Neural network-based estimates of Southern Ocean net community production from in-situ O_2/Ar and satellite observation: A methodological study

Chueh-Hsin Chang[1,2], Nathaniel C. Johnson[3,4], Nicolas Cassar[2]

Biogeosciences

Sep 5th, 2013 submitted

Jan 30th 2014 revised MS

[1] Research Center for Environmental Changes, Academia Sinica, Taipei 11529, Taiwan.

[2] Division of Earth and Ocean Sciences, Nicholas School of the Environment, Duke University, Durham, NC 27708, USA.

[3] International Pacific Research Center, SOEST, University of Hawai'i at Manoa, Honolulu, HI 96822, USA.

[4] Scripps Institution of Oceanography, La Jolla, CA 92037, USA.

Abstract

Southern Ocean organic carbon export plays an important role in the global carbon cycle, yet its basin-scale climatology and variability are uncertain due to limited coverage of *in situ* observations. In this study, a neural network approach based on the self-organizing map (SOM) is adopted to construct weekly gridded ($1^{o}\times1^{o}$) maps of organic carbon export for the Southern Ocean from 1998 to 2009. The SOM is trained with *in situ* measurements of O_2/Ar-derived net community production (NCP) that are tightly linked to the carbon export in the mixed layer on timescales of 1–2 weeks, and six potential NCP predictors: photosynthetically available radiation (PAR), particulate organic carbon (POC), chlorophyll (Chl), sea surface temperature (SST), sea surface height (SSH), and mixed layer depth (MLD). This non-parametric approach is based entirely on the observed statistical relationships between NCP and the predictors, and therefore is strongly constrained by observations.

A thorough cross-validation yields three retained NCP predictors, Chl, PAR, and MLD. Our constructed NCP is further validated by good agreement with previously published independent *in situ* derived NCP of weekly or longer temporal resolution through real-time and climatological comparisons at various sampling sites. The resulting November–March NCP climatology reveals a pronounced zonal band of high NCP roughly following the Subtropical Front in the Atlantic, Indian and western Pacific sectors, and turns southeastward shortly after the dateline. Other regions of elevated NCP include the upwelling zones off Chile and Namibia, Patagonian Shelf, Antarctic coast, and areas surrounding the Islands of Kerguelen, South Georgia, and Crozet. This basin-scale NCP climatology closely resembles that of the satellite POC field and observed air-sea CO_2 flux. The long-term mean area-integrated NCP south of 50^{o}S from our dataset, 17.9 mmol C $m^{-2}d^{-1}$, falls within the range of 8.3–24 mmol C $m^{-2}d^{-1}$ from other model estimates. A broad agreement is found in the basin-wide NCP climatology among various models but with significant spatial variations, particularly in the Patagonian Shelf. Our approach provides a comprehensive view of the Southern Ocean NCP climatology and a potential opportunity to further investigate interannual and intraseasonal variability.

1. Introduction

The Southern Ocean plays an important role in the global carbon cycle. The current annual global ocean uptake of atmospheric carbon dioxide (CO_2) is about 2 petagrams (Pg) of carbon, half of which is taken up by the vast Southern Ocean south of 30°S [*Takahashi et al.*, 2012]. Atmospheric CO_2 absorbed by the ocean can be transferred from the surface to the deep ocean via various physical, chemical and biological mechanisms associated with the solubility and biological pumps [*Volk and Hoffert*, 1985; *Carlson et al.*, 2010].

Biological carbon export from the ocean surface is a function of various processes, including net community production (NCP), which reflects the metabolic balance between gross primary production (GPP) and community respiration [*Codispoti et al.*, 1986; *Minas et al.*, 1986]. It describes the net rate at which CO_2 is transformed to particulate and dissolved organic carbon (POC and DOC). For the present study, we use NCP estimates derived from *in situ* measurements of the elemental ratio of O_2/Ar. The O_2/Ar method measures biological O_2 supersaturation in the mixed layer [*Craig and Hayward*, 1987], and yields NCP estimates over the O_2 residence timescale (1–2 weeks) [*Reuer et al.*, 2007; *Cassar et al.*, 2007, 2009, 2011]. On this timescale, the NCP derived from this method is tightly linked to the export of organic carbon from the mixed layer at steady state, under the assumptions that both vertical mixing of O_2-depleted waters from below and accumulation of POC and DOC in the mixed layer are negligible [*Cassar et al.*, 2009, 2011; *Jonsson et al.*, 2013]. Although we use NCP and carbon export production interchangeably in this study, it should be noted that under some circumstances, the assumption of steady-state is violated [*Hamme et al.*, 2012; *Jonsson et al.*, 2013].

While *in situ* O_2/Ar measurements shed new light on the NCP distribution and variability, the Southern Ocean remains seriously undersampled. The difficulty in obtaining a large-scale picture of the carbon export owes to the unavailability of direct satellite measurements. In addition, NCP is highly variable in space and time and cannot be derived from linear interpolation between *in situ* measurements. Field experiments also reveal that the plankton ecosystem and CO_2 flux variability are not dominated by just one single mechanism but by a confluence of several processes that shift in relative importance over time and space [*Banse*, 1996; *Abbott et al.*, 2000, 2001; *Cassar et al.*, 2011; *Tortell et al.*, 2012], which are difficult to capture in biogeochemical models.

An alternative strategy is to use a data-driven modeling approach. We may achieve a more comprehensive characterization of temporal and spatial variability of NCP by examining the statistical relationships between NCP and physical as well as biogeochemical properties that potentially have impacts on carbon export. In addition to mixed layer depth (MLD) and light (i.e., photosynthetically available radiation (PAR)) [*Cassar et al.*, 2011], POC, Chl, sea surface temperature (SST), and sea surface height (SSH) are likely important factors regulating or correlated with NCP in the Southern Ocean. POC production is the dominant form of NCP in the Southern Ocean [*Ogawa et al.*, 1999; *Wiebinga and de Baar,* 1998; *Kaehler et al.*, 1997; *Hansell and Carlson*, 1998; *Sweeney et al.*, 2000; *Schlitzer*, 2002; *Ishii et al.*, 2002; *Allison et al.*, 2010], and Chl concentration is commonly used to estimate net primary production (NPP) from satellites [*Behrenfeld and Falkowsky*, 1997; *Moore and Abbott*, 2000; *Campbell et al.*, 2002; *Carr et al.*, 2006; *Bissinger et al.*, 2008; *Friedrichs et al.*, 2009; *Saba et al.*, 2011; *Friedland et al.*, 2012; *Nevison et al.*, 2012; *Olonscheck et al.*, 2013]. SST has been used to derive export and export efficiency based on the relationship with NPP and through its influence on heterotrophic activity [*Laws et al.*, 2000; *Laws* 2004; *Laws et al.*, 2011]. SSH yields information on oceanic eddies, fronts, and nutrient transport that are crucial to spatial variation of biological activity [*Abbott et al.*, 2000, 2001; *Glorioso et al.*, 2005; *Kahru et al.*, 2007; *Gruber et al.*, 2011].

Advances in remote sensing and statistical algorithms now permit satellite data-driven modeling of NCP. Satellite-borne sensors have accumulated records for a decade or longer of PAR, POC, Chl, SST, and SSH of sufficient resolution and coverage in space and time. Southern Ocean MLD products became available in recent years from Argo float profiles [*Wong* 2005; *Sallée et al.*, 2006; *Schneider and Bravo*, 2006; *Dong et al.*, 2008] as well as from high resolution ocean general circulation models (OGCMs) [*Aoki et al.*, 2007a; *Sterl et al.*, 2012]. In this study, we combine the *in situ* NCP measurements from 60 crossings spanning more than a decade with gridded datasets of NCP predictors, PAR, POC, Chl, SST, SSH, and MLD, to generate weekly, gridded maps of NCP estimates over the Southern Ocean from 1998 through 2009. We generate these NCP predictions through the use of self-organizing map (SOM) analysis, a type of clustering approach that has arisen in the field of artificial neural networks

[*Kohonen*, 2001]. SOM analysis has gained in popularity in the atmospheric and ocean sciences over the past decade, with applications in categorizing atmospheric teleconnection patterns [*Reusch et al.*, 2005; *Johnson et al.*, 2008; *Johnson and Feldstein*, 2010; *Johnson*, 2013], and in generating maps of pCO_2 for the North Atlantic [*Friedrich and Oschlies*, 2009; *Telszewski et al.*, 2009] and for the global ocean [*Sasse et al.*, 2013].

In the present application, we follow the general approach of *Friedrich and Oschlies* [2009] and *Telszewski et al.* [2009], whereby we use the SOM with the combined purpose of cluster analysis and nonlinear, nonparametric regression between a set of predictors and NCP. Under this approach, which we describe more thoroughly in section 3, we allow the data to determine the potentially complex relationships between the predictors and NCP. Thus, the predictor/NCP relationships are unconstrained by any pre-conceived, uncertain functional forms and are determined entirely from the observed data, which contrasts previous studies of Southern Ocean NCP. Nevertheless, we find that our estimates of NCP agree broadly with previous estimates while also providing additional information on temporal and spatial variability. The remainder of the paper is organized as follows. In section 2 we describe the data used in the study. Section 3 provides a description of the SOM methodology for generating weekly NCP maps and for calculating error estimates. In section 4 we present our results, noting some of the most salient features from the constructed NCP dataset. Section 5 provides a discussion and conclusions.

2. Data

We make extensive use of gridded data products and cruise measurements in the Southern Ocean domain poleward of 20^{o}S and for the period between 1998 and 2009. The gridded and research cruise data are described below.

2.1. Gridded predictor data

We consider six gridded data products, PAR, POC, Chl, SST SSH, and MLD, as potential predictors of NCP for use in the SOM analysis and for the generation of weekly NCP maps, as described more thoroughly in section 3.

We utilize satellite PAR and POC from the Moderate Resolution Imaging Spectroradiometer flown on the Aqua satellite (MODIS-Aqua) 8-day mean 9 km for the period 10 July 2002−30 December 2009. The weekly averaged Chl are constructed from the daily 9-km maps of Sea-viewing Wide Field-of-view Sensor (SeaWiFS), version 5.2 for the period 7 January 1998−26 December 2007 [*O'Reilly et al.*, 1998]. For SST, we use NOAA Optimum Interpolation 0.25^o Daily SST Blended with Advanced Very High Resolution Radiometer (AVHRR) and AMSR version 2 data (OI SST) [*Reynolds et al.*, 2007] for the period 7 January 1998−19 August 2009. The weekly SSH anomaly maps are obtained from the Archiving, Validation and Interpretation of Satellite Oceanographic Data (AVISO) on about a 1/3^o × 1/3^o grid [*Ducet et al.*, 2000] from 7 January 1998 to 22 July 2009. To determine the absolute SSH, we added the AVISO SSH anomaly to the sea level climatology of Niiler and Maximenko [*Niiler et al.*, 2003; *Maximenko et al.*, 2009]. We choose this particular SSH climatology because it has high spatial resolution [*Sokolov and Rintoul*, 2007].

Because the coverage of Argo float profiles is not homogenous [*Akoi et al.*, 2007a], and the available gridded Argo data are either of coarser resolutions or shorter time periods (http://www.argo.ucsd.edu/Gridded_fields.html), we choose the MLD of the high resolution OGCM for the Earth Simulator (OFES) [*Masumoto et al.*, 2004; *Sasaki et al.*, 2006, 2008]. The OFES is an eddy-resolving quasi-global (75^{o}N −75^{o}S) ocean model based on the Geophysical Fluid Dynamics Laboratory Modular Ocean Model version 3 (GFDL MOM3) with 0.1^o

horizontal resolution and 54 vertical levels. It provides MLD at 0.1-degree spatial resolution every three days. The model captures realistic upper ocean dynamics, including eddies and heat balance [*Sasaki and Nonaka*, 2006; *Taguchi et al.*, 2007; *Scott et al.*, 2008; *Zhuang et al.*, 2010; *Yoshida et al.*, 2010; *Sasaki et al.*, 2011; *Chang et al.*, 2012], and has been used to investigate the Southern Ocean dynamical variability [*Aoki et al.*, 2007a, 2007b, 2010; *Sasaki and Schneider*, 2008; *Thompson et al.*, 2010; *Thompson and Richards*, 2011]. In the present study, we use the MLD from the OFES simulation forced by the QuikSCAT satellite wind field from 28 July 1999 to 28 October 2009.

For the interpolation of the predictor data to the daily ship track locations, all gridded data are first interpolated to daily resolution. Although we interpolate all predictor data to the daily ship tracks, sub-weekly variability is missing from those predictors of original temporal resolutions of 7-8 days. For the generation of weekly NCP maps, all gridded predictor data are interpolated to a common 1^{o} x 1^{o} latitude-longitude grid poleward of 20^{o}S at weekly temporal resolution.

2.2. Research cruise data

In the SOM analysis described below, the predictand of interest is an estimate of NCP from an extensive set of published data obtained from 41 research cruises in the Southern Ocean between 1999 and 2009 [*Reuer et al.*, 2007; *Cassar et al.*, 2007, 2011]. Figure 1a shows our ship tracks with time of the cruises color-coded in months. We see that the ship tracks mainly cover regions of high chlorophyll (see Figure 2c) during the growing season between November and March. The histogram of the ship track NCP distribution is shown in Figure 1b. From visual inspection, we also exclude spuriously large NCP outliers exceeding 180 mmol C $m^{-2}d^{-1}$. Figure 1c provides a detailed distribution of NCP below the outlier threshold. For all available ship track data, which are sampled unevenly in time, we calculate the daily mean NCP, latitude, and longitude. We then linearly interpolate all available daily gridded predictor data to the ship track locations. Negative NCP values are possibly due to net heterotrophy or measurements contaminated by the upwelling of oxygen-undersaturated water. Because we are unable to estimate this potential bias, we exclude all days with negative NCP values prior to the SOM analysis. Overall, we retain 401 days of ship track data for the SOM analysis. All NCP and predictor data are standardized for the SOM analysis. Owing to the skewness of the NCP, Chl, MLD and POC data, we perform a log10 transformation to these variables prior to the standardization. As a result, the SOM analysis is applied to all predictor and predictand data that have approximately Gaussian distributions with a mean of zero and a standard deviation of one.

In this study, the growing season is defined as November through March. Unless otherwise noted, all units are converted to mmol C $m^{-2}d^{-1}$ for carbon export by division with a molar photosynthetic quotient for NCP of 1.4 O_2/CO_2 [*Laws,* 1991].

