Dyurgerov and Carter, 2004).

Key et al. (2004) interpolated bottle carbon measurements collected between 1985 and 1999 to diagnose  $1^{\circ}\times1^{\circ}$  global climatologies for  $C_{\rm T}$  and  $A_{\rm T}$  on 33 depth surfaces (GLODAP-v1.1). Comparison between the GLODAP-v1.1 0m distributions and SOMLO predictions reveals good general agreement between the two approaches (Fig. 2.28 and 2.29). However, the average annual-mean SOMLO  $C_{\rm T}$  concentration between 65°N and 77°S is 14 µmol kg<sup>-1</sup> higher than the GLODAP-v1.1 average of 2033 µmol kg<sup>-1</sup>. In particular,  $C_{\rm T}$  concentrations in the Southern Ocean and equatorial Pacific are where the largest discrepancies are found. This observation could either reflect the uptake of anthropogenic CO<sub>2</sub> that was not accounted for by the GLODAP study (Key et al., 2004), or result from a 30% improvement in Southern Ocean data coverage since 1999. However, these observed discrepancies are most likely caused by spatiotemporal biases within the GLODAP dataset.



**Figure 2.28:** Global distributions of (**a**) bottle  $C_{\rm T}$  measurements corrected to the year 2000, (**b**) annual-mean ocean surface SOMLO  $C_{\rm T}$  predictions for the nominal year of 2000, and (**c**) GLODAP-v1.1 0m  $C_{\rm T}$  distribution of Key et al. (2004). Values are given in µmol kg<sup>-1</sup>.



**Figure 2.29**: Global distributions of (**a**) bottle  $A_T$  measurements, (**b**) annual-mean ocean surface SOMLO  $A_T$  predictions, and (**c**) GLODAP-v1.1 0m  $A_T$  distribution of Key et al. (2004). Values are given in  $\mu$ mol kg<sup>-1</sup>.

#### 2.11.3 Seasonal variability

Comparison between seasonal amplitudes in predicted  $C_{\rm T}$  and  $A_{\rm T}$  reveals variability in  $A_{\rm T}$  is on average, 24 µmol kg<sup>-1</sup> lower than  $C_{\rm T}$  (Fig. 2.30). This is likely related to the much stronger influence of biogenic CaCO<sub>3</sub> formation on  $A_{\rm T}$  relative to organic matter synthesis, coupled to CaCO<sub>3</sub> only representing ~10% of total organic carbon export (see *Sect. 1.2.1*). Despite this, large amplitudes in  $A_{\rm T}$  are found in regions where strong seasonal variability in riverine inputs drives large seasonal shifts in salinity and nutrient concentrations, for example, in the Bay of Bengal (George et al., 1994) and Amazonian outflow (Cooley et al., 2007).

For  $C_{\rm T}$ , a more pronounced spatial pattern of variability is found, which is characterized by weak amplitudes in the permanently stratified subtropical gyres, and stronger variability in high latitudes via the combination of biological CO<sub>2</sub> consumption during summertime and enhanced upwelling of carbon enriched sub-surface waters in wintertime (Sarmiento and Gruber, 2006). The observed stronger seasonal variability in the northern subpolar region (North of ~30°N) relative to the Southern Hemisphere (South of ~45°S), has been explained by weaker biological production in the Southern Ocean due to low micro-nutrient supply, in particular that of iron (Boyd, 2009).

Monthly SOMLO-predicted  $C_{\rm T}$  and  $A_{\rm T}$  climatologies for the surface ocean are presented in Figs. 2.31 and 2.32.



**Figure 2.30:** Seasonal amplitudes for SOMLO predicted (**a**)  $C_{\rm T}$  and (**b**)  $A_{\rm T}$  ocean surface distributions, given in µmol kg<sup>-1</sup>. Calculated as the difference between the maximum and minimum predictions in each 1°×1° grid cell. See Figs. 2.31 and 2.32 for monthly distributions.



**Figure 2.31:** Monthly SOMLO-predicted ocean surface  $C_{\rm T}$  distributions [µmol kg<sup>-1</sup>].



**Figure 2.32:** Monthly SOMLO-predicted ocean surface  $A_{\rm T}$  distributions [µmol kg<sup>-1</sup>].

## 2.12 Conclusion

In this study, a newly synthesized global carbon  $C_T/A_T$  mixed-layer bottle database (~33,000) was exploited to investigate two different empirical approaches that diagnose mixed-layer carbon dynamics from standard hydrographic parameters. Using independent data as a test, the traditional multiple-linear regression approach constrains the global  $C_T$  system to within 15.6 µmol kg<sup>-1</sup>, and 10.4 µmol kg<sup>-1</sup> for  $A_T$ . By then deploying a new non-linear neural network based approach, the global predictive skill was improved by 2.7 to 3.0 µmol kg<sup>-1</sup> for  $C_T$ , or 19.4% over the MLR, and 0.7 to 1.4 µmol kg<sup>-1</sup> for  $A_T$ , or ~10%. In particular, regions of known complexity and importance to carbon cycling like the Southern Ocean, North Atlantic and equatorial Pacific are where the new non-linear approach excels, reducing errors by up to 30% over traditional linear approaches. The neural network technique was further tested at the BATS and HOT hydrographic time-series sites and found to predict both seasonal and inter-annual variability of carbon very well, while the traditional MLR approach does not.

The predictive skill of the neural network approach is shown to be spatially and temporally robust, making the model a powerful tool for diagnosing carbon dynamics in the ocean. In reality, the intensity of a sampling regime needed to constrain seasonal to inter-annual variability for carbon is so great that it will always be difficult to achieve on a global scale. I have demonstrated here, that applying the SOMLO model to the WOA09 climatologies can capture coherent spatial and temporal carbon fields over the global open-ocean surface. This new constraint will immediately advance our understanding of oceanic carbon variability and allow us to probe important issues relating to carbon dynamics in a high  $CO_2$  world.

Chapter 3.

# Implications of natural CO<sub>2</sub> variability for ocean acidification

## Abstract

Approximately one-third of anthropogenic CO<sub>2</sub> emissions produced over the last 250 years have been absorbed into the world's oceans. Although this ecosystem service has largely mediated human-induced climate change, oceanic uptake of CO2 lowers seawater pH and carbonate ion concentrations, with potentially detrimental impacts for multiple marine ecosystems. Detecting ocean acidification trends and predicting the onset of threshold limits for marine organisms requires accurate knowledge of natural variability in the ocean carbonate system. Here, a first bottle-based estimate of monthly pH and aragonite saturation-state distributions is presented in the global open-ocean for the nominal year of 2000. This new information is then used to investigate the influence of seasonal variability for aragonite under-saturation onset and detecting ocean acidification trends. On a global scale, seasonal variability brings forward aragonite under-saturation by ~19 years on average relative to annual-mean estimates, exposing of an additional  $24 \times 10^6$  km<sup>2</sup> of ocean surface to at least month-long under-saturation conditions by centuries end. For detecting ocean acidification trends, seasonality will likely hamper efforts to discern robust anthropogenic signals by ~40 years in most of the subtropics, increasing to 60 years in higher latitudes.

## 3.1 Introduction

Rising atmospheric CO<sub>2</sub> concentrations via fossil-fuel emissions is inducing oceanic CO<sub>2</sub> uptake. Several tracer-based techniques estimate that one-third of anthropogenic CO<sub>2</sub> emissions released since the industrial revolution have been absorbed by the oceans (Khatiwala et al., 2009; McNeil et al., 2003; Sabine et al., 2004). The immediate impact of this additional CO<sub>2</sub> is a shift in the oceans chemical composition, resulting in lower seawater pH and carbonate ion ( $CO_3^{2-}$ ) concentrations – commonly referred to as 'ocean acidification' (Zeebe et al., 2008).

Recent studies have identified detrimental impacts for multiple marine ecosystems as a consequence of ocean acidification. For example, a reduction in the availability of carbonate ions affects the ability of marine calcifying organisms to form and preserve their calcium carbonate (CaCO<sub>3</sub>) shells and skeletons (e.g., Gattuso et al., 1998; Langdon and Atkinson, 2005; Riebesell et al., 2000), while lower pH disrupts physiological processes (e.g., Michaelidis et al., 2007; Wootton et al., 2008).

The extent to which ocean acidification will affect marine organisms has been the focus of major research over recent years, but remains largely unknown (Fabry et al., 2008). Some laboratory and mesocosm studies suggest a 0.2 to 0.3 drop in pH will inhibit marine calcification (e.g., Riebesell et al., 2000; Ries et al., 2009). To contextualize this change, the average ocean surface pH has already decreased by ~0.1 units since preindustrial times (Feely et al., 2004), and is projected to drop by a further 0.3-0.4 units by 2100 (Orr et al., 2005). A reduction of this magnitude is estimated to reduce subtropical coral reef calcification by 60% relative to preindustrial times (Friedrich et al., 2012).

The saturation state of CaCO<sub>3</sub> minerals also provides a metric for understanding the implications of carbonate ion depletion for marine calcifiers. Calculated from the

product of in situ  $Ca^{2+}$  and  $CO_3^{2-}$  concentrations divided by the apparent stoichiometric solubility product ( $K_{sp}^*$ ) for either aragonite (Ar) or calcite (Ca):

$$\Omega_{Ar} = [Ca^{2+}][CO_3^{2-}]/K_{sp(Ar)}^*$$

$$\Omega_{Ca} = [Ca^{2+}][CO_3^{2-}]/K_{sp(Ca)}^*$$
(3.1)

Ocean regions where  $\Omega_{Ar}$  or  $\Omega_{Ca}$  is under-saturated (<1) represents corrosive seawater conditions for CaCO<sub>3</sub>, resulting in enhanced dissolution and reduced calcification rates (Fabry et al., 2008).

In addition to understanding the effects of ocean acidification, one of the greatest challenges facing the ocean acidification community is detecting robust secular trends in pH and  $\Omega_{Ar}$ . However, strong variability on seasonal to inter-annual and longer time-scales hampers our ability to discern robust ocean acidification trends.

Despite significant efforts over recent years to establish a global carbonate measurement network, such a large-scale initiative remains very limited, resulting in a near unknown level of understanding regarding regional natural higher-frequency dynamics (Monteiro and Co-Authors, 2010). This represents a critical gap in our ability to probe important questions relating to the influence of natural variability on future ocean acidification, and validate numerical models from which the majority of our understanding is derived.

#### 3.1.1 Onset of aragonite under-saturation

Although global measurements indicate that the contemporary ocean surface is supersaturated with respect to CaCO<sub>3</sub> minerals, model studies suggest the entire Southern Ocean and some parts of the subarctic Pacific will become under-saturated in the more soluble form of CaCO<sub>3</sub>, aragonite, by 2100 (Orr et al., 2005). However, these modelbased estimates are derived from annual-mean projections and do not account for the influence of natural variability inducing earlier under-saturation conditions. McNeil and Matear (2008) demonstrated how strong  $CO_3^{2-}$  seasonality in the Southern Ocean leads to earlier aragonite under-saturation by ~20 years relative to the model based annualmean predictions. More recent data-based studies in Australia's Great Barrier Reef reconfirm the importance of understanding natural variability to accurately predict future ocean acidification onset (Shaw et al., 2012, 2013). Here, this understanding is expanded on by evaluating the influence of natural carbonate variability for aragonite under-saturation across the global ocean.

#### 3.1.2 Detecting Ocean acidification trends

Detecting robust secular trends in pH and  $\Omega_{Ar}$  provides important information on the oceans evolving carbonate system and risks for marine ecosystems. Despite a growing global network of measurements, the ability to detect robust ocean acidification trends is obscured by natural variability on seasonal to inter-annual time-scales. By comparing the magnitude of anthropogenic change to natural noise, referred to as the signal-to-noise ratio (SNR), it is possible to quantify the influence of natural variability for detecting robust secular trends. In general, a lower SNR requires a longer measurement period to identify a robust trend.

Friedrich et al. (2012) employed a numerical model to constrain global detection-times for human-induced  $\Omega_{Ar}$  trends based on a SNR approach. In their study, they computed SNR as the ratio between anthropogenic  $\Omega_{Ar}$  change since preindustrial times (1750), and the seasonal peak-to-peak amplitude in model based estimates. By assuming a SNR exceedance factor of 2 as the detection-limit, their results suggest anthropogenic  $\Omega_{Ar}$  trends were first detectable prior to 1900 in subtropical regions, and within the last 50 years in the equatorial Pacific and high latitudes. However, accurate carbonate measurements only began in the early 1990's. Here, the SOMLO-predicted  $C_{T}$  and  $A_{T}$  distributions are used to calculate monthly pH and  $\Omega_{Ar}$  distributions that are then projected between 1990 and 2100 to estimate ocean acidification detection-times. By prescribing a baseline year of 1990 for anthropogenic change, detection-time

estimates are applicable for current global repeat measurement programs that began in the early 1990s (i.e. WOCE followed by CLIVAR), and provide important insights for establishing new time-series measurement programs.

Detection-times for physical climate trends (e.g. sea surface temperature) have been well established from direct observations and statistical approaches due to large availability of data (e.g. Leroy et al., 2008). These methodologies are designed for timeseries datasets with sampling frequencies greater than the time-scales for natural variability. While there exists some ocean carbon time-series with sufficient temporal measurements to apply these techniques (e.g., Bermuda Atlantic and Hawaiian Ocean time-series sites), the global measurement network for pH and  $\Omega_{Ar}$  consists primarily of one time samples, and is therefore ill suited for the global application of these statistical approaches.

In this chapter, data-based monthly pH and  $\Omega_{Arg}$  global distributions for the nominal year of 2000 are first presented and then projected between 1990 and 2100 based on a constant air-sea  $pCO_2$  gradient. This new ocean surface constraint is then used to explore the influence of natural seasonal variability for aragonite undersaturation onset and detection-times for ocean acidification trends.

## 3.2 Diagnosing global pH and $\Omega_{Ar}$ distributions

The oceans inorganic carbon system can be fully constrained by knowing any two measurements within its inorganic carbon constituents; partial pressure of CO<sub>2</sub> (pCO<sub>2</sub>),  $C_{\rm T}$ ,  $A_{\rm T}$  or pH (Dickson et al., 2007). Since  $C_{\rm T}$  and  $A_{\rm T}$  are the only two carbon parameters that are conservative with respect to changes in state (i.e. temperature and pressure), it is typical for numerical models to only employ  $C_{\rm T}$  and  $A_{\rm T}$  as state variables from which the remaining carbon parameters are computed. Ocean surface pH and  $\Omega_{\rm Ar}$  distributions were therefore computed here using SOMLO-predicted monthly  $C_{\rm T}$  and  $A_{\rm T}$  climatologies, along with the World Ocean Atlas 2009 (WOA09) temperature, salinity,

phosphate and silicate ocean surface (0m) climatologies (Antonov et al., 2010; Garcia et al., 2010b; Locarnini et al., 2010). All calculations were conducted using the CO2SYS program developed by Pierrot et al. (2006) and the carbonic acid dissociation constants of Mehrbach et al (1973), as refitted by Dickson and Millero (1987),  $K_{SO_4}$  dissociation constant of Dickson (1990a), boric acid dissociation constant of Dickson (1990b), and the free pH scale.

#### 3.2.1 Spatiotemporal variability

Global mean pH (free scale) and  $\Omega_{Ar}$  for the ocean surface is estimated to be 8.17 and 2.8 respectively. This is in good agreement with a previous data-based pH estimate of 8.1 for the nominal year of 1994 (Orr, 2011). Annual-mean pH and  $\Omega_{Ar}$  distributions exhibit a high degree of similarity to mixed-layer bottle-derived values (Fig. 3.1 and 3.2). In particular, regions that exhibit strong gradients in  $\Omega_{Ar}$  via the influence of surface thermoclines are mirrored in the bottle-derived distribution (~40° North and South), while strong upwelling of acidic deep-waters depleted in  $CO_3^{2-}$  ions in the eastern equatorial Pacific is evident in both distributions. The SOMLO-predicted global  $\Omega_{Ar}$  distributions also reconfirms that the contemporary ocean surface is supersaturated with respect to aragonite, showing 99.5% of predicted  $\Omega_{Ar}$  values are greater or equal to 1.



**Figure 3.1:** Global distributions of (**a**) bottle-derived pH measurements corrected to the year 2000 and (**b**) annual-mean SOMLO-derived pH predictions for the nominal year of 2000 (see *Appendix F* for monthly distribution plots).



**Figure 3.2:** Global distributions of (**a**) bottle-derived  $\Omega_{Ar}$  measurements corrected to the year 2000 and (**b**) annual-mean SOMLO-derived  $\Omega_{Ar}$  predictions for the nominal year of 2000 (see *Appendix F* for monthly distribution plots).

Seasonal amplitudes for pH are found to range from  $0.05\pm0.03$  in (sub)tropical regions (35°S to 35°N) to  $0.12\pm0.14$  in higher latitudes, with a global mean of 0.08 (Fig. 3.3a). Since ocean surface pH is estimated to have decreased by 0.1 over the last ~250 years (Feely et al., 2004), any anthropogenic trend over the past 20 years will be heavily obscured by natural variability.

Comparison between seasonal pH and  $\Omega_{Ar}$  amplitudes reveals a somewhat consistent spatial pattern (Figs. 3.3), with weak variability found in subtropical regions that strengths towards the higher latitudes. However, a global correlation of 0.7 indicates that the magnitude and direction differs distinctly over 30% of the ocean. In the Southern Ocean for example (South of 40°S), seasonal variability for pH exhibits a much stronger relative signal than  $\Omega_{Ar}$ , while between 20°S to 40°S the amplitudes are stronger for  $\Omega_{Ar}$  compared to pH. Given that marine calcifying organisms are typically more sensitive to changes in either pH or  $\Omega_{Ar}$  (Doney et al., 2009; and references within), spatiotemporal differences in pH and  $\Omega_{Ar}$  illustrates the importance to discern secular trends in both carbonate parameters. Furthermore, this somewhat independent variability between pH and  $\Omega_{Ar}$  are in contrast with that found in models (Matsumoto and McNeil, 2012; McNeil and Matear, 2007).



**Figure 3.3:** Seasonal amplitudes for SOMLO-derived (**a**) pH and (**b**)  $\Omega_{Ar}$  distribution. Calculated as the difference between the maximum and minimum predictions in each  $1^{\circ} \times 1^{\circ}$  grid cell (see *Appendix F* for seasonal distribution figures).

#### 3.2.2 Error analysis

The approach employed here to constrain pH and  $\Omega_{Ar}$  distributions includes both random and systematic sources of error. Random errors are introduced through the use SOMLO predicted  $C_T$  and  $A_T$  distributions that have global uncertainties of ±10.9 and ±9.2 µmol kg<sup>-1</sup> respectively (see *Sect. 2.8.2*). To estimate the associated uncertainty in SOMLO-derived pH and  $\Omega_{Ar}$  distributions, the independently predicted  $C_T$  and  $A_T$ concentrations (see *Sect. 2.8.2*) were used to calculate the carbonate parameters via CO2SYS, and then compared to the corresponding bottle-derived measurements to compute global residual standard errors for pH and  $\Omega_{Ar}$  (RSE; Eq. 2.4). Uncertainty in global pH and  $\Omega_{Ar}$  predictions are estimated to be ±0.03 and ±0.15 respectively (Fig. 3.4).

Sources of systematic error include biases in SOMLO-derived pH and  $\Omega_{Ar}$  predictions. By exploring the distribution of independent residual errors (Fig. 3.4c, d), the global bias in SOMLO-derived predictions is found to be 0.005 and 0.02 µmol kg<sup>-1</sup> for pH and  $\Omega_{Ar}$  respectively.

Finally, it's important to acknowledge that uncertainties and biases in the WOA09 objectively analyzed maps will influence the SOMLO-derived pH and  $\Omega_{Ar}$  distributions. However, since uncertainty estimates in the WAO09 products are yet to be quantified, the resulting error in SOMLO predictions cannot be accounted for at this time.



**Figure 3.4:** SIT predictions versus bottle-derived measurements for (a) pH and (b) aragonite saturation-state ( $\Omega_{Ar}$ ).  $r^2$  and N are the r-squared correlation and number of measurements respectively, and red-line is the y = x straight line. Panels (c) and (d) present the residual error density distributions for pH and  $\Omega_{Ar}$  respectively. Comparatively small residual means of 0.005 and 0.02 for pH and  $\Omega_{Ar}$  indicates a negligible bias in SOMLO-derived predictions.

## 3.3 Projecting past and future carbonate distributions

Global pH and  $\Omega_{Ar}$  distributions were predicted for each year between 1990 and 2100 by assuming a constant air-sea *p*CO<sub>2</sub> difference. In this approach, annual-mean *C*<sub>T</sub> concentrations were first computed between 1990 and 2100 using SOMLO-predicted annual-mean *p*CO<sub>2</sub> (see *Sect. 4.4*) and *A*<sub>T</sub> ocean surface concentrations, where ocean surface *p*CO<sub>2</sub> tracked the Representative Concentration Pathways 8.5 (RCP8.5) atmospheric CO<sub>2</sub> projections (Meinshausen et al., 2011; downloaded from http://www.pik-potsdam.de/~mmalte/rcps/). Seasonal variability in *C*<sub>T</sub> was then added to the annual-mean predictions to reconstruct monthly pH and  $\Omega_{Ar}$  distributions from 1990 to 2100. Note that reconstructed seasonal amplitudes were initially constant, however, as ocean carbon chemistry changed with additional CO<sub>2</sub> input (i.e., changes in the Revelle factor), the amplitudes in calculated pH and  $\Omega_{Ar}$  changed; as has been observed in a previous model study by Rodgers et al. (2008), and can be seen in Fig. 3.6. All calculations were performed using the *seacarb* R package of Lavigne H. and Gattuso J.-P. (2012).

RCP atmospheric  $CO_2$  concentrations are a combination of historical observations/emission estimates between 1765 and 2005 and model-based emission projections from 2005 to 2100, which account for socioeconomic change (Meinshausen et al., 2011). The suite of RCP scenarios are defined as four families: RCP3, RCP4.5, RCP6 and RCP8.5; where RCP8.5 representing the highest emissions scenario (Fig. 3.5). Comparison between global  $CO_2$  emission estimates and RCP projections between 2005 and 2012, indicate that emissions are tracking the RCP8.5 scenario (Peters et al., 2012).



**Figure 3.5:** Atmospheric  $CO_2$  projections between 1990 and 2100 as determined by the Representative Concentration Pathways (RCP) project (Meinshausen et al., 2011). Concentrations are for dry air given in parts per million (ppm).

#### 3.3.1 Influence of air-sea pCO<sub>2</sub> disequilibrium

One source of uncertainty in projected carbon concentrations relates to the air-sea  $pCO_2$  disequilibrium signal, whereby ocean surface  $pCO_2$  growth rates diverge from the atmosphere in some ocean regions (mainly high latitudes) due to natural (e.g. upwelling and seasonal sea-ice coverage) and anthropogenic influences (e.g. temperature change and shifting wind patterns). In the southern Indian Ocean for example, a faster oceanic  $pCO_2$  growth rate relative to atmospheric  $CO_2$  is suggested to be driven by enhanced upwelling of  $CO_2$  enriched deep-waters (Metzl, 2009), while time-series stations in the subtropics indicate that ocean surface  $pCO_2$  is closely tracking atmospheric  $CO_2$  (Bates, 2007; Santana-Casiano et al., 2007).

McNeil and Matear (2008) investigate how future  $pCO_2$  disequilibrium influences pH and  $\Omega_{Ar}$  predictions in the Southern Ocean by comparing steady-state predictions to estimates from an ocean carbon model (Matear, 2007). They found steady-state pH and  $\Omega_{Ar}$  predictions lagged by an average of ~8 years due to an average slower oceanic  $pCO_2$  growth rate of ~0.48 µatm yr<sup>-1</sup> over the coming century. Since observed air-sea  $pCO_2$  disequilibrium signals are typically no greater than ±0.48 µatm yr<sup>-1</sup> (Table 3.1), the projected basin-scale pH and  $\Omega_{Ar}$  values presented here will be likely dampened by up to ~8 years in high latitudes, while (sub)tropical regions will be subject to little, or no disequilibrium influence. It is important to note that on a regional scale, disequilibrium has been estimated to dampen the signal by up to 30 years in some parts of the Southern Ocean (McNeil et al., 2010).

<b>`</b>		<b>)</b> - 1					
				pC	O <sub>2</sub> growth rate		
Region	Latitude	Longitude	Period	Ocean surface	Atmosphere	atm-ocean	Authors
			(Sub)tropic	S			
BATS	31°43'N	64°10'W	1983 - 2005	$1.67 \pm 0.28$	$1.78 \pm 0.02$	-0.11	(Bates, 2007)
HOT	22°44'N	158°W	1989 - 2001	$2.46\pm0.28$	$1.48\pm0.05$	0.98	(Dore et al., 2003)
ESTOC	29°10'N	15°30'W	1995 - 2004	$1.55\pm0.43$	$1.6\pm0.7$	-0.05	(Santana-Casiano et al., 2007)
Eq. Pacific	$N_{o}S - S_{o}S$	$170^{\circ}W - 120^{\circ}W$	1981 - 2004	$1.13\pm0.31$	$1.6\pm0.3$	-0.47	(Feely et al., 2006)
Eq. Pacific	$N_{o}S - S_{o}S$	West of 170°W	1981 - 2004	$1.91\pm0.22$	$1.6\pm0.3$	0.31	(Feely et al., 2006)
Eq. Pacific	$N_{o}S - S_{o}S$	$144^{\circ}\mathrm{E} - 160^{\circ}\mathrm{W}$	1985 - 2004	$1.56\pm0.2$	1.5	0.06	(Ishii et al., 2009)
N Pacific	$10^{\circ}N - 40^{\circ}N$		1970 - 2003	$1.33\pm0.47^{\rm a}$	1.5	-0.17	(Takahashi et al., 2006)
N Pacific	$10^{\circ}N - 42^{\circ}N$	$120^\circ\mathrm{E}-105^\circ\mathrm{W}$	1996 - 2005	$1.8\pm0.6$	1.9	-0.1	(Lenton et al., 2012)
N Atlantic	$15^{\circ}N - 40^{\circ}N$		1970 - 2007	$1.83\pm0.28^{\rm a}$	1.5	0.33	(Takahashi et al., 2009)
N Atlantic permanently stratified	Defined by Sar	miento et al. (2004)	1981 - 2009	$1.6 \pm 0.12$	1.65	-0.05	(McKinley et al., 2011)
N Atlantic seasonally stratified	Defined by Sar	miento et al. (2004)	1981 - 2009	$1.67\pm0.15$	1.65	0.02	(McKinley et al., 2011)
			North Pacif				
Western subarctic	$42^{\circ}N - 50^{\circ}N$	$150^{\circ}\mathrm{E} - 170^{\circ}\mathrm{E}$	1995 - 2003	$1.6\pm1.7$	1.9	-0.3	(Lenton et al., 2012)
	$40^{\circ}N - 60^{\circ}N$	$170^{\circ}\mathrm{E} - 120^{\circ}\mathrm{W}^{\mathrm{c}}$	1970 - 2003	$1.13\pm0.43^{\mathrm{a}}$	1.5	-0.37	(Takahashi et al., 2006)
Bering Sea			1970 - 2003	$-1.34 \pm 0.4^{a}$	1.5	-2.84	(Takahashi et al., 2006)
			North Atlan	tic			
North Atlantic drift	44°N - 58°N	$42^{\circ}W - 10^{\circ}W$	1981 - 1998	$1.9\pm1.46^{a}$	1.5	0.4	(Lefèvre et al., 2004)
Seasonally stratified	Defined by Sar	miento et al. (2004)	1981 - 2009	$1.52\pm0.22$	1.65	-0.13	(McKinley et al., 2011)
	$40^{\circ}N - 70^{\circ}N$		1970 - 2003	$1.73\pm0.46^{a}$	1.5	0.23	(Takahashi et al., 2009)
			Southern Oco	an			
South Western Indian Ocean	20°S - 55°S		1991 - 2007	$2.11 \pm 0.07$	1.72	0.39	(Metzl, 2009)
Indian and Pacific	45°S - 62°S		1995 - 2008	$2.2 \pm 0.2$	1.9	0.3	(Lenton et al., 2012)
<sup>a</sup> Growth rate and uncertainty ar	e calculated as	he mean and first s	tandard deviat	ion, respectivel	y, using quote	d rates withi	n the defined region.

