

This discussion paper is/has been under review for the journal Biogeosciences (BG).
Please refer to the corresponding final paper in BG if available.

A novel method for diagnosing seasonal to inter-annual surface ocean carbon dynamics from bottle data using neural networks

T. P. Sasse¹, B. I. McNeil¹, and G. Abramowitz¹

¹Climate Change Research Centre, Faculty of Science, University of New South Wales, Sydney, Australia

Received: 8 October 2012 – Accepted: 17 October 2012 – Published: 1 November 2012

Correspondence to: T. P. Sasse (t.sasse@unsw.edu.au)

Published by Copernicus Publications on behalf of the European Geosciences Union.

BGD

9, 15329–15380, 2012

Diagnosing seasonal to inter-annual surface ocean carbon dynamics

T. P. Sasse et al.

Title Page

Abstract

Introduction

Conclusions

References

Tables

Figures

⏪

⏩

◀

▶

Back

Close

Full Screen / Esc

Printer-friendly Version

Interactive Discussion



Abstract

The ocean's role in modulating the observed $1\text{--}7\text{PgCyr}^{-1}$ inter-annual variability in atmospheric CO_2 growth rate is an important, but poorly constrained process due to sparse spatio-temporal ocean carbon measurements. Here, we investigate and develop a non-linear empirical approach to predict inorganic CO_2 concentrations (total carbon dioxide (C_T) and total alkalinity (A_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 spatio-temporal 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 ad-hoc geographic and temporal partitioning of carbon data to constrain complex global carbon dynamics in the mixed-layer. Synthesising a new global C_T/A_T carbon bottle dataset consisting of $\sim 33\,000$ measurements in the open ocean mixed-layer, we develop a neural network based approach 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; Kohonen, 1988). After the initial SOM analysis, which includes geographic constraints, we apply a local linear optimizer to the neural network which considerably enhances the predictive skill of the new approach. We call this new approach SOMLO, or self-organizing multiple linear output. Using independent bottle carbon data, we compare a traditional MLR analysis to our SOMLO approach to capture the spatial C_T and A_T distributions. We find the SOMLO approach improves predictive skill globally by 19% for C_T , with a global capacity to predict C_T to within $10.9\ \mu\text{mol kg}^{-1}$ ($9.2\ \mu\text{mol kg}^{-1}$ for A_T). The non-linear 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 between 25–40% over traditional linear methods. We further test the SOMLO technique using the Bermuda Atlantic time-series (BATS) and Hawaiian ocean

BGD

9, 15329–15380, 2012

Diagnosing seasonal to inter-annual surface ocean carbon dynamics

T. P. Sasse et al.

Title Page

Abstract

Introduction

Conclusions

References

Tables

Figures

⏪

⏩

◀

▶

Back

Close

Full Screen / Esc

Printer-friendly Version

Interactive Discussion

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.

1 Introduction

5 The oceans role in modulating rising atmospheric carbon dioxide (CO₂) levels has been found to be very important (Khatiwala et al., 2012; Sabine et al., 2004). A variety of data-based estimates suggest net oceanic uptake for CO₂ to be $2.1 \pm 1.0 \text{ Pg C yr}^{-1}$ (1 Pg = 10^{15} g) since the year 2000, or about 25–30 % of anthropogenic CO₂ emissions over that period (Jacobson et al., 2007; Khatiwala et al., 2009; Manning and
10 Keeling, 2006; McNeil et al., 2003; Mikaloff Fletcher et al., 2006; Takahashi et al., 2009). Between 1990 and 2009, atmospheric CO₂ accumulation rates vary between 1–7 Pg C yr⁻¹, 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 CO₂ uptake has advanced, our shorter-term understanding (seasonal to
15 inter-annual) of ocean carbon dynamics remains poorly constrained due to data limitations.

Atmospheric CO₂ observations, inversion techniques and ocean models suggest a large range for inter-annual variability in oceanic CO₂ uptake ($0.1\text{--}1.5 \text{ Pg C yr}^{-1}$) (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 of ocean
20 carbon has come about sporadically, dominated by regional time-series measurement programs (e.g. Bermuda Atlantic time-series (BATS) and Hawaiian Ocean time-series (HOT)). 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 CO₂ in a high-CO₂ world will be limited, particularly since some evidence
25 suggests the ability for the ocean to absorb CO₂ has slowed since the late 1980's as

Diagnosing seasonal to inter-annual surface ocean carbon dynamics

T. P. Sasse et al.

Title Page

Abstract

Introduction

Conclusions

References

Tables

Figures



Back

Close

Full Screen / Esc

Printer-friendly Version

Interactive Discussion



a consequence of decadal-scale trends in winds and oceanic circulation (Le Quéré et al., 2010; Sarmiento et al., 2010).

Standard hydrographic measurements in the ocean (temperature, salinity, dissolved oxygen and nutrients) are sampled and analysed 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_2 ($p\text{CO}_2$), total dissolved carbon dioxide (C_T), total alkalinity (A_T) or pH. National and international efforts to survey the global oceanic C_T and A_T distribution has amounted to approximately 330 000 bottle measurements taken sporadically over the past 30 years. However, our ability to globally understand natural seasonal C_T and A_T dynamics has been hindered due to the large spatio-temporal limitations of this accumulated dataset (Key et al., 2004).

Autonomous $p\text{CO}_2$ measuring devices mounted mainly onto commercial shipping vessels has resulted in a global network of approximately 6.4 million ocean surface $p\text{CO}_2$ measurements (Takahashi et al., 2012). This $p\text{CO}_2$ 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_2 variability within the ocean. However, the global $p\text{CO}_2$ dataset cannot inform us on some very important processes and biogeochemical dynamics that modulate atmospheric CO_2 . The oceans biological carbon export flux has been estimated to be between 11–16 Pg C yr^{-1} from satellite chlorophyll *a* measurements (Falkowski et al., 2000), some 5–8 times the net oceanic CO_2 absorption from the atmosphere. Small changes in the biological carbon flux 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

BGD

9, 15329–15380, 2012

Diagnosing seasonal to inter-annual surface ocean carbon dynamics

T. P. Sasse et al.

Title Page

Abstract

Introduction

Conclusions

References

Tables

Figures

⏪

⏩

◀

▶

Back

Close

Full Screen / Esc

Printer-friendly Version

Interactive Discussion



Diagnosing seasonal to inter-annual surface ocean carbon dynamics

T. P. Sasse et al.

Title Page

Abstract

Introduction

Conclusions

References

Tables

Figures

⏪

⏩

◀

▶

Back

Close

Full Screen / Esc

Printer-friendly Version

Interactive Discussion

constraints like the $p\text{CO}_2$ climatology. Secondly, without equivalent A_T or C_T measurements, $p\text{CO}_2$ 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, spatio-temporal deficiencies of the $p\text{CO}_2$ dataset in regions like the Southern Ocean, introduces uncertainties in the direct evaluation of short-term variability. To understand seasonal to inter-annual variability in these regions requires methods that have better spatio-temporal coverage than is constrained by historical $p\text{CO}_2$ sampling. Here, we seek to diagnose seasonal to inter-annual C_T and A_T concentrations in the mixed-layer that provides independent, but important additional constraints to the global sea-surface $p\text{CO}_2$ climatology.

To varying degrees, concentrations of C_T and A_T are influenced by the solubility of CO_2 , biological processes, vertical and lateral water transport and direct CO_2 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_2 (Weiss, 1974). Information on nitrate (N), silicate (Si), phosphate (P) and dissolved oxygen (DO) variations provide insight into the biological influences on oceanic inorganic carbon (Anderson and Sarmiento, 1994). From this, it should be implicit that we can derive empirical relationships between these standard hydrographical parameters and the carbon constituents. If a robust empirical relationship is established, the order of magnitude more in-situ measurements of these standard hydrographic parameters (Boyer et al., 2009) would give us new constraints on seasonal to inter-annual global carbon dynamics in the mixed-layer.

The use of the global sea-surface $p\text{CO}_2$ dataset would be ideal to develop such empirical algorithms. However, these continuous $p\text{CO}_2$ measurements generally have no coinciding biogeochemical information (i.e. DO or nutrients) that could help establish an empirical relationship. Some have used satellite Chlorophyll *a* measurements to

help constrain ocean surface $p\text{CO}_2$ 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_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_T and A_T (see Table 1).

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 geographically, hydrographically and temporally in an attempt to improve the ability of linear approaches to model the non-linear relationship between inorganic carbon and the standard hydrographic parameters. Here we use a non-linear empirical modelling approach to avoid this ad-hoc partitioning and show that it delivers considerable improvements in predictability. We use a self-organizing map (SOM; Kohonen, 1988) to classify or cluster measurements of hydrographic parameters into groups and then establish the relationship between these parameters and C_T/A_T separately for each group. SOMs have already been found to be well suited in extracting features of the ocean surface $p\text{CO}_2$ dataset in the North Atlantic using a combination of modelled and remotely sensed parameters to constrain the system, (Friedrich and Oschlies, 2009a, b; Lefèvre et al., 2005; Telszewski et al., 2009).

To contextualise this work, we firstly explore the use of the traditional MLR approach to diagnose global seasonal carbon dynamics in the ocean. To do this, we employ the MLR approach on a newly synthesised C_T/A_T bottle dataset of $\sim 33\,000$ mixed-layer samples. Next, we present our SOM-based approach to diagnose seasonal carbon dynamics on a global scale, which better accounts for non-linearities that would limit the ability of the MLR approach. To compare the MLR and our SOM approach, we

BGD

9, 15329–15380, 2012

Diagnosing seasonal to inter-annual surface ocean carbon dynamics

T. P. Sasse et al.

Title Page

Abstract

Introduction

Conclusions

References

Tables

Figures

⏪

⏩

◀

▶

Back

Close

Full Screen / Esc

Printer-friendly Version

Interactive Discussion

develop a global, independent test to assess the models skill. Finally, we use the BATS and HOT in-situ time-series as an explicit test for our new approach.

2 Global carbon measurements and training dataset

The extraordinary effort to collate and synthesis the bottle hydrographic and biogeochemical data has been conducted by several groups; including GLODAP (Global Ocean Data Anlysis Project) (Key et al., 2004), CARINA (CARbon dioxide IN the Atlantic Ocean) (CARINA Group, 2009b, a, 2010) and PACIFICA (PACific ocean Interior CARbon) (<http://pacifica.pices.jp/>).

Precision in measuring bottle C_T and A_T samples has consistently improved over the past 30 yr as a result of advances in techniques and apparatus (Bradshaw et al., 1981; Johnson et al., 1987). However, it was not until the introduction of standard operating procedures and certified reference materials (Department of Energy, 1994; Dickson, 2003, 2007) that the quality consistency of independent laboratory measurements was achieved and is currently estimated to be $\pm 2 \mu\text{mol kg}^{-1}$ (Dickson et al., 2007). To account for any systematic measurement biases between cruises, a secondary Quality Control (QC) method was incorporated by the project groups to identify and smooth out any inconsistencies, as outlined in (Tanhua et al., 2010). The internal consistency of the CARINA C_T/A_T dataset has been estimated to $\pm 2.5 \mu\text{mol kg}^{-1}$ (Tanhua et al., 2010). More recent additional measurements we included in the global dataset underwent a 1st QC check to remove measurements flagged as bad or questionable.

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. We refined the global data to be within the mixed-layer (Supplement A), non-coastal (Supplement B) and data post 1980 due to large uncertainties in early measuring techniques. The final number of usable C_T/A_T discrete measurements in the global mixed-layer was $\sim 33\,000$.

BGD

9, 15329–15380, 2012

Diagnosing seasonal to inter-annual surface ocean carbon dynamics

T. P. Sasse et al.

Title Page

Abstract

Introduction

Conclusions

References

Tables

Figures

◀

▶

◀

▶

Back

Close

Full Screen / Esc

Printer-friendly Version

Interactive Discussion

Whilst the spatial coverage of the refined data is consistent over all major ocean basins (Fig. 1a), there are approximately 45% less wintertime measurements than summer (Fig. 1b) which we examine as a potential cause for bias when applying our approach.