3. Methodology

We construct weekly $1^o \times 1^o$ NCP maps between 1998 and 2009 over the Southern Ocean by calculating NCP from weekly maps of up to six of the gridded predictor variables described in the previous section. For these calculations, we assume that NCP has a potentially complex, nonlinear relationship with these six predictors:

$$NCP = f(PAR, POC, Chl, SST, SSH, MLD) \quad (1)$$

We understand that some of the predictors are not independent, and the information provided by these predictors might be redundant. However, in consideration of variable predictor data availability, as discussed below, such information overlap would be useful in compensation of missing predictors. In order to approximate this functional relationship, we use an artificial neural network approach, self-organizing maps (SOMs), similar to that used by *Friedrich and Oschlies* [2009] and *Telszewski et al.* [2009] for generating maps of the North Atlantic pCO_2. The method of self-organizing maps combines elements of cluster analysis with nonlinear, nonparametric regression [*Kohonen*, 2001]. This particular approach is advantageous for the present purpose because the methodology does not assume a pre-defined functional form between predictor and predictand; rather, the methodology relies on an *unsupervised learning* procedure whereby the potentially complex predictor/predictand relationships are determined entirely by the data used to construct the SOM through a process called training. In addition, the methodology readily handles one or more missing predictors when generating NCP maps, which is a useful property given the limited coverage of satellite predictor data over the Southern Ocean for some periods. The approach used here differs from previous SOM studies [*Friedrich and Oschlies*, 2009; *Telszewski et al.*, 2009] in that we perform a thorough validation analysis to determine an optimal combination of SOM parameters and predictors and to provide estimates of error for weekly NCP predictions. Below we include a brief description of the SOM methodology and descriptions of the procedures for generating NCP maps and calculating error estimates. Additional discussion is found in the Supplementary Methods section of the supporting material, and a more thorough description of the SOM methodology can be found in the appendix of *Johnson et al.* [2008].

3.1. Self-organizing map methodology and NCP dataset construction

In the present application, the SOM is trained with the seven-dimensional (six predictors and the predictand, NCP) daily ship track data, where each daily observation is treated as a seven-dimensional data vector. The NCP mapping is accomplished in two steps: (1) SOM training with ship track data to determine the predictor/NCP clusters, and (2) assignment of weekly gridded predictor data to the best-matching SOM clusters and the concomitant assignment of the associated cluster NCP values to the corresponding grid. The first step generates *K* clusters, where the user specifies the number *K*, that describe prototypical combinations of predictor and NCP values (the method for determining *K* is described below in section 3.2). In the second stage, for each grid and week the available predictor data are combined into a data vector of up to six dimensions; then this data vector is mapped to the best-matching SOM cluster on the basis of minimum Euclidean distance. The NCP value associated with that best-matching cluster, which is determined in step 1, is then assigned to that particular grid and week. This process is repeated for each available grid and week to construct weekly NCP maps.

As mentioned above, the SOM approach has the advantage of readily handling data even when one or more predictors are missing during both the training and NCP mapping stages. Due to limitations of satellite data coverage and differences in the starting and ending dates of the predictor datasets, most ship track days and weekly grids have at least one missing predictor value. In particular, the large cloud cover over the Southern Ocean, which typically exceeds 70% south of 40^{o}S during the growing season [*Warren et al.*, 1988], significantly impairs satellite retrieval of POC and Chl. Table 1 shows the availability of each variable in both the ship track data used to train the SOM and the gridded weekly data used to construct the NCP maps. Some variables such as SST, MLD, and SSH have good spatial and temporal coverage, whereas others are more sparse. Even though POC and Chl are among those of the lower data availability, an improvement is apparent from their relatively high coverage of 40–60%, in contrast to the large cloud cover (> 70% on average), which is a result of interpolation of the predictor data (7- or 8-day $1^{o} \times 1^{o}$) onto daily ship track locations as well as the weekly grids. Overall, only approximately 30% of all ship track days have all six predictor values available. For cases when one or more predictor values are missing, the SOM algorithm finds the best-matching clusters on the basis of minimum Euclidean distance, just as in the usual case, except that all dimensions corresponding to missing data are ignored. In the process of assigning NCP values to the weekly gridded data, the cluster dimension corresponding to NCP is ignored in every case of cluster assignment because NCP is excluded from the predictor data. The NCP value of the best-matching cluster is then assigned to the corresponding grid. Thus, this particular application of SOM analysis essentially represents a method of imputation for missing data.

3.2. SOM parameter determination and error estimation

Each SOM analysis requires a number of specifications to be chosen prior to the analysis such as the type of neighborhood function, type of lattice (usually hexagonal or rectangular), number of rows and columns in the lattice (with the total number of neurons equal to the number of rows multiplied by the number of columns), and the final neighborhood radius, which describes how connected the neurons are to their neighbors in the lattice at the end of training. The readers are referred to *Liu et al.* [2006] for a description of the neighborhood function and lattice. In practice, the performance of the SOM analysis tends to be most sensitive to the chosen number of neurons and to the final neighborhood radius. If the number of neurons (i.e., clusters) is too large and/or the final neighborhood radius is too small, then the clusters may be fit too closely to the training data, and the statistical model may be overfit for NCP prediction. In contrast, if the number of neurons is too small and/or the final neighborhood radius is too large, then the statistical model may not capture the range of NCP variability accurately.

In order to determine an appropriate number of neurons, final neighborhood radius, and predictor combination, we consider several factors, including predictor availability, prior knowledge of Southern Ocean NCP, and a set of cross-validation tests. For the cross-validation tests, we specify K and the final neighborhood radius, partition the NCP and predictor data into training and validation sets, train the SOM with the training set, and then evaluate NCP predictions with respect to the withheld validation data. We evaluate a large number of SOM parameter combinations by calculating the mean absolute error (MAE), root–mean-square error (RMSE), and mean fractional error (MFE) of the predicted NCP. A more thorough description of the cross-validation tests is found in the Supplementary Methods of the supporting material.

Several predictor/parameter combinations emerge as candidates for the NCP construction from the cross-validation tests with their errors on the lower end of the estimates (in mmol C $m^{-2}d^{-1}$): MAE ~ 5.3–9, RMSE ~ 8.4–13, MFE ~ 27–51%. However, because these error estimates only apply to the ship track NCP of limited spatial coverage over the Southern Ocean, we also consider additional criteria based on prior expectations of NCP variability. We examine the temporal evolution of the monthly, area-integrated NCP south of 40°S as well as south of 50°S, and the spatial distribution of NCP climatology for the growing season. Although the true temporal evolution of the area-integrated NCP south of 40°S and 50°S is uncertain, a decline in NCP is expected as the season comes to an end. Therefore, we exclude the candidates that show an increase of NCP from February to March. The rest of the candidate NCP constructions show similar climatological features, except for a few that produce unexpectedly high mean NCP in regions of relative minima in both POC and Chl. Because these regions are outside the ship track coverage, we believe the unexpected high NCP estimates to be the result of overfitting to the ship tracks, which target high NCP regions. Consequently, we exclude these candidate NCP constructions. Ultimately, we choose the NCP construction based on a SOM with 12 rows, 8 columns (i.e., 96 total neurons), a final neighborhood radius of 1, and three predictors (Chl, PAR, MLD) because this combination exemplifies low mean errors with the weekly MAE = 8 mmol C $m^{-2}d^{-1}$, RMSE = 12 mmol C $m^{-2}d^{-1}$ and MFE = 48%, and a reasonable climatological NCP that is broadly consistent with previous studies, as described more thoroughly in Section 4.

We emphasize that this choice of predictor set does not mean that the other predictors are unimportant for NCP variability; rather, the combination of redundancy of predictor information (e.g., positive correlations among POC, Chl and SSH) and variations in data availability suggest that these other predictors do not add sufficient independent information to improve NCP predictions on weekly timescales. Interestingly, as we discuss further below, the NCP climatology has a stronger relationship with the POC than Chl climatology even though POC is not included in the final SOM analysis. This strong correspondence between mean NCP and POC despite the omission of POC as a final predictor should only strengthen the conclusion that POC plays a pivotal role in the spatial variability of NCP in the Southern Ocean.

3.3. Bootstrap NCP dataset constructions

In addition to measurement error and random NCP variability unaccounted by the predictors, which are captured in the error estimates in section 3.2, another important source of error for the long-term mean is the limited data coverage used to construct the SOM. Because we are constructing a gridded NCP dataset over a large domain based on a limited number of research cruise measurements, a small number of measurements may have a disproportionate influence on the regional NCP constructions. To provide a quantitative measure of how this limitation impacts the uncertainty in the NCP climatology constructions, we use a bootstrap approach to construct 100 additional NCP datasets. From these 100 datasets, the NCP climatology variance for a particular location provides an indication of the sensitivity of the NCP estimates to this particular, limited ship track dataset used to train the SOM.

For each of these 100 bootstrap NCP datasets, we perform the NCP construction in the same way as described above but with one distinction: the SOM is trained with resampled ship track data. The resampling procedure, which follows conventional bootstrap procedures, is performed as follows. A bootstrap ship track dataset is constructed by randomly selecting research cruise numbers from 1 through 41 with replacement, and then placing the ship track data from the randomly chosen cruise number into the bootstrap dataset. This process of randomly choosing cruise numbers and placing the corresponding data into the bootstrap ship track dataset is repeated until the bootstrap ship track dataset has the same number of daily NCP and predictor observations as in the original ship track dataset. The SOM is then trained on the bootstrap ship track data with the same parameter and predictor combination as discussed above, and then the bootstrap NCP dataset is constructed based on this modified SOM.

4. Results

In this section, the basin-scale features in the constructed NCP dataset are first described. In the absence of the basin-scale NCP observation, we compare the spatial pattern of the constructed NCP with the satellite-measured POC and Chl, because the former accounts for 80–90% of NCP in the Southern Ocean [*Hansell and Carson* 1998; *Allison et al.*, 2010], and the latter is often used to derive phytoplankton biomass and NPP. Secondly, we examine the regional values of constructed NCP by comparison with those reported in the literature. We also include the 95% bootstrap confidence intervals and seasonal standard deviation to indicate the uncertainty and temporal variability, respectively. For comparison with the *in situ* data that are not on our data grids, a 1^o–2^o spatial average is taken of the constructed NCP surrounding the point of observation. We note that exact agreement is not expected, given that the *in situ* derived NCP used for comparison were obtained by various methods that access different temporal and spatial scales of carbon export and that sometimes include different processes. Because our data are resolved on weekly timescales, we only perform comparisons with measurements of weekly or longer temporal resolution.

Thirdly, we show a dominant basin-scale NCP distribution emerged from various models, with discussion of the discrepancies. In addition, because the biological pump is one of the main mechanisms that drive atmospheric CO_2 into the ocean [e.g., *Volk and Hoffert*, 1985; *Carlson et al.*, 2010] we compare our NCP to the observed air-sea CO_2 flux of the Southern Ocean. To convert from annual means to daily range, we assume that the growing season varies from 90 to 120 days [*Heywood and Whitaker*, 1984; *Sweeny et al.*, 2000; *Reuer et al.*, 2007; *Racault et al.*, 2012] because of the large spatial and temporal variability in its duration for the Southern Ocean [*Lizotte, et al.*, 2001; *Racault et al.*, 2012; *Borrione and Schlitzer* 2013]. The total area south of 50^oS is approximated to be $45.7\times10^6 km^2$ [*Moore and Abbot*, 2000].

4.1. Southern Ocean NCP climatology

Figure 2a shows the spatial distribution of 12-year growing season NCP climatology (1998–2009), superimposed with the major Southern Ocean fronts [*Orsi et al.*, 1995]. An elongated zonal band of high NCP (> 22 mmol C $m^{-2}d^{-1}$) is seen approximately following the Subtropical Front (STF ~ 40°S), where macronutrient-rich subantarctic water converges with the macronutrient-poor subtropical water [*Takahashi et al.*, 2012]. It stretches from the southwest Atlantic, across the south Indian Ocean to the western South Pacific, then splits with the STF east of the dateline around 170°W, and turns southeastward to about 120°W. A sharp NCP gradient exists north of the front, with very low NCP throughout most of the subtropics except near large landmasses. Elevated NCP (≥ 20 mmol C $m^{-2}d^{-1}$) is seen along the Southern Boundary (SBdy), the southernmost limit of the Antarctica Circumpolar Circulation (ACC), and along the Antarctic coast, including the Ross Sea and Amundsen Sea, where strong CO_2 sinks have recently been observed [*Arrigo et al.*, 2008; *Tortell et al.*, 2012]. Between the STF and SBdy, we also see high NCP (> 25 mmol C $m^{-2}d^{-1}$) in the complex region off the southeastern South America between the Río de la Plata and the Falkland Island, including the Patagonian Shelf and Brazil-Malvinas Confluence (BMC) zone, and in the vicinity of the Crozet Islands (48°E–60°E, 42°S–49°S), Kerguelen Island (67°E–95°E, 45°S–55°S), and South Georgia (34°W–42°W, 50°S–55°S).

As discussed in the previous section, one of the limitations of this study is the limited availability and spatial coverage of NCP observations used in the SOM analysis for generalizing the relationships between NCP and each of the predictors. As a quantitative indication of how this limitation may impact the constructed NCP climatology, Figure 3 shows the standard deviation, σ_{boot}, of the growing season mean NCP from the one hundred bootstrap NCP datasets (Section 3.3), superimposed with the locations of each ship track. Largest uncertainties (σ_{boot} ~ 6–7 mmol C $m^{-2}d^{-1}$) are found over the Patagonian Shelf, where ship track coverage is lacking. In regions of dense ship track coverage, the uncertainty is generally lower (σ_{boot} < 3 mmol C $m^{-2}d^{-1}$). The regions of high NCP bootstrap climatology standard deviation provide an indication of where targeted measurements in future studies may help to reduce the uncertainty of Southern Ocean NCP estimates.

The climatologies of POC (2002–2009) and Chl (1998–2007) are shown in Figure 2b and Figure 2c for comparison. Chl is in log scale due to its strong positive skewness. Overall, the regions of high mean NCP (Figure 2a) correspond well with regions of high mean POC and Chl.

The area-weighted, centered pattern correlation coefficients are 0.66 and 0.33 for climatological NCP versus POC and versus $\log_{10}$(Chl), respectively. On the basis of these pattern correlations, the climatological POC and Chl fields are able to explain, in a linear sense, 44% and 11% of the climatological NCP field, respectively. However, the temporal correlation between NCP and POC and $\log_{10}$ Chl in the daily ship track data are only 0.20 and 0.23, respectively, which suggests that POC and Chl alone only explain a small percentage of the NCP variability on daily and shorter timescales. The low correlation between NCP and Chl on shorter timescales is consistent with *Reuer et al.* [2007]. Despite this weak linear relationship in the ship track data, we seem to be able to tease out a clear link in the Southern Ocean climatology.

4.2. Regional NCP evaluation

In the following, we compare the regional NCP between the constructed data and independent *in situ* estimates available in the literature. For the NCP values reported for the period prior to our data availability, we provide the climatological values from our data at the sites and calendar days of the measurements. For those reported for the period overlapping our data period (1998–2009), we carry out a real-time comparison.

First we compare our NCP climatology with that off the southeast coast of the South Island of New Zealand, derived from the Munida time series (171.5°E 45.85°S) using a ^{13}C-based diagnostic box model [*Brix et al.*, 2013]. Their reported NCP climatology, 14.6–22.3 mmol C m^{-2}d^{-1}, is in strong agreement with our constructed climatology of 22.1 mmol C m^{-2}d^{-1}, with a 95% bootstrap confidence interval of 16.7–25.2 mmol C m^{-2}d^{-1}, for the identical time period of mid October–March 1998–2009. Table 2 shows the area-integrated NCP south of 50°S from various models and their corresponding periods. The 12-year climatology (1998–2009) of our constructed NCP is 17.9 mmol C m^{-2}d^{-1} with the 95% bootstrap confidence interval of 13.9–21.4 mmol C m^{-2}d^{-1}, whereas the values of other studies range from 8.3 to 24 mmol C m^{-2}d^{-1}, most of which are encompassed by the 95% bootstrap confidence interval.