**Table 3.1:** Summary of observed air-sea *p*CO<sub>2</sub> growth rates.

111 11

## 3.4 Estimating the onset of aragonite under-saturation

The data-based  $\Omega_{Ar}$  predictions between 2000 and 2100 are used here to estimate the onset year for aragonite under-saturation within the 21<sup>st</sup> century (i.e.  $\Omega_{Ar} = 1$ ). To illustrate the influence of seasonal variability, future  $\Omega_{Ar}$  estimates at two unique sites are presented; one located in the Southern Ocean and the other in the North Atlantic (Fig. 3.6). Strong seasonality in the North Atlantic brings forward the onset of aragonite under-saturation by approximately 27 years relative to the annual-mean estimate (black line; Fig. 3.6a), while weaker variability in the Southern Ocean brings forward under-saturation by 13 years (Fig. 3.6b). It's important to emphasize that this onset starts over one-month, and then eventually extends to be permanent over all months. As much as seasonality brings forward the initial onset of aragonite under-saturation, it also delays the permanent onset (Fig. 3.6b). For the Southern Ocean, seasonality delays the permanent onset by ~18 years (Fig. 3.6b).



**Figure 3.6:** Future aragonite under-saturation ( $\Omega_{Ar}$ <1) onset at locations in the (a) north Atlantic and (b) Southern Ocean. The influence of seasonal variability accelerates under-saturation conditions by 27 and 13 years relative to annual-mean estimates (black line) in the north Atlantic and Southern Ocean, respectively. The red *a*, *b*, and *c* represent the month-long, annual-mean and permanent onset respectively.

On a global scale, annual-mean results suggest aragonite under-saturation will start to occur by 2040 in high latitude surface waters, while tropical and temperate regions (~40°S to ~40°N) will remain super-saturated beyond centuries end (Fig 3.7a). When seasonality is taken into consideration, aragonite under-saturation is brought forward by a global average of 19 years (Fig. 3.7b); with the North Pacific and North Atlantic experiencing the greatest influence, where corrosive surface conditions are accelerated by 35 and 25 years, respectively (Fig. 3.7c).



Figure 3.7: Estimated aragonite under-saturation onset years for (a) annual-mean, (b) one-month, and (b) the temporal difference between the annual-mean and one-month estimates.

Although there has been a wealth of studies investigating temporal onset of aragonite under-saturation (e.g., McNeil and Matear, 2008; Orr et al., 2005; Shaw et al., 2013), seasonality also presents significant implications for the geographical extent of under-saturation. By centuries end, seasonal variability will increase the latitudinal extent of ocean regions exposed to at least month-long corrosive aragonite conditions. On average, seasonality extends the global-mean by about  $4.8^{\circ}$  degrees relative to annual-mean estimates (Fig. 3.8a). This extension translates to an additional ~24×10<sup>6</sup> km<sup>2</sup> of ocean surface area (or 7.6% of total open-ocean area) exposed to at least one month of aragonite under-saturation in 2100 (Fig. 3.8b). This new insight poses a significant threat for multiple calcifying organisms that inhabit these regions. Pteropods for example, are a calcifying zooplankton group that comprise up to 30% of total zooplankton around the Prince Edward Islands (PEI; Fig. 3.8a), and have been found partially dissolved just below the  $\Omega_{Ar}$  saturation horizon (Hunt et al., 2008).



**Figure 3.8: (a)** Area exposed to at least monthly (blue) and annual-mean (red) aragonite under-saturation conditions within the  $21^{st}$  century. The latitudinal extension (blue region) represents  $\sim 24 \times 10^6$  km<sup>2</sup>. (b) Evolution of ocean surface area exposed to undersaturation conditions for at least monthly (blue line) and annual-mean (red line). At centuries end, the difference between the two estimates is  $\sim 24 \times 10^6$  km<sup>2</sup>. The area labeled PEI represents the pteropod study region of Hunt et al. (2008) around the Prince Edward Islands.

The implication of these results are not limited to the high latitudes, strong seasonal  $\Omega_{Ar}$  variability in some subtropical regions (30°S-30°N; see Fig. 3.3b) will likely bring forward the onset of aragonite under-saturation by similar temporal periods

beyond 2100. Since these regions are rich with calcifying coral reef ecosystems, the influence of seasonality presents a major threat to multiple marine ecosystems.

It should be emphasized that the aragonite under-saturation estimates presented here are based on the assumption of a constant air-sea  $pCO_2$  gradient and RCP8.5 atmospheric CO<sub>2</sub> projections. However, the timeframe for under-saturation could be further accelerated or dampened depending on future CO<sub>2</sub> emissions and the air-sea  $pCO_2$  disequilibrium signal.

### 3.5 Detection-times for ocean acidification

Detection-times for robust ocean acidification trends were estimated here using a similar SNR approach as the model-based study of Friedrich et al. (2012). In this approach, SNR were first calculated using annual-mean pH and  $\Omega_{Ar}$  values from 1990 to 2100 and seasonal amplitudes for the nominal year of 2000. An SNR detection-limit of 2 was then applied to estimate the first year for robust anthropogenic trend detection from a 1990 baseline.

To illustrate this approach, Figs. 3.9a, b presents the seasonal pH and  $\Omega_{Ar}$  cycles at a location in the North Atlantic. Peak-to-Peak amplitudes are used here as a metric for natural noise in the system. Embedded within this natural noise is an anthropogenic signal that is constrained by calculating the difference between annual-mean pH and  $\Omega_{Ar}$ values in 1990 to values in 5 year increments up until 2100 (Fig. 3.9c,d). Finally, SNR are calculated as the ratio between the anthropogenic signal and natural noise via

$$SNR = \frac{X_{1990annual mean} - X_{Yr annual mean}}{X_{peak-to-peak amplitude}}$$
(3.2)

where X represents either pH and  $\Omega_{Ar}$ . An SNR detection-limit of 2 is then applied to finally estimate the first year after 1990 for which a robust anthropogenic trend can be detected (Fig. 3.9e, f). For this example, the first year for robust anthropogenic trend detection is estimated to be 2020 and 2075 for pH and  $\Omega_{Ar}$  respectively.



**Figure 3.9:** Example of the detection-time approach at a location in the north Atlantic (40.5°N and 30.5°W). Panels (a) and (b) illustrate the seasonal pH and  $\Omega_{Ar}$  cycles used to constrain the magnitude of natural noise (red dashed lines), panels (c) and (d) presents the anthropogenic change relative to the year 1990 in 5 year increments, and panels (e) and (f) are the calculated signal-to-noise ratios (SNR, see Eq. 3.2). By applying an SNR detection-limit of 2, the first year after 1990 for robust anthropogenic trend detection is estimated (red dashed lines). In this case, robust anthropogenic pH and  $\Omega_{Ar}$  trends will be detectable by 2020 and 2075, respectively.
#### 3.5.1 Application to the global ocean

Applying this method over the global ocean reveals distinct zonal bands for robust anthropogenic detection-years (Fig. 3.10). In the (sub)tropics for example (40°S to 30°N), the average detection-year for pH and  $\Omega_{Ar}$  is 2033 and 2048 respectively, while the first year for robust detection in higher latitudes is estimated to be 2050 and 2067 respectively. From a global perspective, the average detection-year for anthropogenic pH is ~15 years earlier than  $\Omega_{Ar}$ , indicating that pH is a more suitable carbonate parameter for detecting ocean acidification trends within the 21<sup>st</sup> century.

On a global scale, the earliest detection-years are found in the (sub)tropical Atlantic and eastern Pacific, indicating that these regions are the most economical for detecting ocean-acidification trends (Fig. 3.10). Although it is fortunate that some carbon time-series sites are located in these regions (e.g., BATS, HOT and ESTOC; see Fig. 3.10), any residing calcifying organisms will experience the largest degree of change due to high SNR (i.e. the ocean-acidification signal is much larger than natural variability). Beyond the subtropics, stronger seasonality delays ocean acidification detection-times by ~17 years, indicating that substantially larger investments of time and resources will be required to capture robust trends in higher latitudes relative to the subtropics.



**Figure 3.10:** Global estimates for the first year robust anthropogenic trends can be detected from a 1990 baseline for (a) pH and (b) aragonite saturation state. Regions with no estimate (i.e. white) indicate robust anthropogenic trends will not be detectable within the 21<sup>st</sup> century. The BATS and HOT time-series locations are shown in panel (a).

It's important to note that pH and  $\Omega_{Ar}$  experience natural variability on seasonal to inter-annual and longer time-scales. Although time-series measurements show seasonal variability dominates the natural noise signal (Bates, 2007; Keeling et al., 2004), the exclusion of inter-annual variability translates to detection-year estimates that are likely earlier than the required measurement period to detect a robust secular trend. In addition, observational evidence of air-sea  $pCO_2$  disequilibrium introduces further uncertainty in detection-times estimated here, which assumes steady-state.

For monthly distribution plots see Appendix F.

## 3.6 Conclusion

Ocean surface pH and carbonate ion concentrations continue to decline as a result of anthropogenic  $CO_2$  invasion. An increasing number of studies are identifying ocean acidification as a growing concern for multiple marine organisms (e.g., Caldeira and Wickett, 2003; Fabry et al., 2008; Koch et al., 2013). In this study, a data-based approach was used to investigate the influence of natural seasonal variability for aragonite under-saturation onset and detecting robust ocean acidification trends.

This study estimates seasonal variability will bring forward the onset of aragonite under-saturation by a global average of 19 years relative to annual-mean estimates. For multiple calcifying organisms, this results represents the exposure to corrosive conditions much earlier than was previous suggested by studies based on annual-mean estimates (e.g., Orr et al., 2005). In particular, the North Pacific and North Atlantic are regions where seasonality will have the largest influence, bring forward aragonite under-saturation by no less than 25 years relative to annual-mean predictions. In the Southern Ocean, the results presented here are consistent with a similar study by McNeil and Matear (2008), whereby natural variability accelerates aragonite under-saturation by ~16 years.

In addition to the temporal influence, natural seasonal variability increases the latitudinal extent for aragonite under-saturation by ~ $4.8^{\circ}$  relative to the annual-mean in 2100. The resulting exposure of an additional  $24 \times 10^{6}$  km<sup>2</sup> to at least month-long corrosive aragonite condition by centuries end, will present detrimental implications for multiple marine ecosystems (e.g., pteropods). Despite this, seasonality will also delay

the permanent onset of ocean acidification, which will reduce the severity for calcifying organisms that can tolerate short exposures to corrosive conditions.

Natural variability impedes our ability to detect robust ocean acidification trends, which are crucial for understanding future implications for marine ecosystems. Using a baseline of 1990 and an SNR technique, the first year for robust ocean acidification detection is estimated to be 2033 and 2048 in (sub)tropical regions for pH and  $\Omega_{Ar}$  respectively, while smaller SNRs in higher latitudes increase the average detection year by ~17 years. This analysis also found that the detection-period for pH is on average 15 years earlier than  $\Omega_{Ar}$  on a global scale. This is due to the faster rate of increase in SNR for pH relative to  $\Omega_{Ar}$  (i.e, the ratio between anthropogenic change since 1990 and natural seasonal variability is increasing at a faster rate for pH). For organisms that more sensitive to changes in pH, the implication of this result will be the exposure to adverse conditions much sooner than other organisms that are more sensitive to changes in  $\Omega_{Ar}$ . In addition, measurement programs can be dramatically shortened in regions where high correlations between pH and  $\Omega_{Ar}$  exist.

Chapter 4.

# A new constraint on global air-sea CO<sub>2</sub> fluxes using bottle carbon data

This chapter is an edited version of a paper published in Geophysical Research Letters (GRL).

Sasse, T. P., McNeil, B. I., and Abramowitz, G. (2013): A new constraint on global airsea CO<sub>2</sub> fluxes using bottle carbon data, Geophysical Research Letters, 40, 1594-1599 DOI: 10.1002/grl.50342

## Abstract

Numerous tracer-based techniques show that the ocean has absorbed about 70ppmequivalent of fossil fuel CO<sub>2</sub> since the industrial revolution, dramatically mediating climate change. Despite this, our understanding of spatial and temporal patterns of oceanic CO<sub>2</sub> fluxes remains poorly constrained, since methods either require the use of uncertain transport models or suffer from interpolation biases from regionally sparse underway oceanic CO<sub>2</sub> measurements. Here, these issues are addressed by presenting a new observationally-derived ocean surface climatology for the partial pressure of CO<sub>2</sub>  $(pCO_2)$  that provides an independent constraint on contemporary air-sea CO<sub>2</sub> fluxes. The approach uses a neural network, trained on ~17,800 bottle-derived measurements of  $pCO_2$ , to diagnose monthly  $pCO_2$  levels from standard ocean hydrographic data. Although the pattern of contemporary air-sea CO<sub>2</sub> fluxes are generally consistent with the independent "underway"  $pCO_2$  data network, the new results show a strong shift in the magnitude of oceanic sources and sinks of CO<sub>2</sub>. In particular, the contemporary Southern Hemisphere ocean CO<sub>2</sub> uptake of 0.93 PgC yr<sup>-1</sup>, driven by a prominent CO<sub>2</sub> sink in the sub-polar region (25°-60°S), is five times the magnitude of the Northern Hemisphere oceanic sink (0.18 PgC yr<sup>-1</sup>). Globally, the bottle-derived results suggest a net anthropogenic open-ocean CO $_2$  sink of 1.55 $\pm$ 0.32 PgC yr^-1 for the nominal year of 2000.

# 4.1 Introduction

Understanding how the ocean modulates atmospheric carbon dioxide ( $CO_2$ ) on higherfrequency (seasonal to inter-annual) scales is important since the ocean is absorbing up to one-third of anthropogenic  $CO_2$  emissions based on a range of data-based and modelling estimates (Wanninkhof et al., 2013 and references therein). Although our constraint on the global air-sea  $CO_2$  flux has improved over recent years, large uncertainties remain, particularly in understanding higher-frequency regional air-sea flows. Indirect methods have been useful in providing insights for regional flows using inversions of atmospheric CO<sub>2</sub> (Gurney et al., 2008) or ocean interior data (Gruber et al., 2009), however, they require the use of uncertain transport models and data synthesis methods. We have therefore solely relied on global "underway"  $pCO_2$ measurements and their synthesis (Takahashi et al., 2009; herein after referred to as T-09) as the only direct data-based constraint for contemporary air-sea CO<sub>2</sub> fluxes. Although the underway  $pCO_2$  data network has given us tremendous insight into the distributions of contemporary air-sea CO<sub>2</sub> fluxes, considerable uncertainties remain. The Northern Hemispheric oceans for example, are well constrained for oceanic  $pCO_2$ due to autonomous sampling undertaken mainly by commercial ships of opportunity. In the Southern Hemisphere however, where ships of opportunity are sparse, large spatial and temporal data-gaps exist in the underway  $pCO_2$  network. Where underway  $pCO_2$ data is sparse, simplistic interpolation schemes are required, contributing to large regional uncertainty in constraining contemporary air-sea CO<sub>2</sub> fluxes in the ocean. Here, a new and independent monthly ocean surface  $pCO_2$  distribution is presented using an observationally-derived empirical technique that diagnoses  $pCO_2$  from biogeochemical information. This new  $pCO_2$  climatology provides an additional data-based constraint on spatial and temporal patterns of contemporary air-sea CO<sub>2</sub> fluxes throughout the ocean.

## 4.2 Global training dataset

For over twenty years, global oceanographic measurement programs like the World Ocean Circulation Experiment (WOCE) and Climate Variability and Predictability (CLIVAR), have collected and analysed hundreds of thousands of in situ bottle carbon measurements of total dissolved carbon dioxide ( $C_T$ ) and total alkalinity ( $A_T$ ), along with standard hydrographic parameters (SHP; temperature, salinity, dissolved oxygen

and nutrients) (Key et al., 2004, 2010). In situ concentrations of  $C_{\rm T}$  and  $A_{\rm T}$  allows  $p{\rm CO}_2$  to be calculated using well known dissociation constants of CO<sub>2</sub> in seawater (Dickson et al., 2007), thereby providing a global-scale independent  $p{\rm CO}_2$  dataset (Fig. 4.1).

Bottle-derived  $pCO_2$  concentrations were calculated here via the CO2SYS program developed by Pierrot et al. (2006) using bottle measurements of  $C_{\rm T}$ ,  $A_{\rm T}$ , temperature, salinity and silicate/phosphate where available. Selection of the Mehrbach (1973) equilibrium constants for CO<sub>2</sub> chemistry, as refitted by Dickson and Millero (1987), was based on comparison studies between measured and computed  $pCO_2$  concentrations using different equilibrium constants (Lee et al., 2000a; McNeil et al., 2007; Millero et al., 2002; Wanninkhof et al., 1999), and maintained consistency with carbon calculations used in the GLODAP and CARINA products (Key et al., 2004; Pierrot et al., 2010). Accuracy in bottle-derived  $pCO_2$  concentrations has been estimated to be ±8µatm (Lee et al., 2000a; Wanninkhof et al., 1999).





Although spatiotemporal coverage of the global bottle-derived  $pCO_2$  dataset is too sparse on its own, the coinciding SHP provides powerful biogeochemical information that can help diagnose  $pCO_2$  where only SHP data exist. Empirical predictions of CO<sub>2</sub> from hydrographic properties have been successfully deployed to quantify decadal anthropogenic CO<sub>2</sub> accumulation (e.g., McNeil et al., 2001b; Wallace, 1995; Wanninkhof et al., 2010) and regional contemporary air-sea CO<sub>2</sub> fluxes (e.g., Bates et al., 2006; McNeil et al., 2007). If a robust empirical relationship can be established, it can be applied to much larger hydrographic datasets that have been objectively analyses to monthly  $1^{\circ} \times 1^{\circ}$  climatologies (e.g. Locarnini et al., 2010), thereby providing an independent data-based constraint on monthly ocean surface *p*CO<sub>2</sub> distributions and air-sea CO<sub>2</sub> fluxes.

## 4.3 Neural network overview

The model used here couples a neural network clustering algorithm with a principlecomponent regression (PCR; see Appendix D) to derive the empirical relationship between mixed-layer  $pCO_2$  and SHP. In this approach, the algorithm captures largerscale ocean dynamics via clustering data into "biogeochemical fingerprints" in a selforganizing map (SOM; Kohonen, 1988). In brief, the SOM approach utilizes bottlederived  $pCO_2$  measurements and SHP distribution information, along with geographical constraints, to iteratively cluster the bottle measurements into a set of J neurons based on similarities and homogeneity within the dataset. Using an algorithm that employs discrete clustering is appealing, as it removes the need for any ad-hoc data partitioning to help empirically constrain the system. This has led to application of SOMs in a wide range of disciplines (e.g., Abramowitz, 2005; Hsu et al., 2002; Telszewski et al., 2009). After the SOM routine has clustered the multi-dimensional dataset, PCRs are derived between  $pCO_2$  and the SHP using data within each neuron, each of which can be thought of as a local-scale optimizer that follows the global non-linear optimization analysis performed by the SOM. To then predict  $pCO_2$  using any independent set of SHP measurements, a similarity measure is first used to determine which neuron best represents the SHP measurements, then the  $pCO_2$  value is predicted using the regression

parameters established with training data of that neuron. This approach is called SOMLO: self-organizing multiple-linear output (see *Sect. 2.6* for more details).

## 4.4 Application to the global dataset

Training the SOMLO model is conducted using bottle-derived  $pCO_2$  measurements within the mixed-layer where coinciding SHP exist. The global  $pCO_2$  training dataset is further refined to be post-1980, due to large uncertainties in early measuring techniques, and excluded coastal margins to mitigate terrigenous biases on coastal samples (see *Sects. 2.2.1* and *2.2.2* for details). The final dataset of usable mixed-layer measurements to train the global model (22,688 samples) was derived from 293 cruises (see *Auxiliary B* for list of cruises) and data from the Bermuda Atlantic (BATS; Bates, 2007) and Hawaiian Ocean (HOT; Keeling et al., 2004) time-series stations.

In order to account for the influence of oceanic uptake of anthropogenic CO<sub>2</sub>, all bottle-derived pCO<sub>2</sub> samples were normalized to the nominal year of 2000 in a similar way to T-09. This was achieved by assuming rates of change in mixed-layer pCO<sub>2</sub> were in equilibrium with the observed rate at the Mauna-Loa CO<sub>2</sub> measuring site (Tans, P., NOAA/ESRL, www.esrl.noaa.gov/gmd/ccgg/trends and Keeling, R., Scripps Institute of Oceanography scrippsco2.ucsd.edu/). Although there are regions that are known to break this assumption (e.g. some high latitude regions (Lenton et al., 2012)), a sensitivity analysis was performed by training and testing the SOMLO model using data without anthropogenic corrections. Comparison between the two approaches revealed a small difference of 1.6 µatm (or 6.6%), which suggests an insignificant impact on the models ability to predict surface-ocean pCO<sub>2</sub>, and therefore the final air-sea flux results (see *Sect. 4.5.2* for more details).

The optimal parameter combination and SOM size was determined by employing the same approach as outline in *Sect. 2.7.1*. In this SOMLO analysis, the SHP parameter set which captured global  $pCO_2$  with the highest skill was a combination of temperature, salinity, dissolved oxygen and phosphate. The inclusion of geographical information (*n*-vector; see *Sect. 2.7.1*) in classifying the dataset into 49 neurons also enhanced the global skill of the SOMLO technique by ~9%. Due to missing phosphate or dissolved oxygen measurements in some bottle samples, the final number of usable data-points to train the optimal model was 17,753.

It is worth noting that some previous studies derived linear relationships between "underway"  $pCO_2$  and either sea surface temperature and/or salinity (e.g., Cosca et al., 2003; Lefèvre and Taylor, 2002; Nakaoka et al., 2006). Here, the addition of biogeochemical information (i.e. dissolved oxygen and nutrients) improves the predictive skill of SOMLO by ~5 µatm (or 17%).

## 4.5 Testing the new approach

To independently test the model, each cruise and time-series dataset were individually excluded during the SOMLO training process, and then used as an independent dataset to predict  $pCO_2$  concentrations (SIT; see *Sect. 2.4*). Of the 17,753 bottle measurements used to independently test the SOMLO model, 876 had a residual error greater than  $\pm$ 50µatm. Further investigation of these samples revealed that 43% (403) were located within 300 km of a major coastline. Since these anomalous coastal predictions are likely the result of terrogenious influences perturbing local biogeochemical processes, these anomalous coastal measurements were excluded from further SOMLO skill analysis. Comparison between the remaining independent predictions and bottle-derived  $pCO_2$  measurements indicates SOMLO estimates follow a normal distribution with a mean bias of +0.08 µatm and residual standard error (RSE, see Eq. 2.4) of 22.5 µatm (*N*=17,350; Fig. 4.2).



**Figure 4.2:** (a) SIT  $pCO_2$  predictions versus bottle-derived  $pCO_2$  concentrations and (b) residual error density distribution.  $r^2$ , N and RSE represent the r-squared correlation, number of measurements and residual standard error, respectively.

### 4.5.1 Probing the model for seasonal bias

Partitioning the global independent predictions by season and calculating the residual standard errors (RSE; Eq. 2.4) allowed the influence of seasonal data bias in SOMLO predictions to be investigated (Fig. 4.3). Although there are twice as many summertime measurements as winter, no global seasonal bias from the independent testing of SOMLO was found.



**Figure 4.3:** RSE values for  $pCO_2$  SIT predictions partitioned by season. Austral seasons are defined as summer (December to February), autumn (March to May), winter (June to August) and spring (September to November). Boreal hemisphere seasons differ by 6 months. Number of measurements in each season is represented in the brackets.

Further scrutinizing the independent SOMLO predictions on a regional-scale reveals no strong seasonal bias over most of the ocean (Table 4.1). This provides additional confidence in SOMLO predictions for any given set of temperature, salinity, dissolved oxygen and phosphate measurements in the global open-ocean mixed-layer. Exceptions are however found in the high north autumn/winter and tropical autumn months (Table 4.1), where RSE values range between 30-40 µatm. However, caution needs to be exercised when computing RSE from a small group of samples, since only a few measurements with large residual errors can have a strong influence due to the squared residual error term in the RSE equation (see Eq. 2.4). For example, the exclusion of 13 samples (or 2%) in the high north autumn reduces the RSE by  $\sim$ 7 µatm (or to 22.5 µatm). Similar results were found in the high north winter and tropical autumn.

**Table 4.1:** RSE values of SIT predictions partitioned into 5 regions and by season. Seasons are defined by summer (December to February), autumn (March to May), winter (June to August) and spring (September to November) for Southern Hemisphere data. Northern Hemisphere seasons differ by 6 months.

		RSE [ $\mu$ mol kg <sup>-1</sup> ] ( $N^a$ )								
Region	Latitudinal band	Summer	Summer Autumn		Spring	Total $N$				
High North	$44^{\circ}N - 70^{\circ}N$	25.89 (969)	29.27 (688)	40.21 (303)	23.38 (526)	2486				
Temperate North	$18^{\circ}N - 44^{\circ}N$	24.95 (1101)	16.52 (1016)	15.91 (1124)	19.41 (952)	4193				
Tropical	$18^{\circ}\mathrm{S} - 18^{\circ}\mathrm{N}$	21.60 (722)	36.92 (744)	18.39 (570)	23.06 (1179)	3215				
Temperate South	$18^{\circ}\text{S} - 44^{\circ}\text{S}$	22.34 (854)	19.24 (380)	12.63 (526)	16.14 (1165)	2925				
Southern Ocean	South of 44°S	22.17 (1357)	15.86 (1261)	10.97 (367)	22.41 (870)	3855				
Global		23.85 (5353)	24.05 (4293)	19.43 (2890)	21.32 (4814)	17350				

<sup>a</sup> Number of measurements

### 4.5.2 Influence of anthropogenic correction

To evaluate the models sensitivity to anthropogenic  $pCO_2$  corrections, the optimal SOMLO configuration was trained and tested using data that was not corrected for anthropogenic CO<sub>2</sub> uptake (via the SIT approach). The global residual standard error (24.1 µatm) was found to be 1.6 µatm higher than the optimal  $pCO_2$  model trained and tested using correct  $pCO_2$  data (22.5 µatm) (Table 4.2). This small discrepancy between the two approaches indicates an insignificant influence of anthropogenic corrections to the model's ability to predict mixed-layer  $pCO_2$ , which is likely related to the small anthropogenic corrections of ~1.8 µatm yr<sup>-1</sup> compared to large natural seasonal variability of ~60 µatm.