2.1 Normalization of C_T measurements

Global atmospheric CO_2 concentrations during the 1980's, 1990's and 2000's have increased at 1.60 ± 0.56 , 1.47 ± 0.66 and 1.90 ± 0.38 ppm per-year, respectively (Thomas Conway and Pieter Tans, NOAA/ESRL, www.esrl.noaa.gov/gmd/ccgg/trends). Mixed-layer measurements of C_T were corrected for temporal anthropogenic CO_2 uptake to the reference year 2000 by calculating the change in mixed-layer C_T in equilibrium with the atmospheric CO_2 increase using observed Revelle factors (Supplement C). This approach is somewhat equivalent to that of T-09 where all $p\text{CO}_2$ measurements values were corrected to the year 2000 using a rate of $1.5 \mu\text{atm yr}^{-1}$.

There are regions of the ocean where upwelling and sea-ice inhibit air-sea gas exchange, resulting in considerable CO_2 disequilibrium (e.g. Southern Ocean, equatorial Pacific). The anthropogenic CO_2 correction technique used here, like those for T-09 and (Lee et al., 2000), will be biased in these regions. However, by performing a test using no anthropogenic CO_2 correction (Supplement D), we demonstrate the very low impact this anthropogenic correction has to our final result. This is in part due to the large natural fingerprint of C_T ($\pm 50 \mu\text{mol kg}^{-1}$) relative to the small changes ($\sim 1 \mu\text{mol kg}^{-1} \text{yr}^{-1}$) resulting from anthropogenic CO_2 uptake.

3 Testing algorithm skill: a Global Independent Test (GIT) approach

Most empirical studies report statistical errors calculated as the residual standard error (RSE) from linear regressions. For example, C_T in the Indian Ocean was reported to be predicted to within $\pm 5 \mu\text{mol kg}^{-1}$ using a suite of hydrographic parameters

BGD

9, 15329–15380, 2012

Diagnosing seasonal to inter-annual surface ocean carbon dynamics

T. P. Sasse et al.

Title Page

Abstract

Introduction

Conclusions

References

Tables

Figures

⏪

⏩

◀

▶

Back

Close

Full Screen / Esc

Printer-friendly Version

Interactive Discussion



(Bates et al., 2006), $\pm 8 \mu\text{mol kg}^{-1}$ for the Southern Ocean (McNeil et al., 2007) and $\pm 7 \mu\text{mol kg}^{-1}$ for a global dataset (Lee et al., 2000). However, an independent dataset not used in the regressions is needed to accurately report true statistical uncertainty for any empirical approach.

Here, we developed a “Global Independent Test” (GIT) approach in order to compare the MLR and NN empirical approaches consistently. The GIT method evaluates the algorithms 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. 1) of the independent predictions (y_{indp}) then provides a better and accurate estimate of the algorithms skill.

$$\text{RSE} = \sqrt{\frac{\sum (y_{\text{pred}} - y_{\text{in-situ}})^2}{N - 2}} \quad (1)$$

The reason we independently test each individual cruise dataset rather than a randomly selected subset of the data, is due to similar concentrations of carbon and auxiliary measurements within adjacent 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 models true skill, as the remaining two measurements with a very similar “biogeochemical fingerprint” will be used to train the algorithm. The prediction of an entire independent cruise is a more robust measure of the algorithms skill.

BGD

9, 15329–15380, 2012

Diagnosing seasonal to inter-annual surface ocean carbon dynamics

T. P. Sasse et al.

Title Page

Abstract

Introduction

Conclusions

References

Tables

Figures

⏪

⏩

◀

▶

Back

Close

Full Screen / Esc

Printer-friendly Version

Interactive Discussion



4 Traditional MLR approach

4.1 Method description

Multi-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 ,

$$y = \beta_0 + \sum_{n=1}^N \beta_n x_n \quad (2)$$

Where β_0 and β_n represent the intercept and empirically derived coefficients, respectively. Multi-collinearity between predictor variables (MCL) and the distribution and constant variance of the residual errors are both issues that may affect the predictive and diagnostic ability of a MLR. To minimise the effect of these issues, the empirical relationships between C_T/A_T and the standard hydrographic parameters were constrained using a forward stepwise robust MLR routine (Supplement E). In brief, the routine constrains the optimal model configuration through the incorporation of statistical tests to evaluate parameter significance and the necessity of interaction terms to reduce the effects of MCL. This MLR normalization 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 affect the model.

4.2 Ad-hoc vs. universal MLR

To investigate the application of the traditional MLR method, we compared the skill of using one single regression globally (universal MLR) to an ad-hoc approach that partitions the dataset into regions (ad-hoc MLR). We based the ad-hoc approach on

BGD

9, 15329–15380, 2012

Diagnosing seasonal to inter-annual surface ocean carbon dynamics

T. P. Sasse et al.

Title Page

Abstract

Introduction

Conclusions

References

Tables

Figures

⏪

⏩

◀

▶

Back

Close

Full Screen / Esc

Printer-friendly Version

Interactive Discussion



dividing the global carbon dataset on the geographical and temporal guidelines outlined by Lee et al. (2000, 2006) and Bates et al. (2006). In this way, the global dataset was subset into 5 geographic regions to constrain the A_T system, and 11 geographic regions, 8 of which were subjected to further separation into summer and winter months to constrain C_T , (see Fig. 2). The universal method simply uses the entire global dataset without division.

4.3 MLR results

When universally applying the traditional MLR on the $\sim 33\,000$ global mixed-layer C_T measurements, the statistical regression RSE is $15.1\ \mu\text{mol kg}^{-1}$ when using T, S, DO, P, N and Si as predictors (Table 3). If applying the ad-hoc geographical and temporal separations, the statistical regression RSE reduces to $13.2\ \mu\text{mol kg}^{-1}$. However, when the independent test (GIT) is used to evaluate the regressions, errors increase to be $16\ \mu\text{mol kg}^{-1}$ for the ad-hoc approach and $15.6\ \mu\text{mol kg}^{-1}$ for the global regression. For A_T , optimal predictors were found to be T, S, S^2 , DO, P and Si, while a global MLR algorithm captured the signal to within $11\ \mu\text{mol kg}^{-1}$ using the GIT. All empirical relationships for the global and ad-hoc MLR models can be found in Supplement G.

The MLR approach and results gives us a framework to attempt to develop a better method that captures any potential non-linear biases that are contributing to errors of $C_T \pm 16\ \mu\text{mol kg}^{-1}$ and $A_T \pm 11\ \mu\text{mol kg}^{-1}$ on a global scale.

5 Neural network approach

5.1 Overview of the neural network approach

A self-organizing map (SOM) is an algorithm that uses an iterative approach to classify multi-dimensional data into discrete groups, or neurons, usually arranged in a 2-dimensional grid. Using an algorithm that employs discrete clustering is appealing, as

BGD

9, 15329–15380, 2012

Diagnosing seasonal to inter-annual surface ocean carbon dynamics

T. P. Sasse et al.

Title Page

Abstract

Introduction

Conclusions

References

Tables

Figures

⏪

⏩

◀

▶

Back

Close

Full Screen / Esc

Printer-friendly Version

Interactive Discussion



it removes the need for the type of ad-hoc partitioning we discussed in Sect. 4.2. This has led to application of SOMs in a wide range of disciplines (Abramowitz, 2005; Hsu et al., 2002; Pöllä et al., 2009).

Figure 3 illustrates the routine of SOM training and prediction. For a training dataset of P samples consisting of predictor variables \mathbf{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 we represent a node 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 \mathbf{x} . The $y = f(\mathbf{x})$ input-output mapping is then completed by performing a linear regression between \mathbf{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 $(\mathbf{x}_1, \dots, \mathbf{x}_q, \dots, \mathbf{x}_Q)$. First, each \mathbf{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 $j(\mathbf{x}_q)$. Then the regression parameters for $j(\mathbf{x}_q)$ are used to predict y_q .

Here we explore two variants to this approach. The first, as described above, uses a multiple linear regression at each neuron, which we describe here as self-organizing multiple linear output (SOMLO). The second takes the mean of all response values belonging to a node, which we’ll call self-organizing map mean (SOMM). We now describe both in more detail.

5.2 Initialization of the model constraints

For our implementation, the input-output pairs (\mathbf{x}_p, y_p) , $1 \leq p \leq P$ in the training dataset are some subset of $\mathbf{x} = (T, S, DO, N, Si, P)$, and $y = C_T$ or A_T . To ensure each predictor variable has an equal opportunity to define the features of the SOM map during the training routine, we zero-mean and scale the variables by their standard deviation so that their distribution and range is similar. For nitrate, phosphate and silicate, due to the exponential distribution of their measurements, we first \log_{10} scale their measurements.

BGD

9, 15329–15380, 2012

Diagnosing seasonal to inter-annual surface ocean carbon dynamics

T. P. Sasse et al.

Title Page

Abstract

Introduction

Conclusions

References

Tables

Figures

⏪

⏩

◀

▶

Back

Close

Full Screen / Esc

Printer-friendly Version

Interactive Discussion



The J neuron SOM we use here is structured in a hexagonal topology for the current study (Fig. 3). Careful consideration needs to be exercised when defining the size of the SOM map, 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 (ω) of length equal to the number of input variables (maximum 6), and whose values are randomly selected from the input variable range.

5.3 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 (τ), the input data samples are individually presented to the SOM map 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

$$distance(\mathbf{x}_p, \omega_j) = \left[\sum_{n=1}^N (x_{p,n} - \omega_{j,n})^2 \right]^{0.5} \quad (3)$$

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 those neurons in its topological neighbourhood in the SOM are then adjusted towards the value of the current sample value (\mathbf{x}_p) via

$$\omega_j(\tau + 1) = \omega_j(\tau) + h_{j,j(\mathbf{x}_p)} (\mathbf{x}_p - \omega_j(\tau)) \quad (4)$$

In this expression, $h_{j,j(\mathbf{x}_p)}$ determines the extent to which a node’s weight is brought closer to the current sample value (termed a “learning rate”, $h \leq 1$). It also determines the size of the neighbourhood around the winning node that receives a significant adjustment. We use

$$h_{j,j(\mathbf{x}_p)} = \eta(\tau) \exp \left(-\frac{d_{j,j(\mathbf{x}_p)}}{2\sigma^2(\tau)} \right) \quad (5)$$

Diagnosing seasonal to inter-annual surface ocean carbon dynamics

T. P. Sasse et al.

Title Page

Abstract

Introduction

Conclusions

References

Tables

Figures

⏪

⏩

◀

▶

Back

Close

Full Screen / Esc

Printer-friendly Version

Interactive Discussion



where, $d_{j,j(x_p)}$ represents the discrete distance in the SOM map topology between the winning neuron $j(x_p)$ 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
5 adjusts most of the neurons to finish with only adjusting the winning neuron. Similarly, the learning rate $\eta(\tau)$ decreases with iteration, so that regional features of the SOM map gradually develop as iterations continue.

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
10 constraint beyond the hydrographic information (T, S, DO etc.) in clustering the global dataset into the set of J neurons. For more detail see Supplement H.

5.4 Completing the input-output mapping

We complete the $y = f(x)$ in one of two ways. First, the mean of all output values y_p for
15 samples belonging to a node is used – the SOMM. Alternatively, we use MLRs with the training data assigned to the winning neuron to establish this relationship (see Fig. 3). Here we use MLRs after the SOM training through the application of a principal component regression (PCR) and a forward stepwise robust MLR based on hypothesis tests (see Supplement E and F for details). To ensure confidence in regression coefficients,
20 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.

5.5 Predicting with the SOMLO/SOMM system

For any independent input data vector (x_q), we can predict the output value (y_q) using
25 the SOM map trained above via a two-step process. First, determine which neuron

Diagnosing seasonal to inter-annual surface ocean carbon dynamics

T. P. Sasse et al.

Title Page

Abstract

Introduction

Conclusions

References

Tables

Figures



Back

Close

Full Screen / Esc

Printer-friendly Version

Interactive Discussion



in the SOM map each new data sample is closest to using the distance measure in Sect. 5.3 (Eq. 3). 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 training data.

6 Application to the global ocean

6.1 Optimization of the global model

To converge on the optimal SOMLO approach for the ocean carbon mixed-layer dataset, we employed a two phase process. Firstly, three unique subsets of ocean carbon data were extracted to ascertain which hydrographic parameter combination worked best. In the second phase we applied the global independent test (GIT) technique to make an out-of-sample assessment of the global skill of the model.