In the Indian and Pacific sectors of the Southern Ocean, spanning from 1976 through 1997, most of the *in situ* derived NCP are reasonably close to our values, given the uncertainty of the climatology as given by the bootstrap interval and the temporal variability as given by the seasonal standard deviation (Table 3.1). One exception is the high-end values over the southern and southwestern Ross Sea, derived using a seasonal DIC budget approach by *Sweeney et al.* [2000]. These regions are mostly located south of 75°S (Regions I and II in *Sweeney et al.* [2000]). The reason why our model may underestimate the NCP in the region may be due to the poor coverage of the predictors. The predictor, MLD, covers only quasi-global domain with the southern boundary at 75°S. The other two predictors, Chl and PAR, have only less than 60% coverage in the area during the growing season. In Table 3.2, we examine the Atlantic sector. Because the *in situ* derived NCP are collected during the period 1998–2009, we provide the real-time NCP from our dataset for comparison. Our values agree well with previously reported values. For example, both our study and previous measurements determine relatively low NCP values near the Atlantic Polar Frontal Zone (PFZ) during March of 2008 (middle rows) but much higher values around 37 mmol C $m^{-2}d^{-1}$ in the Atlantic-India sector in December of 2006 (bottom row).

Originating from upstream shallow sediments, iron carried by ocean currents can fuel productivity in the waters downstream, leading to phytoplankton blooms. Such an island mass effect has been recorded near the Islands of Kerguelen, South Georgia, and Crozet [*Bakker et al.*, 2007; *Jouandet et al.*, 2008; *Jones et al.*, 2012]. Here we determine if our NCP data reproduce the island mass effect. Both upstream (outside the bloom) and downstream (inside the bloom) values are listed in Table 3.3 Our data capture the upstream and downstream differences in all three island regions. The downstream (inside the bloom) values, however, are smaller than those reported for the Kerguelen Island and South Georgia, possibly due in part to area averaging over a coarse grid in our dataset.

4.3 Basin-scale climatology comparison

4.3.1. Model comparison

Figures 4 and 5b show the basin-scale export rate estimates from two different models, one based on inverse modeling (GCM fitting to observation) [*Schlitzer* 2002], and the other from a satellite NPP-export model calibrated to atmospheric O_2/N_2 measurements [*Nevison et al.*, 2012]. Because only the January climatology is available from *Nevison et al.* [2012], we include our January map in Figure 5a for comparison. We see that the spatial patterns in these two models are in broad agreement with the climatology of our data (Figs 2a and 5a), including regions of high carbon export along the zonal band between 40^{o}S and 60^{o}S, in the coastal upwelling zones off Chile and Namibia, as well as on the Patagonian Shelf. The main difference between the present study (Figure 2a) and the inverse modeling result of Schlitzer [2002] (Figure 4) is that in the latter, the climatology is smoother, the zonal band of high export rate is displaced more to the south, and the high export region off the coast of Chile is more spread out. The broader and smoother features in *Schlitzer* [2002] are likely due to the coarser spatial resolutions available at the time of the study.

The two January climatologies generally differ by less than 10 mmol C m^{-2}d^{-1} throughout the Southern Ocean, except a large discrepancy greater than 100 mmol C m^{-2}d^{-1} over the Patagonian Shelf (Figure 5c). Closer inspection reveals a wide range of NCP values in this region among models: 28–30 mmol C m^{-2}d^{-1} in our January map (Figure 5a), 150–400 mmol C m^{-2}d^{-1} in that of *Nevison et al.* [2012] (Figure 5b), 40–140 mmol C m^{-2}d^{-1} in *Westberry et al.* [2012] (not shown), and 50–60 mmol C m^{-2}d^{-1} in Schlitzer's annual mean [2002] (Figure 4). High daily NCP of 70–90 mmol C m^{-2}d^{-1} has been measured over periods of 3–4 days in this region [*Schloss et al.*, 2007], although it is unclear if such high values over short time periods are representative of the monthly climatology.

The Patagonian Shelf region is known to exhibit highly variable biological activity owing to its uncertain relationships between phytoplankton communities and NCP [*Schloss et al.*, 2007], complicated bathymetry, complex ocean dynamics [*Bianchi et al.*, 2005; *Romero et al.*, 2006; *Rivas*, 2006; *Garcia et al.*, 2008], and multiple sources of iron, including atmospheric dust [*Erickson et al.*, 2003; *Gassó et al.*, 2010; *Signorini et al.*, 2009*; Boyd et al.*, 2012] and ocean upwelling, sediment resuspension, and shelf transport [*Garcia et al.*, 2008; *Signorini et al.*, 2009; *Painter et al.*, 2010]. However, the scarcity of *in situ* measurements of longer timescales has hindered the progress in establishing a reliable regional climatology and has made model validation challenging.

4.3.2. Air-sea CO_2 flux

We now compare our NCP with the air-sea CO_2 flux obtained from the monthly climatological maps of *Takahashi et al.* [2009] for the growing season. In Figure 6a, a zonal band of high CO_2 flux is seen between 40°S and 60°S, similar to the zonal belt of CO_2 flux reported for the February climatology [*Takahashi et al.*, 2012]. Albeit with a much coarser resolution (5° lon × 4° lat), the pattern of high CO_2 flux is in good agreement with the high NCP band (see Figure 2a), outlined by the contour of NCP = 16 mmol C $m^{-2}d^{-1}$ in Figure 6a. The small CO_2 flux compared with NCP could result from CO_2 outgassing due to warming during the growing season, dampening the biologically-driven CO_2 uptake [*Takahashi et al.*, 2012].

Figure 6b shows the monthly mean of the area-integrated NCP, CO_2 flux, and SST south of 50°S from October to March. The NCP starts to increase steadily from 0.6 Pg C yr^{-1} in October until it reaches a peak around 1 Pg C yr^{-1} in December, with a gentle decline from January to March. Similarly, the CO_2 flux changes from 0.2 Pg C yr^{-1} out of the ocean in October to 0.2 Pg C yr^{-1} into the ocean in December, peaks at 0.4–0.5 Pg C yr^{-1} in January and February, lagging the NCP peak by 1−2 month, and declines thereafter. We see that the large difference between the CO_2 flux and NCP seems coincident with the fast increase in SST from October to January. This difference becomes smaller as the SST increase slows down from January to March. Overall, this large imbalance in the early growing season is suggestive of the dominance of the warming-induced CO_2 outgassing, but further investigation is warranted.

5. Discussion and conclusions

In this study we have described the methodology and general features of a 1998–2009 Southern Ocean NCP dataset constructed through a neural network approach. This effort represents the first attempt to construct such a dataset over the Southern Ocean or any large basin entirely on the basis of observed relationships between NCP measurements and NCP predictors. This approach is based on a self-organizing map analysis that assumes no parametric functional form between NCP and the predictors. Overall, we find that our constructed NCP dataset is in good agreement with previously published independent *in situ* derived NCP values of weekly or longer temporal resolution through real-time as well as climatological comparisons at different sampling sites (Tables 2–3). One exception is the region south of 75°S, where the predictor coverage is poor (Section 4.2).

The growing season climatology of our constructed NCP reveals a pronounced zonal band of high NCP that approximately follows the STF between 40°S and 60°S in the Atlantic, Indian and western Pacific sectors, and turns southeastward shortly after the dateline (Figure 2a). Other regions of elevated NCP include area along the SBdy and Antarctic coast, the complex region of Patagonian Shelf and BMC zone, as well as the coastal upwelling zones off Chile and Namibia. This elongated zonal band resembles the observed air-sea CO_2 flux (Figure 6a). The CO_2 flux is generally smaller than the NCP in early growing season (Figure 6b). This difference may result from the rapid temperature increase in the upper ocean during this period, which reduces the CO_2 solubility and possibly results in CO_2 outgassing partially countering the NCP-driven CO_2 uptake (section 4.3.2). However, additional investigation into this hypothesis is necessary in future studies.

The NCP climatological pattern is generally consistent with the expected NCP climatology based on the inverse model of *Schlitzer* [2002] (Figure 4) and the carbon export model of *Nevison et al.* [2012] (Figure 5b) with significant regional variations. The largest discrepancy appears in the Patagonian Shelf, where the estimated climatology ranges from 30 to 400 mmol C m^{-2} d^{-1} among models (section 4.3.1). Additional field campaigns targeting NCP measurements in this region would help to reduce this uncertainty.

The similarity in the climatological spatial distributions of NCP, POC and Chl is readily seen but with notable differences, as evidenced by the pattern correlations of 0.33 and 0.66 between NCP versus $\log_{10}$(Chl) and NCP versus POC respectively (Figure 2, Section 4.1). The low correlation between NCP and Chl may be due to the nonlinear relationship between Chl and phytoplankton biomass, as the Chl concentration depends on both phytoplankton biomass and cellular pigmentation, which adjusts to growth conditions [*Geider et al.*, 1996, 1997, 1998; *Behrenfeld and Boss*, 2003; *Brown et al.*, 2003; *Le Bouteiller et al.*, 2003; *Behrenfeld et al.,* 2005; Armstrong, 2006; *Schultz*, 2008; *Westberry et al.*, 2008; *Wang et al.*, 2009]. Another possibility is that the standard ocean-color to Chl algorithm is not well calibrated for the Southern Ocean, as shown in recent studies [*Mitchell and Kahru*, 2009; *Kahru and Mitchell*, 2010; *Johnson et al.*, 2013].

The fact that the NCP and POC climatologies bear stronger resemblance is consistent with the previous findings that POC production is the largest contributor to NCP in the Southern Ocean [*Ogawa et al.*, 1999; *Wiebinga and de Baar,* 1998; *Kaehler et al.*, 1997; *Hansell and Carlson*, 1998; *Sweeney et al.*, 2000; *Schlitzer*, 2002; *Ishii et al.*, 2002; *Allison et al.*, 2010]. We elaborate further by multiplying POC with MLD to arrive at a quantity we define as POC inventory (mmol C m^{-2}), and then by comparing POC inventory with NCP in Figure 7a. We use the monthly, $3^o\times3^o$ bin-averaged MLD product (2002–2009) derived from the Argo float profiles based on a temperature criterion [*Kara et al.*, 2000] for this calculation (*http://apdrc.soest.hawaii.edu/*). We see that the overall pattern of POC inventory is similar to the NCP distribution.

Although the NCP and POC climatologies correspond well, some spatial variations of the POC-NCP relationship are evident. Such variations may result from true physical differences in the POC-NCP relationships and/or to errors related to the NCP estimates and satellite-derived POC estimates [*Gardner et al.*, 2003; *Stramski et al.*, 2008]. To explore further, we show in Figure 7b the scatter plot of NCP against POC, sorted by latitude bands, for each of the Southern Ocean grid points. This figure demonstrates that although there is a positive correlation between NCP and POC, the relationship appears not to be a simple linear relationship, with variations across different latitude bands. For example, the relationship between POC and NCP appears to be stronger for lower mean POC values at lower latitudes but weaker at higher latitudes poleward of 60°S. If this variation is not due to measurement artifacts, this plot suggests that there may be some regions with high mean POC but relatively low NCP, and vice versa.

Another possibility, however, is that there may be errors in the ship track NCP estimates in some regions characterized by strong vertical mixing of O_2-undersaturated waters to the surface, as pointed out by *Reuer et al.* [2007]. Although we excluded all negative NCP estimates from the SOM analysis, which correspond to regions of upwelled, O_2-undersaturated water, it is possible that this vertical mixing effect still remains in some non-excluded, positive estimates of NCP if the biological productivity of O_2 is strong enough. When this effect occurs, some of the low NCP/high POC regions may have a negative bias from O_2-undersaturated upwelled water. Future investigations into these particular regions are needed to determine to what degree this anomalous low NCP/high POC behavior represents a physical process, a bias, or some combination of the two.

Strong correspondence between POC and NCP in the Southern Ocean on longer timescales suggests that as satellite POC observation becomes available for a longer time period, it can provide a direct view of carbon export variability with a reasonable amount of uncertainty. However, on shorter timescales, the correspondence between NCP and POC is weaker, as evidenced by the correlation of 0.20 in the daily ship track data (section 4.1) and pattern correlations of less than 0.5 in the monthly snapshots (supplementary material). In addition, a major obstacle in monitoring POC variability from satellite is cloud cover, as the Southern Hemisphere belt (between 30°S and 65°S) is among the cloudiest regions on the planet [*Haynes et al.*, 2011]. Therefore, evaluation of NCP variability across a range of timescales requires consideration of the relationships between NCP and multiple variables, as in the present dataset.

Although these results suggest promise in providing insight into Southern Ocean NCP mean state and variability, substantial uncertainty in the NCP construction remains. On weekly timescales, uncertainty due to NCP variance unexplained by the predictors likely dominates, as we estimated a MAE and RMSE of 6.76 and 11.4 mmol C m^{-2} d^{-1}, respectively, based on the ship track data. For longer time averages, i.e. seasonal to decadal, errors in NCP measurements and limitations in ship track coverage likely dominate the uncertainty. As discussed above, efforts to remove possible biases related to the vertical mixing of O_2-undersaturated water would reduce NCP measurement errors.

Regarding the latter source of uncertainty, additional field campaigns to measure Southern Ocean NCP, particularly in several data sparse regions, possibly would lead to improved NCP constructions. Through a bootstrap approach to constructing several overlapping NCP datasets, we have quantified the variance in NCP climatology owing to the limitations of ship track coverage. This analysis has identified several regions where bootstrap climatology variance is high but the number of NCP observations is low or zero (see Figure 3). This finding suggests that targeted measurements in these particular regions may help to constrain the relationships between NCP and each of the predictors, thus resulting in reduced uncertainty in the Southern Ocean NCP climatology and variability.

Notwithstanding these limitations, the dataset we present provides a new opportunity to investigate large-scale variability of NCP and its connections to the Southern Ocean carbon cycle in ways previously not possible in an observation-based dataset. A recent study suggests that a global algorithm for determining NCP may not capture regional NCP differences effectively [*Li and Cassar*, 2013]. The variable relationships between ocean color, Chl concentration, and depth-integrated productivity in different ocean regions [*Campbell et al.*, 2002; *Emerson et al.*, 2008] have challenged the NPP models, with particular difficulty, for example, in regions of extreme Chl [*Carr et al.*, 2006] and coastal waters [*Saba et al.*, 2011]. Our data-driven approach may provide guidance to help correct for biases in the NPP models. Our constructed dataset also may offer the opportunity to investigate interannual NCP variability, even if only for a period of a little more than a decade (see supporting material for a preliminary example). As more NCP measurements and validation data become available, this dataset shall be continually refined, with the hope that applications expand as errors are reduced.