**Table 4.2:** Global skill comparison between optimal SOMLO models trained and tested using data either corrected, or not corrected for anthropogenic CO<sub>2</sub> uptake.

Optimal SOMLO model	Global SIT residual					
trained and tested using:	standard error [µatm]					
Data not corrected	24.1					
Corrected data	22.5					

#### 4.5.3 How well does SOMLO capture temporal variability at BATS?

To illustrate SOMLO's ability to diagnose temporal  $pCO_2$  concentrations using SHP information, the 18-year Bermuda Atlantic time-series dataset was excluded during the training of SOMLO, and then used to independent predict the BATS bottle-derived  $pCO_2$  time-series (Fig. 4.4). Although there is a ~20% underestimate of the peak summertime  $pCO_2$  levels, SOMLO is able to reconstruct the 18-year seasonal pattern within its uncertainty range (blue shaded) at a location where no hydrographic information was used to train the global SOMLO model. It is important to emphasize that BATS data was used to train the final SOMLO model.



**Figure 4.4:** In situ and independently predicted annual  $pCO_2$  cycle at the BATS hydrographic station between 1989 to 2006 (Bates, 2007). Annual-mean cycle and natural variability of the bottle-derived  $pCO_2$  measurements (black points) over the 18 years are represented by the red line and red shading respectively, while the blue line and blue shading represents annual-mean cycle and uncertainty in the independent SOMLO  $pCO_2$  predictions. The SOMLO predictions do not use any in situ  $CO_2$  measurements from BATS, but are based only on standard hydrographic properties (temperature, salinity, dissolved oxygen and phosphate) measured at BATS over the 18 years.

# 4.6 Diagnosing monthly pCO<sub>2</sub> climatologies

The world ocean atlas 2009 project (WOA09) has objectively analysed millions of SHP measurements taken over a fifty year period to produce monthly  $1^{\circ} \times 1^{\circ}$  SHP climatologies (Antonov et al., 2010; Garcia et al., 2010a, 2010b; Locarnini et al., 2010). Using the SOMLO model, these SHP monthly surface-ocean climatologies are exploited to estimate monthly *p*CO<sub>2</sub> distributions for the nominal year of 2000.

Large scale features in the estimated annual-mean  $pCO_2$  distribution (Fig. 4.5a) are consistent with our broad understanding of CO<sub>2</sub> rich waters in the eastern equatorial Pacific via upwelling (Feely et al., 2002), and lower  $pCO_2$  levels via solubility-drivers in temperate regions. In the subtropical regions ( $20^{\circ}-35^{\circ}$  North and South), seasonal amplitudes vary by 20 µatm around a mean of 50 µatm (Fig. 4.5b), with lowest concentrations found during winter months reflecting the dominate influence of cooler temperatures on CO<sub>2</sub> solubility (See Fig. 4.6 for monthly  $pCO_2$  distributions); as has been observed by previous studies (e.g., Bates et al., 1996; González-Dávila et al., 2003; Keeling et al., 2004; Metzl et al., 1998). Moving poleward from the subtropics ( $35^{\circ}-50^{\circ}$  North and South), seasonal amplitudes vary by 30 around a mean of 60 µatm. This larger seasonal signature is driven by a complex interplay of biological and physical-solubility processes (Bates et al., 1996; Brix et al., 2013). Finally, large seasonal cycles in high latitude regions ( $>50^{\circ}$ ) are typically driven by biological production in summertime and enhanced deep water ventilation during wintertime (McNeil et al., 2007).



**Figure 4.5: (a)** Annual-mean SOMLO  $pCO_2$  predictions for the nominal year of 2000 and (b) seasonal variability in  $pCO_2$  calculated as the difference between the maximum and minimum predictions in each  $1^{\circ} \times 1^{\circ}$  grid cell (see Fig. 4.6 for monthly distribution plots).



**Figure 4.6:** Monthly SOMLO-predicted ocean surface *p*CO<sub>2</sub> distributions [µatm]. 138

#### 4.6.1 Evaluating SOMLO-predictions in the subtropical North Atlantic

ESTOC (European station for time-series in the ocean at the Canary Islands) is a high frequency measurement program of  $pCO_2$  and auxiliary parameters that has been ongoing since September 1995 (González-Dávila et al., 2010). The site is located in the subtropical North Atlantic about 100 km North of the island of Gran Carnaria (29.07°N, 15.83°W) (Fig, 4.7a). Seasonal variability in ocean  $pCO_2$  is characterized by minimum concentrations in wintertime and maximum values in summertime via the dominate influence of seasonal temperature variability (González-Dávila et al., 2003). Since the ESTOC dataset was not available for inclusion in the global bottle dataset, nor is it listed as a contributor to the World Ocean Database 2009 (Boyer et al., 2009), it provides an ideal test-bed to evaluate SOMLOs ability to capture seasonal  $pCO_2$  via the WOA09.

Figure 4.7b shows the monthly mean and spread of ocean-surface  $pCO_2$  measurements normalized to year 2000 at ESTOC versus SOMLO  $pCO_2$  predictions using the WOA09 hydrographic information at 29.5°N and 15.5°W (see Fig. 4.7a). All  $pCO_2$  measurements were normalized to the year 2000 using the observed 1.55 µatm yr<sup>-1</sup> rate of increase at the ESTOC time-series site (Santana-Casiano et al., 2007). Although SOMLO predictions are ~20 µatm lower than the monthly-mean at ESTOC, which is likely related to data limitations in the local region, both the magnitude and structure of the seasonal  $pCO_2$  cycle is well constrained within the uncertainty of SOMLO predictions (red-shade; ±22.5 µatm). This illustrates SOMLOs ability to capture seasonal carbon structures using WOA09 hydrographic values at locations where no information was used to either train the SOMLO model or constrain the WOA09 climatologies.



**Figure 4.7: (a)** Location of the ESTOC site (29.07°N, 15.83°W) and WOA09 1°×1° grid-cell at 29.5°N and 15.5°W, where black points represents bottle measurements used to train SOMLO. (b) Monthly mean and spread of  $pCO_2$  measurements at ESTOC (black line and shade) versus monthly SOMLO predictions using the WOA09 hydrographic information at (29.5°N, 15.5°W) (red line and shade).

## 4.6.2 Comparison to T-09 ocean surface pCO<sub>2</sub> climatology

Comparison between the two independent  $pCO_2$  climatologies reveals a global correlation of 0.64 between the bottle-derived SOMLO- $pCO_2$  climatology and the "underway" T-09  $pCO_2$  climatology re-gridded to a 1°×1° resolution (Fig. 4.8a). This indicates that over most of the ocean, these two independent data-based approaches confirm the general spatiotemporal pattern of  $pCO_2$  in the ocean. However, over one-third of the ocean, the magnitude of  $pCO_2$  concentrations differs distinctly, particularly in the Southern Ocean and equatorial Pacific (Fig. 4.8b).



**Figure 4.8: (a)** Climatological  $pCO_2$  concentrations of T-09 (Takahashi et al., 2009) (re-gridded to a 1°×1° resolution) versus SOMLO predicted values. *corr(X,Y)* and *N* represent the correction co-efficient and number of data points respectively. (b) Annual-mean ocean surface  $pCO_2$  difference between SOMLO-predictions and "underway" T-09 (defined as SOMLO minus T-09; given in µatm).

To investigate the observed regional discrepancies between SOMLO-predictions and T-09, the distribution of residual errors using the independent SOMLO predictions are examined within the Southern Ocean and equatorial Pacific (Fig. 4.9b,d). In the equatorial Pacific, the independent testing of SOMLO reveals  $pCO_2$  predictions are biased by -3.42 µatm (N = 1655), and 0.055 µatm for the Southern Ocean (N = 3941). These SOMLO related regional biases are insignificant in comparison to the mean differences from T-09 in the Southern Ocean (~12 µatm) and equatorial Pacific (~ -21 µatm) (Fig. 4.9a,c). This suggests there must be other reasons for the large regional differences between the two methods.

In the equatorial Pacific, the discrepancy may result from the exclusion of almost 200,000 ocean surface  $pCO_2$  measurements collected during El Niño years from the T-09 synthesis (1982-1983, 1986-1987, 1991-1994, 1997-1998, 2002-2003 and 2004-2005). In contrast, of the 1655 equatorial Pacific bottle measurements used to train the global SOMLO model, 60% were collected during these El Niño years. However, El Niño events are characterized by weakened easterly trade winds in the western and central equatorial, which results in a depression of CO<sub>2</sub> enriched subsurface water upwelling and thus lower ocean surface  $pCO_2$  concentrations relative to non-El Niño years (Cosca et al., 2003; Feely et al., 2006). Since the equatorial Pacific  $pCO_2$  distribution of T-09 (18°S to 18°N) is on average 21 µatm lower than SOMLO estimates (Fig. 4.7a), the T-09 exclusion of El Niño data is unlikely related to the observed discrepancy.



Figure 4.9: Comparison between SOMLO-predicted and T-09 pCO<sub>2</sub> climatology values in (a) equatorial Pacific (18°S to 18°N) and (c) Southern Ocean (South of 44°S). Panels (b) and (d) present the residual error distributions for independent SOMLO predictions in the equatorial Pacific and Southern Ocean, respectively.

## 4.6.3 Why not use the $C_{T}$ and $A_{T}$ climatologies to compute $pCO_{2}$ ?

When any two constituents within the oceans inorganic carbon system are known (i.e.  $pCO_2$ ,  $C_T$ ,  $A_T$  or pH), the remaining parameters can be calculated using well known dissociation constants of CO<sub>2</sub> in seawater (Dickson et al., 2007). This permits global  $pCO_2$  distributions to be computed using the SOMLO-predicted  $C_T$  and  $A_T$  concentrations. However, uncertainties in global SOMLO predictions of ±10.9 and ±9.2

µmol kg<sup>-1</sup> for  $C_{\rm T}$  and  $A_{\rm T}$  respectively (see *Sect. 2.8.2*), would introduce uncertainties into calculated  $p{\rm CO}_2$  concentrations. To quantify the associated uncertainty using this approach, the independent  $C_{\rm T}$  and  $A_{\rm T}$  predictions were used to compute  $p{\rm CO}_2$ concentrations from which the RSE could be calculated and compared to the  $p{\rm CO}_2$ -SOMLO model (Eq. 2.4). When computing  $p{\rm CO}_2$  using the independent  $C_{\rm T}$  and  $A_{\rm T}$ predictions, the global RSE was found to be 25.0 µatm, which is 2.4 µatm (or 10.6%) higher than the independent testing of SOMLO using bottle-derived  $p{\rm CO}_2$  (Table 4.3). This indicates that predicting global  $p{\rm CO}_2$  concentrations via the  $p{\rm CO}_2$ -SOMLO model is a more accurate approach.

**Table 4.3:** Comparison between global RSE values computed using SIT predictions of the  $pCO_2$ -SOMLO model and computed  $pCO_2$  values via the optimal  $C_T$  and  $A_T$  SOMLO SIT predictions.

Method to predict $pCO_2$	RSE [µatm]	Number of			
		measurements			
Calculated using $C_{\rm T}$ and $A_{\rm T}$	25.0	16455			
Predicted via optimal <i>p</i> CO <sub>2</sub> -SOMLO model	22.6	16455			

# 4.7 Computing air-sea CO<sub>2</sub> exchange

Exchange of CO<sub>2</sub> across the air-sea interface  $(F_{CO_2})$  is a function of the air-sea  $pCO_2$ difference  $(\Delta pCO_2)$  and gas transfer coefficient  $(k_g)$ ; which is typically expressed as the product of gas transfer velocity  $(k_w)$  and solubility of CO<sub>2</sub> in seawater  $(\alpha)$  via

$$F_{\rm CO_2} = k_{\rm g} \Delta p \rm CO_2 = k_{\rm w} \alpha (p \rm CO_2^{\rm sea} - p \rm CO_2^{\rm air})$$
(4.1)

where the air-sea gradient follows convention resulting in oceanic  $CO_2$  out-gassing being expressed as a positive value. Air-sea flux estimates presented in this study were calculated using the widely used quadratic  $k_w$ -wind parameterization of Wanninkhof (1992), in conjunction with the CCMP (Cross-Calibrated Multi-Platform) wind field product of Atlas et al. (2011). The Wanninkhof (1992) equation was recently updated by Wanninkhof et al. (2013) to be

$$k_{\rm w} = 0.251 \left(\frac{\rm Sc}{660}\right)^{-0.5} \langle U^2 \rangle \tag{4.2}$$

where Sc is the Schmidt number of CO<sub>2</sub> in seawater (kinetic viscosity in water/diffusion coefficient of CO<sub>2</sub> in water) calculated via a third-order polynomial in seawater temperature (Wanninkhof, 1992), and  $\langle U^2 \rangle$  represents the time-mean of the second moment in wind speed at a height of 10m above sea level. All 1°×1°  $\langle U^2 \rangle$  values were interpolated from the 4°×5° squared CCMP 10m wind speeds averaged over the years 1987 to 2008; as constrained by Wanninkhof et al. (2012). Solubility of CO<sub>2</sub> in seawater was calculated using the temperature and salinity dependent function of Weiss (1974).

Atmospheric  $pCO_2$  concentrations were computed using:

$$pCO_2^{air} = XCO_2(P^{baro} - P^{WV})$$
(4.3)

where  $XCO_2$  is the molar fraction of  $CO_2$  in dry-air,  $P^{\text{baro}}$  is the sea-surface barometric pressure, and  $P^{\text{wv}}$  is the water vapour pressure for a given sea surface temperature and salinity. Atmospheric  $pCO_2$  distributions for the nominal year of 2000 were calculated using the GLOBALVIEW-CO<sub>2</sub> weekly  $XCO_2$  product, which was derived using atmospheric  $CO_2$  measurements collected in the year 2000 and averaged into zonal bands (GLOBALVIEW-CO<sub>2</sub>: 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], 2011), and the NCEP/NCAR reanalysis 1981-2010 monthly mean sea-surface barometric pressure (Kalnay et al., 1996) provided by the NOAA/OAR/ESRL PSD, Boulder, Colorado, USA, from their website at http://www.esrl.noaa.gov/psd/.

To account for seasonal ice formation inhibiting air-sea  $CO_2$  exchange, all flux estimates were multiplied by monthly mean ice cover percentage estimates (ICP)

$$F_{\rm CO_2} = F_{\rm CO_2} \left( 1 - \rm{ICP}/100 \right) \tag{4.4}$$

ICP values were interpolated using the monthly mean NCEP/DOE 2 Reanalysis product (Fig. 4.10) (provided by the NOAA/OAR/ESRL PSD, Boulder, Colorado, USA, from their web site at http://www.cdc.noaa.gov/). Open water conditions (i.e. ICP = 0) were assumed in all grid cells where ICP was less than 10%. The inclusion of seasonal ice coverage decreased the global open-ocean uptake estimate by 13% (or 0.16 PgC yr<sup>-1</sup>).



**Figure 4.10:** Ice cover percentage distributions for January (JAN), April (APR), July (JUL) and October (OCT). White regions indicate no ice coverage.

# 4.8 Air-sea CO<sub>2</sub> flux pattern

The annual contemporary flux distribution shows large CO<sub>2</sub> out-gassing in the equatorial region, and strong uptake in the sub-polar regions (25°-60°), in particular throughout the Southern hemisphere (Fig. 4.11 and 4.12). For the equatorial Pacific Ocean (18°S-18°N), SOMLO predictions estimates a +1.1 PgC yr<sup>-1</sup> out-gassing, more than double that estimated using T-09 pCO<sub>2</sub> values (Fig. 4.11). However, relatively sparse bottle measurements in the eastern equatorial Pacific and large inter-annual flux variability in the equatorial Pacific (Feely et al., 2002, 2006) make it difficult to

establish whether SOMLO flux estimates are a true reflection of the long-term trend in this region. For the Southern Ocean, where there is better spatiotemporal data coverage and weaker El-Niño/Sothern-Oscillation influence, SOMLO results suggest a contemporary sink for atmospheric  $CO_2$  of -0.81 PgC yr<sup>-1</sup>, which is much larger than the T-09 based estimate of -0.28 PgC yr<sup>-1</sup>, but consistent with a linear empirical approach (McNeil et al., 2007) and estimates from some ocean biogeochemical models (Lenton et al., 2013).

See Fig. 4.13 for monthly SOMLO-derived contemporary air-sea  $CO_2$  flux distributions.



**Figure 4.11:** Annual contemporary air-sea  $CO_2$  flux distribution (i.e. flux estimates where integrated over the 12 months in each  $1^{\circ} \times 1^{\circ}$  grid cell; see Fig. 4.13 for monthly contemporary flux distributions). This yields a global open-ocean contemporary air-sea  $CO_2$  flux of 1.10 PgC yr<sup>-1</sup>.



Figure 4.12: Regional and global open-ocean contemporary air-sea CO<sub>2</sub> flux estimates for the bottle-derived analysis in comparison to flux estimates using the T-09 underway  $pCO_2$  climatologies (see *Appendix G* for table of regional flux estimates).<sup>a</sup> Equatorial Indian Ocean region covers the latitudinal band 18°S to 24.5°N.



**Figure 4.13:** Monthly SOMLO-derived contemporary air-sea  $CO_2$  flux distributions, given in grams of carbon per meter<sup>2</sup> per month [gC m<sup>-2</sup> month<sup>-1</sup>].

#### 4.8.1 Zonal flux distribution

Scrutinizing the latitudinal distribution of air-sea  $CO_2$  fluxes reveals a slight amplification in SOMLO-derived estimates compared to the "underway" based T-09 estimates (Fig. 4.14). These zonally integrated air-sea flux estimates are a combination of biological and physical-solubility influences on ocean surface  $pCO_2$ . Despite this slight amplification, the high degree of similarity between these two independent distributions indicates that we currently have a good data-based constraint on the openocean air-sea  $CO_2$  flux distribution, which therefore provides an important benchmark for numerical models.



**Figure 4.14:** Zonally integrated annual contemporary air-sea  $CO_2$  fluxes using SOMLO-predicted (red line) and T-09  $pCO_2$  estimates (blue line). The step like appearance of T-09 likely reflects the courser  $4^{\circ} \times 5^{\circ} pCO_2$  resolution used by T-09, which was re-gridded here to a  $1^{\circ} \times 1^{\circ}$  resolution.

# 4.9 Oceanic CO<sub>2</sub> uptake estimates and uncertainty

Integrating SOMLO-derived  $CO_2$  flux estimates over the global open-ocean suggests a contemporary uptake of 1.10 PgC yr<sup>-1</sup> for the year 2000. To quantify errors in contemporary  $CO_2$  flux estimates, both systematic and random sources of error in

SOMLO-predicted  $pCO_2$  distributions are evaluated. The systematic bias of +0.08 µatm, calculated using an independent sub-sample test (see *Sect. 4.5*), translates to a 0.02 PgC yr<sup>-1</sup> over-estimation in the global contemporary flux estimate. For random errors in diagnosing surface-ocean  $pCO_2$ , both the first standard deviation ( $\sigma$ ) in the residual error distribution (22.5 µatm) along with quoted uncertainty in bottle-derived  $pCO_2$  (±8 µatm) were used. To constrain the net variance in  $\Delta pCO_2$  requires the additional consideration of uncertainty in atmospheric  $pCO_2$  estimates, which has been estimated to be ±0.2 µatm with no known systematic offset (Takahashi et al., 2009). Assuming these uncertainty estimates are all one  $\sigma$  around a normal distribution, the net variance in  $\Delta pCO_2$  is calculated to be 930.29 µatm<sup>2</sup> ( = (22.5 + 8)<sup>2</sup> + 0.2<sup>2</sup>). The corresponding variance in flux estimates for any 1°×1° grid cell (*i*) can be calculated using

$$\operatorname{var}(F_i) = (k_i \alpha_i)^2 \operatorname{var}(\Delta p \operatorname{CO}_2)$$
(4.5)

When integrating flux estimates over grid cells (1...*i*...*I*), the net variance for uncorrelated uncertainties is the sum of individual variances:

$$\operatorname{var}(F) = \sum_{i=1}^{I} \operatorname{var}(F_i) \tag{4.6}$$

For the SOMLO-derived global integrated flux estimate, net variance due to random uncertainties in  $\Delta p CO_2$  is  $1.04 \times 10^{-4} (PgC yr^{-1})^2$ , which translates to a standard deviation of  $\pm 0.03$  PgC yr<sup>-1</sup> within a 99.7% confidence interval.

A second, potentially significant source of uncertainty in predicted ocean surface  $pCO_2$  climatology relates to reliability in the WOA09 objectively analysed products. Parameter sensitivity tests identified temperature and phosphate as the two most important parameters for capturing ocean surface  $pCO_2$  in the global SOMLO model. Despite high confidence in the global WOA09 temperature climatology (Locarnini et al., 2010), sparse in situ phosphate measurements in some ocean regions (e.g. Southern Ocean) contributes to uncertainty in the WOA09 interpolated monthly distributions

(Garcia et al., 2010b), and therefore the surface-ocean  $pCO_2$  predictions. As uncertainty estimates in WOA09 objectively analysed products remains elusive, quantifying the related uncertainty in SOMLO-derived flux estimates is currently unachievable.

Combining both atmospheric and ocean surface  $pCO_2$  sources of error, along with estimated uncertainties relating to *k*-wind parameterization (±0.2 PgC yr<sup>-1</sup>) and windspeeds (±0.15 PgC yr<sup>-1</sup>) (Wanninkhof et al., 2013), the net uncertainty in the SOMLO-derived global contemporary flux estimate is calculated to be ±0.25 PgC yr<sup>-1</sup> (=  $(0.02^2 + 0.03^2 + 0.2^2 + 0.15^2)^{0.5}$ ).

### 4.9.1 Net-contemporary air-sea CO<sub>2</sub> flux

The SOMLO-derived open-ocean contemporary CO<sub>2</sub> uptake of  $1.10\pm0.25$  PgC yr<sup>-1</sup> for the year 2000 is similar to the estimate of  $1.21\pm0.59$  PgC yr<sup>-1</sup> derived using T-09 *p*CO<sub>2</sub> distributions; where the uncertainty estimate is taken from the most recent error analysis by Wanninkhof et al. (2013). These *p*CO<sub>2</sub>-constrained global contemporary air-sea CO<sub>2</sub> fluxes are a combination of natural, anthropogenic and the non-steady-state CO<sub>2</sub> flux signals (see *Sect. 1.5*). By including an estimated natural steady-state CO<sub>2</sub> out-gassing of 0.45±0.2 PgC yr<sup>-1</sup> from organic matter deposition from rivers (Jacobson et al., 2007b), the estimate for the net-contemporary open-ocean CO<sub>2</sub> sink is 1.55±0.32 PgC yr<sup>-1</sup> for the year 2000.

## 4.9.2 Coastal air-sea CO<sub>2</sub> exchange

Coastal margins account for approximately ~7% of the global ocean surface area and plays host to some of the oceans most biogeochemically active regions (Borges and Co-Authors, 2010). Since flux estimates presented here are for open-ocean regions only, it's important to account for coastal air-sea  $CO_2$  exchange when comparing flux estimates derived via different methods that capture both the open- and coastal- oceanic flux signals. Wanninkhof et al. (2013) scaled-up T-09 contemporary  $CO_2$  fluxes in 23 ocean regions to estimate a coastal-oceanic sink of 0.2 PgC yr<sup>-1</sup> for the nominal year of 2000,

which is consistent with other data-based estimates that suggests the coastal contemporary CO<sub>2</sub> sink ranges between 0.2 to 0.4 PgC yr<sup>-1</sup> (Borges et al., 2005; Cai et al., 2006; Chen and Borges, 2009). A coastal uptake of 0.2 PgC yr<sup>-1</sup> was therefore added to arrive at a final global net-contemporary oceanic CO<sub>2</sub> sink of  $1.75 \pm 0.32$  PgC yr<sup>-1</sup> for the nominal year of 2000.

## 4.10 Conclusion

Inversions of ocean interior data (Gloor et al., 2003; Mikaloff-Fletcher et al., 2006), atmospheric CO<sub>2</sub> inversions (Gurney et al., 2008), atmospheric O<sub>2</sub>/N<sub>2</sub> (Ishidoya et al., 2012; Manning and Keeling, 2006), indirect tracer-based techniques (Khatiwala et al., 2009; McNeil et al., 2003) and ocean general circulation models participating in the RECCAP project (REgional Carbon Cycle Assessment and Processes), suggest ocean anthropogenic CO<sub>2</sub> uptake ranges between 1.9-2.5 PgC yr<sup>-1</sup> for the year 2000 (Table 4.4), somewhat higher than both the underway  $pCO_2$  estimate of 1.83±0.6 PgC yr<sup>-1</sup> (without the under-sampling correction) (Wanninkhof et al., 2013) and the new bottlederived  $pCO_2$  estimate presented here of 1.75±0.32 PgC yr<sup>-1</sup>. It's important to emphasise however, that the new technique constrains the oceans net-contemporary CO<sub>2</sub> uptake, which includes the net effect of both the anthropogenic and non-steadystate CO<sub>2</sub> flux. The discrepancy between the net-contemporary CO<sub>2</sub> constraint and other anthropogenic CO<sub>2</sub> estimates could be due to a range of issues including natural variability in the oceanic CO<sub>2</sub> sink, discrepancies in quantifying the coastal air-sea CO<sub>2</sub> budget, uncertainties in the riverine out-gassing signal, under-sampling biases, or uncertain transport models. Aside from providing a new constraint on the global netcontemporary oceanic CO<sub>2</sub> sink, the SOMLO technique also provides a new independent way to diagnose ocean surface  $pCO_2$  distributions, which will be important in helping understand any future changes in the efficiency of the oceanic CO<sub>2</sub> sink.