6.1.1 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, we perform tests that exclude the variables one at a time (Fig. 5). These test the ability of the models to capture three unique independent datasets that each represent about 10 % of the global carbon dataset (Table 4). As an example, Fig. 4 presents the spatial distribution of the T1 independent dataset, constituting 11.4 % of the global training 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.

Salinity was found to be the most important parameter for capturing the mixed-layer carbon signal, followed by temperature then nutrients (Fig. 5). The final optimal parameter set and SOM neuron size using the three independent tests were (SOPSi, 25) and

Diagnosing seasonal to inter-annual surface ocean carbon dynamics

T. P. Sasse et al.

Title Page

Abstract

Introduction

Conclusions

References

Tables

Figures



Back

Close

Full Screen / Esc

Printer-friendly Version

Interactive Discussion



(TSPO, 56) for the global A_T and C_T models, respectively (Fig. 6). For C_T the SOMLO model incorporating PCR constrained the system with a higher skill than the robust MLR, whilst A_T was better constrained using the robust MLR model.

The addition of phosphate beyond temperature, salinity and dissolved oxygen improves the prediction of C_T by $\sim 27\%$ or $5.1 \mu\text{mol kg}^{-1}$ (Fig. 6). Without air-sea gas exchange modulating its behaviour, phosphate likely provides clearer constraints on organic matter production and respiration than dissolved oxygen alone. The redundancy of nitrate for both C_T and A_T (Fig. 6) is likely due to the near constant stoichiometric uptake rate of phosphate and nitrate by photosynthesising organisms. The preference of phosphate over nitrate may be a result of the continual production of organic matter by nitrogen-fixers after the nitrate pool is completely depleted (Gruber and Sarmiento, 1997). Furthermore, the re-naming of samples where only “nitrate + nitrite” was listed to nitrate in the GLODAP and CARINA products, (Key et al., 2004), may serve to introduce additional biases in using nitrate.

Precipitation and dissolution of calcium carbonates (CaCO_3) affects the concentration of A_T twice as much as C_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_3 cycling which influences A_T .

Salinities significant importance in constraining the A_T system is likely due to the known high correlation between these two parameters (Millero et al., 1998), whereas the addition of temperature to the parameter set is redundant, as pointed out by some earlier studies (e.g. McNeil et al., 2007).

6.1.2 Importance of geography in the model

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

BGD

9, 15329–15380, 2012

Diagnosing seasonal to inter-annual surface ocean carbon dynamics

T. P. Sasse et al.

Title Page

Abstract

Introduction

Conclusions

References

Tables

Figures

⏪

⏩

◀

▶

Back

Close

Full Screen / Esc

Printer-friendly Version

Interactive Discussion



**Diagnosing seasonal
to inter-annual
surface ocean carbon
dynamics**

T. P. Sasse et al.

[Title Page](#)[Abstract](#)[Introduction](#)[Conclusions](#)[References](#)[Tables](#)[Figures](#)[Back](#)[Close](#)[Full Screen / Esc](#)[Printer-friendly Version](#)[Interactive Discussion](#)

by those in some parts of the North Pacific Ocean, despite there being little ocean inter-connectedness between these two carbon datasets on shorter time-scales. Spatial length-scales 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 in the training of the SOM (Supplement I), but were excluded as predictor parameters in the linear regressions. We found that introducing geographical information to be a powerful addition in improving the skill of the method for C_T by 16 % or $2.2 \mu\text{mol kg}^{-1}$, however there was little improvement for A_T (Fig. 7). The optimal SOMLO configuration additionally incorporates longitude and n-vector geographical inputs in constraining A_T and C_T , respectively.

To better understand and visualize why geography is important, we compare the spatial distribution of neurons for C_T models trained with only biogeochemical information, and both biogeochemical and geographical information (Fig. 8a–b). To illustrate the spatial distribution of the assigned neurons for the global carbon dataset, we plot the neurons using different colours. Here, each colour represents a neuron, while shades of colours indicate close similarity in the weighting vectors. The broad regions of similarity that are captured when the SOM is constrained only by biogeochemical properties include the Southern Ocean, sub-tropical gyres, North Pacific and North Atlantic (Fig. 8a). 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. 8a). 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 sub-tropical convergence zone, but is able to constrain the classification of data to be within each ocean basin (Fig. 8b). 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.

6.1.3 SOMM/SOMLO comparison

Optimal model configurations were tested with neuron sizes extending up to 2500 to explore the ability of the SOMM model in constraining the three independent datasets (Fig. 9). Using all data the SOMM model converged on an RSE value of $16 \mu\text{mol kg}^{-1}$ in constraining C_T . Although the SOMM is powerful in constraining complex non-linear datasets, the relatively sparse carbon dataset limits the ability of the mean-mode of the SOM to predict C_T on a global scale.

We found using a local multiple-linear optimizer (i.e. the MLR) in addition to the global SOM optimizer to significantly improve the models ability to constrain global C_T by $\sim 27\%$ or $4.4 \mu\text{mol kg}^{-1}$. Similar findings are found for the A_T model.

6.2 Measuring the improvement over traditional MLR

To evaluate the skill of the two independent approaches used here (MLR versus SOMLO), we tabulated the results of each technique based on the global independent test (GIT) divided into 5 geographical regions and evaluated globally (Table 5). The SOMLO approach improves the predictive skill of C_T by between 11–30% for all 5 regions (Table 5). In particular, known complex dynamical regions with global CO_2 importance like the equatorial Pacific, Southern Ocean and North Atlantic are where the non-linear SOMLO approach excelled, improving the prediction of C_T by between 23–30% (or $4\text{--}6 \mu\text{mol kg}^{-1}$). From a global point of view, SOMLO improves the predictive skill of C_T in the mixed-layer by $\sim 19\%$.

For A_T , the SOMLO improves the detection only marginally by about 7–10% (or $0.7\text{--}1.4 \mu\text{mol kg}^{-1}$). This is most likely a result of the carbonate system being less prone to

BGD

9, 15329–15380, 2012

Diagnosing seasonal to inter-annual surface ocean carbon dynamics

T. P. Sasse et al.

Title Page

Abstract

Introduction

Conclusions

References

Tables

Figures

⏪

⏩

◀

▶

Back

Close

Full Screen / Esc

Printer-friendly Version

Interactive Discussion



non-linearities and complexity, thereby limiting the benefits of SOMLO, since it better constrains more complex non-linear systems.

6.3 SOMLO regional error assessment

To further scrutinise the spatial skill of the SOMLO model, absolute values of the global independent test (GIT) residual errors were interpolated around the in-situ samples using VG gridding software in the Ocean Data View (ODV) program (Schlitzer, R.: Ocean Data View, <http://odv.awi.de>, 2011). We further separated the model skill into 14 different regions (Table 6; for map of these regions see Supplement J). Although the Arctic Ocean, Bay of Bengal and Sea of Okhotsk are regions not well constrained by the technique, the majority of the ocean maintains a relatively homogenous residual error range (Fig. 10 and Table 6). The 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.

Through the identification and removal of coastal anomalies (see Supplement K for details) and unconstrained regions (Arctic Ocean and Bay of Bengal), the final estimate for the global open-ocean accuracy for C_T and A_T is 10.9 and 9.2 $\mu\text{mol kg}^{-1}$, respectively.

To investigate skewness, we plot the SOMLO global independent test predictions versus the in-situ measurements (Fig. 11a, c). For C_T , skewness is limited ($R^2 = 0.98$), giving us confidence in the models ability to accurately capture the concentrations of C_T and A_T for any given set of temperature, salinity, dissolved oxygen, (silicate for A_T), and phosphate measurements in the open ocean mixed-layer.

Finally, we found no strong seasonal bias in our SOMLO predictions (Fig. 12).

BGD

9, 15329–15380, 2012

Diagnosing seasonal to inter-annual surface ocean carbon dynamics

T. P. Sasse et al.

Title Page

Abstract

Introduction

Conclusions

References

Tables

Figures

⏪

⏩

◀

▶

Back

Close

Full Screen / Esc

Printer-friendly Version

Interactive Discussion



7 Application to the Bermuda Atlantic and Hawaiian Ocean time-series sites

The SOMLO technique was trained on a global C_T and A_T dataset that consisted mostly of sporadic one-time cruises in time. To test how well seasonal to inter-annual variability is captured using our technique, we use carbon time-series data from the BATS and HOT stations as a test-bed.

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

Located in the Sargasso Sea, the BATS hydrographic site is a high frequency measurement program of carbon and auxiliary parameters that has been ongoing since 1989. To test the global SOMLO model in reconstructing the BATS seasonal cycle, we firstly re-trained the global algorithm without using the BATS 1989–2007 carbon time-series dataset. We then use the measured monthly hydrographic properties between 1987–2007 to independently predict C_T and A_T concentrations at the BATS site and finally compare our predicted carbon values to the in-situ measurements to investigate the skill of the technique. We also independently predict C_T/A_T values with the traditional MLR approach as a further test.

Figure 13a–b shows the measured versus predicted C_T and A_T annual cycles at BATS. Within the uncertainty of the SOMLO prediction, both the magnitude and structure of the seasonal C_T cycle at BATS is well constrained, capturing 90 % of the signal (Fig. 13a). For a global MLR approach, the seasonal cycle is overestimated significantly by $\sim 50\%$. For A_T , the small seasonality is captured by both techniques (Fig. 13b).

To gain better insight into how the SOMLO substantially improves the prediction of the BATS seasonal cycle from the traditional MLR analysis, we investigate the neuron distribution for C_T in the North West Atlantic (Fig. 14). 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 we use the spatial boundaries of 30° N to 70° N and 40° W to 85° W that were also used by Lee et al. (2000) in their MLR approach. The traditional MLR explicitly uses all carbon data within the

Diagnosing seasonal to inter-annual surface ocean carbon dynamics

T. P. Sasse et al.

Title Page

Abstract

Introduction

Conclusions

References

Tables

Figures



Back

Close

Full Screen / Esc

Printer-friendly Version

Interactive Discussion



prescribed region, whilst the SOMLO approach partitions the data into neurons without any prior geographic constraints. The benefit in this approach is that when we are applying the SOMLO 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.

7.2 How well does SOMLO capture inter-annual signals?

Inter-annual variability of C_T at BATS is captured to within the uncertainty of the SOMLO technique over the 18 yr period (Fig. 15). This illustrates a new potentially 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 of C_T , we predict the C_T signal at the HOT time-series as reported by (Brix et al., 2004). The SOMLO prediction captures the smoothed inter-annual trend-line at the HOT site to within 85 % (Fig. 16).

The BATS and HOT comparisons provide additional confidence that the SOMLO approach provides good constraints on both seasonal and inter-annual variability of C_T , so that it could be used on a wider scale to help understand the oceans role in modulating atmospheric CO_2 .

8 Comparison to previous techniques

It's important to emphasise reported error estimates of previous empirical studies to those calculated here. RSE values presented by previous empirical studies (see Table 1) are calculated from the regressions residual error rather than independent tests as done here, so direct comparisons between previous studies and our results are not valid. We use the global independent test (GIT, see Sect. 3) in order to accurately report the differences between our results and previous traditional MLR results.

Diagnosing seasonal to inter-annual surface ocean carbon dynamics

T. P. Sasse et al.

Title Page

Abstract

Introduction

Conclusions

References

Tables

Figures



Back

Close

Full Screen / Esc

Printer-friendly Version

Interactive Discussion



Diagnosing seasonal to inter-annual surface ocean carbon dynamics

T. P. Sasse et al.

Title Page

Abstract

Introduction

Conclusions

References

Tables

Figures

⏪

⏩

◀

▶

Back

Close

Full Screen / Esc

Printer-friendly Version

Interactive Discussion

We conduct two sets of calculations as shown in Table 7. The first set of calculations ($RSE_{(MLR-old)}$) involves taking the regressions from a suite of prior work (Bates et al., 2006; Lee et al., 2000, 2006; McNeil et al., 2007) and applying it to the new larger dataset within each region. The second set of calculations ($RSE_{(GIT)}$) involved developing our own set of regressions using the same geographical and temporal boundaries and predictors as the previous authors within the much larger dataset. Using the global independent test (GIT), the skill of the models were calculated (RSE) and could then be directly compared to our SOMLO values (see Table 7).