Acknowledgements. We wish to thank Michael Bender and O_2/Ar collaborators for providing the *in situ* observations, R. Schlitzer, C. Nevison, T. Westberry, J. K. Moore, and J. Boutin for sharing the model data or estimates, R. Schlitzer, J. Dunne, P. J. le B. Williams, V. Valsala, J. McCreary, R. Furue, J.-Y. Yu, and H.-H. Hsu for valuable discussion, and C. Chou for encouragement and support to the study. N. C. J. was supported through a grant from the

NOAA Climate Test Bed program. N. C. was partly supported by an Alfred P. Sloan Fellowship. We acknowledge the following data sources: the NASA Goddard Earth Science Data and Information Services Center (GES DISC) Distributed Active Archive Center (DAAC) for the production and distribution of SeaWiFS Chl (the SeaWiFS Project) and MODIS POC and PAR (the MODIS mission) data (http://oceancolor.gsfc.nasa.gov/); the CLS Space Oceanography Division as part of the Environment and Climate EU ENACT project for providing the AVISO altimeter data (available at AVISO site); the NOAA National Climatic Data Center (NCDC) for providing the OI SST data (http://www.ncdc.noaa.gov/oa/climate/research/sst/oi-daily.php); the Asia-Pacific Data Research Center of the IPRC for providing the Niller and Maximenko SSH climatology (http://apdrc.soest.hawaii.edu/datadoc/mdot.php), and production of ARGO MLD data (http://apdrc.soest.hawaii.edu/dods/public_data/Argo_Products); Japan Agency for Marine-Earth Science and Technology (JAMSTEC) for providing the OFES MLD data (http://www.jamstec.go.jp/esc/ofes/eng/); and Lamont-Doherty Earth Observatory (LDEO) for providing the air-sea CO_2 flux (http://cdiac.ornl.gov/oceans/LDEO_Underway_Database/). The SOM constructed NCP from 1998–2009 reported in this study will be available at http://etta.renci.duke.edu/moncp/c/moncp.html.

References

Abbott, M. R., Richman, J. G., Letelier, R. M. and Bartlett, J. S.: The spring bloom in the Antarctic Polar Frontal Zone as observed from a mesoscale array of bio-optical sensors, *Deep-Sea Res., Part II, 47*, 3285-3314, 2000.

Abbott, M., Richman, J. G., Nahorniak, J. S., and Barksdale, B. S.: Meanders in the Antarctic Polar Frontal Zone and their impact on phytoplankton, *Deep-Sea Res., Part II*, *48*, 3891-3912, 2001.

Allison, D. B., Stramski, D., and Mitchell, B. G.: Seasonal and interannual variability of particulate organic carbon within the Southern Ocean from satellite ocean color observations, *J. Geophys. Res.*, *115*, doi:10.1029/2009JC005347, 2010.

Aoki, S., Hariyama, M., Mitsudera, H., Sasaki, H., and Sasai, Y.: Formation regions of Subantarctic Mode Water detected by OFES and Argo profiling floats, *Geophys. Res. Lett.*, *34*, doi:10.1029/2007GL029828, 2007a.

Aoki, S., et al.: A series of cyclonic eddies in the Antarctic Divergence off Adélie Coast, *J. Geophys. Res.*, *112*, doi:10.1029/2006JC003712, 2007b.

Aoki, S., Sasai, Y., Sasaki, H., Mitsudera, H., and Williams, G. D.: The cyclonic circulation in the Australian-Antarctic basin simulated by an eddy-resolving general circulation model, *Ocean Dynamics*, *60*, 743-757, DOI:10.1007/s10236-009-0261-y, 2010.

Armstrong, R. A.: Optimality-based modeling of nitrogen allocation and photoacclimation in photosynthesis, *Deep-Sea Res., Part II*, *53*, 513-531, 2006.

Arrigo, K. R., van Dijken, G., and Long, M.: Coastal Southern Ocean: A strong anthropogenic CO2 sink, *Geophys. Res. Lett.*, *35*, doi:10.1029/2008GL035624, 2008.

Bakker, D. C. E., Nielsdottir, M. C., Morris, P. J., Venables, H. J., and Watson, A. J.: The island mass effect and biological carbon uptake for the subantarctic Crozet Archipelago, *Deep-Sea Res., Part II*, *54*, 2174-2190, doi:10.1016/j.dsr2.2007.06.009, 2007.

Banse, K.: Low seasonality of low concentrations of surface chlorophyll in the Subantarctic water ring: underwater irradiance, iron, or grazing?, *Prog. Oceanogr.*, *37*, 241-291, 1996.

Behrenfeld, M. J., and Falkowski, P. G.: Photosynthetic rates derived from satellite-based chlorophyll concentration, *Limnol. Oceanogr.*, *42*, 1-20, 1997.

Behrenfeld, M. J., and Boss, E.: The beam attenuation to chlorophyll ratio: An optical index of phytoplankton physiology in the surface ocean?, *Deep Sea Res.*, *Part I*, *50*, 1537-1549, 2003.

Behrenfeld, M. J., Boss, E., Siegel, D. A., and Shea, D. M.: Carbon-based ocean productivity and phytoplankton physiology from space, *Global Biogeochem. Cycles*, *19*, doi:10.1029/2004GB002299, 2005.

Bianchi, A. A., Bianucci, L., Piola, A. R., Pino, D. R., Schloss, I., Poisson, A., and Balestrini, C. F.: Vertical stratification and air-sea CO_2 fluxes in the Patagonian shelf, *J. Geophys. Res.*, *110*, doi:10.1029/2004JC002488, 2005.

Bissinger, J. E., Montagnes, D. J. S., Sharples, J., and Atkinson, D.: Predicting marine phytoplankton maximum growth rates from temperature: Improving on the eppley curve using quantile regression, *Limnol. Oceanogr.*, *53*, 487-493, 2008.

Borrione, I. and Schlitzer, R.: Distribution and recurrence of phytoplankton blooms around South Georgia, Southern Ocean, *Biogeosciences*, *10*, 217-231, doi:10.5194/bg-10-217-2013, 2013.

Boutin, J., and Merlivat, L.: New in situ estimates of carbon biological production rates in the Southern Ocean from CARIOCA drifter measurements, *Geophys. Res. Lett.*, *36*, L13608, doi:10.1029/2009GL038307.1, 2009.

Boyd, P. W., Arrigo, K. R., Strzepek, R., and van Dijken, G. L.: Mapping phytoplankton iron utilization: Insights into Southern Ocean supply mechanisms, *J. Geophys. Res.*, *117*, C06009, doi:10.1029/2011JC007726, 2012.

Brix, H., Currie, K. I., and Mikaloff Fletcher, S. E.: Seasonal variability of the carbon cycle in subantarctic surface water in the South West Pacific, *Global Biogeochem. Cycles*, *27*, doi:10.1002/gbc.20023, 2013.

Brown, S. L., Landry, M. R., Neveux, J., and Dupouy, C.: Microbial community abundance and biomass along a 180 degrees transect in the equatorial Pacific during an El Nino-Southern Oscillation cold phase, *J. Geophys. Res.*, *108*, doi:10.1029/2001JC000817, 2003.

Campbell, J., et al.: Comparison of algorithms for estimating ocean primary production from surface chlorophyll, temperature, and irradiance, *Global Biogeochem. Cycles.*, *16*, doi:10.1029/2001GB001444, 2002.

Carlson, A. C., Bates, N. R., Hansell, D. A., and Steinberg, D. K.: Carbon Cycle, *Marine Chemistry & Geochemistry: A derivative of the Encyclopedia of Ocean Sciences*, Academic Press, 680pp, 2010.

Carr, M. E., et al.: A comparison of global estimates of marine primary production from ocean color, *Deep-Sea Res.*, *Part II*, *53*, 741-770, 2006.

Cassar, N., Bender, M., Barnett, B., Fan, S., Moxim, W. J., Levy, H., and Tilbrook, B.: The Southern Ocean biological response to aeolian iron deposition, *Science*, *317*, 1067-1070, 2007.

Cassar, N., Barnett, B., Bender, M., Kaiser, J., Hamme, R.C., and Tilbrook, B.: Continuous High-Frequency Dissolved O_2/Ar Measurements by Equilibrator Inlet Mass Spectrometry, *Anal Chem*, *81*, 1855-1864, 2009.

Cassar, N., et al.: The influence of iron and light on net community production in the Subantarctic and Polar Frontal Zones, *Biogeosciences*, *8*, 227-237, DOI 10.5194/bg-8-227-2011, 2011.

Chang, C.-H., Xie, S.-P., Schneider, N., Qiu, B., Small, J., Zhuang, W., Taguchi, B., Sasaki, H., and Lin, X.: East Pacific ocean eddies and their relationship to subseasonal variability in Central American wind jets, *J. Geophys. Res.*, *117*, doi:10.1029/2011JC007315, 2012.

Codispoti, L. A., Friedrich, G. E., and Hood, D. W.: Variability in the inorganic carbon system over the southeastern Bering Sea shelf during spring 1980 and spring-summer 1981, *Cont. Shelf Res.*, *5*, 133–160, 1986.

Craig, H. and Hayward, T.: Oxygen Supersaturation in the Ocean: Biological Versus Physical Contributions, *Science*, *235*, 199-202, DOI: 10.1126/science.235.4785.199, 1987.

Dong, S., Sprintall, J., Gille, S. T., and Talley, L.: Southern Ocean mixed-layer depth from Argo float profiles, *J. Geophys. Res.*, *113*, doi:10.1029/2006JC004051, 2008.

Ducet, N., P. Le Traon, Y., and Reverdin, G.: Global high-resolution mapping of ocean circulation from TOPEX/Poseidon and ERS-1 and-2, *J. Geophys. Res.*, *105*, 19,477-19,498, doi:10.1029/2000JC900063, 2000.

Dunne, J. P., Sarmiento, J. L., and Gnanadesikan, A.: A synthesis of global particle export from the surface ocean and cycling through the ocean interior and on the seafloor, *Global Biogeochem. Cycles*, *21*, doi:10.1029/2006GB002907, 2007.

Emerson, S., Stump, C., and Nicholson, D.: Net biological oxygen production in the ocean: Remote in situ measurements of O_2 and N_2 in surface waters, *Global Biogeochem. Cycles*, *22*, doi:10.1029/2007GB003095, 2008.

Erickson III D. J., Hernandez, J. L., Ginoux, P., Gregg, W. W., McClain, C., and Christian, J.: Atmospheric iron delivery and surface ocean biological activity in the Southern Ocean and Patagonian region, *Geophys. Res. Lett.*, *30*, doi:10.1029/2003GL017241, 2003.

Fitzpatrick, M. F., and Warren, S. G.: The relative importance of clouds and sea ice for the solar energy budget of the Southern Ocean, J. Climate, 20, 941-954, DOI: 10.1175/JCLI4040.1, 2007.

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

Friedrichs, M. A. M., et al.: Assessing the uncertainties of model estimates of primary productivity in the tropical Pacific Ocean, *J. Mar. Syst.*, *76*, 113-133, doi:10.1016/j.jmarsys.2008.05.010, 2009.

Friedland, K. D., et al.: Pathways between Primary Production and Fisheries Yields of Large Marine Ecosystems, *PLoS ONE*, *7*, doi:10.1371/journal.pone.0028945, 2012.

Frouin, R., Franz, B. A., and Werdell, P. J.: The SeaWiFS PAR product. p. 46-50. In Algorithm Updates for the Fourth SeaWiFS Data Reprocessing, ed. by S. B. Hooker and E. R. Firestone, NASA/TM-2003-206892, 22, 2003.

Garcia, V. M. T., Garcia, C. A. E., Mata, M. M., Pollery, R. C., Piola, A. R., Signorini, S. R., McClain, C. R., and Iglesias-Rodriguezg, M. D.: Environmental factors controlling the phytoplankton blooms at the Patagonia shelf-break in spring, *Deep-Sea Res.*, *Part I*, *55*, 1150-1166, doi:10.1016/j.dsr.2008.04.011, 2008.

Gardner, W. D., Richardson, M. J., Carlson, C. A., Hansell, D., and Mishonov, A. V.: Determining true particulate organic carbon: bottles, pumps and methodologies, *Deep-Sea Res. Part II*, *50*, 655-674, 2003.

Gassó, S., Stein, A., Marino, F., Castellano, E., Udisti, R., and Ceratto, J.: A combined observational and modeling approach to study modern dust transport from the Patagonia desert to East Antarctica, *Atmos. Chem. Phys.*, *10*, 8287-8303, doi:10.5194/acp-10-8287-2010, 2010.

Geider, R. J., Macintyre, H. L., and Kana, T. M.: A dynamic model of photoadaptation in phytoplankton, *Limnol. Oceanogr.*, *41*, 1-15, 1996.

Geider, R. J., Macintyre, H. L., and Kana, T. M.: Dynamic model of phytoplankton growth and acclimation: responses of the balanced growth rate and the chlorophyll a:carbon ratio to light, nutrient-limitation and temperature, *Mar. Ecol.-Prog. Ser.*, *148*, 187-200, 1997.

Geider, R. J., Macintyre, H. L., and Kana, T. M.: A dynamic regulatory model of phytoplanktonic acclimation to light, nutrients, and temperature, *Limnol. Oceanogr.*, *43*, 679-694, 1998.

Glorioso, P.D., Piola, A.R., and Leben, R.R.: Mesoscale eddies in the Subantarctic Front-Southwest Atlantic, *Sci. Mar.*, *69* (Suppl. 2), 7-15, doi:10.3989/scimar.2005.69s2, 2005.

Gruber, N., Lachkar, Z., Frenzel, H., Marchesiello, P., Münnich, M., McWilliams, J. C., Nagai, T., and Plattner, G.-K.: Eddy-induced reduction of biological production in eastern boundary upwelling systems, *Nature Geoscience*, *4*, 787-792, doi:10.1038/ngeo1273, 2011.

Hamme, R. C., et al.: Dissolved O_2/Ar and other methods reveal rapid changes in productivity during a Lagrangian experiment in the Southern Ocean, *J. Geophys. Res.*, *117*, doi:10.1029/2011JC007046, 2012.

Hansell, D. A. and Carlson, C. A.: Net community production of dissolved organic carbon. *Global Biogeochem. Cycles*, *12*, 443-453, 1998.

Haynes, J. M., Jakob, C., Rossow, W. B., Tselioudis, G., and Brown, J.: Major Characteristics of Southern Ocean Cloud Regimes and Their Effects on the Energy Budget, *J. Climate*, *24*, 5061-5080, doi: 10.1175/2011JCLI4052.1, 2011.

Heywood, R. B. and Whitaker, T. M.: The Antarctic marine flora. In R. M. Laws (ed.), *Antarctic ecology*, *Vol. 2*, pp. 373-419, Academic Press, London, 1984.

Ishii, M., Inoue, H. Y., Matsueda, H., and Tanoue, E.: Close coupling between seasonal biological production and dynamics of dissolved inorganic carbon in the Indian Ocean sector and the western Pacific Ocean sector of the Antarctic Ocean, *Deep-Sea Res.*, *Part I*, *45*, 1187-1209, 1998.

Ishii, M., Inoue, H. Y., and Matsueda, H.: Net community production in the marginal ice zone and its importance for the variability of the oceanic pCO_2 in the Southern Ocean south of Australia. *Deep-Sea Res.*, *Part II*, *49*, 1691-1706, 2002.

Johnson, N. C., Feldstein, S. B., and Tremblay, B.: The continuum of Northern Hemisphere teleconnection patterns and a description of the NAO shift with the use of self-organizing maps. *J. Climate*, *21*, 6354-6371, DOI: 10.1175/2008JCLI2380.1, 2008.

Johnson, N. C. and Feldstein, S. B.:The Continuum of North Pacific Sea Level Pressure Patterns: Intraseasonal, Interannual, and Interdecadal Variability, *J. Climate*, *23*, 851-867, DOI: 10.1175/2009JCLI3099.1, 2010.

Johnson, N. C.: How many ENSO flavors can we distinguish? *J. Climate*, *26*, 4816-4827, doi:10.1175/JCLI-D-12-00649.1, 2013.