Authors	This study (Wanninkhof et al., 2013)		(Mikaloff-Fletcher et al., 2006)	(Gloor et al., 2003)	(Khatiwala et al., 2009)	(McNeil et al., 2003)		(Gurney et al., 2004)	(Peylin et al., 2013)	(Manning and Keeling, 2006)	(Ishidoya et al., 2012)		(Wanninkhof et al., 2013)	lded in quadrature.
Comment	2, <i>difference</i> without under-sampling correction	or Measurements	Mean of 10 models	Mean of 3 models	Green function	CFCs	heric Measurements	Mean of 12 models	Mean of 11 models			OGCM output	Median of 6 models	d to their quoted fluxes. Errors were ad
Time period	<b>irect air-sea pCO</b> Nominal 2000 Nominal 2000	on Ocean interio	Nominal 1995	Nominal 1990	2000 - 2010	1990 – 1999	based on Atmospl	1992 - 1996	2001 - 2004	1993 - 2003	2000 - 2010	imates bases on (	1990 - 2007	carbon was added
Estimate [PgC yr <sup>-1</sup> ]	D -1.75 ± 0.32 -1.83 ± 0.6	Based	$-2.2 \pm 0.32$	$-1.8 \pm 0.4$	$-2.3 \pm 0.6$	$-2.0 \pm 0.4$	Methods l	$-1.79 \pm 0.8^{a}$	$-2.08 \pm 0.53^{a}$	$-2.2 \pm 0.6$	$-2.5\pm0.7$	Est	$-1.9 \pm 0.3$	riverine derived
Flux signal	Net-contemporary Net-contemporary		Steady-state anthro. flux	Steady-state anthro. flux	Steady-state anthro. flux	Steady-state anthro. flux		Net-contemporary	Net-contemporary	Steady-state anthro. flux	Steady-state anthro. flux		Anthropogenic	assing 0.45±0.2 PgC yr <sup>-1</sup> of
Method	SOMLO-predicted T-09 pCO <sub>2</sub>		Ocean Inversion	Ocean Inversion	Transient Tracers	Transient Tracers		Atm. Inversions	Atm. Inversions	O <sub>2</sub> /N <sub>2</sub> ratio	O <sub>2</sub> /N <sub>2</sub> ratio		<b>RECCAP</b> models	<sup>a</sup> A steady-state out-g

**Table 4.4:** Global anthropogenic (anthro.) or net-contemporary air-sea CO2 flux estimates of SOMLO and other independent methods. A negative value represents oceanic uptake.MethodFlux signalEstimateTime periodComment
# Chapter 5.

# **Closing remarks**

## 5.1 Summary of findings

Despite the ocean's importance in absorbing anthropogenic CO<sub>2</sub>, our ability to detect changes and diagnose spatiotemporal carbon distributions is hampered due to large spatiotemporal data deficiencies. In this thesis, I have developed a novel non-linear empirical technique, called SOMLO (Self-Organizing Multiple-Linear Output), that captures mixed-layer inorganic carbon concentrations ( $C_T$ ,  $A_T$  and pCO<sub>2</sub>) from standard hydrographic parameters which have a much greater global data coverage (SHP; temperature, salinity, dissolved oxygen and nutrients). By then applying the model to monthly SHP climatologies has provided a first data-based constraint on all inorganic carbon variables, with immediate implications for advancing our understanding of ocean carbon dynamics. The specific focus in this thesis was on diagnosing patterns of air-sea CO<sub>2</sub> exchange and investigating the influence of natural seasonal variability for future ocean acidification onset and detection. In *Chapter 2*, the SOMLO approach was presented and deployed on ~33,000 mixed-layer bottle measurements collected between 1980 and 2010 to capture global  $C_{\rm T}$  and  $A_{\rm T}$  concentrations. This new dataset represents a ~45% increase in the global measurement network used to derive the empirical relationships in previous global studies (Lee et al., 2000b; 2006). Developing and applying a systematic independent test approach (see *Sect. 2.4*) demonstrated significant improvements in SOMLO predictions over traditional linear ad-hoc approaches; improving global predictive skill by 19.4% for  $C_{\rm T}$ . In particular, the new non-linear method improved accuracy in the most complex and dynamically important ocean regions (i.e. equatorial Pacific and Southern Ocean), reducing errors by up to 30% over traditional approaches. On a global scale, SOMLO is able to predict  $C_{\rm T}$  and  $A_{\rm T}$  to within ±10.9 and ±9.2 µmol kg<sup>-1</sup> respectively. No seasonal bias was found in SOMLO predictions.

To further scrutinize the model's temporal skill, the 18 year Bermuda Atlantic (BATS) and 20 year Hawaiian Ocean (HOT) hydrographic time-series stations were used as independent test-beds. Hydrographic information was shown to capture 90% of the seasonal cycle and inter-annual variability at these locations. Finally, the new model was used to diagnose monthly open-ocean  $C_{\rm T}$  and  $A_{\rm T}$  distributions for the nominal year of 2000 using the World Ocean Atlas 2009 objectively analyzed SHP surface (0m) climatologies (WOA09; Antonov et al., 2010; Garcia et al., 2010a, 2010b; Locarnini et al., 2010).

The new ocean surface carbon distributions were used in *Chapter 3* to better investigate the influence of natural variability on future ocean acidification onset and detection. Under business-as-usual conditions, seasonality was found to bring forward the onset of aragonite under-saturation globally by about 19 years relative to annual-mean estimates, exposing an additional  $24 \times 10^6$  km<sup>2</sup> (7.6% of total open-ocean area) to at least month-long corrosive aragonite conditions by centuries end.

Strong seasonal variability relative to human-induced changes will also hamper our ability to detect robust trends in aragonite saturation state and pH. To quantify detection-times for ocean acidification trends, a similar signal-to-noise approach used by the model-based study of Freidrich et al. (2012) was employed. For repeat measurement programs that began in the early 1990's, seasonality will obscure anthropogenic trends in pH and  $\Omega_{Ar}$  by ~45 and ~60 years in (sub)tropical regions (40°S to 30°N), while stronger variability in higher latitudes will lengthen this periods by ~15 years.

In *Chapter 4*, the new SOMLO technique was applied to bottle-derived  $pCO_2$  measurements to diagnose monthly ocean surface  $pCO_2$  distributions for the nominal year of 2000 via the WOA09. Comparison to the Takahashi et al. (2009) "underway"  $pCO_2$  climatology (herein after referred to as T-09) revealed broad consistencies between the two independent data-based constraints; showing a global correlation of 0.64. However, large discrepancies were found in the Southern Ocean and equatorial Pacific, which are likely related to spatiotemporal measurement biases coupled to strong seasonal to inter-annual variability; particularly in the equatorial Pacific (Feely et al., 2002).

The bottle-derived  $pCO_2$  distributions where then used to diagnose patterns of contemporary air-sea CO<sub>2</sub> exchange via the flux equation of Wanninkhof et al. (2013) and CCMP 10m wind fields of Atlas et al. (2011). This new and independent constraint suggests a much stronger CO<sub>2</sub> out-gassing in the equatorial Pacific and large uptake signal in the Southern Ocean relative to the T-09 "underway"  $pCO_2$  climatology. Despite this, the integrated open-ocean CO<sub>2</sub> sink was found to be consistent with the global underway  $pCO_2$  database, suggesting a contemporary oceanic CO<sub>2</sub> uptake of 1.10±0.25 PgC yr<sup>-1</sup> for the nominal year of 2000.

By finally accounting for the stead-state natural out-gassing of riverine-derived carbon ( $0.45\pm0.2$  PgC yr<sup>-1</sup>; Jacobson et al. (2007a)) and coastal CO<sub>2</sub> uptake (0.2 PgC yr<sup>-1</sup>; Wanninkhof et al. (2012)), the net-contemporary global oceanic CO<sub>2</sub> sink was estimated to be  $1.75\pm0.32$  PgC yr<sup>-1</sup> for the nominal year of 2000. Although this estimate is slightly lower than other anthropogenic flux estimates (see Table 4.3), it is consistent with other studies that suggest the global oceanic sink is decreasing (e.g., McNeil and Matear, 2012; Sarmiento et al., 2010).

#### 5.2 Significance and way forward

#### 5.2.1 Understanding ocean carbon dynamics

Predicting how the ocean will respond to a high  $CO_2$  world requires a vast spatiotemporal carbon measurement network from which to diagnose drivers of variability and validate numerical model output. Despite a significant growth in global measurements over recent decades, large spatial and temporal data deficiencies limit our ability to achieve these goals.

In this thesis, I have diagnosed monthly ocean surface carbon distributions for the nominal year of 2000 using a novel empirical method. This new data-based constraint presents an important tool for advancing our understanding of ocean carbon dynamics and validating model output. For example, the thermal influence on oceanic  $pCO_2$  can be quantified following the method outlined by Takahashi et al. (2002), while other studies have applied decompositions to separate the influence of temperature, salinity,  $C_T$  and  $A_T$  on oceanic  $pCO_2$  variability (e.g., Lenton et al., 2012; McNeil and Matear, 2008). From a more fundamental level, the distribution of weighting vectors in the self-organizing map and multiple-linear regression coefficients contain valuable information regarding the importance and magnitude of each predictor variable in driving the ocean carbon system. This type of statistical analysis is one possible focus for future research. Beyond diagnosing carbon climatologies, the SOMLO model provides a powerful tool for predicting mixed-layer carbon concentrations from SHP. As the global network of SHP measurements increases exponentially due to autonomous measuring devices (e.g. the Argo array of temperature/salinity profiling floats), empirical carbon predictions will become increasingly important for understanding the oceans evolving carbon system and inter-annual variability. Other SHP time-series stations, such as the array of hydrographic stations around Australia (Critchley et al., 2009), presents attractive opportunities to apply the SOMLO model to capture seasonal to inter-annual inorganic carbon variability, thereby providing a platform for diagnosing drivers of variability, and identifying potential future implications for marine ecosystems.

#### 5.2.2 Constraining biogeochemical provinces

Defining biogeochemical provinces is an important aspect when grouping sporadic measurements to detect robust secular trends (e.g., in the "underway"  $pCO_2$  network). Previous approaches have solely relied on physical processes (mixed-layer depth and upwelling) and somewhat subjective geographical partitioning to establish a global province array (Longhurst et al., 1995; Sarmiento et al., 2004). Here, the ability of the SOM model in clustering measurements with similar "biogeochemical fingerprints" has been demonstrated. This presents the possibility to apply the neural network algorithm to global distributions of biogeochemical provinces. In doing so, it could provide a more accurate method for detecting secular trends in sporadic measurements by reducing the influence of measurements that are not consistent with the larger "biogeochemical fingerprint" within a province.

#### 5.2.3 Influence of natural seasonal variability

Traditional studies investigating future aragonite under-saturation have solely focused on temporal aspects (e.g., McNeil and Matear, 2008; Orr et al., 2005). Here, I have presented a first study probing the influence of natural variability in a geographical context. This revealed a significant spatial expansion  $(24 \times 10^6 \text{ km}^2)$  of at least monthlong corrosive aragonite conditions relative to annual-mean estimates by centuries end. For multiple calcifying organisms that inhabit these regions, for example some pteropod species (Hunt et al., 2008), this represents a major challenge in preserving and secreting their aragonite shells and skeletons over the coming century. Since oceanic CaCO<sub>3</sub> plays an important role in the global carbon cycle, the implications of ocean acidification presented here provides new insight into the future system, thereby allowing us to better understand and model the changing carbon cycle.

The estimated detection-times will also aid in the direction of available funds and resources when determining the location and experimental set up for measurement programs aimed at detecting ocean acidification trends.

It's important to note that air-sea  $pCO_2$  disequilibrium will likely affect the future pH and  $\Omega_{Ar}$  predictions presented here, which assume steady-state. In order to account for this influence, ocean general circulation models will be incorporated into future work to better predict ocean acidification properties.

#### 5.2.4 A new constraint on air-sea CO<sub>2</sub> flux

Despite significant advances in diagnosing air-sea  $CO_2$  flows over the last decade, uncertain transport models and spatiotemporal biases in "underway"  $pCO_2$ measurements contribute to large uncertainties in regional estimates. These issues are addressed in this thesis by presenting a new observationally-derived ocean surface climatology for the partial pressure of  $CO_2$  ( $pCO_2$ ), which provides an independent constraint on the patterns of contemporary air-sea  $CO_2$  fluxes. When integrated globally, the contemporary air-sea  $CO_2$  flux estimate is consistent with the global "underway"  $pCO_2$  network. However, the new results suggests a strong shift in the magnitude of oceanic sources and sinks, particularly in the Southern Ocean and equatorial Pacific. Although the cause of these discrepancies is likely related to 162 spatiotemporal data limitations, the exponential increase in satellite and autonomous measurements will dramatically enhance the ability of empirical models, like SOMLO, to constrain the patterns of air-sea  $CO_2$  fluxes over the coming years.

Empirical  $pCO_2$  predictions will also become increasing important for capturing changes in the oceanic  $CO_2$  sink, and therefore better informing us on the ocean's future role in mediating human-induced climate change.

#### 5.3 Acknowledgements

This work would not be possible without the efforts of all captains, crew, technicians and researchers who measured and synthesised the global carbon and hydrographic datasets used in this study. In particular, I would thank the researchers involved in the Global Ocean Data Analysis Project (GLODAP; Key et al., 2004), CARbon dioxide IN the Atlantic project (CARINA; Key et al., 2010) and the PACIFic ocean Interior CArbon project (PACIFICA; Suzuki et al., 2013). All training of the SOMLO model was conducted using the R-statistical project software (R Development Core Team: R: A Language and Environment for Statistical Computing, http://www.R-project.org, 2012), along with the kohonen, ggplot2 and akima R packages developed by Wehrens and Buydens (2007), Wickham (2009) and Akima et al. (akima: Interpolation of irregularly spaced data, http://CRAN.R-project.org/package=akima, 2012), respectively. I would also like to thank the developers of the Ocean Data Viewer (ODV) program (Schlitzer, R.: Ocean Data View, http://odv.awi.de, 2011), from which many of the figures presented in this thesis were developed.

## References

Abramowitz, G. (2005), Towards a benchmark for land surface models, Geophys. Res. Lett., 32, L22702, DOI: 10.1029/2005gl024419

Amante, C., and Eakins, B. W. (2009), ETOPO1 1 Arc-Minute Global Relief Model: Procedures, Data Sources and Analysis, NOAA Technical Memorandum NESDIS NGDC-24, 19 pp, March

Anderson, L. A., and Sarmiento, J. L. (1994), Redfield ratios of remineralization determined by nutrient data analysis, Global Biogeochem. Cycles, 8, 65-80, DOI: 10.1029/93gb03318

Andres, R. J., Boden, T. A., Bréon, F. M., Ciais, P., Davis, S., Erickson, D., Gregg, J. S., Jacobson, A., Marland, G., Miller, J., Oda, T., Olivier, J. G. J., Raupach, M. R., Rayner, P., and Treanton, K. (2012), A synthesis of carbon dioxide emissions from fossil-fuel combustion, Biogeosciences, 9, 1845-1871, DOI: 10.5194/bg-9-1845-2012

Antonov, J. I., Seidov, D., Boyer, T. P., Locarnini, R. A., Mishonov, A. V., Garcia, H. E., Baranova, O. K., Zweng, M. M., and Johnson, D. R. (2010), World Ocean Atlas 2009, Volume 2: Salinity, in, edited by: Levitus, S., NOAA Atlas NESDIS 69, U.S. Government Printing Office, Washington, D.C., 184 pp.

Archer, D., Kheshgi, H., Maier, and Reimer, E. (1997), Multiple timescales for neutralization of fossil fuel CO<sub>2</sub>, Geophys. Res. Lett., 24, 405-408, DOI: 10.1029/97gl00168

Archer, D., Eby, M., Brovkin, V., Ridgwell, A., Cao, L., Mikolajewicz, U., Caldeira, K., Matsumoto, K., Munhoven, G., Montenegro, A., and Tokos, K. (2009), Atmospheric Lifetime of Fossil Fuel Carbon Dioxide, Annual Review of Earth and Planetary Sciences, 37, 117-134, DOI: 10.1146/annurev.earth.031208.100206

Arrigo, K. R., Pabi, S., van Dijken, G. L., and Maslowski, W. (2010), Air-sea flux of  $CO_2$  in the Arctic Ocean, 1998-2003, J. Geophys. Res., 115, G04024, DOI: 10.1029/2009jg001224

Atlas, R., Hoffman, R. N., Ardizzone, J., Leidner, S. M., Jusem, J. C., Smith, D. K., and Gombos, D. (2011), A Cross-calibrated, Multiplatform Ocean Surface Wind Velocity Product for Meteorological and Oceanographic Applications, Bulletin of the American Meteorological Society, 92, 157-174, DOI: 10.1175/2010bams2946.1

Baker, D. F., Law, R. M., Gurney, K. R., Rayner, P., Peylin, P., Denning, A. S., Bousquet, P., Bruhwiler, L., Chen, Y. H., Ciais, P., Fung, I. Y., Heimann, M., John, J.,

Maki, T., Maksyutov, S., Masarie, K., Prather, M., Pak, B., Taguchi, S., and Zhu, Z. (2006), TransCom 3 inversion intercomparison: Impact of transport model errors on the interannual variability of regional CO<sub>2</sub> fluxes, 1988-2003, Global Biogeochem. Cycles, 20, GB1002, DOI: 10.1029/2004gb002439

Bates, N. R., Michaels, A. F., and Knap, A. H. (1996), Seasonal and interannual variability of oceanic carbon dioxide species at the U.S. JGOFS Bermuda Atlantic Time-series Study (BATS) site, Deep Sea Research Part II: Topical Studies in Oceanography, 43, 347-383, DOI: 10.1016/0967-0645(95)00093-3

Bates, N. R., Pequignet, A. C., and Sabine, C. L. (2006), Ocean carbon cycling in the Indian Ocean: 1. Spatiotemporal variability of inorganic carbon and air-sea CO<sub>2</sub> gas exchange, Global Biogeochem. Cycles, 20, GB3020, DOI: 10.1029/2005gb002491

Bates, N. R. (2007), Interannual variability of the oceanic  $CO_2$  sink in the subtropical gyre of the North Atlantic Ocean over the last 2 decades, J. Geophys. Res., 112, C09013, DOI: 10.1029/2006jc003759

Behrenfeld, M. J., O/'Malley, R. T., Siegel, D. A., McClain, C. R., Sarmiento, J. L., Feldman, G. C., Milligan, A. J., Falkowski, P. G., Letelier, R. M., and Boss, E. S. (2006), Climate-driven trends in contemporary ocean productivity, Nature, 444, 752-755, DOI: 10.1038/nature05317

Bender, M. L., Ho, D. T., Hendricks, M. B., Mika, R., Battle, M. O., Tans, P. P., Conway, T. J., Sturtevant, B., and Cassar, N. (2005), Atmospheric  $O_2/N_2$  changes, 1993-2002: Implications for the partitioning of fossil fuel CO<sub>2</sub> sequestration, Global Biogeochem. Cycles, 19, GB4017, DOI: 10.1029/2004gb002410

Berelson, W. M., Balch, W. M., Najjar, R., Feely, R. A., Sabine, C., and Lee, K. (2007), Relating estimates of CaCO<sub>3</sub> production, export, and dissolution in the water column to measurements of CaCO<sub>3</sub> rain into sediment traps and dissolution on the sea floor: A revised global carbonate budget, Global Biogeochem. Cycles, 21, GB1024, DOI: 10.1029/2006gb002803

Boden, T. A., Marland, G., and Andres, R. J. (2012), Global, Regional, and National Fossil-Fuel CO<sub>2</sub> Emissions, DOI: 10.3334/CDIAC/00001\_V2012

Borges, A., and Co-Authors: A Global Sea Surface Carbon Observing System: Inorganic and Organic Carbon Dynamics in Coastal Oceans, Proceedings of OceanObs'09: Sustained Ocean Observations and Information for Society (Vol. 2), Venice, Italy, 21-25 September 2009, 2010.

Borges, A. V., Delille, B., and Frankignoulle, M. (2005), Budgeting sinks and sources of CO<sub>2</sub> in the coastal ocean: Diversity of ecosystems counts, Geophys. Res. Lett., 32, L14601, DOI: 10.1029/2005gl023053

Bowman, K. P., and Cohen, P. J. (1997), Interhemispheric Exchange by Seasonal Modulation of the Hadley Circulation, Journal of the Atmospheric Sciences, 54, 2045-2059, DOI: 10.1175/1520-0469(1997)054<2045:iebsmo>2.0.co;2

Boyce, D. G., Lewis, M. R., and Worm, B. (2010), Global phytoplankton decline over the past century, Nature, 466, 591-596, DOI: 10.1038/nature09268

Boyd, P. W., and Doney, S. C. (2002), Modelling regional responses by marine pelagic ecosystems to global climate change, Geophys. Res. Lett., 29, 1806, DOI: 10.1029/2001gl014130

Boyd, P. W., and Trull, T. W. (2007), Understanding the export of biogenic particles in oceanic waters: Is there consensus?, Progress In Oceanography, 72, 276-312, DOI: 10.1016/j.pocean.2006.10.007

Boyd, P. W. (2009), Ocean iron cycle, in: Surface ocean-lower atmosphere processes, edited by: Le Quéré, C., and Saltzman, E. C., American Geophysical Union, Washinton, DC, 161-179 pp.

Boyer, T. P., Antonov, J. I., Baranova, O. K., Garcia, H. E., Johnson, D. R., Locarnini, R. A., Mishonov, A. V., O'Brien, T. D., Seidov, D., Smolyar, I. V., and Zweng, M. M. (2009), World Ocean Database 2009, in, edited by: Levitus, S., Ed., NOAA Atlas NESDIS 66, U.S. Gov. Printing Office, Wash., D.C., 216 pp.

Bradshaw, A. L., Brewer, P. G., Shafer, D. K., and Williams, R. T. (1981), Measurements of total carbon dioxide and alkalinity by potentiometric titration in the GEOSECS program, Earth and Planetary Science Letters, 55, 99-115, DOI: 10.1016/0012-821x(81)90090-x

Brix, H., Gruber, N., and Keeling, C. D. (2004), Interannual variability of the upper ocean carbon cycle at station ALOHA near Hawaii, Global Biogeochem. Cycles, 18, GB4019, DOI: 10.1029/2004gb002245

Brix, H., Currie, K. I., and Mikaloff Fletcher, S. E. (2013), Seasonal Variability of the Carbon Cycle in Sub-Antarctic Surface Water in the South West Pacific, Global Biogeochem. Cycles, n/a-n/a, DOI: 10.1002/gbc.20023

Brydon, D., Sun, S., and Bleck, R. (1999), A new approximation of the equation of state for seawater, suitable for numerical ocean models, J. Geophys. Res., 104, 1537-1540, DOI: 10.1029/1998JC900059

Byrne, R. H., Mecking, S., Feely, R. A., and Liu, X. (2010), Direct observations of basin-wide acidification of the North Pacific Ocean, Geophys. Res. Lett., 37, DOI: 10.1029/2009GL040999

Cai, W., Dai, M., and Wang, Y. (2006), Air-sea exchange of carbon dioxide in ocean margins: A province-based synthesis, Geophys. Res. Lett., 33, L12603, DOI: 10.1029/2006gl026219

Caldeira, K., and Wickett, M. E. (2003), Oceanography: Anthropogenic carbon and ocean pH, Nature, 425, 365-365, DOI: 10.1038/425365a

Cameron, D. R., Lenton, T. M., Ridgwell, A., Shepherd, J. G., Marsh, R., and Yool, A. (2005), A factorial analysis of the marine carbon cycle and ocean circulation controls on atmospheric CO<sub>2</sub>, Global Biogeochem. Cycles, 19, GB4027, DOI: 10.1029/2005gb002489

CARINA Group (2009a), Carbon in the Atlantic Ocean Region - the CARINA project: and Version Results Data. 1.0. http://cdiac.ornl.gov/ftp/oceans/CARINA/CARINA Database/CARINA.ATL.V1.0/, Carbon Dioxide Information Analysis Center, Oak Ridge National Laboratory, U.S. Oak Ridge. DOI: Department of Energy, Tennessee. 10.3334/CDIAC/otg.CARINA.ATL.V1.0

CARINA Group (2009b), Carbon in the Arctic Mediterranean Seas Region - the **CARINA** project: Results and Data. Version 1.2. http://cdiac.ornl.gov/ftp/oceans/CARINA/CARINA Database/CARINA.AMS.V1.2/, Carbon Dioxide Information Analysis Center, Oak Ridge National Laboratory, U.S. Department Energy, Oak Ridge, Tennessee, of DOI: 10.3334/CDIAC/otg.CARINA.AMS.V1.2

CARINA Group (2010), Carbon in the Southern Ocean Region - the CARINA project: Version Results and Data, 1.1, http://cdiac.ornl.gov/ftp/oceans/CARINA/CARINA Database/CARINA.SO.V1.1/, Carbon Dioxide Information Analysis Center, Oak Ridge National Laboratory, U.S. Department of Energy, Oak Ridge, Tennessee, DOI: 10.3334/CDIAC/otg.CARINA.SO.V1.1

Chen, C.-T. A., and Borges, A. V. (2009), Reconciling opposing views on carbon cycling in the coastal ocean: Continental shelves as sinks and near-shore ecosystems as sources of atmospheric CO<sub>2</sub>, Deep Sea Research Part II: Topical Studies in Oceanography, 56, 578-590, DOI: 10.1016/j.dsr2.2009.01.001

Chen, L., Xu, S., Gao, Z., Chen, H., Zhang, Y., Zhan, J., and Li, W. (2011), Estimation of monthly air-sea  $CO_2$  flux in the southern Atlantic and Indian Ocean using in-situ and remotely sensed data, Remote Sensing of Environment, 115, 1935-1941, DOI: 10.1016/j.rse.2011.03.016

Chierici, M., Olsen, A., Johannessen, T., Trinañes, J., and Wanninkhof, R. (2009), Algorithms to estimate the carbon dioxide uptake in the northern North Atlantic using shipboard observations, satellite and ocean analysis data, Deep Sea Research Part II: Topical Studies in Oceanography, 56, 630-639, DOI: 10.1016/j.dsr2.2008.12.014 Cooley, S. R., Coles, V. J., Subramaniam, A., and Yager, P. L. (2007), Seasonal variations in the Amazon plume-related atmospheric carbon sink, Global Biogeochem. Cycles, 21, GB3014, DOI: 10.1029/2006gb002831

Cosca, C. E., Feely, R. A., Boutin, J., Etcheto, J., McPhaden, M. J., Chavez, F. P., and Strutton, P. G. (2003), Seasonal and interannual CO<sub>2</sub> fluxes for the central and eastern equatorial Pacific Ocean as determined from fCO<sub>2</sub>-SST relationships, J. Geophys. Res., 108, 3278, DOI: 10.1029/2000jc000677

Cotrim da Cunha, L., Buitenhuis, E. T., Le Quéré, C., Giraud, X., and Ludwig, W. (2007), Potential impact of changes in river nutrient supply on global ocean biogeochemistry, Global Biogeochem. Cycles, 21, GB4007, DOI: 10.1029/2006gb002718

Critchley, G. W., Abell, G., Allen, S., Berry, K., Bonham, P., Clementson, L., Coman, F., Richardson, A., Robert, S., Terhell, D., Thompson, P., Tilbrook, B., Seuront, L., Pritchard, T., Bainbridge, S., Skuza, M., Steinberg, C., Magierowski, R., Roberts, K., and Tattersall, K. (2009), Interated Marine Observing System, edited by: Critchley, G. W., CSIRO Marine and Atmospheric Research Laboratories, 102 pp.

de Boyer, M. C., Madec, G., Fischer, A. S., Lazar, A., and Iudicone, D. (2004), Mixed layer depth over the global ocean: An examination of profile data and a profile-based climatology, J. Geophys. Res., 109, C12003, DOI: 10.1029/2004jc002378

Department of Energy (1994), Handbook of methods for the analysis of the various parameters of the carbon dioxide system in sea water. Version 2, edited by: Dickson, A. G., and Goyet, C., ORNL/CDIAC-74, Carbon Dioxide Inf. and Anal. Cent., Oak Ridge, Natl. Lab., Oak Ridge, Tenn.