The SOMLO, as shown at BATS/HOT, improves the predictive skill of C_T and A_T in most regions by between (10–40%). Globally for C_T , the SOMLO reduces the error by 28% beyond the MLR method that was used to conduct the only global analysis (Lee et al., 2000).

9 Conclusions

Here, we have exploited the global carbon C_T/A_T mixed-layer bottle database ($\sim 33\,000$) 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 C_T system to $15.6\ \mu\text{mol kg}^{-1}$ and A_T to $10.4\ \mu\text{mol kg}^{-1}$. We then deploy a new non-linear neural network based approach that improves the predictive skill by $2.7\text{--}3\ \mu\text{mol kg}^{-1}$ for C_T , or $\sim 19\%$ over the MLR, and $0.7\text{--}1.4\ \mu\text{mol kg}^{-1}$ for A_T or $\sim 10\%$. 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 35% over traditional linear approaches. We further test our neural network technique and find it to predict both seasonal and inter-annual variability of carbon at BATS and HOT very well.

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. We demonstrate here, that the use of non-linear empirical techniques on a global scale could potentially advance our understanding of oceanic carbon variability, particularly in a future where the amount of autonomous hydrographic data is increasing exponentially.

Supplementary material related to this article is available online at:
<http://www.biogeosciences-discuss.net/9/15329/2012/bgd-9-15329-2012-supplement.pdf>.

Acknowledgements. We thank all captains, crew and researchers who helped to collect, analysis and synthesis the in-situ bottle dataset used in this study. All training of the 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), with kohonen, ggplot2 and akima packages developed by Wehrens and Buydens (2007), Wickham (2009) and Akima H. et al.: akima: Interpolation of irregularly spaced data, <http://CRAN.R-project.org/package=akima>, 2012, respectively. We also thank the developers of the Ocean Data Viewer (ODV) program (Schlitzer, R.: Ocean Data View, <http://odv.awi.de>, 2011), from which many figures presented in this paper were developed.

References

- Abramowitz, G.: Towards a benchmark for land surface models, *Geophys. Res. Lett.*, 32, L22702, doi:10.1029/2005gl024419, 2005.
- Anderson, L. A. and Sarmiento, J. L.: Redfield ratios of remineralization determined by nutrient data analysis, *Global Biogeochem. Cycles*, 8, 65–80, doi:10.1029/93gb03318, 1994.
- Bates, N. R., Pequignet, A. C., and Sabine, C. L.: Ocean carbon cycling in the Indian Ocean: 1. Spatiotemporal variability of inorganic carbon and air-sea CO₂ gas exchange, *Global Biogeochem. Cycles*, 20, GB3020, doi:10.1029/2005gb002491, 2006.

Diagnosing seasonal to inter-annual surface ocean carbon dynamics

T. P. Sasse et al.

Title Page

Abstract

Introduction

Conclusions

References

Tables

Figures



Back

Close

Full Screen / Esc

Printer-friendly Version

Interactive Discussion



Diagnosing seasonal to inter-annual surface ocean carbon dynamics

T. P. Sasse et al.

[Title Page](#)
[Abstract](#)
[Introduction](#)
[Conclusions](#)
[References](#)
[Tables](#)
[Figures](#)
[Back](#)
[Close](#)
[Full Screen / Esc](#)
[Printer-friendly Version](#)
[Interactive Discussion](#)


- 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.: Atmospheric O₂/N₂ changes, 1993–2002: Implications for the partitioning of fossil fuel CO₂ sequestration, *Global Biogeochem. Cycles*, 19, GB4017, doi:10.1029/2004gb002410, 2005.
- 5 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.: *World Ocean Atlas database 2009*, edited by: Levitus, S., NOAA Atlas NESDIS 66, US Gov. Printing Office, Wash. DC, 216 pp., 2009.
- Bradshaw, A. L., Brewer, P. G., Shafer, D. K., and Williams, R. T.: Measurements of total carbon dioxide and alkalinity by potentiometric titration in the GEOSECS program, *Earth Planet. Sc. Lett.*, 55, 99–115, doi:10.1016/0012-821x(81)90090-x, 1981.
- 10 Brix, H., Gruber, N., and Keeling, C. D.: Interannual variability of the upper ocean carbon cycle at station ALOHA near Hawaii, *Global Biogeochem. Cycles*, 18, GB4019, doi:10.1029/2004gb002245, 2004.
- 15 CARINA Group: Carbon in the Arctic Mediterranean Seas Region – the CARINA project: Results and Data, Version 1.2., Carbon Dioxide Information Analysis Center, Oak Ridge National Laboratory, US Department of Energy, Oak Ridge, Tennessee doi:10.3334/CDIAC/otg.CARINA.AMS.V1.2, 2009a.
- CARINA Group: Carbon in the Atlantic Ocean Region – the CARINA project: Results and Data, Version 1.0, Carbon Dioxide Information Analysis Center, Oak Ridge National Laboratory, US Department of Energy, Oak Ridge, Tennessee doi: 10.3334/CDIAC/otg.CARINA.ATL.V1.0, 2009b.
- 20 CARINA Group: Carbon in the Southern Ocean Region – the CARINA project: Results and Data, Version 1.1, Carbon Dioxide Information Analysis Center, Oak Ridge National Laboratory, US Department of Energy, Oak Ridge, Tennessee, doi: 10.3334/CDIAC/otg.CARINA.SO.V1.1, 2010.
- 25 Chen, L., Xu, S., Gao, Z., Chen, H., Zhang, Y., Zhan, J., and Li, W.: Estimation of monthly air-sea CO₂ flux in the southern Atlantic and Indian Ocean using in-situ and remotely sensed data, *Remote Sens. Environ.*, 115, 1935–1941, doi:10.1016/j.rse.2011.03.016, 2011.
- 30 Chierici, M., Olsen, A., Johannessen, T., Trinañes, J., and Wanninkhof, R.: Algorithms to estimate the carbon dioxide uptake in the northern North Atlantic using shipboard observations, satellite and ocean analysis data, *Deep Sea Research Part 2: Topical Studies in Oceanography*, 56, 630–639, doi:10.1016/j.dsr2.2008.12.014, 2009.

Diagnosing seasonal to inter-annual surface ocean carbon dynamics

T. P. Sasse et al.

Title Page

Abstract

Introduction

Conclusions

References

Tables

Figures

⏪

⏩

◀

▶

Back

Close

Full Screen / Esc

Printer-friendly Version

Interactive Discussion



Department of Energy: 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., 1994.

5 Dickson, A. G., Afghan, J. D., and Anderson, G. C.: Reference materials for oceanic CO₂ analysis: a method for the certification of total alkalinity, *Mar. Chem.*, 80, 185–197, doi:10.1016/s0304-4203(02)00133-0, 2003.

Dickson, A. G., Sabine, C. L., Christian, J. R.(Eds.): Guide to best practices for ocean CO₂ measurements, PICES Special Publication 3, 191, 2007.

10 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.: 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, 2000.

15 Friedrich, T. and Oschlies, A.: Neural network-based estimates of North Atlantic surface pCO₂ from satellite data: A methodological study, *J. Geophys. Res.*, 114, C03020, doi:10.1029/2007jc004646, 2009a.

Friedrich, T. and Oschlies, A.: Basin-scale pCO₂ maps estimated from ARGO float data: A model study, *J. Geophys. Res.*, 114, C10012, doi:10.1029/2009jc005322, 2009b.

20 Gruber, N. and Sarmiento, J. L.: Global patterns of marine nitrogen fixation and denitrification, *Global Biogeochem. Cycles*, 11, 235–266, doi:10.1029/97gb00077, 1997.

Hsu, K., Gupta, H. V., Gao, X., Sorooshian, S., and Imam, B.: 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, 2002.

25 Jacobson, A. R., Mikaloff Fletcher, S. E., Gruber, N., Sarmiento, J. L., and Gloor, M.: 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, 2007.

Johnson, K. M., Sieburth, J. M., Williams, P. J. I., and Brändström, L.: Coulometric total carbon dioxide analysis for marine studies: Automation and calibration, *Marine Chemistry*, 21, 117–133, doi:10.1016/0304-4203(87)90033-8, 1987.

30 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.: A global ocean carbon climatology: Results

Diagnosing seasonal to inter-annual surface ocean carbon dynamics

T. P. Sasse et al.

[Title Page](#)
[Abstract](#)
[Introduction](#)
[Conclusions](#)
[References](#)
[Tables](#)
[Figures](#)
[Back](#)
[Close](#)
[Full Screen / Esc](#)
[Printer-friendly Version](#)
[Interactive Discussion](#)


from Global Data Analysis Project (GLODAP), Global Biogeochem. Cycles, 18, GB4031, doi:10.1029/2004gb002247, 2004.

Khatiwala, S., Primeau, F., and Hall, T.: Reconstruction of the history of anthropogenic CO₂ concentrations in the ocean, *Nature*, 462, 346–349, doi:10.1038/nature08526, 2009.

5 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.: Global ocean storage of anthropogenic carbon, *Biogeosciences Discuss.*, 9, 8931–8988, doi:10.5194/bgd-9-8931-2012, 2012.

Kirchman, D. L.: *Processes in Microbial Ecology*, Oxford University Press, 368 pp., 2012.

10 Kohonen, T.: *Self-organization and associative memory*, Springer-Verlag Berlin Heidelberg New York, Also Springer Series in Information Sciences, Vol. 8, 312 pp., 1988.

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.: Two decades of ocean CO₂ sink and variability, *Tellus B*, 55, 649–656, doi:10.1034/j.1600-0889.2003.00043.x, 2003.

15 Le Quéré, C., Takahashi, T., Buitenhuis, E. T., Rödenbeck, C., and Sutherland, S. C.: Impact of climate change and variability on the global oceanic sink of CO₂, *Global Biogeochem. Cycles*, 24, GB4007, doi:10.1029/2009gb003599, 2010.

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

25 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.: 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, 2006.

Lefèvre, N., Watson, A. J., and Watson, A. R.: A comparison of multiple regression and neural network techniques for mapping in situ pCO₂ data, *Tellus*, 57, 375–384, doi:10.1111/j.1600-0889.2005.00164.x, 2005.

30 Manning, A. C. and Keeling, R. F.: 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, 2006.

Diagnosing seasonal to inter-annual surface ocean carbon dynamics

T. P. Sasse et al.

Title Page

Abstract

Introduction

Conclusions

References

Tables

Figures

⏪

⏩

◀

▶

Back

Close

Full Screen / Esc

Printer-friendly Version

Interactive Discussion

- McKinley, G. A., Fay, A. R., Takahashi, T., and Metzl, N.: Convergence of atmospheric and North Atlantic carbon dioxide trends on multidecadal timescales, *Nat. Geosci.*, 4, 606–610, doi:10.1038/ngeo1193, 2011.
- McNeil, B. I.: Diagnosing coastal ocean CO₂ 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, 2010.
- McNeil, B. I., Matear, R. J., Key, R. M., Bullister, J. L., and Sarmiento, J. L.: Anthropogenic CO₂ uptake by the ocean based on the global chlorofluorocarbon data set, *Science*, 299, 235–239, doi:10.1126/science.1077429, 2003.
- McNeil, B. I., Metzl, N., Key, R. M., Matear, R. J., and Corbiere, A.: An empirical estimate of the Southern Ocean air-sea CO₂ flux, *Global Biogeochem. Cycles*, 21, GB3011, doi:10.1029/2007gb002991, 2007.
- 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.: Inverse estimates of anthropogenic CO₂ uptake, transport, and storage by the ocean, *Global Biogeochem. Cycles*, 20, GB2002, doi:10.1029/2005gb002530, 2006.
- Millero, F. J., Lee, K., and Roche, M.: Distribution of alkalinity in the surface waters of the major oceans, *Marine Chemistry*, 60, 111–130, doi:10.1016/s0304-4203(97)00084-4, 1998.
- 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.: Variability of global net air-sea CO₂ fluxes over the last three decades using empirical relationships, *Tellus B*, 62, 352–368, doi:10.1111/j.1600-0889.2010.00498.x, 2010.
- 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.: Sensitivity of inverse estimation of annual mean CO₂ sources and sinks to ocean-only sites versus all-sites observational networks, *Geophys. Res. Lett.*, 33, L05814, doi:10.1029/2005gl025403, 2006.
- Pöllä, M., Honkela, T., and Kohonen, T.: Bibliography of Self-Organizing Map (SOM) Papers: 2002–2005 Addendum, Technical report, TKK-ICS-R23, 2009.
- Rayner, P. J., Law, R. M., Allison, C. E., Francey, R. J., Trudinger, C. M., and Pickett-Heaps, C.: Interannual variability of the global carbon cycle (1992–2005) inferred by inversion of

Diagnosing seasonal to inter-annual surface ocean carbon dynamics

T. P. Sasse et al.

Title Page

Abstract

Introduction

Conclusions

References

Tables

Figures

◀

▶

◀

▶

Back

Close

Full Screen / Esc

Printer-friendly Version

Interactive Discussion

atmospheric CO₂ and δ¹³CO₂ measurements, *Global Biogeochem. Cycles*, 22, GB3008, doi:10.1029/2007gb003068, 2008.