Johnson, R., Strutton, P. G., Wright, S. W., McMinn, A., and Meiners, K. M.: Three improved satellite chlorophyll algorithms for the Southern Ocean, *J. Geophys. Res.*, doi: 10.1002/jgrc.20270, in press, 2013.

Jones, E. M., Bakker, D. C. E., Venables, H. J., and Watson, A. J.: Dynamic seasonal cycling of inorganic carbon downstream of SouthGeorgia, Southern Ocean, *Deep-Sea Res.*, *Part II*, *59-60*, 25-35, doi:10.1016/j.dsr2.2011.08.001, 2012.

Jonsson, B. F., Doney, S. C., Dunne, J., and Bender, M.: Evaluation of the Southern Ocean O_2/Ar-based NCP estimates in a model framework, *J. Geophys. Res. Biogeosci.*, *118*, doi:10.1002/jgrg.20032, 2013.

Jouandet, M. P., Blain, S., Metzl, N., Brunet, C., Trull, T.W., and Obernosterer, I.: A seasonal carbon budget for a naturally iron-fertilised bloom over the Kerguelen Plateau in the Southern Ocean, *Deep-Sea Res.*, *Part II*, *55*, 856–867, doi:10.1016/j.dsr2.2007.12.037, 2008.

Kaehler, P., Bjornsen, P. K., Lochte, K., and Antia, A.: Dissolved organic matter and its utilization by bacteria during spring in the Southern Ocean, *Deep-Sea Res.*, *Part II*, *44*, 341-353, 1997.

Kahru, M., Mitchell, B. G., Gille, S. T., Hewes, C. D., and Holm-Hansen, O.: Eddies enhance biological production in the Weddell-Scotia Confluence of the Southern Ocean, *Geophys. Res. Lett.*, *34*, doi:10.1029/2007GL030430, 2007.

Kahru, M. and Mitchell, B. G.: Blending of ocean colour algorithms applied to the Southern Ocean, Kahru, M., and B.G. Mitchell, Blending of ocean colour algorithms applied to the Southern Ocean, *Remote Sensing Letters*, *1*, 119-124, DOI:10.1080/01431160903547940, 2010.

Kara, A. B., Rochford, P. A, and Hurlburt, H. E.: An optimal definition for ocean mixed layer depth, *J. Geophys. Res.*, *105*, 16,803-16,821, 2000.

Kohonen, T.: *Self-Organizing Maps*, Springer, 501 pp, 2001.

Laws, E. A.: Photosynthetic quotients, new production and net community production in the open ocean, *Deep Sea Res.*, *Part A*, *38*, 143-167, 1991.

Laws, E. A., Falkowski, P. G., Smith, W. O. Jr., Ducklow, H. W., and McCarthy, J. J.: Temperature effects on export production in the open ocean, *Global Biogeochem. Cycles*, *14*, 1231-1246, doi:10.1029/1999GB001229, 2000.

Laws, E. A.: Export flux and stability as regulators of community composition in pelagic marine biological communities: Implications for regime shifts, *Prog. Oceanogr.*, *60*, 343-354, doi:10.1016/j.pocean.2004.02.015, 2004.

Laws, E. A., D'Sa, E., and Naik, P.: estimates of sea surface temperature and primary production, *Limnol. Oceanogr.*: *Methods*, *9*, 593-601, DOI 10.4319/lom.2011.9.593, 2011.

Le Bouteiller, A., Leynaert, A., Landry, M. R., Le Borgne, R., Neveux, J., Rodier, M., Blanchot, J., and Brown, S. L.: Primary production, new production, and growth rate in the equatorial Pacific: Changes from mesotrophic to oligotrophic regime, *J. Geophys. Res.*, *108*, doi:10.1029/2001JC000914, 2003.

Li, Z., and Cassar N.: Estimation of net community production (NCP) using O_2/Ar measurements and satellite observations, paper presented at *45th International Liege Colloquium*, 13-17 May 2013, Liege, Belgium, 2013.

Liu, Y., Weisberg, R. H., and Mooers, C. N. K.: Performance evaluation of the self-organizing map for feature extraction, *J. Geophys. Res.*, *111*, doi:10.1029/2005JC003117, 2006.

Lizotte, M. P.: The Contributions of Sea Ice Algae to Antarctic Marine Primary Production, *Amer. Zool.*, *41(1)*, 57-73, doi: 10.1093/icb/41.1.57, 2001.

Masumoto, Y., et al.: A fifty-year eddy-resolving simulation of the world ocean: Preliminary outcomes of OFES (OGCM for the Earth Simulator), *J. Earth Simulator, 1,* 35–56, 2004.

Maximenko, N., Niiler, P., Rio, M.-H., Melnichenko, O., Centurioni, L., Chambers, D., Zlotnicki, V., and Galperin, B.: Mean dynamic topography of the ocean derived from satellite and drifting buoy data using three different techniques, *J. Atmos. Oceanic Technol.*, *26*, 1910-1919,doi:10.1175/2009JTECHO672.1, 2009.

Minas, H. J., Minas, M., and Packard, T. T.: Productivity in upwelling areas deduced from hydrographic and chemical fields. *Limnol. Oceanogr.*, 31, 1180-1204, 1986.

Minas, H. J., and Minas, M.: Net community production in "high nutrient-low chlorophyll" waters of the tropical and Antarctic Oceans: grazing vs. iron hypothesis, *Oceanol. Acta*, *15*, 145-162, 1992.

Mitchell, B. G., and Kahru, M.: Bio-optical algorithms for ADEOS-2 GLI, *Journal of The Remote Sensing Society of Japan*, *29*, 80-85, 2009.

Moore, J. K., and Abbott, M. R.: Phytoplankton chlorophyll distributions and primary production in the Southern Ocean, *J. Geophys. Res.*, *105*, 28,709-28,722, DOI: 10.1029/1999JC000043, 2000.

Moore, J. K., Lindsay, K., Doney, S. C., Long, M. C., and Misumi, K.:, Marine Ecosystem Dynamics and Biogeochemical Cycling in the Community Earth System Model (CESM1(BGC)): Comparison of the 1990s with the 2090s under the RCP 4.5 and RCP 8.5 Scenarios, *J. Climate* (Accepted), 2013.

Moore, T. S., DeGrandpre, M. D., Sabine, C. L., Hamme, R. C., Zappa, C. J., McGillis, W. R., Feely, R. A., and Drennan, W. M.: Sea surface pCO_2 and O_2 in the Southern Ocean during the austral fall, 2008, *J. Geophys. Res.*, *116*, doi:10.1029/2010JC006560, 2011.

Morris, P. J., Sanders, R., Turnewitsch, R., and Thomalla, S.: ^{234}Th-derived particulate organic carbon export from an island-induced phytoplankton bloom in the Southern Ocean, *Deep-Sea Res.*, *Part II*, *54*, 2208-2232, doi:10.1016/j.dsr2.2007.06.002, 2007.

Nevison, C. D., Keeling, R. F., Kahru, M., Manizza, M., Mitchell, B. G., and Cassar, N.: Estimating net community production in the Southern Ocean based on atmospheric potential oxygen and satellite ocean color data, *Global Biogeochem. Cycles*, *26*, doi:10.1029/2011GB004040, 2012.

Niiler, P. P., Maximenko, N. A., and McWilliams, J. C.: Dynamically balanced absolute sea level of the global ocean derived from near-surface velocity observations, *Geophys. Res. Lett.*, *30*, doi:10.1029/2003GL018628, 2003.

Ogawa, H., Fukuda, R., and Koike, I.: Vertical distributions of dissolved organic carbon and nitrogen in the Southern Ocean, *Deep-Sea Res.*, *Part II*, *46*, 1809-1826, 1999.

Olonscheck, D., Hofmann, M, Worm, B., and Schellnhuber, H. J.: Decomposing the effects of ocean warming on chlorophyll a concentrations into physically and biologically driven contributions , *Environ. Res. Lett.*, *8*, 014043, doi:10.1088/1748-9326/8/1/014043, 2013.

O'Reilly, J. E., Maritorena, S., Mitchell, B. G., Siegel, D. A., Carder, K. L., Garver, S. A., Kahru, M., and McClain, C.: Ocean Color Chlorophyll Algorithms for SEAWIFS, *Marine Science Faculty Publications*, Paper 6, 1998.

Orsi, A. H., Whitworth, T. III, and Nowlin, W. D. Jr.: On the meridional extent and fronts of the Antarctic Circumpolar Current, *Deep-Sea Res.*, *Part I*, *42*, 641-673, 1995.

Painter, S. C., Poulto, A. J. n, Allen, J. T., Pidcock, R., and Balch, W. M.: The COPAS'08 expedition to the Patagonian Shelf: Physical and environmental conditions during the 2008 coccolithophore bloom, *Cont. Shelf Res.*, *30(18)*, 1907-1923, doi:10.1016/j.csr.2010.08.013, 2010.

Pollard, R., Tréguer, P., and Read, J. F.: Quantifying nutrient supply to the Southern Ocean, *J. Geophys. Res.*, *111*, doi:10.1029/2005JC003076, 2006.

Racault, M. F., Le Quéré, C., Buitenhuis, E., Sathyendranath, S., and Platt, T.: Phytoplankton phenology in the global ocean, *Ecol. Indic.*, *14*, 152-163, doi: 10.1016/j.ecolind.2011.07.010, 2012.

Reuer, M. K., Barnett, B. A., Bender, M. L., Falkowski, P. G., and Hendricks, M. B.: New estimates of Southern Ocean biological production rates from O_2/Ar ratios and the triple isotope composition of O_2, *Deep-Sea Res.*, *Part I*, *54*, 951-974, 2007.

Reusch, D. B., Alley, R. B., and Hewitson, B. C.:North Atlantic climate variability from a self-organizing map perspective. *J. Geophys. Res.*, *112*, doi:10.1029/2006JD007460, 2007.

Reynolds, R. W., Smith, T. M., Liu, C., Chelton, D. B., Casey, K. S., and Schlax, M. G.: Daily High-Resolution-Blended Analyses for Sea Surface Temperature, *J. Climate*, *20*, 5473-5496, DOI: 10.1175/2007JCLI1824.1, 2007.

Rivas, A. L.: Quantitative estimation of the influence of surface thermal fronts over chlorophyll concentration at the Patagonian shelf, *J. Mar. Syst.*, *63*, 183-190, doi:10.1016/j.jmarsys.2006.07.002, 2006.

Romero, S. I., Piola, A. R., Charo, M., and Garcia, C. A. E.: Chlorophyll-a variability off Patagonia based on SeaWiFS data, *J. Geophys. Res.*, *111*, doi:10.1029/2005JC003244, 2006.

Rubin, S. I., Takahashi, T., Chipman, D. W., and Goddard, J. G.: Primary productivity and nutrient utilization ratios in the Pacific sector of the Southern Ocean based on seasonal changes in seawater chemistry, *Deep-Sea Res.*, *Part I*, *45*, 1211-1234, 1998.

Saba, V. S., et al.: An evaluation of ocean color model estimates of marine primaryvproductivity in coastal and pelagic regions across the globe, *Biogeosciences*, *8*, 489-503, 2011, doi:10.5194/bg-8-489-2011, 2011.

Sallée, J., Wienders, N., Speer, K., and Morrow, R.: Formation of subantarctic mode water in the southeastern Indian Ocean, *Ocean Dyn.*, *56*, 525-542, 2006.

Sasaki, H., Sasai, Y., Nonaka, M., Masumoto, Y., and S. Kawahara: An eddy-resolving simulation of the quasi-global ocean driven by satellite-observed wind field: Preliminary outcomes from physical and biological fields, *J. Ear. Sim., 6*, 35-49, 2006.

Sasaki, H. and Nonaka, M.: Far-reaching Hawaiian Lee Countercurrent driven by wind-stress curl induced by warm SST band along the current, *Geophys. Res. Lett., 33*, doi:10.1029/2006GL026540, 2006.

Sasaki, H., Nonaka, M., Masumoto, Y., Sasai, Y., Uehara, H., and Sakuma, H.: An eddy-resolving hindcast simulation of the quasi-global ocean from 1950 to 2003 on the Earth Simulator, *High-Resolution Numerical Modeling of the Atmosphere and Ocean*, *Chapter 10*, 157–186, Springer, New York, 2008.

Sasaki, Y. N., Minobe, S., Schneider, N., Kagimoto, T., Nonaka, M., and Sasaki, H.: Decadal sea level variability in the South Pacific in a global eddy-resolving ocean model hindcast, *J. Phys. Oceanogr.*, *38*, 1764-1779, 2008.

Sasaki, Y. and Schneider, N.: Interannual to decadal Gulf Stream variability in an eddy-resolving ocean model. *Ocean Modelling*, *39*, 209-219, 2011.

Sasse, T. P., McNeil, B. I. and Abramowitz, G.: A new constraint on global air-sea CO_2 fluxes using bottle carbon data, *Geophys. Res. Lett.*, doi: 10.1002/grl.50342, 2013.

Schlitzer, R.: Carbon export fluxes in the Southern Ocean: results from inverse modeling and comparison with satellite-based estimates, *Deep-Sea Res.*, *Part II*, *49*, 1623-1644, 2002.

Schloss, I. R., et al.: Role of plankton communities in sea-air variations in pCO_2 in the SW Atlantic Ocean, *Mar. Ecol. Prog. Ser.*, *332*, 93-106, 2007.

Schneider, W., and Bravo, L.:, Argo profiling floats document Subantarctic Mode Water formation west of Drake Passage, *Geophys. Res. Lett.*, *33*, doi:10.1029/2006GL026463, 2006.

Schultz, P.: Oberving phytoplankton physiology and ocean ecosystem structure from space, *PhD thesis*, *245 pp*., Princeton University, Princeton, 2008.

Scott, R. B., Arbic, B. K., Holland, C. L., Sen, A. and Qiu, B.: Zonal versus meridional velocity variance in satellite observations and realistic and idealized ocean circulation models, *Ocean Modelling*, *23*, 102-112, doi:10.1016/j.ocemod.2008.04.009, 2008.

Shim, J., Kang, Y. C., Kim, D. and Choi, S.-H.: Distribution of net community production and surface pCO_2 in the Scotia Sea, Antarctica, during austral spring 2001', *Mar. Chem.*, *101*, 1-2, 68-84, DOI: 10.1016/j.marchem.2005.12.007, 2006.

Signorini, S. R., Garcia, V. M. T., Piola, A. R., Evangelista, M., McClain, C. R., Garcia, C. A. E., and Mata, M. M.: Further studies on the physical and biogeochemical causes for large interannual changes in the Patagonian Shelf spring-summer phytoplankton bloom biomass, *NASA/TM-2009-214176*, 43 pp, 2009.

Sokolov, S., and Rintoul, S. R.: On the relationship between fronts of the Antarctic Circumpolar Current and surface chlorophyll concentrations in the Southern Ocean, *J. Geophys. Res.*, *112*, doi:10.1029/2006JC004072, 2007.

Sterl A, et al.: A look at the ocean in the EC-Earth climate model, *Clim. Dyn.*, *39*, 2631-2657, DOI 10.1007/s00382-011-1239-2, 2012.

Stramski, D., et al.: Relationships between the surface concentration of particulate organic carbon and optical properties in the eastern South Pacfic and eastern Atlantic Oceans, *Biogeosciences*, *5*, 171-201, 2008.