Dickson, A. G., and Millero, F. J. (1987), A comparison of the equilibrium constants for the dissociation of carbonic acid in seawater media, Deep Sea Research Part A. Oceanographic Research Papers, 34, 1733-1743, DOI: 10.1016/0198-0149(87)90021-5

Dickson, A. G. (1990a), Standard potential of the reaction:  $AgCl(s) + 1/2H_2(g) = Ag(s) + HCl(aq)$ , and and the standard acidity constant of the ion  $HSO_4^-$  in synthetic sea water from 273.15 to 318.15 K, The Journal of Chemical Thermodynamics, 22, 113-127, DOI: 10.1016/0021-9614(90)90074-Z

Dickson, A. G. (1990b), Thermodynamics of the dissociation of boric acid in synthetic seawater from 273.15 to 318.15 K, Deep Sea Research Part A. Oceanographic Research Papers, 37, 755-766, DOI: 10.1016/0198-0149(90)90004-F

Dickson, A. G. (1990c), The oceanic carbon dioxide system: planning for quality data, US JGOFS News, 2, 2

Dickson, A. G., Afghan, J. D., and Anderson, G. C. (2003), Reference materials for oceanic  $CO_2$  analysis: a method for the certification of total alkalinity, Marine Chemistry, 80, 185-197, DOI: 10.1016/s0304-4203(02)00133-0

Dickson, A. G., Sabine, C. L., and Christian, J. R. (Eds.) (2007), Guide to best practices for ocean CO<sub>2</sub> measurements, PICES Special Publication 3, 191 pp.

Dlugokencky, E., and Tans, P. (2012), Trends in atmospheric carbon dioxide, www.esrl.noaa.gov/gmd/ccgg/trends/, NOAA/ESRL,

Doney, S. C., Fabry, V. J., Feely, R. A., and Kleypas, J. A. (2009), Ocean Acidification: The Other CO<sub>2</sub> Problem, Annual Review of Marine Science, 1, 169-192, DOI: 10.1146/annurev.marine.010908.163834

Dore, J. E., Lukas, R., Sadler, D. W., and Karl, D. M. (2003), Climate-driven changes to the atmospheric CO<sub>2</sub> sink in the subtropical North Pacific Ocean, Nature, 424, 754-757, DOI: 10.1038/nature01885

Durack, P. J., and Wijffels, S. E. (2010), Fifty-Year Trends in Global Ocean Salinities and Their Relationship to Broad-Scale Warming, Journal of Climate, 23, 4342-4362, DOI: 10.1175/2010jcli3377.1

Dyurgerov, M. B., and Carter, C. L. (2004), Observational Evidence of Increases in Freshwater Inflow to the Arctic Ocean, Arctic, Antarctic, and Alpine Research, 36, 117-122, DOI: 10.1657/1523-0430(2004)036[0117:OEOIIF]2.0.CO;2

Egleston, E. S., Sabine, C. L., and Morel, F. M. M. (2010), Revelle revisited: Buffer factors that quantify the response of ocean chemistry to changes in DIC and alkalinity, Global Biogeochem. Cycles, 24, GB1002, DOI: 10.1029/2008gb003407

Etheridge, D. M., Steele, L. P., Langenfelds, R. L., Francey, R. J., Barnola, J. M., and Morgan, V. I. (1996), Natural and anthropogenic changes in atmospheric CO<sub>2</sub> over the last 1000 years from air in Antarctic ice and firn, J. Geophys. Res., 101, 4115-4128, DOI: 10.1029/95jd03410

Fabry, V. J., Seibel, B. A., Feely, R. A., and Orr, J. C. (2008), Impacts of ocean acidification on marine fauna and ecosystem processes, ICES Journal of Marine Science / Journal du Conseil, 65, 414-432, DOI: 10.1093/icesjms/fsn048

Falkowski, P., Scholes, R. J., Boyle, E., Canadell, J., Canfield, D., Elser, J., Gruber, N., Hibbard, K., Högberg, P., Linder, S., Mackenzie, F. T., Moore III, B., Pedersen, T., Rosenthal, Y., Seitzinger, S., Smetacek, V., and Steffen, W. (2000), The Global Carbon Cycle: A Test of Our Knowledge of Earth as a System, Science, 290, 291-296, DOI: 10.1126/science.290.5490.291

Falkowski, P. G., Barber, R. T., and Smetacek, V. (1998), Biogeochemical Controls and Feedbacks on Ocean Primary Production, Science, 281, 200-206, DOI: 10.1126/science.281.5374.200

Faraway, J. J. (2004), Linear Models with R, 1 ed., Chapman and Hall/CRC, 242 pp.

Feely, R. A., Boutin, J., Cosca, C. E., Dandonneau, Y., Etcheto, J., Inoue, H. Y., Ishii, M., Le Quéré, C., Mackey, D. J., McPhaden, M. J., Metzl, N., Poisson, A., and Wanninkhof, R. (2002), Seasonal and interannual variability of  $CO_2$  in the equatorial Pacific, Deep Sea Research Part II: Topical Studies in Oceanography, 49, 2443-2469, DOI: 10.1016/s0967-0645(02)00044-9

Feely, R. A., Sabine, C. L., Lee, K., Berelson, W., Kleypas, J., Fabry, V. J., and Millero, F. J. (2004), Impact of Anthropogenic CO<sub>2</sub> on the CaCO<sub>3</sub> System in the Oceans, Science, 305, 362-366, DOI: 10.1126/science.1097329

Feely, R. A., Takahashi, T., Wanninkhof, R., McPhaden, M. J., Cosca, C. E., Sutherland, S. C., and Carr, M.-E. (2006), Decadal variability of the air-sea CO<sub>2</sub> fluxes in the equatorial Pacific Ocean, J. Geophys. Res., 111, C08S90, DOI: 10.1029/2005jc003129

Feely, R. A., Sabine, C. L., Byrne, R. H., Millero, F. J., Dickson, A. G., Wanninkhof, R., Murata, A., Miller, L. A., and Greeley, D. (2012), Decadal changes in the aragonite and calcite saturation state of the Pacific Ocean, Global Biogeochem. Cycles, 26, DOI: 10.1029/2011GB004157

Forster, P., Ramaswamy, V., Artaxo, P., Berntsen, T., Betts, R. A., Fahey, D. W., Haywood, J., Lean, J., Lowe, D. C., Myhre, G., Nganga, J., Prinn, R. G., Raga, G., Schulz, M., and van Dorland, R. (2007), Changes in Atmospheric Constituents and in Radiative Forcing, Climate Change 2007: The Physical Science Basis. Contribution of Working Group I to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change, edited by: Solomon, S., Qin, D., Manning, M., Chen, Z., Marquis, M., Averyt, K. B., Tignor, M., and Miller, H. L., Cambridge University Press, Cambridge, United Kingdom and New York, USA

Friedlingstein, P., Cox, P., Betts, R., Bopp, L., von Bloh, W., Brovkin, V., Cadule, P., Doney, S., Eby, M., Fung, I., Bala, G., John, J., Jones, C., Joos, F., Kato, T., Kawamiya, M., Knorr, W., Lindsay, K., Matthews, H. D., Raddatz, T., Rayner, P., Reick, C., Roeckner, E., Schnitzler, K. G., Schnur, R., Strassmann, K., Weaver, A. J., Yoshikawa, C., and Zeng, N. (2006), Climate–Carbon Cycle Feedback Analysis: Results from the C<sup>4</sup>MIP Model Intercomparison, Journal of Climate, 19, 3337-3353, DOI: 10.1175/jcli3800.1

Friedlingstein, P., Houghton, R. A., Marland, G., Hackler, J., Boden, T. A., Conway, T. J., Canadell, J. G., Raupach, M. R., Ciais, P., and Le Quere, C. (2010), Update on CO<sub>2</sub> emissions, Nature Geosci, 3, 811-812, DOI: 10.1038/ngeo1022

Friedrich, T., and Oschlies, A. (2009a), Basin-scale pCO<sub>2</sub> maps estimated from ARGO float data: A model study, J. Geophys. Res., 114, C10012, DOI: 10.1029/2009jc005322

Friedrich, T., and Oschlies, A. (2009b), Neural network-based estimates of North Atlantic surface  $pCO_2$  from satellite data: A methodological study, J. Geophys. Res., 114, C03020, DOI: 10.1029/2007jc004646

Friedrich, T., Timmermann, A., Abe-Ouchi, A., Bates, N. R., Chikamoto, M. O., Church, M. J., Dore, J. E., Gledhill, D. K., Gonzalez-Davila, M., Heinemann, M., Ilyina, T., Jungclaus, J. H., McLeod, E., Mouchet, A., and Santana-Casiano, J. M. (2012), Detecting regional anthropogenic trends in ocean acidification against natural variability, Nature Clim. Change, 2, 167-171, DOI: 10.1038/nclimate1372

Gade, K. (2010), A Non-singular Horizontal Position Representation, Journal of Navigation, 63, 395 - 417, DOI: 10.1017/S0373463309990415

Garcia, H. E., Locarnini, R. A., Boyer, T. P., Antonov, J. I., Baranova, O. K., Zweng, M. M., and Johnson, D. R. (2010a), World Ocean Atlas 2009, Volume 3: Dissolved Oxygen Apparent Oxygen Utilization, and Oxygen Saturation, in, edited by: Levitus, S., NOAA Atlas NESDIS 70, U.S. Government Printing Office, Washington, D.C., 344 pp.

Garcia, H. E., Locarnini, R. A., Boyer, T. P., Antonov, J. I., Zweng, M. M., Baranova, O. K., and Johnson, D. R. (2010b), World Ocean Atlas 2009, Volume 4: Nutrients (phosphate, nitrate, silicate), in, edited by: Levitus, S., NOAA Atlas NESDIS 71, U.S. Government Printing Office, Washington, D.C., 398 pp.

Gattuso, J. P., Frankignoulle, M., Bourge, I., Romaine, S., and Buddemeier, R. W. (1998), Effect of calcium carbonate saturation of seawater on coral calcification, Global and Planetary Change, 18, 37-46, DOI: 10.1016/S0921-8181(98)00035-6

George, M. D., Dileep Kumar, M., Naqvi, S. W. A., Banerjee, S., Narvekar, P. V., de Sousa, S. N., and Jayakumar, D. A. (1994), A study of the carbon dioxide system in the northern Indian Ocean during premonsoon, Marine Chemistry, 47, 243-254, DOI: 10.1016/0304-4203(94)90023-X

Gibbs, M. T., Hobday, A. J., Sanderson, B., and Hewitt, C. L. (2006), Defining the seaward extent of New Zealand's coastal zone, Estuarine, Coastal and Shelf Science, 66, 240-254, DOI: 10.1016/j.ecss.2005.08.015

Gloor, M., Gruber, N., Sarmiento, J., Sabine, C. L., Feely, R. A., and Rödenbeck, C. (2003), A first estimate of present and preindustrial air-sea CO<sub>2</sub> flux patterns based on ocean interior carbon measurements and models, Geophys. Res. Lett., 30, 1010, DOI: 10.1029/2002gl015594

González-Dávila, M., Santana-Casiano, J. M., Rueda, M.-J., Llinás, O., and González-Dávila, E.-F. (2003), Seasonal and interannual variability of sea-surface carbon dioxide species at the European Station for Time Series in the Ocean at the Canary Islands (ESTOC) between 1996 and 2000, Global Biogeochem. Cycles, 17, 1076, DOI: 10.1029/2002gb001993

González-Dávila, M., Santana-Casiano, J. M., Rueda, M. J., and Llinás, O. (2010), The water column distribution of carbonate system variables at the ESTOC site from 1995 to 2004, Biogeosciences, 7, 3067-3081, DOI: 10.5194/bg-7-3067-2010

Gruber, N., Sarmiento, J. L., and Stocker, T. F. (1996), An improved method for detecting anthropogenic  $CO_2$  in the oceans, Global Biogeochem. Cycles, 10, 809-837, DOI: 10.1029/96gb01608

Gruber, N., and Sarmiento, J. L. (1997), Global patterns of marine nitrogen fixation and denitrification, Global Biogeochem. Cycles, 11, 235-266, DOI: 10.1029/97gb00077

Gruber, N., Gloor, M., Mikaloff Fletcher, S. E., Doney, S. C., Dutkiewicz, S., Follows, M. J., Gerber, M., Jacobson, A. R., Joos, F., Lindsay, K., Menemenlis, D., Mouchet, A., Müller, S. A., Sarmiento, J. L., and Takahashi, T. (2009), Oceanic sources, sinks, and transport of atmospheric CO<sub>2</sub>, Global Biogeochem. Cycles, 23, GB1005, DOI: 10.1029/2008gb003349

Gurney, K. R., Law, R. M., Denning, A. S., Rayner, P. J., Pak, B. C., Baker, D., Bousquet, P., Bruhwiler, L., Chen, Y.-H., Ciais, P., Fung, I. Y., Heimann, M., John, J., Maki, T., Maksyutov, S., Peylin, P., Prather, M., and Taguchi, S. (2004), Transcom 3 inversion intercomparison: Model mean results for the estimation of seasonal carbon sources and sinks, Global Biogeochem. Cycles, 18, GB1010, DOI: 10.1029/2003gb002111

Gurney, K. R., Baker, D., Rayner, P. J., and Denning, S. (2008), Interannual variations in continental-scale net carbon exchange and sensitivity to observing networks estimated from atmospheric CO<sub>2</sub> inversions for the period 1980 to 2005, Global Biogeochem. Cycles, 22, GB3025, DOI: 10.1029/2007gb003082

Hales, B., Strutton, P. G., Saraceno, M., Letelier, R., Takahashi, T., Feely, R., Sabine, C., and Chavez, F. (2012), Satellite-based prediction of pCO<sub>2</sub> in coastal waters of the eastern North Pacific, Progress In Oceanography, 103, 1-15, DOI: 10.1016/j.pocean.2012.03.001

Hall, T. M., Haine, T. W. N., and Waugh, D. W. (2002), Inferring the concentration of anthropogenic carbon in the ocean from tracers, Global Biogeochem. Cycles, 16, 1131, DOI: 10.1029/2001gb001835

Hartigan, J. A., and Wong, M. A. (1979), Algorithm AS 136: A K-Means Clustering Algorithm, Journal of the Royal Statistical Society. Series C (Applied Statistics), 28, 100-108, DOI: 10.2307/2346830

Haykin, S. (1999), Neural networks: a comprehensive foundation, Prentice Hall, 842 pp.

Hofmann, D. J., Butler, J. H., Dlugokencky, E. J., Elkins, J. W., Masarie, K., Montzka, S. A., and Tans, P. (2006), The role of carbon dioxide in climate forcing from 1979 to 2004: introduction of the Annual Greenhouse Gas Index, Tellus B, 58, 614-619, DOI: 10.1111/j.1600-0889.2006.00201.x

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 B, 55, 378-390, DOI: 10.1034/j.1600-0889.2003.01450.x

Hsu, K., Gupta, H. V., Gao, X., Sorooshian, S., and Imam, B. (2002), Self-organizing linear output map (SOLO): An artificial neural network suitable for hydrologic modeling and analysis, Water Resour. Res., 38, 1302, DOI: 10.1029/2001wr000795

Hunt, B. P. V., Pakhomov, E. A., Hosie, G. W., Siegel, V., Ward, P., and Bernard, K. (2008), Pteropods in Southern Ocean ecosystems, Progress In Oceanography, 78, 193-221, DOI: 10.1016/j.pocean.2008.06.001

Ishidoya, S., Aoki, S., Goto, D., Nakazawa, T., Taguchi, S., and Patra, P. K. (2012), Time and space variations of the  $O_2/N_2$  ratio in the troposphere over Japan and estimation of the global  $CO_2$  budget for the period 2000-2010, Tellus B, 64, DOI: 10.3402/tellusb.v64i0.18964

Ishii, M., Inoue, H. Y., Midorikawa, T., Saito, S., Tokieda, T., Sasano, D., Nakadate, A., Nemoto, K., Metzl, N., Wong, C. S., and Feely, R. A. (2009), Spatial variability and decadal trend of the oceanic CO<sub>2</sub> in the western equatorial Pacific warm/fresh water, Deep Sea Research Part II: Topical Studies in Oceanography, 56, 591-606, DOI: 10.1016/j.dsr2.2009.01.002

Jacobson, A. R., Mikaloff Fletcher, S. E., Gruber, N., Sarmiento, J. L., and Gloor, M. (2007a), A joint atmosphere-ocean inversion for surface fluxes of carbon dioxide: 1. Methods and global-scale fluxes, Global Biogeochem. Cycles, 21, GB1019, DOI: 10.1029/2005gb002556

Jacobson, A. R., Mikaloff Fletcher, S. E., Gruber, N., Sarmiento, J. L., and Gloor, M. (2007b), A joint atmosphere-ocean inversion for surface fluxes of carbon dioxide: 2. Regional results, Global Biogeochem. Cycles, 21, GB1020, DOI: 10.1029/2006gb002703

Jickells, T. D. (1998), Nutrient Biogeochemistry of the Coastal Zone, Science, 281, 217-222, DOI: 10.1126/science.281.5374.217

Johnson, K. M., Sieburth, J. M., Williams, P. J. I., and Brändström, L. (1987), Coulometric total carbon dioxide analysis for marine studies: Automation and calibration, Marine Chemistry, 21, 117-133, DOI: 10.1016/0304-4203(87)90033-8 Joyce, T., and Corry, C. (Eds.) (1994), Requirements for WOCE Hydrographic Programme Data Reporting, 90-1 Rev. 2, WOCE Hydrographic Programme Office, La Jolla, California, 145 pp.

Kalnay, E., Kanamitsu, M., Kistler, R., Collins, W., Deaven, D., Gandin, L., Iredell, M., Saha, S., White, G., Woollen, J., Zhu, Y., Leetmaa, A., Reynolds, R., Chelliah, M., Ebisuzaki, W., Higgins, W., Janowiak, J., Mo, K. C., Ropelewski, C., Wang, J., Jenne, R., and Joseph, D. (1996), The NCEP/NCAR 40-Year Reanalysis Project, Bulletin of the American Meteorological Society, 77, 437-471, DOI: 10.1175/1520-0477(1996)077<0437:tnyrp>2.0.co;2

Kara, A. B., Rochford, P. A., and Hurlburt, H. E. (2003), Mixed layer depth variability over the global ocean, J. Geophys. Res., 108(C3), 3079, DOI: 10.1029/2000jc000736

Keeling, C. D., Brix, H., and Gruber, N. (2004), Seasonal and long-term dynamics of the upper ocean carbon cycle at Station ALOHA near Hawaii, Global Biogeochem. Cycles, 18, GB4006, DOI: 10.1029/2004gb002227

Key, R. M., Kozyr, A., Sabine, C. L., Lee, K., Wanninkhof, R., Bullister, J. L., Feely, R. A., Millero, F. J., Mordy, C., and Peng, T. H. (2004), A global ocean carbon climatology: Results from Global Data Analysis Project (GLODAP), Global Biogeochem. Cycles, 18, GB4031, DOI: 10.1029/2004gb002247

Key, R. M., Tanhua, T., Olsen, A., Hoppema, M., Jutterström, S., Schirnick, C., van Heuven, S., Kozyr, A., Lin, X., Velo, A., Wallace, D. W. R., and Mintrop, L. (2010), The CARINA data synthesis project: introduction and overview, Earth Syst. Sci. Data, 2, 105-121, DOI: 10.5194/essd-2-105-2010

Khatiwala, S., Primeau, F., and Hall, T. (2009), Reconstruction of the history of anthropogenic  $CO_2$  concentrations in the ocean, Nature, 462, 346-349, DOI: 10.1038/nature08526

Khatiwala, S., Tanhua, T., Mikaloff Fletcher, S., Gerber, M., Doney, S. C., Graven, H. D., Gruber, N., McKinley, G. A., Murata, A., Ríos, A. F., Sabine, C. L., and Sarmiento, J. L. (2012), Global ocean storage of anthropogenic carbon, Biogeosciences Discuss., 9, 8931-8988, DOI: 10.5194/bgd-9-8931-2012

Kirchman, D. L. (2012), Processes in Microbial Ecology, Oxford University Press, 368 pp.

Koch, M., Bowes, G., Ross, C., and Zhang, X.-H. (2013), Climate change and ocean acidification effects on seagrasses and marine macroalgae, Global Change Biology, 19, 103-132, DOI: 10.1111/j.1365-2486.2012.02791.x

Kohonen, T. (1988), Self-organization and associative memory, Springer-Verlag Berlin Heidelberg New York. Also Springer Series in Information Sciences, volume 8, 312 pp. Kohonen, T. (2001), Self-Organizing Maps, 3 ed., Number 30 in Springer Series in Information Sciences, Springer-Verlag, Berlin, 501 pp.

Langdon, C., and Atkinson, M. J. (2005), Effect of elevated  $pCO_2$  on photosynthesis and calcification of corals and interactions with seasonal change in temperature/irradiance and nutrient enrichment, J. Geophys. Res.: Oceans, 110, C09S07, DOI: 10.1029/2004jc002576

Lavigne, H., and Gattuso, J. P. (2012), seacarb: seawater carbonate chemistry with R, R package version 2.4.4., http://CRAN.R-project.org/package=seacarb

Law, R. M., Chen, Y.-H., Gurney, K. R., and Modellers, T. (2003), TransCom 3 CO<sub>2</sub> inversion intercomparison: 2. Sensitivity of annual mean results to data choices, Tellus B, 55, 580-595, DOI: 10.1034/j.1600-0889.2003.00053.x

Le Quéré, C., Aumont, O., Bopp, L., Bousquet, P., Ciais, P., Francey, R., Heimann, M., Keeling, C. D., Keeling, R. F., Kheshgi, H., Peylin, P., Piper, S. C., Prentice, I. C., and Rayner, P. J. (2003), Two decades of ocean CO<sub>2</sub> sink and variability, Tellus B, 55, 649-656, DOI: 10.1034/j.1600-0889.2003.00043.x

Le Quéré, C., Rödenbeck, C., Buitenhuis, E. T., Conway, T. J., Langenfelds, R., Gomez, A., Labuschagne, C., Ramonet, M., Nakazawa, T., Metzl, N., Gillett, N., and Heimann, M. (2007), Saturation of the Southern Ocean  $CO_2$  Sink Due to Recent Climate Change, Science, 316, 1735-1738, DOI: 10.1126/science.1136188

Le Quéré, C., Raupach, M. R., Canadell, J. G., and Marland, G. (2009), Trends in the sources and sinks of carbon dioxide, Nature Geosci, 2, 831-836, DOI: 10.1038/ngeo689

Le Quéré, C., Takahashi, T., Buitenhuis, E. T., Rödenbeck, C., and Sutherland, S. C. (2010), Impact of climate change and variability on the global oceanic sink of CO<sub>2</sub>, Global Biogeochem. Cycles, 24, GB4007, DOI: 10.1029/2009gb003599

Lee, K., Millero, F. J., Byrne, R. H., Feely, R. A., and Wanninkhof, R. (2000a), The recommended dissociation constants for carbonic acid in seawater, Geophys. Res. Lett., 27, 229-232, DOI: 10.1029/1999gl002345

Lee, K., Wanninkhof, R., Feely, R. A., Millero, F. J., and Peng, T. H. (2000b), Global relationships of total inorganic carbon with temperature and nitrate in surface seawater, Global Biogeochem. Cycles, 14, 979–994, DOI: 10.1029/1998GB001087

Lee, K., Tong, L. T., Millero, F. J., Sabine, C. L., Dickson, A. G., Goyet, C., Park, G.-H., Wanninkhof, R., Feely, R. A., and Key, R. M. (2006), Global relationships of total alkalinity with salinity and temperature in surface waters of the world's oceans, Geophys. Res. Lett., 33, L19605, DOI: 10.1029/2006gl027207 Lefèvre, N., and Taylor, A. (2002), Estimating pCO<sub>2</sub> from sea surface temperatures in the Atlantic gyres, Deep Sea Research Part I: Oceanographic Research Papers, 49, 539-554, DOI: 10.1016/s0967-0637(01)00064-4

Lefèvre, N., Watson, A. J., Olsen, A., Ríos, A. F., Pérez, F. F., and Johannessen, T. (2004), A decrease in the sink for atmospheric CO<sub>2</sub> in the North Atlantic, Geophys. Res. Lett., 31, L07306, DOI: 10.1029/2003gl018957

Lefèvre, N., Watson, A. J., and Watson, A. R. (2005), A comparison of multiple regression and neural network techniques for mapping in situ pCO<sub>2</sub> data, Tellus, 57, 375-384, DOI: 10.1111/j.1600-0889.2005.00164.x

Lenton, A., Matear, R. J., and Tilbrook, B. (2006), Design of an observational strategy for quantifying the Southern Ocean uptake of CO<sub>2</sub>, Global Biogeochem. Cycles, 20, GB4010, DOI: 10.1029/2005gb002620

Lenton, A., Metzl, N., Takahashi, T., Kuchinke, M., Matear, R. J., Roy, T., Sutherland, S. C., Sweeney, C., and Tilbrook, B. (2012), The observed evolution of oceanic pCO<sub>2</sub> and its drivers over the last two decades, Global Biogeochem. Cycles, 26, GB2021, DOI: 10.1029/2011gb004095

Lenton, A., Tilbrook, B., Law, R., Bakker, D., Doney, S. C., Gruber, N., Hoppema, M., Ishii, M., Lovenduski, N. S., Matear, R. J., McNeil, B. I., Metzl, N., Mikaloff Fletcher, S. E., Monteiro, P., Rödenbeck, C., Sweeney, C., and Takahashi, T. (2013), Sea-air CO<sub>2</sub> fluxes in the Southern Ocean for the period 1990-2009, Biogeosciences Discuss., 10, 285-333, DOI: 10.5194/bgd-10-285-2013

Leroy, S. S., Anderson, J. G., and Ohring, G. (2008), Climate Signal Detection Times and Constraints on Climate Benchmark Accuracy Requirements, Journal of Climate, 21, 841-846, DOI: 10.1175/2007jcli1946.1

Levitus, S., Antonov, J., and Boyer, T. (2005), Warming of the world ocean, 1955-2003, Geophys. Res. Lett., 32, L02604, DOI: 10.1029/2004gl021592

Li, Z., Adamec, D., Takahashi, T., and Sutherland, S. C. (2005), Global autocorrelation scales of the partial pressure of oceanic CO<sub>2</sub>, J. Geophys. Res., 110, C08002, DOI: 10.1029/2004jc002723

Locarnini, R. A., Mishonov, A. V., Antonov, J. I., Boyer, T. P., Garcia, H. E., Baranova, O. K., Zweng, M. M., and Johnson, D. R. (2010), World Ocean Atlas 2009, Volume 1: Temperature, in, edited by: Levitus, S., Ed. NOAA Atlas NESDIS 68, U.S. Government Printing Office, Washington, D.C., 184 pp.