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.: The oceanic sink for anthropogenic CO₂, *Science*, 305, 367–371, 2004.

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

Sarmiento, J. L., Dunne, J., Gnanadesikan, A., Key, R. M., Matsumoto, K., and Slater, R.: A new estimate of the CaCO₃ to organic carbon export ratio, *Global Biogeochem. Cycles*, 16, 1107, doi:10.1029/2002gb001919, 2002.

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

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.: Climatological mean and decadal change in surface ocean pCO₂, and net sea-air CO₂ flux over the global oceans, *Deep Sea Research Part 2, Topical Studies in Oceanography*, 56, 554–577, doi:10.1016/j.dsr2.2008.12.009, 2009.

Takahashi, T., Sutherland, S.C., and Kozyr, A.: *Global Ocean Surface Water Partial Pressure of CO₂ Database: Measurements Performed During 1957–2011 (Version 2011)*, ORNL/CDIAC-160, NDP-088(V2011), Carbon Dioxide Information Analysis Center, Oak Ridge National Laboratory, US Department of Energy, Oak Ridge, Tennessee, doi:10.3334/CDIAC/OTG.NDP088(V2011), 2012.

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

Estimating the monthly $p\text{CO}_2$ distribution in the North Atlantic using a self-organizing neural network, Biogeosciences, 6, 1405–1421, doi:10.5194/bg-6-1405-2009, 2009.

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

5 Wehrens, R. and Buydens, L. M. C.: Self-and Super-organizing Maps in R: The kohonen Package, J. Stat. Softw., 21, 19 pp., 2007.

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

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

BGD

9, 15329–15380, 2012

Diagnosing seasonal to inter-annual surface ocean carbon dynamics

T. P. Sasse et al.

Title Page

Abstract

Introduction

Conclusions

References

Tables

Figures

⏪

⏩

◀

▶

Back

Close

Full Screen / Esc

Printer-friendly Version

Interactive Discussion



Diagnosing seasonal to inter-annual surface ocean carbon dynamics

T. P. Sasse et al.

Table 1. Previous empirical approaches to constrain surface A_T and C_T distributions. T = Temperature, S = Salinity, DO = Dissolved Oxygen, AOU = Apparent Oxygen Utilization, N = Nitrate (NO_3^-), Si = Silicate (SiO_4), P = Phosphate (PO_4^{3-}), Chl a = Chlorophyll a , Lat = Latitude, Long = Longitude.

Study Region	Response	Predictors	N^a	RSE ^b ($\mu\text{mol kg}^{-1}$)	Author
Global	NA_T^c	T	1740	5	Millero et al. (1998)
Global	A_T	T, T^2 , S, S^2 , Long	5692	8.1	Lee et al. (2006)
Indian Ocean	A_T	T, S, N, AOU, Depth, Lat, P	2363	4.5–6.4 ^d	Bates et al. (2006)
Southern Ocean	A_T	S, N, Si	1200	8.1	McNeil et al. (2007)
Arctic Ocean	A_T	T, S	853	26.9, 75	Arrigo et al. (2010)
Global	NC_T^c	T, T^2 , N	~ 4900	7	Lee et al. (2000)
Indian Ocean	C_T	T, S, N, AOU, Depth, Lat, P	2395	4.4–6.0 ^d	Bates et al. (2006)
Southern Ocean	C_T	T, S, DO, N, Si	1032	8	McNeil et al. (2007)
Arctic Ocean	C_T	Chl a , T, S	853	33.4, 61.6, 17.3	Arrigo et al. (2010)

^a Number of measurements used in the study.

^b Residual Standard Error.

^c Salinity normalized concentrations of C_T and A_T ($\times \frac{35}{5}$).

^d Range of RSE values presented for the four monsoonal/inter-monsoonal seasons.

Title Page

Abstract

Introduction

Conclusions

References

Tables

Figures

⏪

⏩

◀

▶

Back

Close

Full Screen / Esc

Printer-friendly Version

Interactive Discussion

Diagnosing seasonal to inter-annual surface ocean carbon dynamics

T. P. Sasse et al.

Title Page

Abstract

Introduction

Conclusions

References

Tables

Figures

⏪

⏩

◀

▶

Back

Close

Full Screen / Esc

Printer-friendly Version

Interactive Discussion

Table 2. Data sources of our global merged dataset.

Source	Number of Measurements
CARINA	12 599
PACIFICA	9690
GLODAP	6674
CLIVAR ^a	1689
AAIW ^b	755
BATS ^c	705
HOT ^d	540
NACP ^e	291
Miscellaneous	192
Total	33 135

^a Climate Variability and Predictability.

^b Antarctic Intermediate Cruise.

^c Bermuda Atlantic time-series.

^d Hawaiian Ocean time-series.

^e North Atlantic Carbon Program.

Diagnosing seasonal to inter-annual surface ocean carbon dynamics

T. P. Sasse et al.

Title Page

Abstract

Introduction

Conclusions

References

Tables

Figures

⏪

⏩

◀

▶

Back

Close

Full Screen / Esc

Printer-friendly Version

Interactive Discussion



Table 3. Universal and ad-hoc MLR results for **(a)** C_T and **(b)** A_T .

Region	Zone ^a	N^b	N cruises ^c	RSE ($\mu\text{mol kg}^{-1}$)			
				Regression		Independent test (GIT)	
				Ad-hoc	Universal	Ad-hoc	Universal
(a)							
Sub-trop	1	5388	109	11.9	17.1	15.2	17.3
Eq Pac	2	752	14	11.3	16.8	18.9	17.7
North Atl	3	4626	69	13.2	15.5	15.5	16.2
North Pac	4	2344	112	17.7	17.2	16.8	17.5
SO	5	7856	75	12.5	12.4	16.4	12.8
Global		20 966	289	13.2	15.1	16.0	15.6
(b)							
Sub-trop	1	4917	94	10.2	10.2	11.0	10.4
Eq Pac	2	513	7	6.9	12.4	9.4	13.0
North Atl	3	3181	53	7.7	10.0	7.9	10.1
North Pac	4	1956	88	14.3	16.4	14.8	16.6
SO	5	6084	58	8.0	9.1	9.4	9.8
Global		16 651	224	9.5	10.8	10.4	11.1

^a Corresponding geographical region in Fig. 2.

^b Number of measurements in the corresponding region.

^c Number of unique cruises/time series in the region.

BGD

9, 15329–15380, 2012

Diagnosing seasonal to inter-annual surface ocean carbon dynamics

T. P. Sasse et al.

[Title Page](#)[Abstract](#)[Introduction](#)[Conclusions](#)[References](#)[Tables](#)[Figures](#)[Back](#)[Close](#)[Full Screen / Esc](#)[Printer-friendly Version](#)[Interactive Discussion](#)

Table 4. Summary of the three independent datasets used to constrain the general configuration of the SOMLO model.

Independent dataset	Number of samples	Percentage of global dataset
T1	3769	11.4
T2	2919	8.8
T3	3391	10.2
Total	10 079	30.4

Diagnosing seasonal to inter-annual surface ocean carbon dynamics

T. P. Sasse et al.

Title Page

Abstract

Introduction

Conclusions

References

Tables

Figures

⏪

⏩

◀

▶

Back

Close

Full Screen / Esc

Printer-friendly Version

Interactive Discussion



Table 5. Skill comparison between the traditional MLR and SOMLO approaches for **(a)** C_T and **(b)** A_T .

Region	Zone ^a	RSE ($\mu\text{mol kg}^{-1}$)		
		Ad-hoc MLR	SOMLO	% Improvement
(a)				
Sub-trop	1	15.2	13.5	11.2
Eq Pac	2	18.9	13.3	29.7
North Atl	3	15.5	11.7	24.5
North Pac	4	16.8	14.3	14.9
SO	5	16.4	12.7	22.6
Global		16.0 (15.6) ^b	12.9	19.4 (17.4) ^b
(b)				
Sub-trop	1	11.0	9.2	16.4
Eq Pac	2	9.4	9.6	0
North Atl	3	8.0	8.5	0
North Pac	4	14.8	14.4	2.7
SO	5	9.4	8.8	6.4
Global		10.4 (11.1) ^b	9.7	6.7 (12.6) ^b

^a Corresponding geographical region in Fig. 2.

^b Universal MLR.

Diagnosing seasonal to inter-annual surface ocean carbon dynamics

T. P. Sasse et al.

Title Page

Abstract Introduction

Conclusions References

Tables Figures

◀ ▶

◀ ▶

Back Close

Full Screen / Esc

Printer-friendly Version

Interactive Discussion

Discussion Paper | Discussion Paper | Discussion Paper | Discussion Paper | Discussion Paper

Table 6. Regional and global SOMLO skill evaluation.

Region	Zone ^a	RSE ^b C _T	RSE ^b A _T	N ^c C _T	N ^c A _T
Arctic Ocean	1	26.6	22.1	782	795
Sup-Polar North Atlantic	2	11.6	9.0	4425	2641
Sub-Tropical North Atlantic	3	9.1	6.6	1481	1254
Equatorial Atlantic	4	13.7	13.0	654	582
Sub-Tropical South Atlantic	5	10.6	8.7	659	551
Sub-Polar North Pacific	6	11.2	14.7	2053	1615
Sub-Tropical North Pacific	7	11.1	8.2	2367	1446
Equatorial Pacific	8	11.2	8.3	1524	802
Sub-Tropical South Pacific	9	12.3	7.7	1824	1404
Sub-Tropical North Indian (Exc. Bay of Bengal)	10	22.1 (13.9)	13.4 (7.5)	143 (111)	168 (136)
Equatorial Indian	11	11.8	7.7	512	500
Sub-Tropical South Indian	12	11.5	5.6	1411	1388
Southern Ocean	13	8.7	8.8	3950	3088
Subantarctic waters	14	9.5	8.5	2250	1474
Global		11.8	10.2	24 035	17 708
Global (below 70° N)		10.9	9.2	23 253	16 913

^a Corresponding geographical region in Fig. J1.

^b Residual Standard Error ($\mu\text{mol kg}^{-1}$).

^c Number of measurements in the region.



Diagnosing seasonal to inter-annual surface ocean carbon dynamics

T. P. Sasse et al.

Title Page	
Abstract	Introduction
Conclusions	References
Tables	Figures
⏪	⏩
◀	▶
Back	Close
Full Screen / Esc	
Printer-friendly Version	
Interactive Discussion	

Table 7. Comparison to previous empirical approaches.

Study	Response Variable	N ^d	RSE (μmol kg ⁻¹)			% improvement	Author
			RSE _(MLR-old)	RSE _(GIT)	SOMLO _(GIT)		
Global ^a	C _T	13 881	22.0	17.8	12.8	28	Lee et al. (2000)
Indian ^b	C _T	2052	15.2	21.4	13.0	39	Bates et al. (2006)
Southern Ocean	C _T	4196	17.3	8.8	9.0	0	McNeil et al. (2007)
Global (exc. North Pacific) ^c	A _T	10 360 (8995)	11.7 (10.3)	10.9 (10.4)	10.7 (9.9)	2	Lee et al. (2006)
Indian ^b	A _T	2042	9.4	11.8	7.1	40	Bates et al. (2006)
Southern Ocean	A _T	4196	10.3	10.3	9.3	10	McNeil et al. (2007)

^a Using only surface data (above 30 m).