Sweeney, C., et al.: Biogeochemical regimes, net community production and carbon export in the Ross Sea, Antarctica, *Deep-Sea Res.*, *Part II*, *47*, 3369-3394, 2000.

Taguchi, B., Xie, S.-P., Schneider, N., Nonaka, M., Sasaki, H., and Sasai, Y.: Decadal variability of the Kuroshio Extension: Observations and an eddy-resolving model hindcast, *J. Clim.*, *20*, 2357-2377, doi:10.1175/JCLI4142.1, 2007.

Takahashi, T., et al.: Climatological mean and decadal changes in surface ocean pCO_2, and net sea-air CO_2 flux over the global oceans, *Deep-Sea Res.*, *Part II*, *56*, 554-577, 2009.

Takahashi, T., Sweeney, C., Hales, B., Chipman, D. W., Newberger, T., Goddard. J. G., Iannuzzi, R. A., and Sutherland. S. C.: The changing carbon cycle in the Southern Ocean, *Oceanogr. 25*(3), 26-37, http://dx.doi.org/10.5670/oceanog.2012.71, 2012.

Telszewski, M., et al.: Estimating the monthly pCO_2 distribution in the North Atlantic using a self-organizing neural network, *Biogeosciences*, *6*, 1405-1421, doi 10.5194/bg-6-1405-2009, 2009.

Thompson, A. F., Haynes, P. H., Wilson, C., and Richards, K. J.: Rapid Southern Ocean front transitions in an eddy-resolving ocean GCM, *Geophys. Res. Lett.*, *37*, doi:10.1029/2010GL045386, 2010.

Thompson, A. F., and Richards, K. J.: Low frequency variability of Southern Ocean jets, *J. Geophys. Res.*, *116*, doi:10.1029/2010JC006749, 2011.

Tortell, P. D., Long, M. C., Payne, C. D., Alderkamp, A.-C., Dutrieux, P., and Arrigo, K. R.: Spatial distribution of pCO2, ?O2/Ar and dimethylsulfide (DMS) in polynya waters and the sea ice zone of the Amundsen Sea, Antarctica, *Deep Sea Res.*, *Part II*, *71-76*, 77-93, doi:10.1016/j.dsr2.2012.03.010, 2012.

Volk, T., and Hoffert, M. I.: Ocean carbon pumps: Analysis of relative strengths and efficiencies in ocean-driven atmospheric CO_2 changes, in The Carbon Cycle and Atmospheric CO_2: Natural Variations Archean to Present, *Geophys. Monogr. Ser.*, vol. 32, edited by E. T.

Sundquist and W. S. Broecker, pp. 99–110, AGU, Washington, D. C., doi:10.1029/GM032p0099, 1985.

Wang, X. J., Behrenfeld, M., Le Borgne, R., Murtugudde, R., and Boss, E.: Regulation of phytoplankton carbon to chlorophyll ratio by light, nutrients and temperature in the Equatorial Pacific Ocean: a basin-scale model, *Biogeosciences*, *6*, 391-404, doi:10.5194/bg-6-391-2009, 2009.

Warren, S. G., Hahn, C. J., London, J., Chervin, R. M., and Jenne, R. L.: Global distribution of total cloud cover and cloud type amounts over the ocean, NCAR Tech. Note TN-317 STR, Boulder, CO, 42 pp. 170 maps, 1988.

Westberry, T., Behrenfeld, M. J., Siegel, D. A., and Boss, E.: Carbon-based primary productivity modeling with vertically resolved photoacclimation, *Global Biogeochem. Cycles*, *22*, doi:10.1029/2007GB003078, 2008.

Westberry, T. K., Williams, P. J. le B., and Behrenfeld, M. J.: Global net community production and the putative net heterotrophy of the oligotrophic oceans, *Global Biogeochem. Cycles*, *26*, doi:10.1029/2011GB004094, 2012.

Wiebinga, C. J. and de Baar, H. J. W.: Determination of the distribution of dissolved organic carbon in the Indian Sector of the Southern Ocean, *Marine Chemistry*, *61*, 185-201, 1998.

Wong, A. P. S.: Subantarctic ModeWater and Antarctic Intermediate Water in the South Indian Ocean based on profiling float data 2000-2004, *J. Mar. Res.*, *63*, 789-812, 2005.

Yoshida, S., Qiu, B., and Hacker, P.: Wind-generated eddy characteristics in the lee of the island of Hawaii, *J. Geophys. Res.*, *115*, doi:10.1029/2009JC005417, 2010.

Zhuang, W., Xie, S.-P., Wang, D., Taguchi, B., Aiki, H., and Sasaki, H.: Intraseasonal variability in sea surface height over the South China Sea, *J. Geophys. Res., 115,* C04010, doi:10.1029/2009JC005647, 2010.

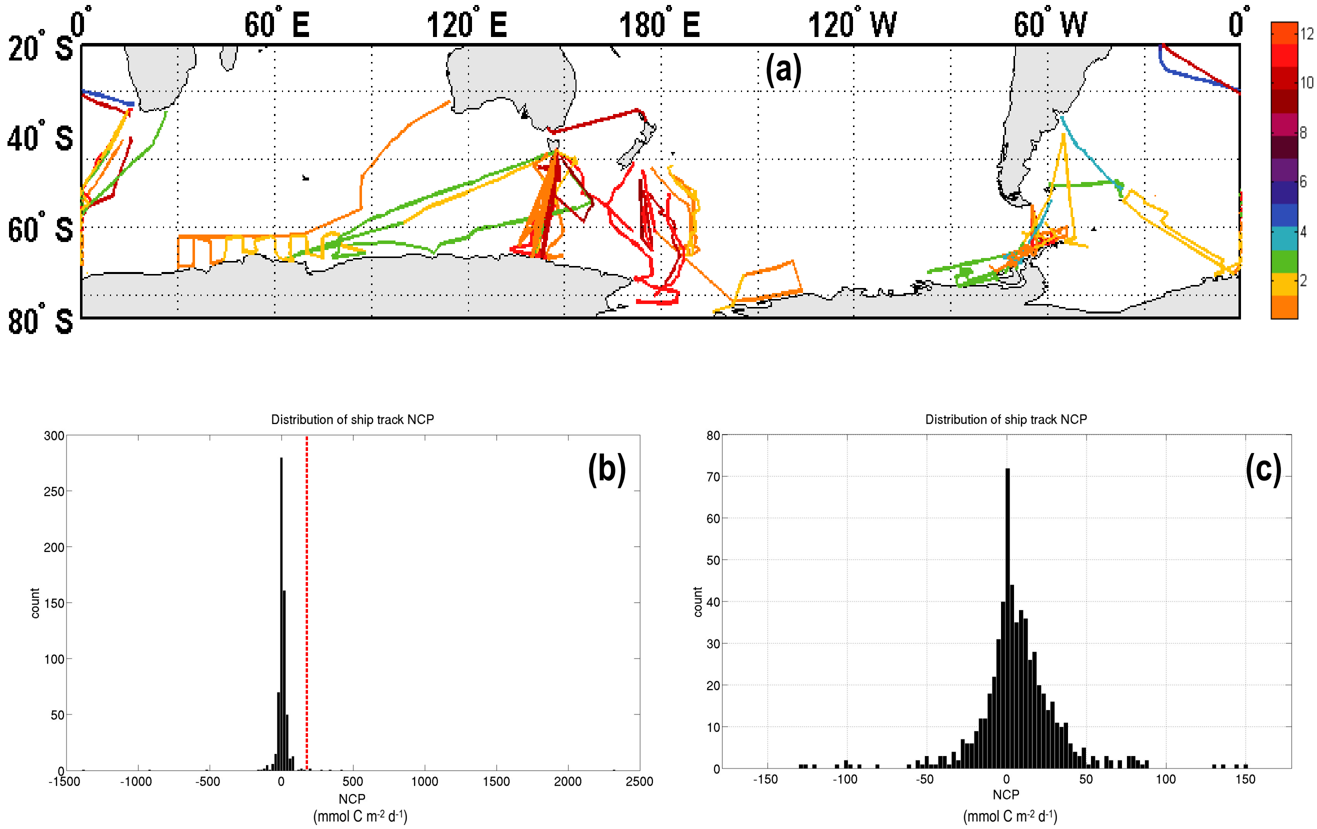

Figure 1. (a) The ship tracks of the *in situ* NCP measurements used in this study. The time of the research cruises is color-coded in months. (b) The histogram of the ship track NCP distribution. The red dashed line marks NCP = 180 mmol C $m^{-2}d^{-1}$. (c) Detailed distribution of NCP below the outlier threshold.

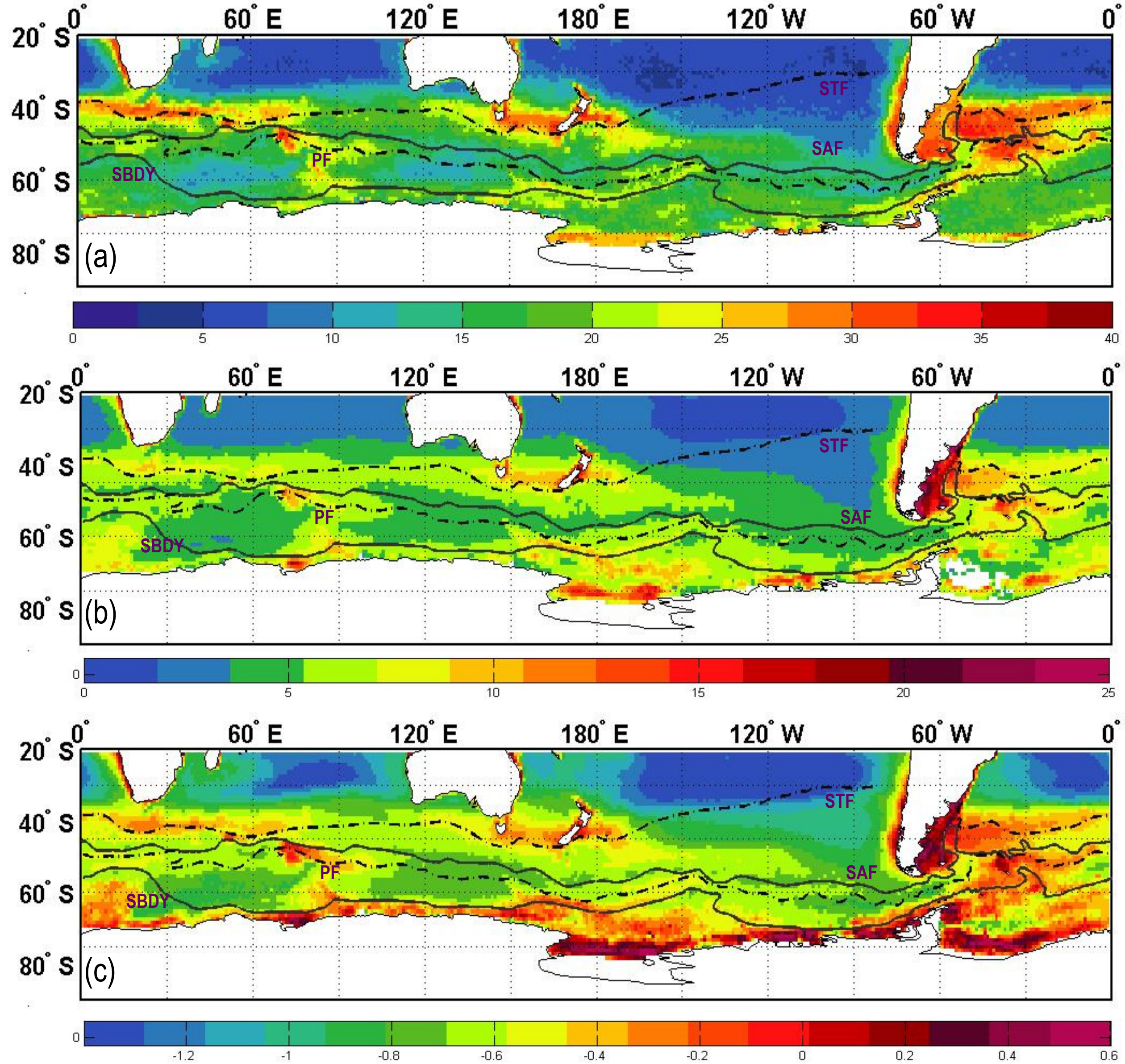

Figure 2. Growing season (November–March) climatologies (in color) of (a) NCP (mmol C $m^{-2}d^{-1}$) for 1998-2009, (b) POC (mmol C m^{-3}) for 2002-2009, and (c) $\log_{10}$ Chl (mg m^{-3}) for 1998-2007. The major Southern Ocean fronts [*Orsi et al.*, 1995] are superimposed (from the north): the Subtropical Front (STF), the Subantarctic Front (SAF), the Polar Front (PF), and the Southern Boundary (SBdy).

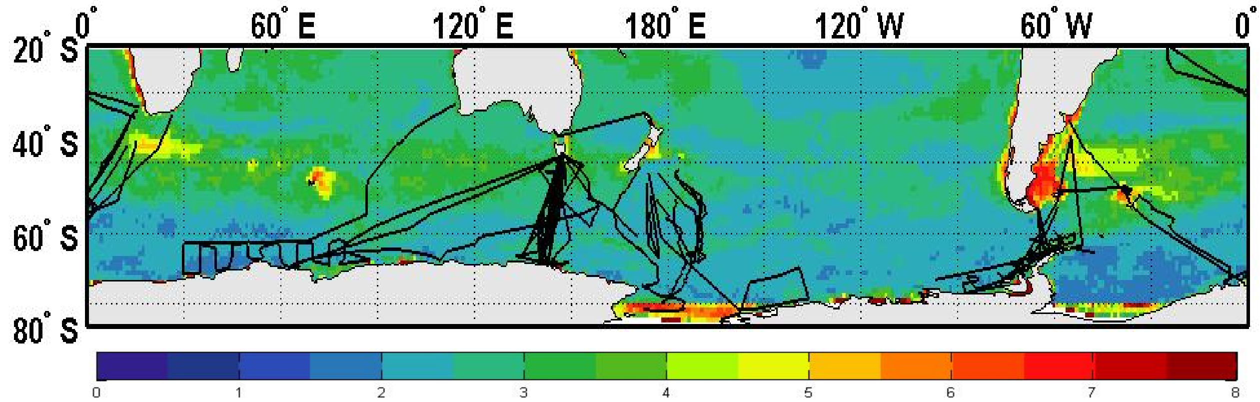

Figure 3. Standard deviation of the growing season mean NCP from the 100 bootstrap NCP datasets (σ_{boot}) (mmol C $m^{-2}d^{-1}$, in color) superimposed with the locations of each ship track (black lines).