Longhurst, A., Sathyendranath, S., Platt, T., and Caverhill, C. (1995), An estimate of global primary production in the ocean from satellite radiometer data, Journal of Plankton Research, 17, 1245-1271, DOI: 10.1093/plankt/17.6.1245

Lovenduski, N. S., Gruber, N., Doney, S. C., and Lima, I. D. (2007), Enhanced CO<sub>2</sub> outgassing in the Southern Ocean from a positive phase of the Southern Annular Mode, Global Biogeochem. Cycles, 21, GB2026, DOI: 10.1029/2006gb002900

Lyman, J. M., Good, S. A., Gouretski, V. V., Ishii, M., Johnson, G. C., Palmer, M. D., Smith, D. M., and Willis, J. K. (2010), Robust warming of the global upper ocean, Nature, 465, 334-337, DOI: 10.1038/nature09043

Manning, A. C., and Keeling, R. F. (2006), Global oceanic and land biotic carbon sinks from the Scripps atmospheric oxygen flask sampling network, Tellus B, 58, 95-116, DOI: 10.1111/j.1600-0889.2006.00175.x

Matear, R. J., and Lenton, A. (2008), Impact of Historical Climate Change on the Southern Ocean Carbon Cycle, Journal of Climate, 21, 5820-5834, DOI: 10.1175/2008jcli2194.1

Matsumoto, K., and McNeil, B. (2012), Decoupled Response of Ocean Acidification to Variations in Climate Sensitivity, Journal of Climate, 26, 1764-1771, DOI: 10.1175/jcli-d-12-00290.1

McKinley, G. A., Fay, A. R., Takahashi, T., and Metzl, N. (2011), Convergence of atmospheric and North Atlantic carbon dioxide trends on multidecadal timescales, Nature Geosci, 4, 606-610, DOI: 10.1038/ngeo1193

McNeil, B. I., Matear, R. J., and Tilbrook, B. (2001a), Does carbon-13 track anthropogenic CO<sub>2</sub> in the southern ocean?, Global Biogeochem. Cycles, 15, 597-613, DOI: 10.1029/2000gb001352

McNeil, B. I., Tilbrook, B., and Matear, R. J. (2001b), Accumulation and uptake of anthropogenic  $CO_2$  in the Southern Ocean, south of Australia between 1968 and 1996, J. Geophys. Res., 106, 31431-31445, DOI: 10.1029/2000jc000331

McNeil, B. I., Matear, R. J., Key, R. M., Bullister, J. L., and Sarmiento, J. L. (2003), Anthropogenic CO<sub>2</sub> uptake by the ocean based on the global chlorofluorocarbon data set, Science, 299, 235, DOI: 10.1126/science.1077429

McNeil, B. I., and Matear, R. J. (2007), Climate change feedbacks on future oceanic acidification, Tellus B, 59, 191-198, DOI: 10.1111/j.1600-0889.2006.00241.x

McNeil, B. I., Metzl, N., Key, R. M., Matear, R. J., and Corbiere, A. (2007), An empirical estimate of the Southern Ocean air-sea CO<sub>2</sub> flux, Global Biogeochem. Cycles, 21, GB3011, DOI: 10.1029/2007gb002991

McNeil, B. I., and Matear, R. J. (2008), Southern Ocean acidification: A tipping point at 450-ppm atmospheric CO<sub>2</sub>, Proceedings of the National Academy of Sciences, 105, 18860-18864, DOI: 10.1073/pnas.0806318105

McNeil, B. I. (2010), Diagnosing coastal ocean CO<sub>2</sub> interannual variability from a 40 year hydrographic time series station off the east coast of Australia, Global Biogeochem. Cycles, 24, GB4034, DOI: 10.1029/2010gb003870

McNeil, B. I., Tagliabue, A., and Sweeney, C. (2010), A multi-decadal delay in the onset of corrosive 'acidified' waters in the Ross Sea of Antarctica due to strong air-sea  $CO_2$  disequilibrium, Geophys. Res. Lett., 37, L19607, DOI: 10.1029/2010gl044597

McNeil, B. I., and Matear, R. J. (2012), The non-steady-state oceanic CO<sub>2</sub> signal: its importance, magnitude and a novel way to detect it, Biogeosciences Discuss., 9, 13161-13186, DOI: 10.5194/bgd-9-13161-2012

Meehl, G. A., Stocker, T. F., Collins, W. D., Friedlingstein, P., Gaye, A. T., Gregory, J. M., Kitoh, A., Knutti, R., Murphy, J. M., Noda, A., Raper, S. C. B., Watterson, I. G., A.J., W., and Zhao, Z.-C. (2007), Global Climate Projections, Climate Change 2007: The Physical Science Basis. Contribution of Working Group I to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change, edited by: Solomon, S., Qin, D., Manning, M., Chen, Z., Marquis, M., Averyt, K. B., Tignor, M., and Miller, H. L., Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA

Mehrbach, C., Culberson, C. H., Hawley, J. E., and Pytkowicz, R. M. (1973), Measurement of the Apparent Dissociation Constants of Carbonic Acid in Seawater at Atmospheric Pressure, Limnology and Oceanography, 18, 897-907

Meinshausen, M., Smith, S. J., Calvin, K., Daniel, J. S., Kainuma, M. L. T., Lamarque, J. F., Matsumoto, K., Montzka, S. A., Raper, S. C. B., Riahi, K., Thomson, A., Velders, G. J. M., and Vuuren, D. P. P. (2011), The RCP greenhouse gas concentrations and their extensions from 1765 to 2300, Climatic Change, 109, 213-241, DOI: 10.1007/s10584-011-0156-z

Melssen, W., Wehrens, R., and Buydens, L. M. C. (2006), Supervised Kohonen networks for classification problems, Chemometrics and Intelligent Laboratory Systems, 83, 99-113, DOI: 10.1016/j.chemolab.2006.02.003

Metzl, N., Louanchi, F., and Poisson, A. (1998), Seasonal and interannual variations of sea surface carbon dioxide in the subtropical Indian Ocean, Marine Chemistry, 60, 131-146, DOI: 10.1016/s0304-4203(98)00083-8

Metzl, N., Brunet, C., Jabaud-Jan, A., Poisson, A., and Schauer, B. (2006), Summer and winter air-sea CO<sub>2</sub> fluxes in the Southern Ocean, Deep Sea Research Part I: Oceanographic Research Papers, 53, 1548-1563, DOI: 10.1016/j.dsr.2006.07.006

Metzl, N. (2009), Decadal increase of oceanic carbon dioxide in Southern Indian Ocean surface waters (1991-2007), Deep Sea Research Part II: Topical Studies in Oceanography, 56, 607-619, DOI: 10.1016/j.dsr2.2008.12.007

Michaelidis, B., Spring, A., and Pörtner, H. (2007), Effects of long-term acclimation to environmental hypercapnia on extracellular acid–base status and metabolic capacity in Mediterranean fish Sparus aurata, Mar Biol, 150, 1417-1429, DOI: 10.1007/s00227-006-0436-8

Mikaloff-Fletcher, S. E., Gruber, N., Jacobson, A. R., Doney, S. C., Dutkiewicz, S., Gerber, M., Follows, M., Joos, F., Lindsay, K., Menemenlis, D., Mouchet, A., Müller, S. A., and Sarmiento, J. L. (2006), Inverse estimates of anthropogenic CO<sub>2</sub> uptake, transport, and storage by the ocean, Global Biogeochem. Cycles, 20, GB2002, DOI: 10.1029/2005gb002530

Mikaloff-Fletcher, S. E., Gruber, N., Jacobson, A. R., Gloor, M., Doney, S. C., Dutkiewicz, S., Gerber, M., Follows, M., Joos, F., Lindsay, K., Menemenlis, D., Mouchet, A., Müller, S. A., and Sarmiento, J. L. (2007), Inverse estimates of the oceanic sources and sinks of natural CO<sub>2</sub> and the implied oceanic carbon transport, Global Biogeochem. Cycles, 21, GB1010, DOI: 10.1029/2006gb002751

Millero, F. J., Lee, K., and Roche, M. (1998), Distribution of alkalinity in the surface waters of the major oceans, Marine Chemistry, 60, 111-130, DOI: 10.1016/s0304-4203(97)00084-4

Millero, F. J., Pierrot, D., Lee, K., Wanninkhof, R., Feely, R. A., Sabine, C. L., Key, R. M., and Takahashi, T. (2002), Dissociation constants for carbonic acid determined from field measurements, Deep Sea Research Part I: Oceanographic Research Papers, 49, 1705-1723, DOI: 10.1016/s0967-0637(02)00093-6

Monteiro, P., and Co-Authors: A Global Sea Surface Carbon Observing System: Assessment of Changing Sea Surface  $CO_2$  and Air-Sea  $CO_2$  Fluxes, Proceedings of OceanObs'09: Sustained Ocean Observations and Information for Society (Vol. 2), Venice, Italy, 21-25 September 2009, 2010.

Nakaoka, S.-I., Aoki, S., Nakazawa, T., Hashida, G., Morimoto, S., Yamanouchi, T., and Yoshikawa-Inoue, H. (2006), Temporal and spatial variations of oceanic pCO<sub>2</sub> and air–sea  $CO_2$  flux in the Greenland Sea and the Barents Sea, 58, DOI: 10.1111/j.1600-0889.2006.00178.x

Olsen, A., Omar, A. M., Bellerby, R. G. J., Johannessen, T., Ninnemann, U., Brown, K. R., Olsson, K. A., Olafsson, J., Nondal, G., Kivimäe, C., Kringstad, S., Neill, C., and Olafsdottir, S. (2006), Magnitude and origin of the anthropogenic CO<sub>2</sub> increase and <sup>13</sup>C Suess effect in the Nordic seas since 1981, Global Biogeochem. Cycles, 20, GB3027, DOI: 10.1029/2005gb002669

Orr, J. C., Fabry, V. J., Aumont, O., Bopp, L., Doney, S. C., Feely, R. A., Gnanadesikan, A., Gruber, N., Ishida, A., Joos, F., Key, R. M., Lindsay, K., Maier-Reimer, E., Matear, R., Monfray, P., Mouchet, A., Najjar, R. G., Plattner, G.-K., Rodgers, K. B., Sabine, C. L., Sarmiento, J. L., Schlitzer, R., Slater, R. D., Totterdell, I. J., Weirig, M.-F., Yamanaka, Y., and Yool, A. (2005), Anthropogenic ocean

acidification over the twenty-first century and its impact on calcifying organisms, Nature, 437, 681-686, DOI: 10.1038/nature04095

Orr, J. C. (2011), Recent and future changes in ocean carbonate chemistry, in: Ocean acidification, edited by: Gattuso, J. P., and Hansson, L., Oxford University Press, 41-65 pp.

Park, G.-H., Wanninkhof, R. I. K., Doney, S. C., Takahashi, T., Lee, K., Feely, R. A., Sabine, C. L., Triñanes, J., and Lima, I. D. (2010), Variability of global net sea–air CO<sub>2</sub> fluxes over the last three decades using empirical relationships, Tellus B, 62, 352-368, DOI: 10.1111/j.1600-0889.2010.00498.x

Patra, P. K., Maksyutov, S., Ishizawa, M., Nakazawa, T., Takahashi, T., and Ukita, J. (2005), Interannual and decadal changes in the sea-air  $CO_2$  flux from atmospheric  $CO_2$  inverse modeling, Global Biogeochem. Cycles, 19, GB4013, DOI: 10.1029/2004GB002257

Patra, P. K., Gurney, K. R., Denning, A. S., Maksyutov, S., Nakazawa, T., Baker, D., Bousquet, P., Bruhwiler, L., Chen, Y.-H., Ciais, P., Fan, S., Fung, I., Gloor, M., Heimann, M., Higuchi, K., John, J., Law, R. M., Maki, T., Pak, B. C., Peylin, P., Prather, M., Rayner, P. J., Sarmiento, J. L., Taguchi, S., Takahashi, T., and Yuen, C.-W. (2006), Sensitivity of inverse estimation of annual mean CO<sub>2</sub> sources and sinks to ocean-only sites versus all-sites observational networks, Geophys. Res. Lett., 33, L05814, DOI: 10.1029/2005gl025403

Peters, G. P., Andrew, R. M., Boden, T., Canadell, J. G., Ciais, P., Le Quere, C., Marland, G., Raupach, M. R., and Wilson, C. (2012), The challenge to keep global warming below 2°C, Nature Clim. Change, advance online publication, DOI: 10.1038/nclimate1783

Petit, J. R., Jouzel, J., Raynaud, D., Barkov, N. I., Barnola, J. M., Basile, I., Bender, M., Chappellaz, J., Davis, M., Delaygue, G., Delmotte, M., Kotlyakov, V. M., Legrand, M., Lipenkov, V. Y., Lorius, C., Pepin, L., Ritz, C., Saltzman, E., and Stievenard, M. (1999), Climate and atmospheric history of the past 420,000 years from the Vostok ice core, Antarctica, Nature, 399, 429-436, DOI: 10.1038/20859

Peylin, P., Law, R. M., Gurney, K. R., Chevallier, F., Jacobson, A. R., Maki, T., Niwa, Y., Patra, P. K., Peters, W., Rayner, P. J., Rödenbeck, C., and Zhang, X. (2013), Global atmospheric carbon budget: results from an ensemble of atmospheric CO<sub>2</sub> inversions, Biogeosciences Discuss., 10, 5301-5360, DOI: 10.5194/bgd-10-5301-2013

Pierrot, D., Lewis, E., and Wallace, D. W. R. (2006), MS Excel Program Developed for CO<sub>2</sub> System Calculations, ORNL/CDIAC-105a. Carbon Dioxide Information Analysis Center, Oak Ridge National Laboratory, U.S. Department of Energy, Oak Ridge, Tennessee., DOI: 10.3334/CDIAC/otg.CO2SYS\_XLS\_CDIAC105a

Pierrot, D., Brown, P., Van Heuven, S., Tanhua, T., Schuster, U., Wanninkhof, R., and Key, R. M. (2010), CARINA TCO<sub>2</sub> data in the Atlantic Ocean, Earth Syst. Sci. Data, 2, 177-187, DOI: 10.5194/essd-2-177-2010

Pöllä, M., Honkela, T., and Kohonen, T. (2009), Bibliography of Self-Organizing Map (SOM) Papers: 2002-2005 Addendum, 236

Quay, P. D., Tilbrook, B., and Wong, C. S. (1992), Oceanic Uptake of Fossil Fuel CO<sub>2</sub>: Carbon-13 Evidence, Science, 256, 74-79, DOI: 10.1126/science.256.5053.74

Raupach, M. R. (2011), Carbon cycle: Pinning down the land carbon sink, Nature Clim. Change, 1, 148-149, DOI: 10.1038/nclimate1123

Rayner, P. J., Law, R. M., Allison, C. E., Francey, R. J., Trudinger, C. M., and Pickett-Heaps, C. (2008), Interannual variability of the global carbon cycle (1992-2005) inferred by inversion of atmospheric  $CO_2$  and  $\delta^{13}CO_2$  measurements, Global Biogeochem. Cycles, 22, GB3008, DOI: 10.1029/2007gb003068

Revelle, R., and Suess, H. E. (1957), Carbon Dioxide Exchange Between Atmosphere and Ocean and the Question of an Increase of Atmospheric  $CO_2$  during the Past Decades, Tellus, 9, 18-27, DOI: 10.1111/j.2153-3490.1957.tb01849.x

Riebesell, U., Zondervan, I., Rost, B., Tortell, P. D., Zeebe, R. E., and Morel, F. M. M. (2000), Reduced calcification of marine plankton in response to increased atmospheric CO<sub>2</sub>, Nature, 407, 364-367, DOI: 10.1038/35030078

Ries, J. B., Cohen, A. L., and McCorkle, D. C. (2009), Marine calcifiers exhibit mixed responses to  $CO_2$ -induced ocean acidification, Geology, 37, 1131-1134, DOI: 10.1130/g30210a.1

Rodgers, K. B., Sarmiento, J. L., Aumont, O., Crevoisier, C., de Boyer Montégut, C., and Metzl, N. (2008), A wintertime uptake window for anthropogenic CO<sub>2</sub> in the North Pacific, Global Biogeochem. Cycles, 22, GB2020, DOI: 10.1029/2006gb002920

Roemmich, D., Riser, S., Davis, R., and Desaubies, Y. (2004), Autonomous profiling floats: Workhorse for broad-scale ocean observations, Marine Technology Society, 38, 31-39, DOI: 10.4031/002533204787522802

Sabine, C. L., Feely, R. A., Gruber, N., Key, R. M., Lee, K., Bullister, J. L., Wanninkhof, R., Wong, C. S., Wallace, D. W. R., and Tilbrook, B. (2004), The oceanic sink for anthropogenic CO<sub>2</sub>, Science, 305, 367-371, DOI: 10.1126/science.1097403

Santana-Casiano, J. M., González-Dávila, M., Rueda, M.-J., Llinás, O., and González-Dávila, E.-F. (2007), The interannual variability of oceanic CO<sub>2</sub> parameters in the northeast Atlantic subtropical gyre at the ESTOC site, Global Biogeochem. Cycles, 21, GB1015, DOI: 10.1029/2006gb002788 Sarmiento, J. L., Orr, J. C., and Siegenthaler, U. (1992), A Perturbation Simulation of CO<sub>2</sub> Uptake in an Ocean General Circulation Model, J. Geophys. Res., 97, 3621-3645, DOI: 10.1029/91jc02849

Sarmiento, J. L., Dunne, J., Gnanadesikan, A., Key, R. M., Matsumoto, K., and Slater, R. (2002), A new estimate of the CaCO<sub>3</sub> to organic carbon export ratio, Global Biogeochem. Cycles, 16, 1107, DOI: 10.1029/2002gb001919

Sarmiento, J. L., and Gruber, N. (2002), Sinks for Anthropogenic Carbon, Physics Today, 55, 30-36, DOI: 10.1063/1.1510279

Sarmiento, J. L., Slater, R., Barber, R., Bopp, L., Doney, S. C., Hirst, A. C., Kleypas, J., Matear, R., Mikolajewicz, U., Monfray, P., Soldatov, V., Spall, S. A., and Stouffer, R. (2004), Response of ocean ecosystems to climate warming, Global Biogeochem. Cycles, 18, GB3003, DOI: 10.1029/2003gb002134

Sarmiento, J. L., and Gruber, N. (2006), Ocean biogeochemical dynamics, Princeton University Press, 526 pp.

Sarmiento, J. L., Gloor, M., Gruber, N., Beaulieu, C., Jacobson, A. R., Mikaloff Fletcher, S. E., Pacala, S., and Rodgers, K. (2010), Trends and regional distributions of land and ocean carbon sinks, Biogeosciences Discuss., 7, 2351-2367, DOI: 10.5194/bg-7-2351-2010

Seitzinger, S. P., Harrison, J. A., Dumont, E., Beusen, A. H. W., and Bouwman, A. F. (2005), Sources and delivery of carbon, nitrogen, and phosphorus to the coastal zone: An overview of Global Nutrient Export from Watersheds (NEWS) models and their application, Global Biogeochem. Cycles, 19, GB4S01, DOI: 10.1029/2005gb002606

Shaw, E. C., McNeil, B. I., and Tilbrook, B. (2012), Impacts of ocean acidification in naturally variable coral reef flat ecosystems, J. Geophys. Res.: Oceans, 117, n/a-n/a, DOI: 10.1029/2011jc007655

Shaw, E. C., McNeil, B. I., Tilbrook, B., Matear, R., and Bates, M. L. (2013), Anthropogenic changes to seawater buffer capacity combined with natural reef metabolism induce extreme future coral reef  $CO_2$  conditions, Global Change Biology, n/a-n/a, DOI: 10.1111/gcb.12154

Suzuki, T., Ishii, M., Aoyama, M., Christian, J. R., Enyo, K., Kawano, T., Key, R. M., Kosugi, N., Kozyr, A., Miller, L. A., Murata, A., Nakano, T., Ono, T., Saino, T., Sasaki, K., Sasano, D., Takatani, Y., Wakita, M., and Sabine, C. (2013), PACIFICA Data Synthesis Project, http://pacifica.pices.jp, ORNL/CDIAC-159, NDP-092. Carbon Dioxide Information Analysis Center, Oak Ridge National Laboratory, U.S. Department of Energy, Oak Ridge, Tennessee., DOI: 10.3334/CDIAC/OTG.PACIFICA\_NDP092

Sweeney, C., Takahashi, T., Gnanadesikan, A., Wanninkhof, R., Feely, R. A., Friedrich, G., Chaves, F., Bates, N. R., Olafsson, J., and Sarmiento, J. L. (2002), Spatial and

temporal variability of surface water pCO<sub>2</sub> and sampling strategies, A large-scale CO<sub>2</sub> observing plan: In situ oceans and atmosphere, edited by: Bender, M., Doney, S., Feely, RA, et al., National Technical Information Service, Springfield, Virginia, USA, 155–175

Takahashi, T., Feely, R. A., Weiss, R. F., Wanninkhof, R. H., Chipman, D. W., Sutherland, S. C., and Takahashi, T. T. (1997), Global air-sea flux of CO<sub>2</sub>: An estimate based on measurements of sea–air pCO<sub>2</sub> difference, Proceedings of the National Academy of Sciences, 94, 8292-8299

Takahashi, T., Sutherland, S. C., Sweeney, C., Poisson, A., Metzl, N., Tilbrook, B., Bates, N. R., Wanninkhof, R., Feely, R. A., and Sabine, C. L. (2002), Global sea-air CO<sub>2</sub> flux based on climatological surface ocean pCO<sub>2</sub>, and seasonal biological and temperature effects, Deep Sea Research Part II: Topical Studies in Oceanography, 49, 1601-1622, DOI: 10.1016/S0967-0645(02)00003-6

Takahashi, T., Sutherland, S. C., Feely, R. A., and Wanninkhof, R. (2006), Decadal change of the surface water  $pCO_2$  in the North Pacific: A synthesis of 35 years of observations, J. Geophys. Res., 111, C07S05, DOI: 10.1029/2005jc003074

Takahashi, T., Sutherland, S. C., Wanninkhof, R., Sweeney, C., Feely, R. A., Chipman, D. W., Hales, B., Friederich, G., Chavez, F., Sabine, C. L., Watson, A., Bakker, D. C. E., Schuster, U., Metzl, N., Yoshikawa-Inoue, H., Ishii, M., Midorikawa, T., Nojiri, Y., Körtzinger, A., Steinhoff, T., Hoppema, M., Olafsson, J., Arnarson, T. S., Tilbrook, B., Johannessen, T., Olsen, A., Bellerby, R. G. J., Wong, C. S., Delille, B., Bates, N. R., and de Baar, H. J. W. (2009), Climatological mean and decadal change in surface ocean pCO<sub>2</sub>, and net sea-air CO<sub>2</sub> flux over the global oceans, Deep Sea Research Part II: Topical Studies in Oceanography, 56, 554-577, DOI: 10.1016/j.dsr2.2008.12.009

Takahashi, T., Sutherland, S. C., and Kozyr, A. (2012), Global Ocean Surface Water Partial Pressure of CO<sub>2</sub> Database: Measurements Performed During 1957-2011 (Version 2011), Carbon Dioxide Information Analysis Center, Oak Ridge National Laboratory, U.S. Department of Energy, Oak Ridge, Tennessee

Tanhua, T., van Heuven, S., Key, R. M., Velo, A., Olsen, A., and Schirnick, C. (2010), Quality control procedures and methods of the CARINA database, Earth Syst. Sci. Data, 2, 35-49, DOI: 10.5194/essd-2-35-2010

Telszewski, M., Chazottes, A., Schuster, U., Watson, A. J., Moulin, C., Bakker, D. C. E., González-Dávila, M., Johannessen, T., Körtzinger, A., Lüger, H., Olsen, A., Omar, A., Padin, X. A., Ríos, A. F., Steinhoff, T., Santana-Casiano, M., Wallace, D. W. R., and Wanninkhof, R. (2009), Estimating the monthly pCO<sub>2</sub> distribution in the North Atlantic using a self-organizing neural network, Biogeosciences, 6, 1405-1421, DOI: 10.5194/bg-6-1405-2009

Vázquez-Rodríguez, M., Pérez, F. F., Velo, A., Ríos, A. F., and Mercier, H. (2012), Observed acidification trends in North Atlantic water masses, Biogeosciences, 9, 5217-5230, DOI: 10.5194/bg-9-5217-2012

Wallace, D. W. R. (1995), Monitoring global ocean inventories, Dev. Panel Background Rep. 5, 54 pp

Wanninkhof, R. (1992), Relationship Between Wind Speed and Gas Exchange Over the Ocean, J. Geophys. Res., 97, 7373-7382, DOI: 10.1029/92jc00188

Wanninkhof, R., Lewis, E., Feely, R. A., and Millero, F. J. (1999), The optimal carbonate dissociation constants for determining surface water  $pCO_2$  from alkalinity and total inorganic carbon, Marine Chemistry, 65, 291-301, DOI: 10.1016/s0304-4203(99)00021-3

Wanninkhof, R., Doney, S. C., Bullister, J. L., Levine, N. M., Warner, M., and Gruber, N. (2010), Detecting anthropogenic CO<sub>2</sub> changes in the interior Atlantic Ocean between 1989 and 2005, J. Geophys. Res., 115, C11028, DOI: 10.1029/2010jc006251

Wanninkhof, R., Park, G. H., Takahashi, T., Sweeney, C., Feely, R., Nojiri, Y., Gruber, N., Doney, S. C., McKinley, G. A., Lenton, A., Le Quéré, C., Heinze, C., Schwinger, J., Graven, H., and Khatiwala, S. (2013), Global ocean carbon uptake: magnitude, variability and trends, Biogeosciences, 10, 1983-2000, DOI: 10.5194/bg-10-1983-2013

Watson, A. J., Schuster, U., Bakker, D. C. E., Bates, N. R., Corbière, A., González-Dávila, M., Friedrich, T., Hauck, J., Heinze, C., Johannessen, T., Körtzinger, A., Metzl, N., Olafsson, J., Olsen, A., Oschlies, A., Padin, X. A., Pfeil, B., Santana-Casiano, J. M., Steinhoff, T., and Telszewski, M. (2009), Tracking the Variable North Atlantic Sink for Atmospheric CO<sub>2</sub>, Science, 326, 1391-1393, DOI: 10.1126/science.1177394

Wehrens, R., and Buydens, L. M. C. (2007), Self-and Super-organizing Maps in R: The kohonen Package, Journal of Statistical Software, 21, 19 pp

Weiss, R. F. (1974), Carbon dioxide in water and seawater: the solubility of a non-ideal gas, Marine Chemistry, 2, 203-215, DOI: 10.1016/0304-4203(74)90015-2

Wetzel, P., Winguth, A., and Maier-Reimer, E. (2005), Sea-to-air CO<sub>2</sub> flux from 1948 to 2003: A model study, Global Biogeochem. Cycles, 19, GB2005, DOI: 10.1029/2004gb002339

Wickham, H. (2009), ggplot2: elegant graphics for data analysis, Springer New York, 214 pp.