^b Only measurements from within our defined mixed-layer were used to constrain new regressions and test previous regressions.

^c 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 0–360°.

^d Number of measurements.



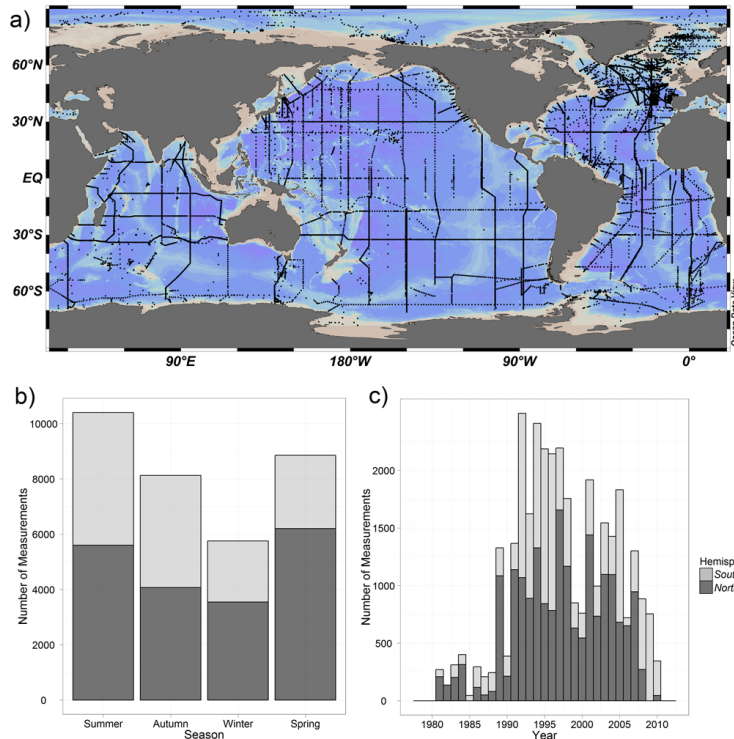


Fig. 1. (a) Global distribution of the training dataset, (b) seasonal and (c) yearly histograms of the training dataset separated into Southern (light shade) and Northern (dark shade) Hemisphere. Southern Hemisphere seasons are defined as Summer (December–February), Autumn (March–May), Winter (June–August) and Spring (September–November), Northern Hemisphere seasons are opposite.

Diagnosing seasonal to inter-annual surface ocean carbon dynamics

T. P. Sasse et al.

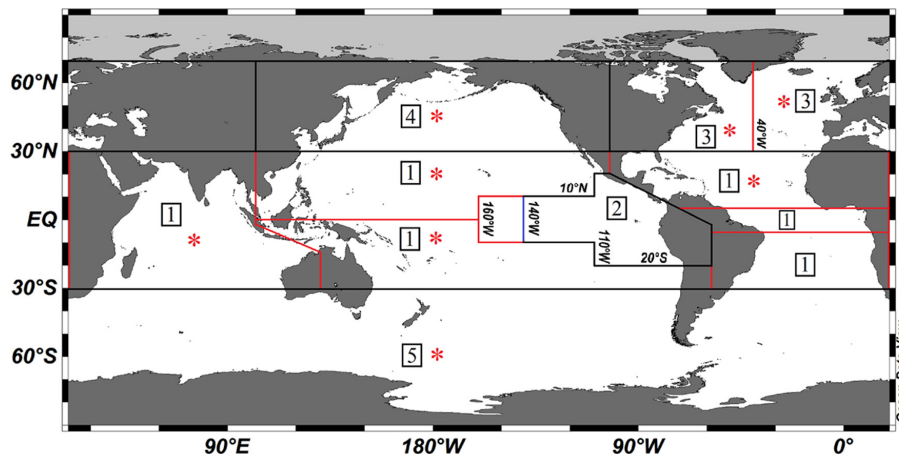


Fig. 2. Spatio-temporal division of the global training dataset for the ad-hoc MLR approach. Black boundaries are common for both C_T and A_T models, whilst red boundaries are for C_T only and blue for A_T only. A red asterisk indicates that MLR's were developed for both Summer (November–April for Austral hemisphere) and winter (May–October for Austral hemisphere) periods to constrain C_T . Boreal Summer/Winter seasons are opposite.

Discussion Paper | Discussion Paper | Discussion Paper | Discussion Paper | Discussion Paper

Title Page

Abstract

Introduction

Conclusions

References

Tables

Figures

⏪

⏩

◀

▶

Back

Close

Full Screen / Esc

Printer-friendly Version

Interactive Discussion



Diagnosing seasonal to inter-annual surface ocean carbon dynamics

T. P. Sasse et al.

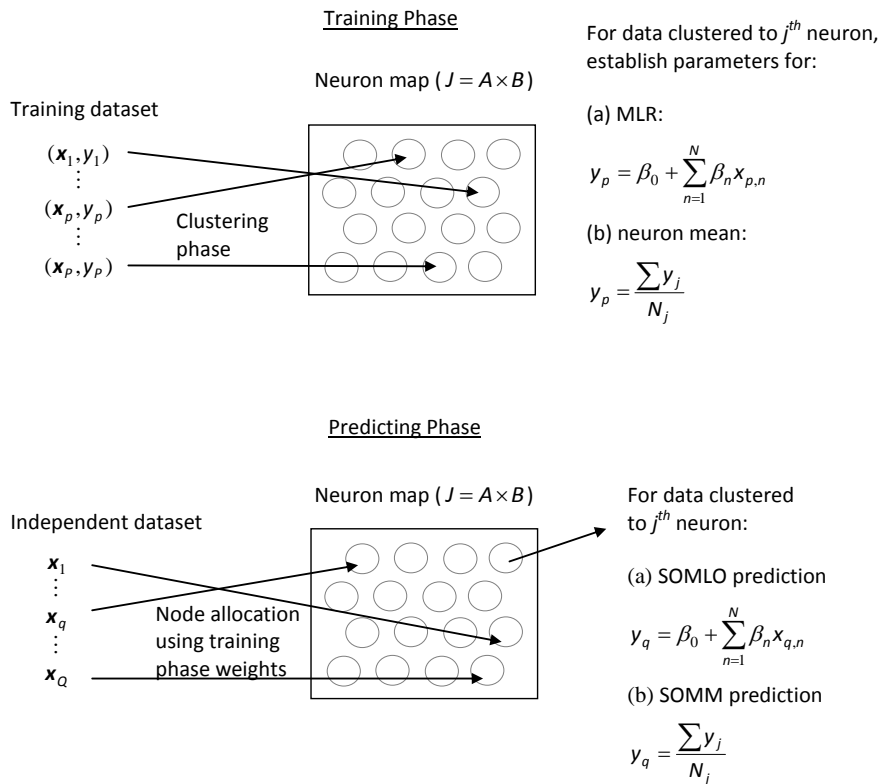


Fig. 3. Schematic diagram of neural network training and prediction phases.

Title Page

Abstract Introduction

Conclusions References

Tables Figures

⏪ ⏩

◀ ▶

Back Close

Full Screen / Esc

Printer-friendly Version

Interactive Discussion



**Diagnosing seasonal
to inter-annual
surface ocean carbon
dynamics**

T. P. Sasse et al.

Title Page

Abstract

Introduction

Conclusions

References

Tables

Figures

◀

▶

◀

▶

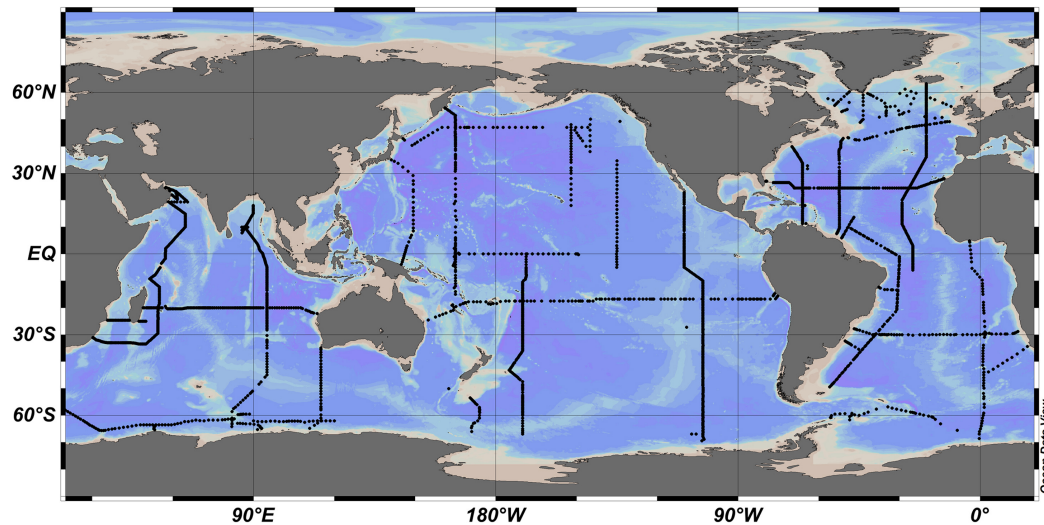
Back

Close

Full Screen / Esc

Printer-friendly Version

Interactive Discussion

**Fig. 4.** T1 independent dataset distribution.

Diagnosing seasonal to inter-annual surface ocean carbon dynamics

T. P. Sasse et al.

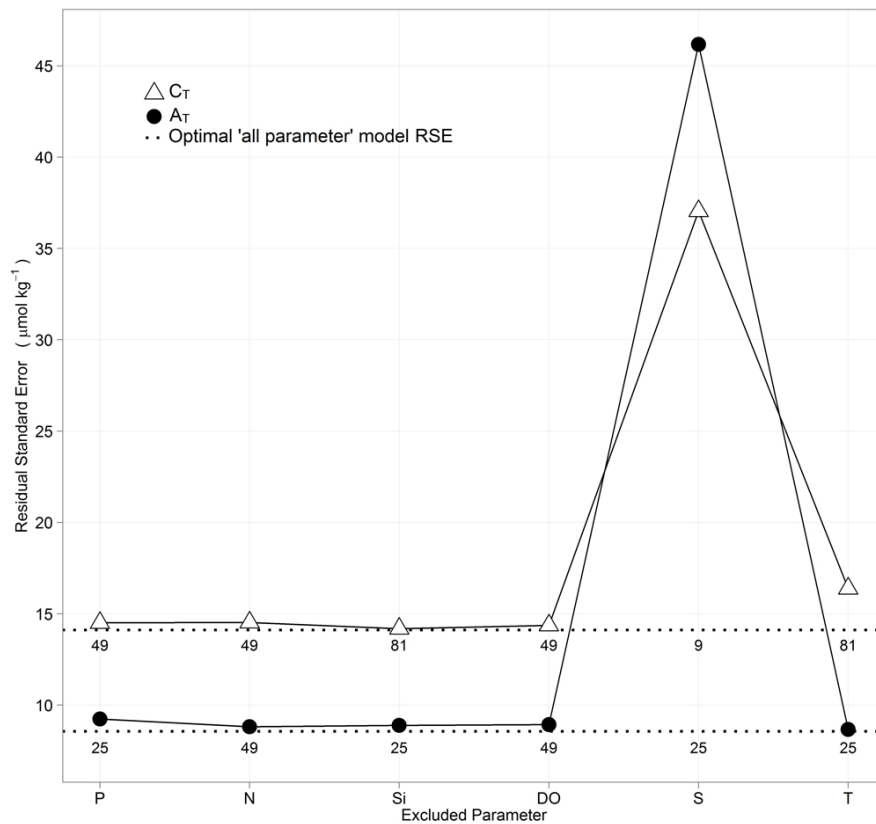


Fig. 5. RSE results for the SOMLO when applied to the three independent datasets. Numbers under the dotted line represents the optimal number of neurons to constrain the system.