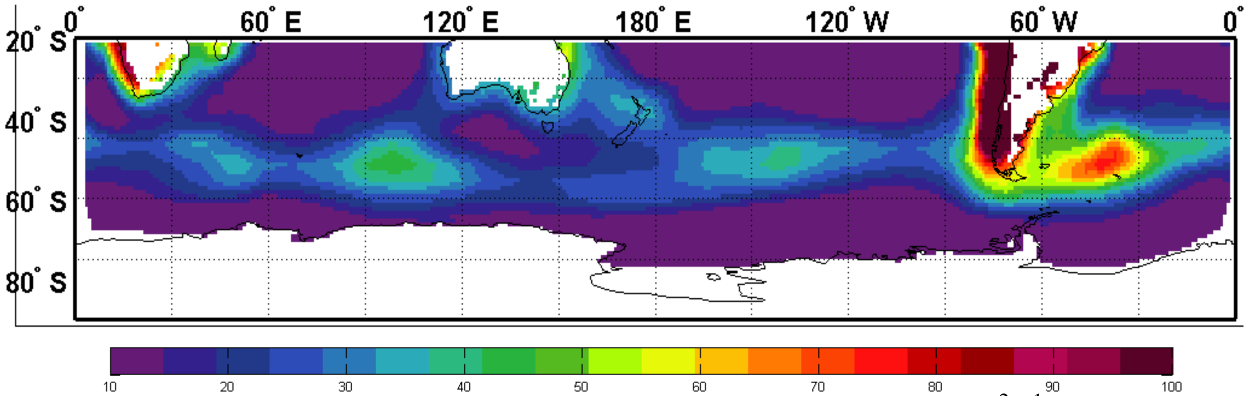

Figure 4. Long-term mean annual POC export of *Schlitzer* [2002] (mmol C $m^{-2}d^{-1}$).

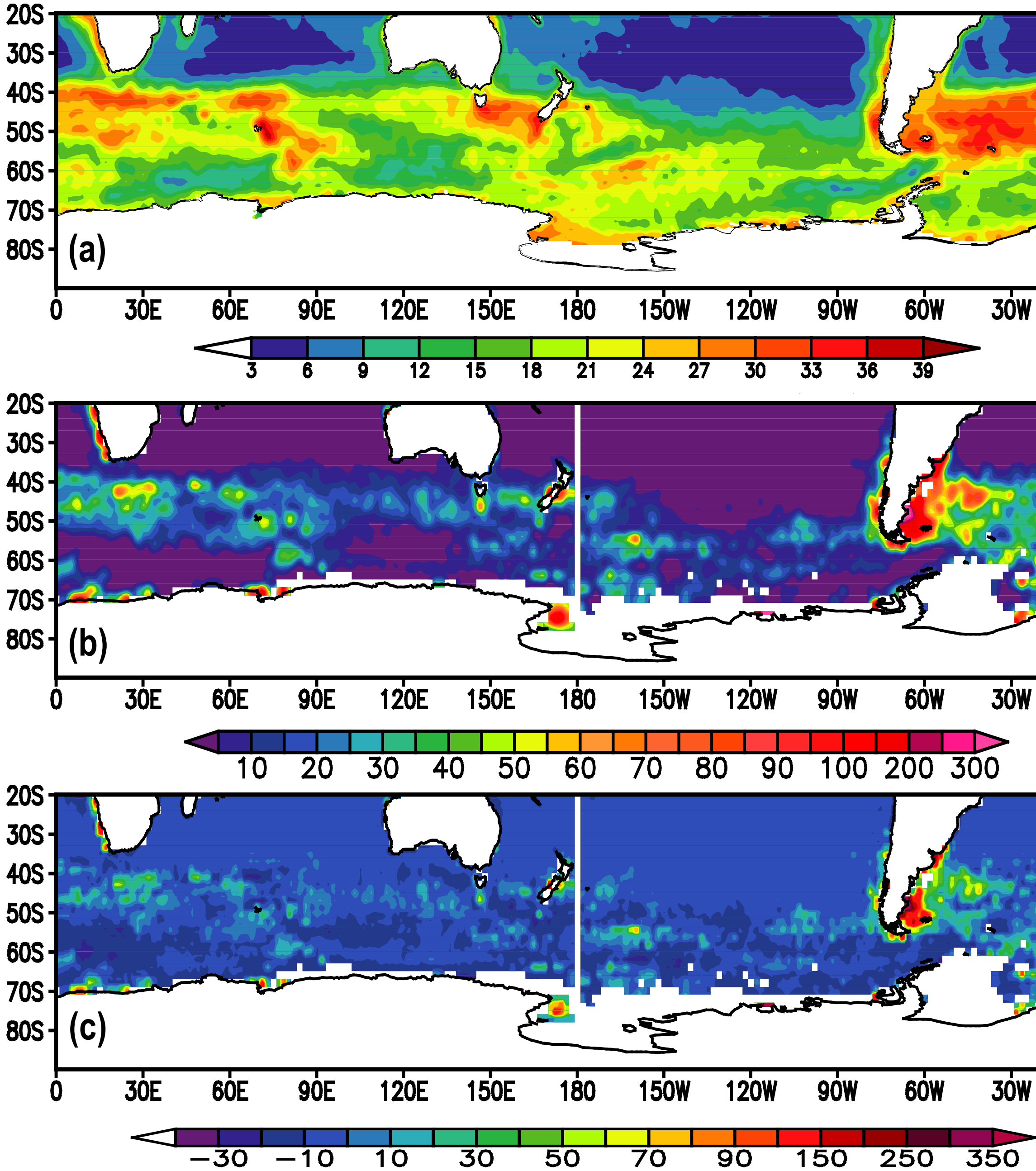

Figure 5. January NCP climatologies: (a) Our constructed NCP (1998-2009); (b) *Nevison et al.* [2012] (1998-2007); (c) Difference between two climatologies [(b) – (a)]. (Unit: mmol C $m^{-2}d^{-1}$). Note that the contour intervals change from 5 and 10 in (b) and (c), respectively, to 50 mmol C $m^{-2}d^{-1}$ for contour values greater than 100 mmol C $m^{-2}d^{-1}$ to accommodate the large values on the high ends.

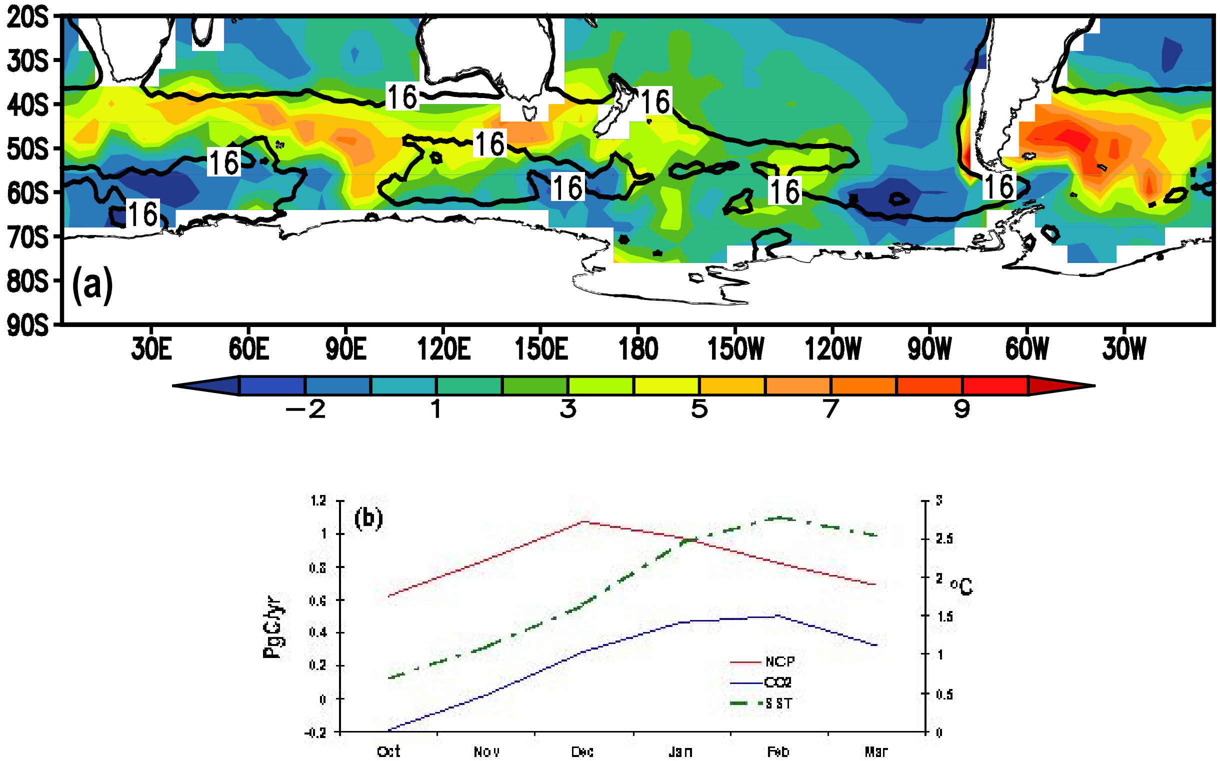

Figure 6. Comparison with air-sea CO_2 flux (a) November–March climatology of air-sea CO_2 flux [*Takahashi et al.*, 2009] (mmol C $m^{-2}d^{-1}$, in color) superimposed with the contour of NCP = 16 mmol C $m^{-2}d^{-1}$. (b) Evolution of monthly mean area-integrated (>50°S) NCP (red), CO_2 flux (blue), and SST (green) from October to March. The left y-axis corresponds with the NCP and CO_2 (Pg/yr) and the right y-axis corresponds with SST (°C). Note that the air-sea CO_2 flux is defined positive into the ocean.

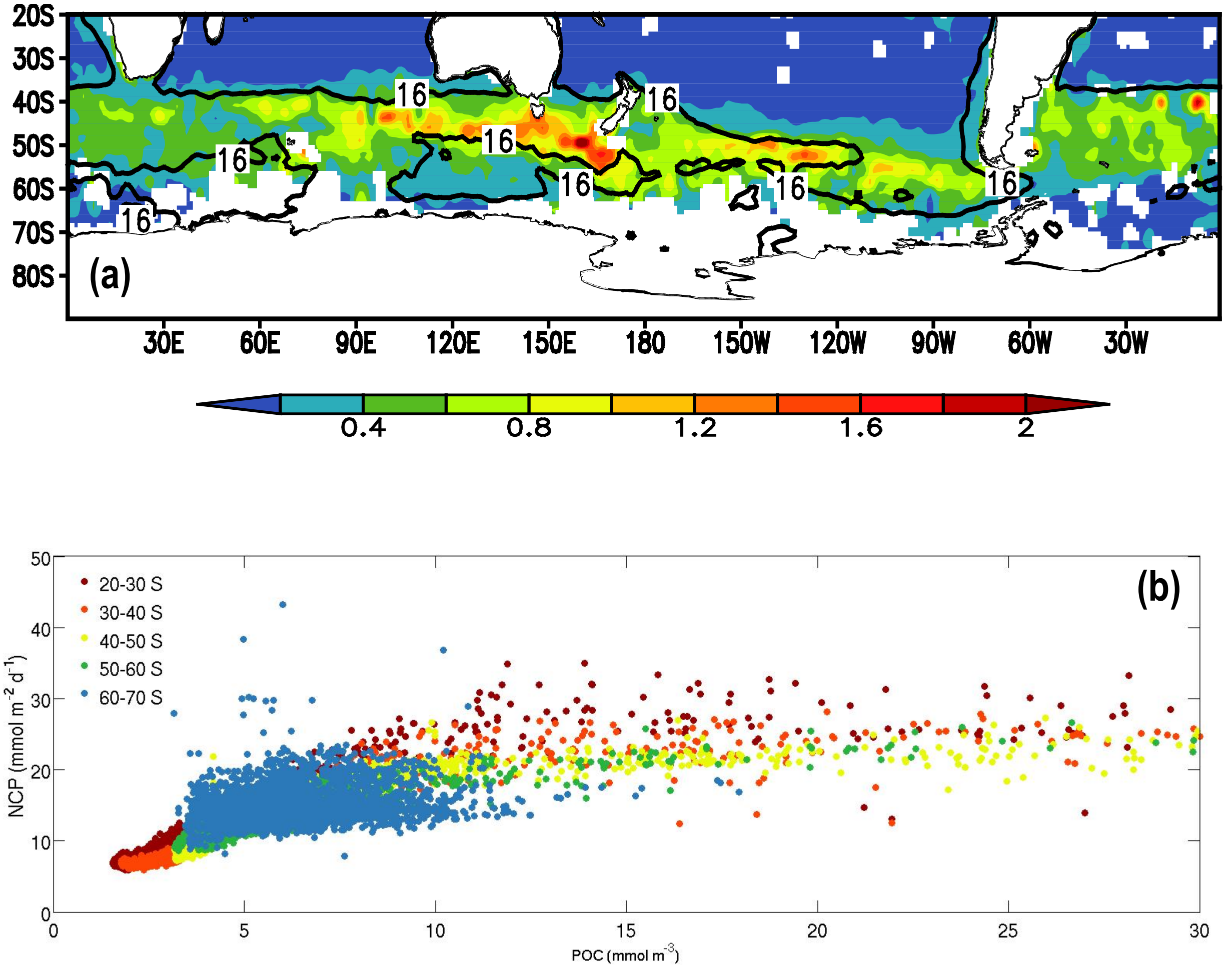

Figure 7. (a) November–March climatology of POC inventory, defined as POC × MLD (mmol C m^{-2}, in color), superimposed with the contour of NCP = 16 mmol C $m^{-2}d^{-1}$. (b) The Scatter plot of NCP (mmol C $m^{-2}d^{-1}$, y-axis) against POC (mmol C m^{-3}, x-axis), sorted by latitudinal bands (color).

Table 1. Availability percentage of predictor variables in the ship track and weekly gridded data used to generate NCP maps.

Variable	Ship track availability (%)	Weekly gridded map availability (%)
SST	97.8	97.0
Chl	42.1	59.9
POC	62.6	40.4
PAR	81.3	45.9
MLD	97.8	83.7
SSH	85.3	78.9

Table 2. Comparison of area-integrated NCP for the Southern Ocean south of 50^{o}S.

Reference	**Method**	**Period**	**NCP (mmol C m^{-2} d^{-1})**	
			Previous studies	This study (95% CI)[1]
Moore et al. [2013]	NCAR CMIP5 coupled carbon-climate simulation	1990s	12–15	
Westberry et al. [2012]	Carbon-based NPP model and empirical relationships of *in vitro* PvR[2]	2004	8.3	
Nevison et al. [2012]	VGPM NPP and export model	1998-2007	12–24	**17.9 (13.9–21.4)**
Dunne et al. [2007]	POC export from empirical equation based on in situ observation	1998-2004	11–14	**1998-2009**
Pollard et al. [2006]	Nutrient drawdown based on Ekman flux divergence	long-term	17–22	
Schlitzer [2002]	3D inverse model	steady-state	15–20	

[1] 95% bootstrap confidence interval.
[2] PvR: photosynthesis versus respiration.

Table 3.1. Climatological comparison of independent *in situ* NCP measurements and the constructed NCP.