Winn, C. D., Li, Y.-H., Mackenzie, F. T., and Karl, D. M. (1998), Rising surface ocean dissolved inorganic carbon at the Hawaii Ocean Time-series site, Marine Chemistry, 60, 33-47, DOI: 10.1016/s0304-4203(97)00085-6

Wootton, J. T., Pfister, C. A., and Forester, J. D. (2008), Dynamic patterns and ecological impacts of declining ocean pH in a high-resolution multi-year dataset, Proceedings of the National Academy of Sciences, 105, 18848-18853, DOI: 10.1073/pnas.0810079105

Zeebe, R. E., and Wolf-Gladrow, D. (2001), CO<sub>2</sub> in Seawater - Equilibrium, Kinetics, Isotopes, Elsevier, 346 pp.

Zeebe, R. E., Zachos, J. C., Caldeira, K., and Tyrrell, T. (2008), Carbon Emissions and Acidification, Science, 321, 51-52, DOI: 10.1126/science.1159124

# Appendix A.

# The ocean carbon system

A more in-depth summary of the oceans inorganic carbon cycle is presented in this appendix. For a complete discussion, see Sarmiento and Gruber (2006), or Bopp and Le Quéré (2009).

## A1.1 Inorganic carbon chemistry

Following the series of chemical reactions presented below, when aqueous carbon dioxide ( $CO_{2(aq)}$ ) reacts with water molecules it yields a diprotic carbonic acid molecule ( $H_2CO_3$ ), which can then dissociate to form a bicarbonate and further to a carbonate ion. Since the hydration of ( $CO_{2(aq)}$ ) is slow compared to the subsequent ionization of  $H_2CO_3$  (Soli and Byrne, 2002), it is difficult to distinguish between these two species. The sum of  $CO_{2(aq)}$  and  $H_2CO_3$  is therefore expressed as a hypothetical species  $H_2CO_3^*$  following Sarmiento and Gruber (2006)

$$\operatorname{CO}_{2(\operatorname{aq})} + \operatorname{H}_2 O \xleftarrow{k_0} \operatorname{H}_2 \operatorname{CO}_3^*$$
 (A1.1)

$$H_2CO_3^* \xleftarrow{} HCO_3^- + H^+$$
(A1.2)

$$HCO_{3}^{-} \xleftarrow{k_{2}} CO_{3}^{2-} + H^{+}$$
(A1.3)

Rapid chemical reactions allows for the assumption of thermodynamic equilibrium between the inorganic carbon species, which are expressed by the following equilibrium relationships:

$$k_0 = \frac{[\mathrm{H}_2 \mathrm{CO}_3^*]}{p \mathrm{CO}_2} \tag{A1.4}$$

$$k_1 = \frac{[\mathrm{H}^+][\mathrm{HCO}_3]}{[\mathrm{H}_2\mathrm{CO}_3^*]}$$
(A1.5)

$$k_2 = \frac{[\mathrm{H}^+][\mathrm{CO}_3^{2-}]}{[\mathrm{HCO}_3^{-}]} \tag{A1.6}$$

where  $pCO_2$  is the partial pressure of  $CO_2$  and the square brackets represent the total concentration in solution (expressed as micro-moles per kilogram of solution [µmol kg<sup>-1</sup>]). These carbonic acid equilibrium constants are well constrained for a given temperature, salinity and pressure (e.g., Weiss 1974, Mehrbach et al., 1973, Dickson and Millero 1987).

The sum of  $H_2CO_3^*$ ,  $HCO_3^-$  and  $CO_3^{2-}$  is referred to as the total dissolved inorganic carbon ( $C_T$ ):

$$C_{\rm T} = [{\rm H}_2 {\rm CO}_3^*] + [{\rm HCO}_3^-] + [{\rm CO}_3^{2-}]$$
(A1.7)

In the literature, the symbols DIC,  $\Sigma CO_2$  and  $TCO_2$  are all used to define  $C_T$ . For a typical ocean surface pH of 8.2,  $C_T$  is found in the approximate ratio of 1:90:9 for  $[H_2CO_3^*]$ : $[HCO_3^-]$ : $[CO_3^{2-}]$  (Sarmiento and Gruber, 2006). Given that oceanic  $pCO_2$  is only influenced by the concentration of  $H_2CO_3^*$ , this speciation represents the oceans ability to absorb atmospheric CO<sub>2</sub> far in excess than solubility alone.

Another important parameter to constrain the ocean carbon system is total alkalinity  $(A_T)$ , which is a measure of the charge balance in the ocean:

$$A_{\rm T} = [\rm HCO_3^{-}] + 2[\rm CO_3^{2-}] + [\rm B(\rm OH)_4^{-}] + [\rm OH^{-}] - [\rm H^{+}]$$
(A1.8)  
76.8% 18.8% 4.2% 0.2%

Percentages shown under the ion species indicates their relative contributions in a typical surface water sample with a pH of 8.2 (as quoted by Sarmiento and Gruber (2006)). The reason why carbonate ions are counted twice is due to their double negative charge.

In the ocean carbon system,  $C_T$ ,  $A_T$ ,  $pCO_2$  and pH can all be measured directly, or if any two parameters are known together with temperature and salinity, all parameters ( $C_T$ ,  $A_T$ ,  $pCO_2$ , pH, [ $CO_2^*$ ], [HCO\_3] and [ $CO_3^{2-}$ ]) can be theoretically computed using the previous set of equations. Ocean carbon models therefore typically employ  $C_{\rm T}$  and  $A_{\rm T}$  as the only state variables, since they are both conservation with respect to changes in state, whereas  $p{\rm CO}_2$  and pH are not.

#### A1.2 Photosynthesis

Photosynthesis, also referred to as primary production, describes the biological consumption of  $CO_2$  and nutrients (nitrate, phosphate, silicate and micro-nutrients such as iron) to form organic matter and oxygen:

$$CO_2 + H_2O + Nutrients \xleftarrow{\text{Light}} Organic Matter + O_2$$
 (A1.9)

where phytoplanktonic organisms utilize light as their energy source and water as the reducing agent (Ulloa and Grob 2009).

The uptake of CO<sub>2</sub>, nutrients and protons during photosynthesis maintains a relatively constant stoichiometric ratio of approximately 106:16:1:19 for carbon, nitrogen (as nitrate, nitrite or ammonia), phosphorus (as phosphate) and protons (Redfield et al., 1963). The net result of photosynthesis is a decrease in  $C_{\rm T}$  and relatively smaller increase in  $A_{\rm T}$  due to proton depletion. Efficiency of photosynthesis is limited to a first degree by the availability of nutrients and light.

Respiration of organic matter is the reverse process of photosynthesis, whereby heterotrophic organisms oxidize organic matter as a source of energy. In the marine ecosystem, a fraction of organic carbon is exported into the deep ocean through settling or advection processes. This transport of organic carbon into the ocean depths, commonly referred to as export production, has been estimated to be between 11-16 PgC yr<sup>-1</sup> using satellite-based chlorophyll *a* measurements (Falkowski et al., 2000, Schlitzer, 2002).

## A1.3 Calcification

Biogenic formation and dissolution of calcium carbonate minerals (calcite and aragonite) from some phytoplankton (e.g. Coccolithophorids) and zooplankton (e.g. pteropods) species is expressed by

$$Ca^{2+} + CO_3^{2-} \leftrightarrow CaCO_3 \tag{A1.10}$$

Combining equation A1.10 with equations A1.1 to A1.3 illustrates how  $CaCO_3$  formation releases  $CO_2$  into the ocean

$$Ca^{2+} + 2HCO_3^{-} \leftrightarrow CaCO_3 + CO_2 + H_2O$$
(A1.11)

A significant fraction of  $CaCO_3$  formed in the ocean surface sinks into the ocean depths where it's either dissolved in the water column, releasing  $C_T$  and  $A_T$ , or is deposited in the sediments. The net result of this process is an export of  $C_T$  and  $A_T$  in a ratio of 1:2 respectively.

#### References

Bopp, L., and Le Quéré, C. (2009), Ocean carbon cycle, in: Surface ocean-lower atmosphere processes, edited by: Le Quéré, C., and Saltzman, E. C., American Geophysical Union, Washington, DC, 181-195, DOI: 10.1029/2008GM000780

Dickson, A. G., and Millero, F. J. (1987), A comparison of the equilibrium constants for the dissociation of carbonic acid in seawater media, Deep Sea Research Part A. Oceanographic Research Papers, 34, 1733-1743, DOI: 10.1016/0198-0149(87)90021-5

Falkowski, P., Scholes, R. J., Boyle, E., Canadell, J., Canfield, D., Elser, J., Gruber, N., Hibbard, K., Högberg, P., Linder, S., Mackenzie, F. T., Moore III, B., Pedersen, T., Rosenthal, Y., Seitzinger, S., Smetacek, V., and Steffen, W. (2000), The Global Carbon Cycle: A Test of Our Knowledge of Earth as a System, Science, 290, 291-296, DOI: 10.1126/science.290.5490.291

Mehrbach, C., Culberson, C. H., Hawley, J. E., and Pytkowicz, R. M. (1973), Measurement of the Apparent Dissociation Constants of Carbonic Acid in Seawater at Atmospheric Pressure, Limnology and Oceanography, 18, 897-907 Redfield, A. C., Ketchum, B. H., and Richards, F. A. (1963), The influence of organisms on the composition of seawater, The Sea, edited by: Hill, M. N., Wiley Interscience, New York, 26-77 pp.

Sarmiento, J. L., and Gruber, N. (2006), Ocean biogeochemical dynamics, Princeton University Press, 526 pp.

Schlitzer, R. (2002), Carbon export fluxes in the Southern Ocean: results from inverse modeling and comparison with satellite-based estimates, Deep Sea Research Part II: Topical Studies in Oceanography, 49, 1623-1644 , DOI: 10.1016/S0967-0645(02)00004-8

Soli, A. L., and Byrne, R. H. (2002),  $CO_2$  system hydration and dehydration kinetics and the equilibrium  $CO_2/H_2CO_3$  ratio in aqueous NaCl solution, Marine Chemistry, 78, 65-73, DOI: 10.1016/S0304-4203(02)00010-5

Ulloa, O., and Grob, C. (2009), Marine Pelagic Ecosystems, in: Surface ocean-lower atmosphere processes, edited by: Le Quéré, C., and Saltzman, E. S., American Geophysical Union, Washinton, DC, 119-137, DOI: 10.1029/2008GM000780

Weiss, R. F. (1974), Carbon dioxide in water and seawater: the solubility of a non-ideal gas, Marine Chemistry, 2, 203-215, DOI: 10.1016/0304-4203(74)90015-2
### Appendix B.

### Summary of cruise data

**Table B1:** Cruise summary of global combined bottle dataset. Dataset name refers to project group responsible for synthesizing the data, including; CARINA (CARbon dioxide IN the Atlantic ocean; CARINA Group, 2009a, 2009b, 2010), PACIFICA (PACIFic ocean Interior Carbon project; http://pacifica.pices.jp) and GLODAP (Global Ocean Data Analysis Project; Key et al., 2004). Dataset file refers to the project file name where the cruise data was taken from. EXPOCODE is the expedition code assigned to each cruise, as defined by the National Oceanographic Data Centre (NODC) of the United States of America, and Year represents the year of first sample.

Dotocot nomo	Datasat fila	EXPOCODE	Voor	Numbe	er of meas	urements
Dataset hanne	Dataset Ine	EAFOCODE	1 Cal	in the s	surface mi	xed-layer.
				$C_{\mathrm{T}}$	$A_{\mathrm{T}}$	$pCO_2$
CARINA	ATL.v1.0	06GA19960613	1996	105	0	0
CARINA	ATL.v1.0	06GA20000506	2000	39	39	39
CARINA	ATL.v1.0	06MT19920316	1992	0	257	0
CARINA	ATL.v1.0	06MT19920509	1992	117	114	113
CARINA	ATL.v1.0	06MT19941012	1994	41	30	30
CARINA	ATL.v1.0	06MT19941115	1994	140	0	0
CARINA	ATL.v1.0	06MT19960613	1996	4	4	4
CARINA	ATL.v1.0	06MT19960910	1996	32	32	32
CARINA	ATL.v1.0	06MT19970107	1997	55	64	54
CARINA	ATL.v1.0	06MT19970515	1997	35	33	31
CARINA	ATL.v1.0	06MT19970707	1997	38	38	37
CARINA	ATL.v1.0	06MT19970815	1997	66	0	0
CARINA	ATL.v1.0	06MT19990610	1999	27	26	26
CARINA	ATL.v1.0	06MT19990711	1999	37	37	37
CARINA	ATL.v1.0	06MT20010507	2001	35	35	35
CARINA	ATL.v1.0	06MT20010717	2001	103	102	101
CARINA	ATL.v1.0	06MT20011018	2001	16	16	16
CARINA	ATL.v1.0	06MT20021013	2002	2	4	2
CARINA	ATL.v1.0	06MT20030723	2003	46	44	44
CARINA	ATL.v1.0	06MT20040311	2004	87	89	87
CARINA	ATL.v1.0	18HU19920527	1992	20	0	0
CARINA	ATL.v1.0	18HU19930405	1993	37	0	0
CARINA	ATL.v1.0	18HU19930617	1993	10	0	0
CARINA	ATL.v1.0	18HU19931105	1993	60	0	0
CARINA	ATL.v1.0	18HU19940524	1994	12	0	0
CARINA	ATL.v1.0	18HU19941012	1994	62	0	0
CARINA	ATL.v1.0	18HU19950419	1995	43	0	0

CARINA	ATL.v1.0	18HU19970509	1997	55	70	37
CARINA	ATL.v1.0	29CS19930510	1993	18	29	18
CARINA	ATL.v1.0	29GD19821110	1982	103	103	103
CARINA	ATL.v1.0	29GD19831201	1983	149	150	149
CARINA	ATL.v1.0	29GD19840218	1984	247	247	247
CARINA	ATL.v1.0	29GD19840711	1984	68	68	68
CARINA	ATL.v1.0	29GD19860904	1986	92	92	92
CARINA	ATL.v1.0	29HE19980730	1998	31	31	31
CARINA	ATL.v1.0	29HE20030408	2003	14	14	14
CARINA	ATL.v1.0	31AN19890420	1989	88	110	75
CARINA	ATL.v1.0	316N19971005	1997	106	101	100
CARINA	ATL.v1.0	316N20010627	2001	41	41	41
CARINA	ATL.v1.0	316N20030922	2003	93	90	89
CARINA	ATL.v1.0	316N20031023	2003	135	126	124
CARINA	ATL.v1.0	32OC19950529	1995	4	0	0
CARINA	ATL.v1.0	33LK19960415	1996	70	70	70
CARINA	ATL.v1.0	33RO19980123	1998	192	192	191
CARINA	ATL.v1.0	33RO20030604	2003	211	173	172
CARINA	ATL.v1.0	33SW20010102	2001	108	108	108
CARINA	ATL.v1.0	33SW20030418	2003	35	35	35
CARINA	ATL.v1.0	35A320010203	2001	336	336	336
CARINA	ATL.v1.0	35A320010322	2001	140	144	140
CARINA	ATL.v1.0	35LU19890509	1989	45	45	45
CARINA	ATL.v1.0	35LU19950909	1995	41	41	41
CARINA	ATL.v1.0	35TH19990712	1999	117	117	117
CARINA	ATL.v1.0	35TH20010823	2001	141	141	141
CARINA	ATL.v1.0	35TH20020611	2002	44	85	44
CARINA	ATL.v1.0	35TH20040604	2004	108	108	108
CARINA	ATL.v1.0	64PE20000926	2000	68	0	0
CARINA	ATL.v1.0	64TR19890731	1989	55	51	50
CARINA	ATL.v1.0	64TR19900417	1990	0	50	0
CARINA	ATL.v1.0	64TR19900701	1990	2	0	0
CARINA	ATL.v1.0	64TR19900714	1990	4	0	0
CARINA	ATL.v1.0	64TR19910408	1991	194	0	0
CARINA	ATL.v1.0	67SL19881117	1988	33	33	33
CARINA	ATL.v1.0	74AB19900528	1990	3	0	0
CARINA	ATL.v1.0	74AB19910501	1991	612	612	612
CARINA	ATL.v1.0	74AB19910614	1991	28	0	0
CARINA	ATL.v1.0	74AB20050501	2005	49	49	46
CARINA	ATL.v1.0	74DI19890511	1989	778	159	159
CARINA	ATL.v1.0	74DI19890612	1989	7	0	0
CARINA	ATL.v1.0	74DI19900425	1990	39	38	38
CARINA	ATL.v1.0	74DI19900515	1990	14	0	0
CARINA	ATL.v1.0	74DI19900612	1990	14	16	14
CARINA	ATL.v1.0	74DI19970807	1997	163	163	163
CARINA	ATL.v1.0	74DI19980423	1998	134	134	134
CARINA	ATL.v1.0	74DI20040404	2004	58	53	51
CARINA	ATL.v1.0	IrmingerSea	1991	411	0	0
CARINA	ATL.v1.0	OMEX2	1997	0	149	0

~		1100 1 -	0.61.010000000	1002	4.7	0	0
C	ARINA	AMS.v1.2	06AQ19930806	1993	45	0	0
C	ARINA	AMS.v1.2	06AQ19960712	1996	79	79	79
C	ARINA	AMS.v1.2	06MT19920701	1992	0	81	0
C	ARINA	AMS.v1.2	18RD19980404	1998	107	92	89
C	ARINA	AMS.v1.2	18RD19990827	1999	63	53	53
C	ARINA	AMS.v1.2	18SN19940724	1994	35	34	34
C	ARINA	AMS.v1.2	18SN19970803	1997	7	7	7
C	ARINA	AMS.v1.2	18SN19970831	1997	1	1	1
C	ARINA	AMS.v1.2	18SN19970924	1997	10	9	9
C	ARINA	AMS.v1.2	316N20020530	2002	112	115	109
C	ARINA	AMS.v1.2	32H120020505	2002	14	11	11
C	ARINA	AMS.v1.2	32H120020718	2002	40	42	40
C	ARINA	AMS.v1.2	32H120040515	2004	12	12	12
C	ARINA	AMS.v1.2	32H120040718	2004	38	35	35
C	ARINA	AMS.v1.2	32L919920715	1992	1	3	1
C	ARINA	AMS.v1.2	32L919930718	1993	15	0	0
C	ARINA	AMS.v1.2	58AA19940224	1994	181	159	154
C	ARINA	AMS.v1.2	58AA19940826	1994	76	69	68
C	ARINA	AMS.v1.2	58AA19961121	1996	30	20	11
C	ARINA	AMS.v1.2	58AA19970225	1997	172	180	169
C	ARINA	AMS.v1.2	58AA19980308	1998	159	165	133
C	ARINA	AMS.v1.2	58AA19991003	1999	1	1	1
C	ARINA	AMS.v1.2	58AA20010527	2001	70	35	33
C	ARINA	AMS.v1.2	58GS20030922	2003	61	72	60
C	ARINA	AMS.v1.2	58JH19920712	1992	21	0	0
C	ARINA	AMS.v1.2	58JH19940723	1994	92	0	0
C	ARINA	AMS.v1.2	58JH19970414	1997	88	88	88
C	ARINA	AMS.v1.2	58JH19980801	1998	77	77	77
C	ARINA	AMS.v1.2	58JM20040724	2004	3	3	3
C	ARINA	AMS.v1.2	58JM20050520	2005	7	7	7
C	ARINA	AMS.v1.2	74JC19960720	1996	54	49	49
C	ARINA	AMS.v1.2	77DN19910726	1991	54	57	47
C	ARINA	AMS.v1.2	77DN20010717	2001	0	62	0
C	ARINA	AMS.v1.2	77DN20020420	2002	169	169	169
C	ARINA	AMS.v1.2	IcelandSea	1991	409	0	0
C	ARINA	SO.v1.1	06AQ19920929	1992	181	184	165
C	ARINA	SO.v1.1	06AQ19980328	1998	197	0	0
C	ARINA	SO.v1.1	09AR19960822	1996	204	182	181
C	ARINA	SO.v1.1	09AR20011029	2001	181	178	175
C	ARINA	SO.v1.1	29HE19951203	1995	45	45	45
C	ARINA	SO.v1.1	29HE19960117	1996	34	34	34
C	ARINA	SO.v1.1	29HE20010305	2001	56	56	56
C	ARINA	SO.v1.1	29HE20020304	2002	34	34	34
C	ARINA	SO.v1.1	33RO20050111	2005	183	176	176
C	ARINA	SO.v1.1	35MF19980121	1998	75	69	69
C	ARINA	SO.v1.1	35MF19980818	1998	103	89	89
C	ARINA	SO.v1.1	35MF19981205	1998	76	71	71
C	ARINA	SO.v1.1	35MF19990104	1999	74	69	69
C	ARINA	SO.v1.1	35MF20000117	2000	74	0	0

CADINIA	00 1 1	25 ME20000710	2000	00	0.0	00	
CARINA	SO.VI.I	35MF20000719	2000	98	98	98	
CARINA	SO.VI.I	35MF20010103	2000	42	42	42	
CARINA	SO.VI.I	35MF20020104	2002	57	57	57	
CARINA	SO.v1.1	35MF20030123	2003	4/	4/	4/	
CARINA	SO.v1.1	35MF20040103	2004	54	54	54	
CARINA	SO.v1.1	49HH19941213	1994	40	40	40	
CARINA	SO.v1.1	49HH20011208	2001	58	0	0	
CARINA	SO.v1.1	49NZ20031106	2003	54	54	54	
CARINA	SO.v1.1	49NZ20031209	2003	69	69	69	
CARINA	SO.v1.1	49ZS19921203	1992	70	68	68	
CARINA	SO.v1.1	58A119890214	1989	21	6	6	
CARINA	SO.v1.1	61TG20020206	2002	29	0	0	
CARINA	SO.v1.1	61TG20030217	2003	17	0	0	
CARINA	SO.v1.1	74AB20020301	2002	95	112	95	
CARINA	SO.v1.1	74DI20041103	2004	76	77	76	
CARINA	SO.v1.1	74DI20041213	2004	69	68	68	
CARINA	SO.v1.1	90AV20041104	2004	82	82	82	
CARINA	SO.v1.1	91AA19971204	1997	140	138	137	
PACIFICA		09FA20010524	2001	209	208	206	
PACIFICA		18DD19890213	1989	15	0	0	
PACIFICA		18DD19920203	1992	32	4	4	
PACIFICA		18DD19920908	1992	7	0	0	
PACIFICA		18DD19930226	1993	8	0	0	
PACIFICA		18DD19930514	1993	1	1	1	
PACIFICA		18DD19940510	1994	5	3	3	
PACIFICA		18DD19950207	1995	2	0	0	
PACIFICA		18DD19950508	1995	2	0	0	
PACIFICA		18DD19960219	1996	9	6	6	
PACIFICA		18DD19960506	1996	8	6	6	
PACIFICA		18DD19960812	1996	7	4	4	
PACIFICA		18DD19970212	1997	10	8	7	
PACIFICA		18DD19970604	1997	5	3	2	
PACIFICA		18DD19970827	1997	4	8	3	
PACIFICA		18DD19980219	1998	4	0	0	
PACIFICA		18DD19990602	1999	9	5	5	
PACIFICA		18DD19990824	1999	2	2	2	
PACIFICA		18DD20000208	2000	3	3	3	
PACIFICA		18DD20000200	2000	8	8	8	
		18DD20000991	2000	3	3	3	
PACIFICA		18DD20000005	2000	21	18	18	
PACIFICA		18DD20010200	2001	21	13	13	
PACIFICA		18DD20040210	2004	10	10	10	
PACIFICA		18DD20040002	2004	10	0	0	
PACIFICA		18DD20040821	2004	22	0	0	
PACIFICA		18DD20030211	2003	22	22	22	
PACIFICA		180020000130	2000	22 14	21 12	21 12	
PACIFICA		180020000919	2000	14	15	15	
PACIFICA		18DD20070208	2007	21	20	20	
PACIFICA		18DD20070530	2007	9	9	9	
PACIFICA		18DD20070814	2007	10	10	10	

PACIFICA	18DD20080131	2008	16	16	16
PACIFICA	18DD20080813	2008	15	12	12
PACIFICA	18EN19870330	1987	19	0	0
PACIFICA	18EN19870717	1987	6	0	0
PACIFICA	18EN19911017	1991	4	0	0
PACIFICA	18PZ19861014	1986	6	0	0
PACIFICA	18PZ19870923	1987	5	0	0
PACIFICA	18PZ19871124	1987	8	0	0
PACIFICA	18PZ19880502	1988	9	0	0
PACIFICA	18PZ19880628	1988	9	0	0
PACIFICA	18PZ19881129	1988	15	0	0
PACIFICA	18PZ19890501	1989	10	0	0
PACIFICA	18PZ19891003	1989	6	0	0
PACIFICA	18PZ19900509	1990	26	0	0
PACIFICA	18PZ19900822	1990	9	0	0
PACIFICA	32MB19900222	1990	71	0	0
PACIFICA	33KB20020923	2002	44	44	44
PACIFICA	33KK20020701	2002	12	12	12
PACIFICA	33RR20030714	2003	32	32	32
PACIFICA	35A319940923	1994	195	190	187
PACIFICA	49EW19981003	1998	8	7	7
PACIFICA	49EW19991007	1999	8	8	8
PACIFICA	49EW20001000	2000	4	5	4
PACIFICA	49FA20000509	2000	49	0	0
PACIFICA	49FA20010518	2001	7	0	0
PACIFICA	49FA20010710	2001	24	0	0
PACIFICA	49FA20020625	2002	22	0	0
PACIFICA	49FA20021031	2002	19	0	0
PACIFICA	49FA20030225	2003	11	0	0
PACIFICA	49FA20030422	2003	20	0	0
PACIFICA	49FA20030625	2003	56	0	0
PACIFICA	49FA20031028	2003	60	0	0
PACIFICA	49FA20040120	2004	56	0	0
PACIFICA	49FA20040421	2004	40	0	0
PACIFICA	49FA20040616	2004	60	0	0
PACIFICA	49FA20050114	2005	52	0	0
PACIFICA	49FA20050419	2005	46	0	0
PACIFICA	49FA20050615	2005	14	0	0
PACIFICA	49FA20060118	2006	41	0	0
PACIFICA	49FA20060429	2006	25	0	0
PACIFICA	49FA20060614	2006	34	0	0
PACIFICA	49FA20070117	2007	20	0	0
PACIFICA	49FA20070424	2007	6	0	0
PACIFICA	49FA20071024	2007	12	0	0
PACIFICA	49FA20080117	2008	18	21	17
PACIFICA	49FA20080422	2008	7	0	0
PACIFICA	49H519960509	1996	6	0	0
PACIFICA	49H520010113	2001	9	0	0
PACIFICA	49H520010508	2001	1	0	0