Title Page

Abstract Introduction

Conclusions References

Tables Figures

◀ ▶

◀ ▶

Back Close

Full Screen / Esc

Printer-friendly Version

Interactive Discussion



Diagnosing seasonal to inter-annual surface ocean carbon dynamics

T. P. Sasse et al.

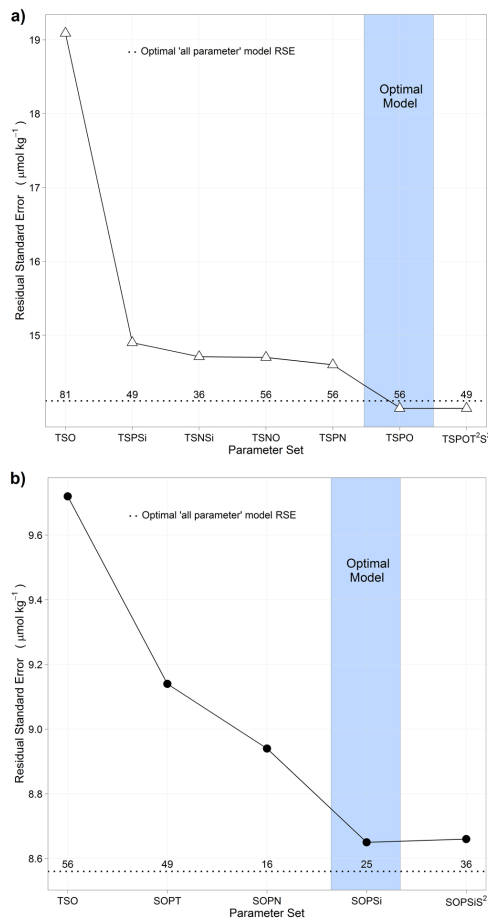


Fig. 6. Optimal RSE values for **(a)** C_T and **(b)** A_T SOMLO models. Numbers above the line represent the optimal number of neurons.

Title Page

Abstract Introduction

Conclusions References

Tables Figures

⏪ ⏩

◀ ▶

Back Close

Full Screen / Esc

Printer-friendly Version

Interactive Discussion



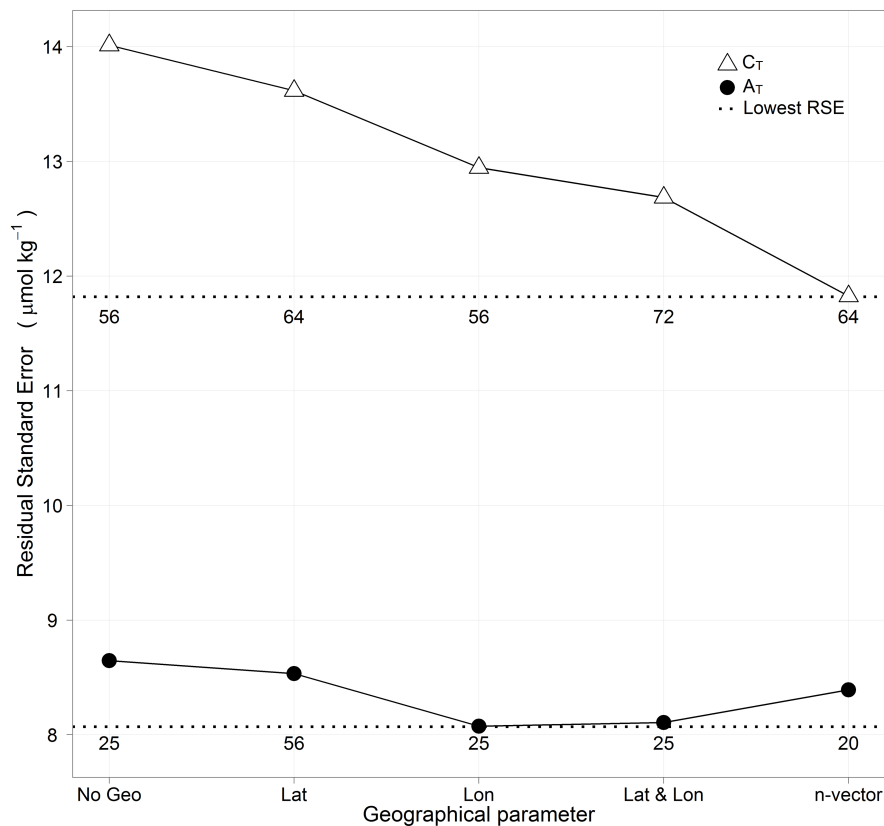


Fig. 7. Skill of optimal SOMLO models with geographical constraints. Numbers below dashed line represent the optimal number of neurons.

Diagnosing seasonal to inter-annual surface ocean carbon dynamics

T. P. Sasse et al.

Title Page

Abstract Introduction

Conclusions References

Tables Figures

⏪ ⏩

◀ ▶

Back Close

Full Screen / Esc

Printer-friendly Version

Interactive Discussion



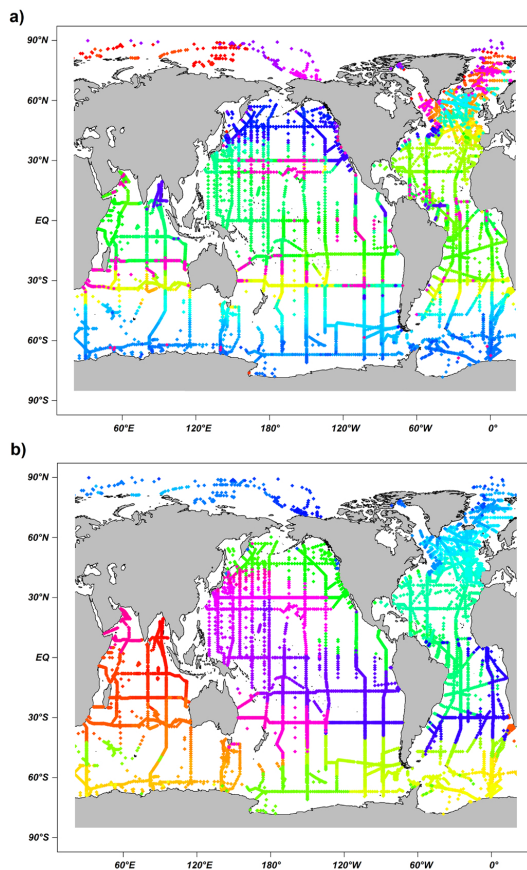


Fig. 8. Distribution of assigned neurons for SOM models trained with **(a)** biogeochemical information only and **(b)** biogeochemical and geographical information.

Diagnosing seasonal to inter-annual surface ocean carbon dynamics

T. P. Sasse et al.

Title Page

Abstract Introduction

Conclusions References

Tables Figures

⏪ ⏩

◀ ▶

Back Close

Full Screen / Esc

Printer-friendly Version

Interactive Discussion



Diagnosing seasonal to inter-annual surface ocean carbon dynamics

T. P. Sasse et al.

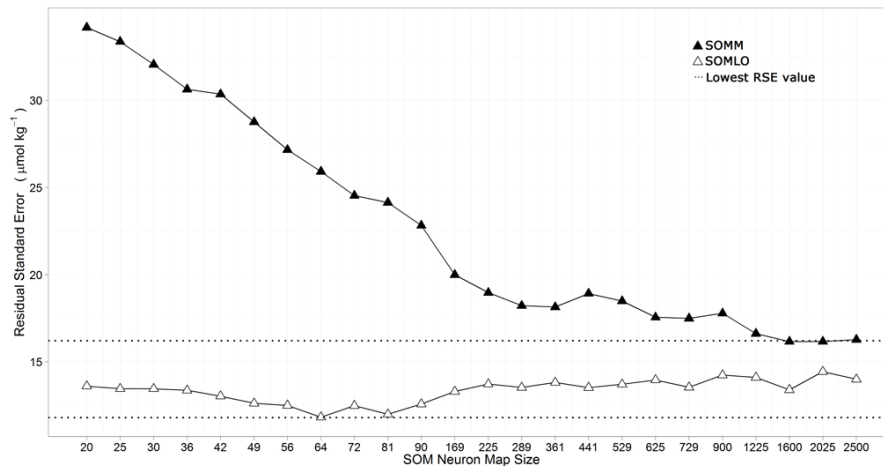


Fig. 9. Skill comparison between the SOMLO and SOMM models in capturing C_T .

Title Page

Abstract Introduction

Conclusions References

Tables Figures

⏪ ⏩

◀ ▶

Back Close

Full Screen / Esc

Printer-friendly Version

Interactive Discussion



Diagnosing seasonal to inter-annual surface ocean carbon dynamics

T. P. Sasse et al.

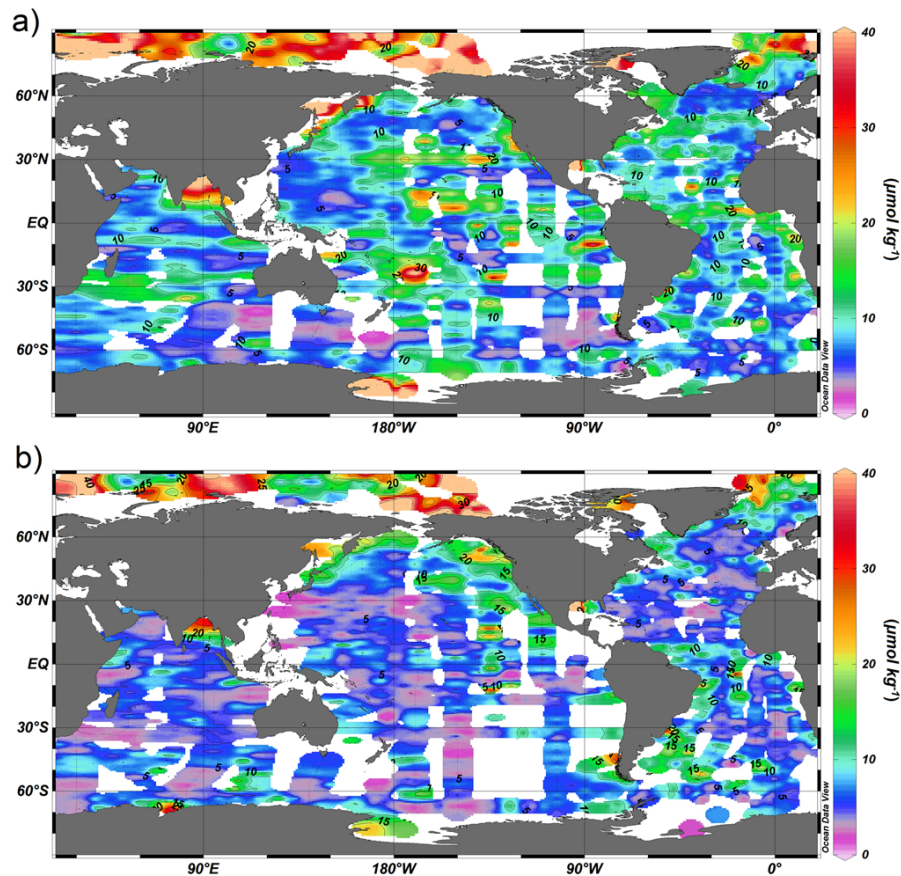


Fig. 10. Distribution of global independent test (GIT) absolute residual errors for (a) C_T and (b) A_T .

[Title Page](#)[Abstract](#)[Introduction](#)[Conclusions](#)[References](#)[Tables](#)[Figures](#)[⏪](#)[⏩](#)[◀](#)[▶](#)[Back](#)[Close](#)[Full Screen / Esc](#)[Printer-friendly Version](#)[Interactive Discussion](#)

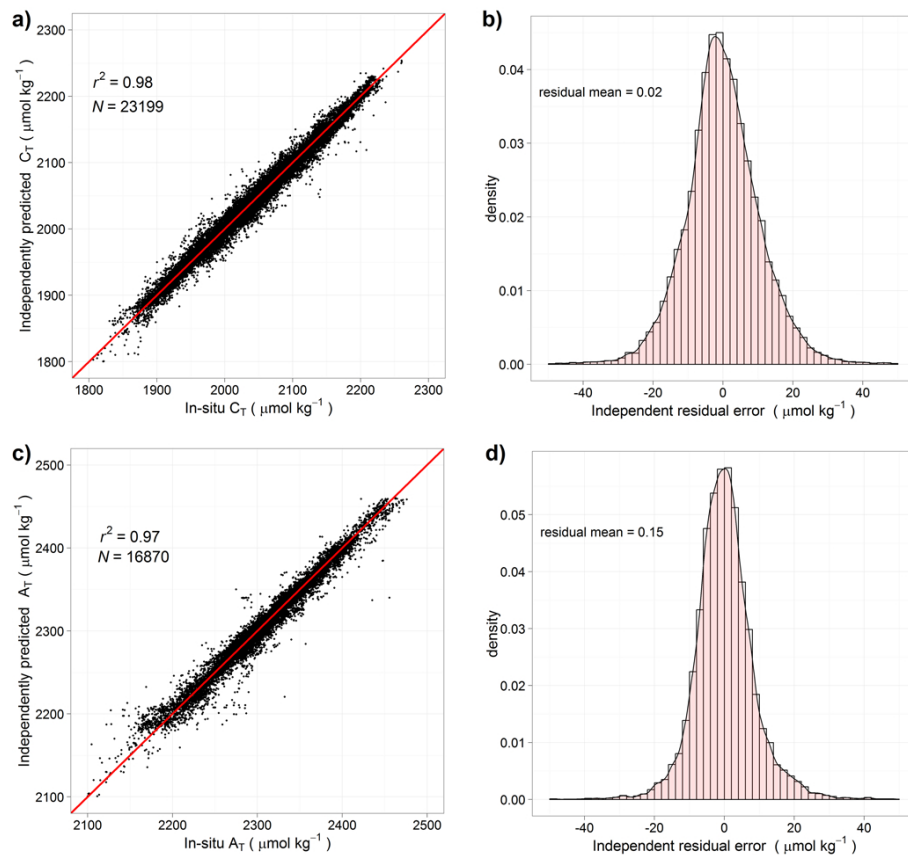


Fig. 11. Global independent test (GIT) predictions versus in-situ measurements and residual error density distribution for optimal **(a–b)** C_T and **(c–d)** A_T SOMLO configurations. $r^2 = r$ -squared correlation and $N =$ number of samples.

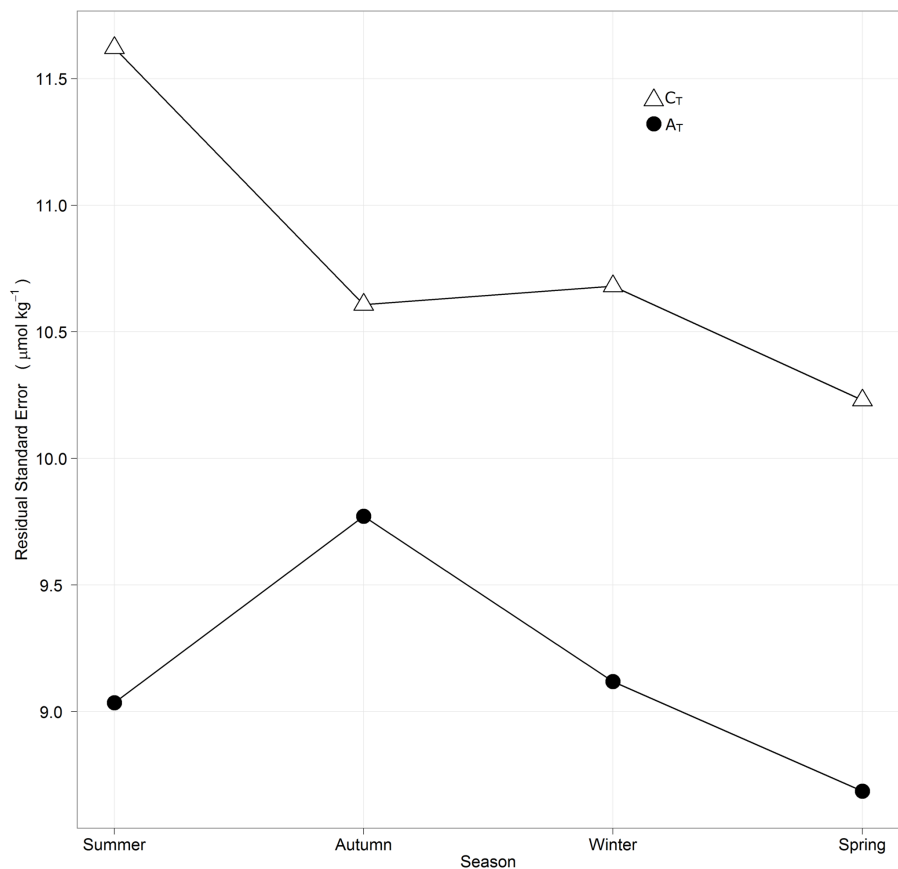


Fig. 12. SOMLO C_T/A_T seasonal independent test RSE values. For austral hemisphere seasons are defined as Summer (December–February), Autumn (March–May), Winter (June–August) and Spring (September–November). Boreal hemisphere seasons differ by 6 months.

Diagnosing seasonal to inter-annual surface ocean carbon dynamics

T. P. Sasse et al.

Title Page

Abstract

Introduction

Conclusions

References

Tables

Figures

◀

▶

◀

▶

Back

Close

Full Screen / Esc

Printer-friendly Version

Interactive Discussion



Diagnosing seasonal to inter-annual surface ocean carbon dynamics

T. P. Sasse et al.

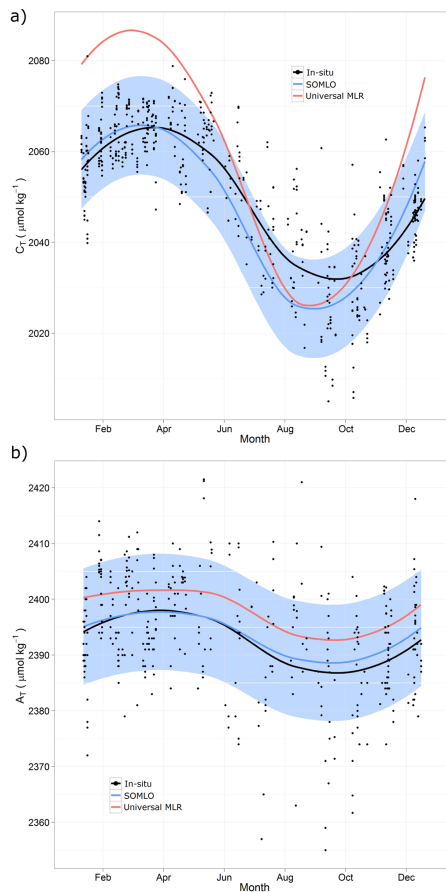
[Title Page](#)[Abstract](#)[Introduction](#)[Conclusions](#)[References](#)[Tables](#)[Figures](#)[◀](#)[▶](#)[◀](#)[▶](#)[Back](#)[Close](#)[Full Screen / Esc](#)[Printer-friendly Version](#)[Interactive Discussion](#)

Fig. 13. BATS in-situ and independently predicted seasonal cycles for (a) C_T and (b) A_T . Blue shaded region illustrates the uncertainty in SOMLO predictions.

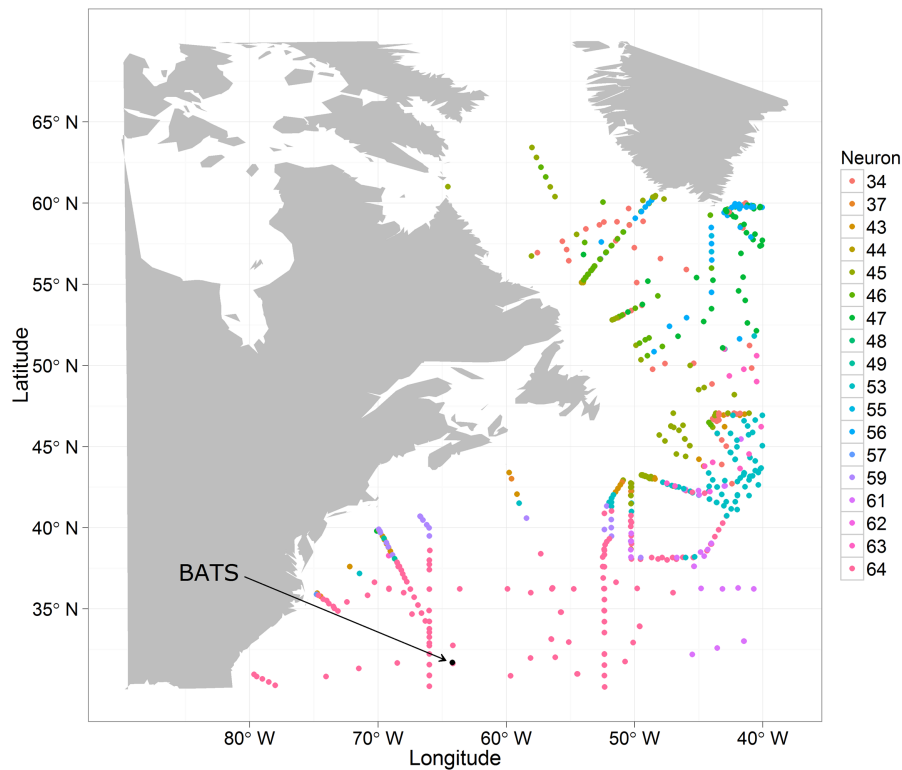


Fig. 14. Distribution of assigned neurons in the Northwest Atlantic region for optimal C_T SOMLO model.

Diagnosing seasonal to inter-annual surface ocean carbon dynamics

T. P. Sasse et al.

Title Page

Abstract Introduction

Conclusions References

Tables Figures

⏪ ⏩

◀ ▶

Back Close

Full Screen / Esc

Printer-friendly Version

Interactive Discussion



Diagnosing seasonal to inter-annual surface ocean carbon dynamics

T. P. Sasse et al.

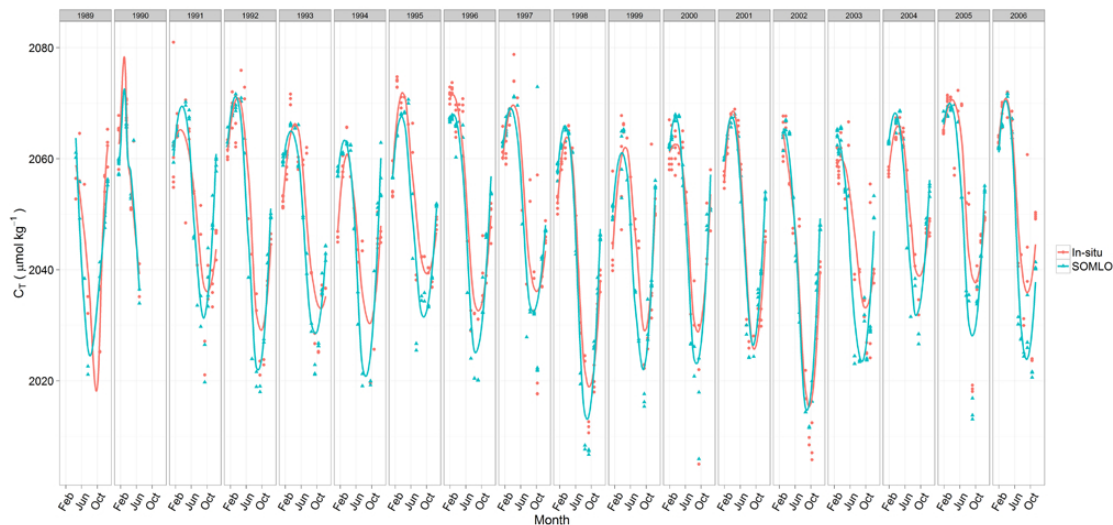


Fig. 15. In-situ and independently predicted BATS C_T measurements partitioned into years with loess line.

Discussion Paper | Discussion Paper | Discussion Paper | Discussion Paper | Discussion Paper

Title Page

Abstract Introduction

Conclusions References

Tables Figures

⏪ ⏩

◀ ▶

Back Close

Full Screen / Esc

Printer-friendly Version

Interactive Discussion



Diagnosing seasonal to inter-annual surface ocean carbon dynamics

T. P. Sasse et al.



Fig. 16. In-situ and independently predicted HOT C_T measurements with loess line.

Title Page

Abstract

Introduction

Conclusions

References

Tables

Figures



Back

Close

Full Screen / Esc

Printer-friendly Version

Interactive Discussion