PACIFICA	49H520020510	2002	5	0	0
PACIFICA	49H520021001	2002	3	0	0
PACIFICA	49H520030115	2003	9	0	0
PACIFICA	49H520030609	2003	1	0	0
PACIFICA	49H520030701	2003	2	0	0
PACIFICA	49H520060510	2006	3	0	0
PACIFICA	49H520060715	2006	3	0	0
PACIFICA	49H520070123	2007	10	0	0
PACIFICA	49H520070508	2007	2	0	0
PACIFICA	49HG19920807	1992	98	101	97
PACIFICA	49HG19930431	1993	35	35	35
PACIFICA	49HG19930807	1993	81	83	81
PACIFICA	49HG19940413	1994	89	87	87
PACIFICA	49HG19940808	1994	35	36	35
PACIFICA	49HG19950414	1995	44	47	44
PACIFICA	49HG19950807	1995	61	65	61
PACIFICA	49HG19960412	1996	19	19	19
PACIFICA	49HG19960807	1996	51	51	51
PACIFICA	49HG19971110	1997	102	0	0
PACIFICA	49HG19980812	1998	33	33	33
PACIFICA	49HG19990804	1999	47	46	46
PACIFICA	49HG20000912	2000	40	40	40
PACIFICA	49HG20010813	2001	39	39	39
PACIFICA	49HH20011127	2001	20	0	0
PACIFICA	49HH20030930	2003	3	0	0
PACIFICA	49HO19920620	1992	5	0	0
PACIFICA	49HO19950615	1995	1	0	0
PACIFICA	49HO19970602	1997	7	0	0
PACIFICA	49HO19970711	1997	1	0	0
PACIFICA	49HO19980601	1998	9	11	6
PACIFICA	49HO19980622	1998	6	8	6
PACIFICA	49HO19980718	1998	33	6	5
PACIFICA	49HO19990601	1999	6	6	6
PACIFICA	49HO19990622	1999	6	7	6
PACIFICA	49HO19990719	1999	31	31	31
PACIFICA	49HO20000601	2000	12	12	12
PACIFICA	49HO20000621	2000	11	10	10
PACIFICA	49HO20000729	2000	13	14	13
PACIFICA	49HO20010604	2001	8	10	8
PACIFICA	49HO20010710	2001	12	12	12
PACIFICA	49NB20030700	2003	4	4	4
PACIFICA	49NB20040500	2004	7	2	2
PACIFICA	49NZ19971111	1997	37	35	35
PACIFICA	49NZ19981030	1998	58	57	57
PACIFICA	49NZ19981222	1999	24	0	0
PACIFICA	49NZ19990508	1999	45	43	43
PACIFICA	49NZ19990911	1999	15	15	15
PACIFICA	49NZ19991013	1999	16	0	0
PACIFICA	49NZ19991120	1999	26	0	0

PACIFICA	49NZ20000105	2000	118	118	118
PACIFICA	49NZ20000509	2000	8	8	8
PACIFICA	49NZ20000803	2000	11	11	11
PACIFICA	49NZ20001227	2001	30	0	0
PACIFICA	49NZ20010604	2001	43	41	41
PACIFICA	49NZ20010723	2001	38	36	36
PACIFICA	49NZ20010828	2001	6	6	6
PACIFICA	49NZ20020107	2002	27	0	0
PACIFICA	49NZ20020822	2002	40	40	38
PACIFICA	49NZ20021011	2002	12	12	12
PACIFICA	49NZ20021113	2002	5	0	0
PACIFICA	49NZ20021217	2002	23	0	0
PACIFICA	49NZ20030220	2003	109	104	101
PACIFICA	49NZ20030521	2003	19	19	19
PACIFICA	49NZ20030803	2003	130	127	126
PACIFICA	49NZ20030909	2003	106	103	103
PACIFICA	49NZ20040327	2004	26	25	25
PACIFICA	49NZ20040807	2004	43	41	41
PACIFICA	49NZ20040901	2004	0	90	0
PACIFICA	49NZ20041013	2004	15	15	15
PACIFICA	49NZ20041117	2004	52	51	51
PACIFICA	49NZ20050228	2005	20	21	20
PACIFICA	49NZ20050525	2005	62	62	62
PACIFICA	49NZ20050913	2005	62	26	25
PACIFICA	49NZ20051031	2005	66	64	64
PACIFICA	49NZ20051127	2005	113	112	112
PACIFICA	49NZ20060526	2006	74	65	64
PACIFICA	49NZ20060801	2006	6	6	6
PACIFICA	49NZ20060821	2006	2	2	2
PACIFICA	49NZ20070216	2007	151	76	76
PACIFICA	49NZ20070724	2007	45	44	44
PACIFICA	49NZ20070904	2007	42	30	29
PACIFICA	49NZ20071008	2007	95	94	94
PACIFICA	49NZ20071122	2007	67	67	67
PACIFICA	49NZ20081011	2008	82	57	56
PACIFICA	49NZ20090410	2009	100	100	100
PACIFICA	49NZ20090521	2009	87	87	87
PACIFICA	49OS19930611	1993	36	39	33
PACIFICA	49OS20020600	2002	3	3	3
PACIFICA	49OS20030300	2003	10	0	0
PACIFICA	49OS20030500	2003	2	0	0
PACIFICA	49OS20030600	2003	2	0	0
PACIFICA	49OS20030800	2003	3	0	0
PACIFICA	49RY19970121	1997	18	0	0
PACIFICA	49RY20000607	2000	29	0	0
PACIFICA	49RY20030425	2003	2	0	0
PACIFICA	49RY20030620	2003	22	0	0
PACIFICA	49RY20030714	2003	34	0	0
PACIFICA	49RY20040114	2004	70	0	0

PACIFICA	49RY20040608	2004	22	0	0
PACIFICA	49RY20040704	2004	23	0	0
PACIFICA	49RY20041021	2004	20	0	0
PACIFICA	49RY20050114	2005	29	0	0
PACIFICA	49RY20050615	2005	28	0	0
PACIFICA	49RY20050708	2005	17	0	0
PACIFICA	49RY20051021	2005	16	0	0
PACIFICA	49RY20060113	2006	38	0	0
PACIFICA	49RY20060607	2006	11	2	0
PACIFICA	49RY20060630	2006	21	0	0
PACIFICA	49RY20061021	2006	20	0	0
PACIFICA	49RY20070118	2007	9	0	0
PACIFICA	49RY20070606	2007	26	0	0
PACIFICA	49RY20070724	2007	22	0	0
PACIFICA	49RY20080124	2008	44	0	0
PACIFICA	49SO19970113	1997	96	54	47
PACIFICA	49SO19970509	1997	16	12	11
PACIFICA	49SO19970820	1997	16	16	16
PACIFICA	49SO19980414	1998	1	1	1
PACIFICA	49SU19980422	1998	2	0	0
PACIFICA	49SU19980709	1998	3	0	0
PACIFICA	49SU19980819	1998	4	0	0
PACIFICA	49SU19981007	1998	5	0	0
PACIFICA	49SU19990121	1999	6	0	0
PACIFICA	49SU19990428	1999	4	0	0
PACIFICA	49SU19990709	1999	4	0	0
PACIFICA	49SU19990830	1999	4	0	0
PACIFICA	49SU19991006	1999	5	0	0
PACIFICA	49SU20000121	2000	7	0	0
PACIFICA	49SU20000507	2000	4	0	0
PACIFICA	49SU20000628	2000	4	0	0
PACIFICA	49SU20000817	2000	4	0	0
PACIFICA	49SU20001011	2000	5	0	0
PACIFICA	49TA19980418	1998	16	0	0
PACIFICA	49TA20010417	2001	8	0	0
PACIFICA	49TA20020108	2002	9	0	0
PACIFICA	49TA20020313	2002	5	0	0
PACIFICA	49TA20020416	2002	3	0	0
PACIFICA	49TA20020705	2002	2	0	0
PACIFICA	49WA19960516	1996	24	24	24
PACIFICA	49WA20021107	2002	2	0	0
PACIFICA	49WA20030905	2003	1	0	0
PACIFICA	49WA20040924	2004	3	0	0
PACIFICA	49WA20060301	2006	13	0	0
PACIFICA	49XK19940106	1994	80	0	0
PACIFICA	49XK19971129	1997	34	0	0
PACIFICA	49K619990523	1999	35	54	34
PACIFICA	90BM19930830	1993	25	21	21
PACIFICA	318M200406	2004	173	163	162

PACIFICA		49K619940107	1994	76	77	71
PACIFICA		316N19920530	1992	95	0	0
PACIFICA		316N19920502	1992	74	0	0
PACIFICA		316N19920713	1992	48	0	0
PACIFICA		49XK19960617	1996	14	2	2
PACIFICA		49RY19940707	1994	46	0	0
PACIFICA		325019931005	1993	42	42	41
PACIFICA		09AR19930311	1993	94	0	0
PACIFICA		322019920816	1992	90	89	88
PACIFICA		325019930705	1993	94	89	88
PACIFICA		31DS19960105	1996	250	188	185
PACIFICA		18DD19940906	1994	29	28	26
PACIFICA		18DD19941013	1994	36	37	34
PACIFICA		316N19921006	1992	192	0	0
PACIFICA		31WT19910831	1991	28	28	27
PACIFICA		325020060213	2006	238	232	229
PACIFICA		31DS19910307	1991	67	0	0
PACIFICA		33RR20050106	2005	155	158	155
PACIFICA		31WT19910716	1991	134	9	9
PACIFICA		31WT19910531	1991	36	27	27
PACIFICA		316N19921204	1992	160	0	0
PACIFICA		325019930515	1993	69	74	61
PACIFICA		33RO20071215	2007	456	414	413
PACIFICA		31DS19940126	1994	302	296	273
PACIFICA		316N19930222	1993	204	0	0
PACIFICA		318M19940327	1994	238	250	237
PACIFICA		325019940124	1994	41	38	32
GLODAP	Atlantic v1.1	06MT18_1	1991	33	0	0
GLODAP	Atlantic v1.1	18HU95011_1	1995	31	0	0
GLODAP	Atlantic v1.1	18HU98023_1	1998	15	15	15
GLODAP	Atlantic v1.1	06MT39_3	1997	20	20	20
GLODAP	Atlantic v1.1	06MT28_1	1994	55	26	26
GLODAP	Atlantic v1.1	06MT15_3	1991	3	0	0
GLODAP	Atlantic v1.1	06MT22_5	1992	50	17	17
GLODAP	Atlantic v1.1	06MT11_5	1990	111	112	110
GLODAP	Atlantic v1.1	06AQANTX_4	1992	143	0	0
GLODAP	Atlantic v1.1	06AQANTXIII_4	1996	254	0	0
GLODAP	Atlantic v1.1	35A3CITHER3_2	1995	35	30	28
GLODAP	Atlantic v1.1	35A3CITHER3_1	1995	40	92	39
GLODAP	Atlantic v1.1	316N142_3	1994	105	105	104
GLODAP	Atlantic v1.1	OACES91_1-2	1991	68	63	59
GLODAP	Atlantic v1.1	OACES93	1993	135	134	126
GLODAP	Atlantic v1.1	3230CITHER2_1-2	1994	153	97	93
GLODAP	Atlantic v1.1	316N151_3	1997	61	78	61
GLODAP	Atlantic v1.1	316N151_4	1997	48	51	40
GLODAP	Atlantic v1.1	316N151_2	1997	126	131	124
GLODAP	Atlantic v1.1	316N147_2	1996	107	116	102
GLODAP	Atlantic v1.1	06AQANTX_7	1992	127	0	0
GLODAP	Atlantic v1.1	35MF103_1	1996	181	183	181

GLODAP	Atlantic v1.1	316N83_a,c	1983	178	175	175
GLODAP	Atlantic v1.1	06AQANTV-2,3	1986	168	161	161
GLODAP	Atlantic v1.1	Somov	1981	0	62	0
GLODAP	Atlantic v1.1	TTONAS_1-7	1981	207	204	203
GLODAP	Atlantic v1.1	TTOTAS_1-3	1982	103	103	103
GLODAP	Atlantic v1.1	29HE06_1-3	1992	43	38	37
GLODAP	Atlantic v1.1	318MSAVE_1-5	1987	469	330	329
GLODAP	Indian v1.1	3175MB95_07	1995	172	172	164
GLODAP	Indian v1.1	320696_3	1996	312	187	181
GLODAP	Indian v1.1	35MFCIVA_1	1993	148	150	145
GLODAP	Indian v1.1	316N145_5	1994	133	110	109
GLODAP	Indian v1.1	316N145_6	1995	128	127	127
GLODAP	Indian v1.1	316N145_7	1995	147	144	139
GLODAP	Indian v1.1	316N145_8	1995	115	120	111
GLODAP	Indian v1.1	316N145_9	1995	264	249	239
GLODAP	Indian v1.1	316N145_10	1995	166	183	147
GLODAP	Indian v1.1	316N145_11,12	1995	161	159	158
GLODAP	Indian v1.1	316N145_13	1995	49	43	43
GLODAP	Indian v1.1	316N145_14	1995	162	156	150
GLODAP	Indian v1.1	INDIGO-1	1985	177	176	176
GLODAP	Pacific v1.1	90KDIOFFE6_1	1992	170	167	167
GLODAP	Pacific v1.1	09AR9404_1	1994	60	58	58
GLODAP	Pacific v1.1	EQ92SPR	1992	222	245	222
GLODAP	Pacific v1.1	EQ92FAL	1992	244	276	238
GLODAP	Pacific v1.1	RR_KIWI_6	1997	136	135	132
GLODAP	Pacific v1.1	RR_KIWI_7	1997	64	59	59
GLODAP	Pacific v1.1	NBP-97_1	1997	90	0	0
GLODAP	Pacific v1.1	NBP-97_3	1997	140	142	138
CLIVAR		58GS20060721	2006	31	26	26
CLIVAR		77DN20050819	2005	59	52	52
CLIVAR		18DD20030212	2003	21	21	21
CLIVAR		18DD20100203	2010	17	15	15
CLIVAR		18DD20100606	2010	10	10	10
CLIVAR		33RO20100308	2010	189	183	183
CLIVAR		33RR20090320	2009	419	418	417
CLIVAR		33RR20080204	2008	196	258	192
CLIVAR		33RR20070204	2007	200	202	199
CLIVAR		33RR20070322	2007	128	123	122
CLIVAR		318M20091121	2009	134	142	128
CLIVAR		318M20100105	2010	127	104	101
CLIVAR		325020080826	2008	65	63	62
CLIVAR		21OR20080102	2008	4	4	4
GOMECC		RB-07-05	2007	63	64	63
NACP		32WC20070511	2007	226	224	223
S.O ss0299		ANZCW0306005204	1999	38	0	0
S.O ss1195		ANZCW0306004894	1995	42	0	0
S.O v697		?199697060	1997	93	69	69
WOCE		1011110060510	1000	10	10	10
		18HU19960512	1990	19	18	18
AAIW		18HU19960512 316N20050821	2005	19 668	18 748	18 661

BATS	1989-2006	702	479	476
НОТ	1988-2008	517	497	474
TOTAL		31438	24392	22688

### References

CARINA Group (2009a), Carbon in the Arctic Mediterranean Seas Region - the project: Results Version CARINA and Data. 1.2. http://cdiac.ornl.gov/ftp/oceans/CARINA/CARINA Database/CARINA.AMS.V1.2/, Carbon Dioxide Information Analysis Center, Oak Ridge National Laboratory, U.S. Tennessee, Department of Energy, Oak Ridge, DOI: 10.3334/CDIAC/otg.CARINA.AMS.V1.2

CARINA Group (2009b), Carbon in the Atlantic Ocean Region - the CARINA project: Results and Data. Version 1.0. http://cdiac.ornl.gov/ftp/oceans/CARINA/CARINA Database/CARINA.ATL.V1.0/, Carbon Dioxide Information Analysis Center, Oak Ridge National Laboratory, U.S. Department of Energy, Oak Ridge, Tennessee. DOI: 10.3334/CDIAC/otg.CARINA.ATL.V1.0

CARINA Group (2010), Carbon in the Southern Ocean Region - the CARINA project: Results and Version Data. 1.1. http://cdiac.ornl.gov/ftp/oceans/CARINA/CARINA Database/CARINA.SO.V1.1/, Carbon Dioxide Information Analysis Center, Oak Ridge National Laboratory, U.S. Department of Energy, Oak Ridge, Tennessee, DOI: 10.3334/CDIAC/otg.CARINA.SO.V1.1

Key, R. M., Kozyr, A., Sabine, C. L., Lee, K., Wanninkhof, R., Bullister, J. L., Feely, R. A., Millero, F. J., Mordy, C., and Peng, T. H. (2004), A global ocean carbon climatology: Results from Global Data Analysis Project (GLODAP), Global Biogeochem. Cycles, 18, GB4031, DOI: 10.1029/2004gb002247

# Appendix C.

# Multiple-Linear Regression equations

**Table C1.** Ad-hoc and universal  $A_{\rm T}$  regression co-efficients ( $\beta$ ) for the equation:  $A_{\rm T} = \beta_0 + \Sigma\beta x + ({\rm Int.})$ . Where  $\beta_0$  and Int. represent the intercept and interaction terms respectively, and x are the predictor parameters. The predictor parameters are defined as temperature (T), salinity (S), dissolved oxygen (DO), silicate (Si) and phosphate (P). For region boundaries see Fig. 2.7.

Region	Zone <sup>a</sup>	Intercent $(B_0)$			B				Int	Int	Int
			H	S	$S^2$	DO	Si	Р			
(Sub)tropical	1	2064.66	-0.3	-47.57	1.54	0.13	-1.12	10.1	I	I	I
Equatorial Pacific	0	1142.6	-1.39	I	0.96	0.14	I	-3.51	I	I	I
North Atlantic	ω	1543.52	-4.78	I	0.64	-0.04	-0.29	-9.04	0.13*S*T	I	I
North Pacific	4	721.6	I	44.31	I	0.09	-7.81	9.97	0.24*S*Si	Ι	I
Southern Ocean	5	7661.04	-1.46	-362.53	5.86	0.54	-12.17	-6.56	0.08*S*T	0.44*S*Si	-0.01*DO*Si
Universal model		1972.44	-12.78	-33.44	1.19	0.16	0.39	6.89	0.37*S*T	I	I
a Correction and inc. according	his rocion	in Eic 27									

Corresponding geographic region in Fig. 2.7

<b>[able C2.</b> Ad-hoc and universal $C_T$ regression co-efficients ( $\beta$ ) for the equation: $C_T = \beta_0 + \Sigma \beta x + (Int.)$ . Where $\beta_0$ and Int. represent the intercept
In interaction terms respectively, and x are the predictor parameters. The predictor parameters are defined as temperature (1), saming (5), $\frac{1}{10000000000000000000000000000000000$
iissolved oxygen (DU), shircate (D) and prosphate (P). Summer and winter months are defined as hovember-April and May-October for the
southern Hemisphere, respectively. Northern Hemisphere is opposite.

Region (See Fig. 2.7)	Intercept $(\beta_0)$			β				Int.	Int.	Int.	Int.
	I	Τ	S	DO	z	Si	Р				
North Pacific summer	1066.9	I	24.16	0.38	5.07	1	I	I	I	I	I
North Pacific winter	868.73	-7.95	36.54	l	Ι	4.73	Ι	-0.01*Si*DO	Ι	I	Ι
Southern Ocean summer	698.74	I	35.84	0.25	I	0.42	83.28	I	I	I	Ι
Southern Ocean winter	1494.14	-48.81	22.3	-0.31	Ι	0.48	I	0.03*T*DO	$0.92^{*}T^{*}S$	Ι	I
Northwest Atlantic summer	1709.15	I	9.13	I	9.14	I	I	I	Ι	I	Ι
Northwest Atlantic winter	1128.79	-5.93	28.71	ļ	17.31	I	Ι	-0.05*N*DO	Ι	I	Ι
Northeast Atlantic summer	1625.17	I	12.35	ļ	6.13	I	I	I	Ι	I	I
Northeast Atlantic winter	1013.95	61.48	32.75	I	-61.93	-2.05	-12.54	0.4*N*Si	-0.3*T*DO	-1.67*T*S	1.78*N*S
Equatorial Pacific	467.55	-7.11	48.4	I	2.34	1.44	38.85	I	Ι	I	Ι
(Sub)tropical North Pacific summer	519.77	-9.98	48.97	ļ	20.25	I	1.92	Ι	Ι	I	Ι
(Sub)tropical North Pacific winter	236.82	-5.32	50.65	0.38	-1.5	I	139.22	I	Ι	I	Ι
(Sub)tropical South Pacific summer	147.12	-4.61	52.63	0.34	7.48	1.67	72.68	Ι	Ι	Ι	Ι
(Sub)tropical South Pacific winter	643.05	-12.01	46.68	I	I	-1.28	107.88	Ι	Ι	I	Ι
Indian Ocean summer	551.82	-6.59	45.21	I	I	Ι	27.16	0.14*S*N	Ι	I	Ι
Indian Ocean winter	1733.55	-1.84	Ι	-4.78	18.13	2.64	67.1	0.17*DO*S	Ι	I	Ι
(Sub)tropical North Atlantic summer	619.34	-8.18	46.94	-0.37	-31.07	Ι	Ι	Ι	Ι	I	Ι
(Sub)tropical North Atlantic winter	765	-7.36	39.89	I	I	-4.88	109.27	Ι	Ι	I	Ι
Equatorial Atlantic	163.5	-8.91	59.32	-0.17	I	Ι	Ι	Ι	Ι	I	Ι
Sub-tropical South Atlantic	2277.89	-6.15	Ι	-7.48	-5.08	Ι	74.92	0.2*DO*S	Ι	Ι	Ι
Universal model	596.77	-8.21	45.5	-0.17	1.12	0.45	17.83	0.01*T*DO	1.52*T*P	I	I

### Appendix D.

### **Principle Component Regression**

Principal Component Regression (PCR) is an empirical approach when multicollinearity exists between predictor variables. The process (outlined in Fig. D1) first calculates the principal components  $(n_1, ..., n_i, ..., n_I)$  of the predictor variables  $(p_1, ..., p_n, ..., p_N)$ . Then a least-squares multiple-linear regression is established between a subset of the principal components and the response variable (y). The subsets begin with just the first principal component, then the first two, through to all principal components. The PCR deemed optimal is simply the regression with the lowest Residual Standard Error (RSE; see Eq. 2.4).

$$(\boldsymbol{p}_{1},...,\boldsymbol{p}_{n},...,\boldsymbol{p}_{N})$$

$$\downarrow PCA$$

$$(\boldsymbol{n}_{1},...,\boldsymbol{n}_{i},...,\boldsymbol{n}_{I})$$

$$\downarrow Constrain PCR's$$

$$\boldsymbol{M}_{1} = \beta_{0} + \beta_{1}\boldsymbol{n}_{I}$$

$$\vdots$$

$$\boldsymbol{M}_{i} = \beta_{0} + \sum_{i=1}^{i}\beta_{i}\boldsymbol{n}_{i}$$

$$\vdots$$

$$\boldsymbol{M}_{I} = \beta_{0} + \sum_{i=1}^{I}\beta_{i}\boldsymbol{n}_{i}$$

$$\vdots$$

$$\boldsymbol{M}_{I} = \beta_{0} + \sum_{i=1}^{I}\beta_{i}\boldsymbol{n}_{i}$$

Figure D1: Principle Component Regression schematic.

### Appendix E.

# Geographic boundaries used to partition the global SIT predictions



**Figure E1:** Geographical boundaries used to partition the global systematic independent test (SIT) predictions.

Appendix F.

### Monthly pH and $\Omega_{\text{Ar}}$ distributions



Figure F1: Monthly SOMLO-derived ocean surface  $\Omega_{Ar}$  distributions for the nominal year of 2000.



Figure F2: Monthly SOMLO-derived ocean surface pH distributions for the nominal year of 2000.

### Appendix G.

## Table of regional air-sea CO<sub>2</sub> flux estimates

**Table G1:** Regional contemporary air-sea  $CO_2$  flux estimates using SOMLO predictedand Takahashi et al. (2009) (referred to as T-09) ocean surface  $pCO_2$  distributions.

Dagion	Latitudinal	Area	Contemporary	air-sea CO <sub>2</sub>
Region	band	$[10^{6} \text{km}^{2}]$	flux [PgC yr <sup>-1</sup> ]	
		_	SOMLO	T-09
High North Atlantic	North of 44°N	11.1	-0.37	-0.30
High North Pacific	North of 44°N	10.6	-0.07	-0.06
Temperate North Atlantic	18°N - 44°N	18.0	-0.12	-0.14
Temperate North Pacific	18°N - 44°N	30.8	-0.21	-0.38
Equatorial Atlantic	18°N - 18°S	20.9	0.13	0.11
Equatorial Pacific	18°N - 18°S	63.0	1.06	0.50
Equatorial Indian	24.5°N - 18°S	24.3	0.11	0.10
Temperate South Atlantic	18°S - 44°S	17.7	-0.20	-0.11
Temperate South Pacific	18°S - 44°S	36.2	-0.32	-0.28
Temperate South Indian	18°S - 44°S	24.4	-0.31	-0.36
Southern Ocean	South of 44°S	61.7	-0.81	-0.28
Northern Hemisphere	North of 0°	124.3	-0.18	-0.64
Southern Hemisphere	South of $0^{\circ}$	194.4	-0.93	-0.57
Global open-ocean		318.7	-1.10±0.25	-1.21±0.59

### References

Takahashi, T., Sutherland, S. C., Wanninkhof, R., Sweeney, C., Feely, R. A., Chipman, D. W., Hales, B., Friederich, G., Chavez, F., Sabine, C. L., Watson, A., Bakker, D. C. E., Schuster, U., Metzl, N., Yoshikawa-Inoue, H., Ishii, M., Midorikawa, T., Nojiri, Y., Körtzinger, A., Steinhoff, T., Hoppema, M., Olafsson, J., Arnarson, T. S., Tilbrook, B., Johannessen, T., Olsen, A., Bellerby, R. G. J., Wong, C. S., Delille, B., Bates, N. R., and de Baar, H. J. W. (2009), Climatological mean and decadal change in surface ocean pCO<sub>2</sub>, and net sea-air CO<sub>2</sub> flux over the global oceans, Deep Sea Research Part II: Topical Studies in Oceanography, 56, 554-577, DOI: 10.1016/j.dsr2.2008.12.009