Reference	**Method**	**Location**	**Time**	**NCP (mmol C $m^{-2}d^{-1}$)**	
				Previous study	This study Climatology (σ)[1] (95% CI)[2]
		Indian sector			
Minas & Minas [1992]	Mass balance based on nutrient drawdown	(65°E 40°–62°S)	Aug/Sep–Feb/Mar 1976–1977	17	18 (±12) (15–24) [3]
Ishii et al. [1998]	Seasonal ΔDIC (temporal difference)	(30°–40°E 67°–68.2°S)	14–17 Feb 1993	14–19	19 (±10) (7.5–21)
		(47.5°–48.8°E 66°–66.5°S)	19 Feb 1993	22–29	20 (±11) (11–24)
		(49.1°–67.8°E 65°–65.7°S)	26–28 Feb 1993	13–17	13(±9.8) (7.6–22)
		(70.6°–77.5°E 67°–69°S)	28 Feb–1 Mar 1993	24–32	26 (±10) (15–37)
		Indian & Pacific sectors			
		(80°–150°E 63°–65°S)	4–13 Mar 1993	20–27	20(±11) (6.8–27)
Rubin et al. [1998]	Seasonal ΔDIC (vertical gradient)	(110°–171°E 67°–70°S)	Winter–late Feb/mid Mar 1992, 1994	6.5–24	18 (±12) (13–21)[3]
Ishii et al. [2002]	Seasonal ΔDIC (temporal difference)	(140°E 64°–65.5°S)	19 Dec 1994–21 Jan 1995	2.5–28	16 (±12) (11–21)
		Ross Sea			
Sweeny et al. [2000]	Seasonal ΔDIC (vertical gradient)	(163°–186°E 74°–78°S)	Mid Oct 1996–Mid Feb 1997	22–64	25 (±10) (16–34)

[1] seasonal standard deviation.
[2] 95% bootstrap confidence interval.
[3] November-March mean.

Table 3.2. Real-time comparison of independent *in situ* NCP measurements and the constructed NCP.

Reference	Method	Location	Time	NCP (mmol C $m^{-2}d^{-1}$)	
				Previous studies	This study (95% CI)[1]
		Scotia Sea			
Shim et al. [2006]	nutrient drawdown	(*52°W 57°–60°S*)	20 Nov–31 Dec 2001	24–29	26 (20–46)
		Atlantic sector			
Hamme et al. [2012]	O_2/Ar	(37°–*38°W 50°–51°S*)	2–9 Mar 2008	11–22	8.4 (0.9–36)
			9–14 Mar 2008	5–13.8	7.6 (1.9–37)
Moore et al. [2011]	Mass balance based on ΔDIC and O_2	(37°–*38°W 50°–51°S*)	9–14 Mar 2008	3.2–6.7	7.6 (1.9–37))
		Atlantic-Indian sector			
Boutin & Merlivat [2009]	Based on mixed layer budget on diurnal time scale	(*16.4°–21.2°E 46.8°–47.8°S*)	28 Nov–30 Dec 2006	30–51	37 (15–51)

[1] 95% bootstrap confidence interval.

Table 3.3. Real-time comparison of island mass effect.

Reference	Method	Time	Location	**NCP** (mmol C $m^{-2}d^{-1}$)	
				Previous study	This study (95% CI)[1]
			Kerguelen		
Jouandet et al. [2008]	seasonal ΔDIC (vertical gradient)	Nov 2004–Feb 2005	Inside bloom (*72^{o}E 50.5^{o}S*)	49–98	44 (21–57)
			Outside bloom (*78^{o}E 52^{o}S*)	17–26	18 (12–40)
			South Georgia		
Jones et al. [2012]	seasonal ΔDIC (vertical gradient)	Nov 2007–Feb 2008	Bloom (*39^{o}–40^{o}W 52^{o}S*)	43	32 (20–40)
			HNLC (*42^{o}W 56^{o}–57^{o}S*)	12	18 (16–36)
			Crozet		
Bakker et al. [2007]	ΔDIC in the upper 100m	8 Nov 2004–16 Jan 2005	Bloom (47^{o}–52^{o}*E 43^{o}–45.5^{o}S*)	33–45	32 (20–40)
			HNLC (47.8^{o}–49^{o}*E 51.5^{o}–52.9^{o}S*)	16–19	15 (13–30)

[1] 95% bootstrap confidence interval.

Supporting Material to Neural network-based estimates of Southern Ocean net community production from in-situ O_2/Ar and satellite observation: A methodological study

S1. Supplementary Methods

S1.1 General Desription

The SOM methodology partitions a potentially large, high-dimensional dataset into a smaller number of representative clusters. In contrast with conventional cluster analysis, these SOM clusters, each of which is associated with a component called a node or neuron, become topologically ordered on a lower-dimensional, typically two-dimensional, lattice so that similar clusters are located close together in the lattice and dissimilar clusters are located farther apart. This topological ordering occurs through the use of a neighborhood function, which acts like a kernel density smoother among a neighborhood of neurons within this low-dimensional lattice. As a result, neighboring neurons within this lattice influence each other to produce smoothly varying clusters that represent the multi-dimensional distribution function of the data used to construct the SOM.

Our approach of determining predictor/predictand SOM clusters is quite similar to that of *Telszewski et al.* [2009] except for one main difference: we incorporate the predictand into the SOM analysis rather than labeling each neuron with an associated NCP value after the SOM has been trained. Thus we combine the first two steps of map generation from *Telszewski et al.* [2009] into a single step. We choose this alternative approach so that the neighborhood function, which smoothes the clusters in the data space, may operate on the NCP as well as the predictor data.

Because the SOM training is limited to the ship track data, one of the main requirements for a successful NCP dataset construction is that the range of predictor values in the ship track data should approximately span the predictor range in the gridded data used to generate the Southern Ocean mapping [*Kohonen,* 2001*; Telszewski et al.,* 2009]. To assess the reasonableness of the range spanned by the training data, we calculate the percentage of gridded predictor data that falls outside of the range of the training data for each of the three retained predictors, Chl, PAR, and MLD. For Chl, approximately 4.4% of the gridded values fall below the ship track minimum (0.04 mg m^{-3}), and 2.0% exceed the ship track maximum (1.56 mg m^{-3}). For PAR, 5.2% of the gridded values fall below the ship track minimum (9.63 μE m^{-2} s^{-1}), and 1.3% exceed the ship track maximum (59.5 μE m^{-2} s^{-1}). For MLD, only 0.3% of the gridded values fall below the ship track minimum (3.5 m), and 0.3% exceed the ship track maximum (595 m). In summary, the range of the predictor values in the ship tracks appears to be fairly representative of that of the Southern Ocean overall during the period of interest, with well over 90% of the gridded values falling within the range of the ship track values. These calculations add support to the appropriateness of using the ship track data to generate a broad Southern Ocean mapping. However, we do note that a higher percentage of gridded predictors fall below the ship track minimum for Chl and PAR, which suggests that there may be at least a slight NCP prediction bias in regions of very low Chl and/or PAR.

As discussed in the main text, the parameters for which the SOM likely is most sensitive are the total number of neurons (i.e., the product of the number of rows and columns in the SOM) and the final neighborhood radius. For that reason, we vary each of those parameters in the cross-validations discussed below. For the sake of completeness, we mention here the other parameter choices used in our study (see *Liu et al.* [2006] and *Johnson et al.* [2008] for a description of these parameters). We use a rectangular lattice and a Gaussian neighborhood function. The neighborhood function uses an initial radius of 4, which gradually shrinks to the final radius of 1. The "rough" training phase uses 20 iterations of the batch algorithm, which is followed by 500 iterations during the fine-tuning phase.

S1.2 Cross-validations

To determine a set of candidate predictor and parameter combinations, we first perform a set of cross-validation tests in the following manner. We identify 39 weeks in the ship track database that have at least five days of NCP data within a seven-day period and then divide these 39 weeks into five validation segments (eight weeks each segment except one with seven weeks). We next perform a five-fold cross-validation for many predictor/parameter combinations, whereby we train the SOM with all ship track data excluding the validation segments and evaluate the prediction of weekly mean NCP for the validation segments in five separate iterations. To minimize the possibility that the data in the validation and training samples are highly correlated and thus leading to over-confident NCP predictions, we add the condition that the data from any particular ship track cannot be split between training and validation samples. We calculate the MAE, RMSE, and MFE of the predicted NCP.

For the SOM parameter combinations we evaluate the following values for the number of rows and columns: 1-6, 8, 10, 12, 14, 18, and 24. We also vary the final neighborhood radius from zero to five. With 12 possible values for the number of rows and columns and six values for the final neighborhood radius, we test 864 possible SOM parameter combinations. In addition, we test all 63 possible predictor combinations to give a total of 54,432 cross-validation tests. We record the parameter combination with the minimum MAE, RMSE, and MFE for each of the 63 predictor combinations.

S2. Interannual NCP variability

To explore the potential use of our constructed dataset to study interannual NCP variability, we present snapshots of November NCP for 2003 and 2004 in Figures S1a and S1b. These results should be interpreted with caution because we have not yet assessed the uncertainty in interannual predictions. In both figures, two large patches of high NCP are seen over southwest Atlantic in the Brazil-Malvinas Confluence zone as well as in the region near southeast Australia and New Zealand, which are marked with blue squares in Figure S1. Our constructed dataset predicts variations between these two years in the two regions. The Australia-New Zealand patch (140^{o}E–170^{o}W, 35^{o}S–46^{o}S) exhibits a distinct southeastward extension in 2003 (Figure S1a), whereas it is zonally confined in 2004 (Figure S1b). Over the Brazil-Malvinas patch (65^{o}W–45^{o}W, 35^{o}S–46^{o}S), the area-averaged NCP decreases from 37 to 27 mmol C m^{-2}d^{-1} from 2003 to 2004. The November maps of POC (Figures S2a, b) and Chl (not shown) also show similar variations for the same years, which support the physical basis for these NCP changes. The pattern correlation between NCP and POC ($\log_{10}$(Chl)) are 0.48 (0.42) and 0.47 (0.39) for 2003 and 2004, respectively.

These large-scale variations in biological productivity plausibly may relate to dominant modes of the ocean-atmosphere interaction and the associated atmospheric teleconnections, as well as ocean current variability. For example, possible contributors include the change from neutral ENSO to El Nino conditions between 2003 and 2004 [*Yu et al.*, 2012], and the pronounced southward shift of the Brazil Current front from the continental shelf observed in 2003 [*Goni et al.*, 2011]. However, more in depth analysis of the mechanisms of variability is reserved for future studies.

One may question whether the constructed NCP dataset can capture intraseasonal and interannual variability, given the fairly weak relationship between daily NCP and POC/Chl in the ship track observations, as reported in the main text, the temporal correlation between daily NCP and POC/$\log_{10}$(Chl) is only 0.20/0.23. Because the residence time of POC and NCP integration time are of similar magnitude, 1−2 weeks in the surface ocean, and POC is the dominant form of NCP in the Southern Ocean, the low correlation between POC and NCP on daily timescales suggests sub-weekly transient processes and/or measurement errors that weaken the POC/NCP relationship.

The weak correlation between NCP and Chl is similar to the value of 0.33 reported in *Reuer et al.* [2007], although *Reuer et al.* [2007] consider area averages in three discrete zones for each of 23 transits rather than discrete points along the ship tracks. However, a substantially improved correlation of 0.62 is achieved in *Reuer et al.* [2007] between the *in situ* NCP and NPP, calculated using the VGPM (Vertically Generalized Productivity Model) of *Behrenfeld and Falkowski* [1997] that accounts for additional predictors (e.g., Chl, SST, and PAR). Given that our SOM-based approach includes additional biogeochemical and physical properties, aside from Chl that is also incorporated in the VGPM NPP estimates of *Reuer et al.* [2007], that our results are constrained by *in situ* observations, and that we find good agreement with previously reported independent, *in situ* NCP measurements (Tables 3.2 and 3.3) through real-time comparisons, we expect that our reconstruction explains a larger fraction of NCP variance on intraseasonal and interannual timescales than indicated by the low POC and Chl correlations. Additional validation tests are required to assess the reliability of the predicted interannual and possibly intraseasonal NCP variability, and relation to plausible physical mechanisms.

Reference

Behrenfeld, M. J., and P. G. Falkowski (1997), Photosynthetic rates derived from satellite-based chlorophyll concentration, *Limnol. Oceanogr.*, *42*, 1-20.

Goni, G.J., F. Bringas, and P.N. DiNezio (2011), Observed low frequency variability of the Brazil Current front, *J. Geophys. Res.*, *116*, doi:10.1029/2011JC007198.

Reuer, M. K., B. A. Barnett, M. L. Bender, P. G. Falkowski, and M. B. Hendricks (2007), New estimates of Southern Ocean biological production rates from O_2/Ar ratios and the triple isotope composition of O_2, *Deep-Sea Res.*, *Part I*, *54*, 951-974.

Telszewski, M., et al. (2009), Estimating the monthly pCO_2 distribution in the North Atlantic using a self-organizing neural network, *Biogeosciences*, *6*, 1405-1421, doi 10.5194/bg-6-1405-2009.

Yu, J.-Y., Y. Zou, S.T. Kim, and T. Lee (2012), The changing impact of El Niño on US winter temperatures, *Geophys. Res. Lett.*, *39*, doi:10.1029/2012GL052483.

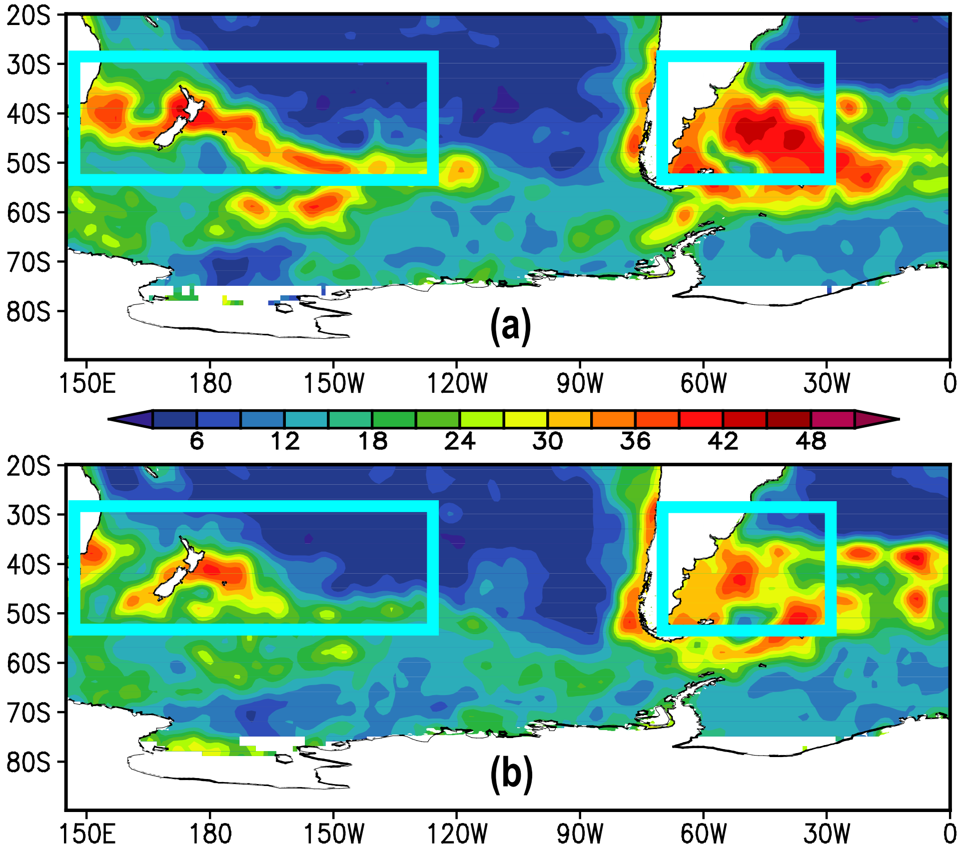

Figure S 1. November NCP (mmol C $m^2 d^{-1}$) for (a) 2003, and (b) 2004. The blue squares mark the two regions discussed in the supporting text.

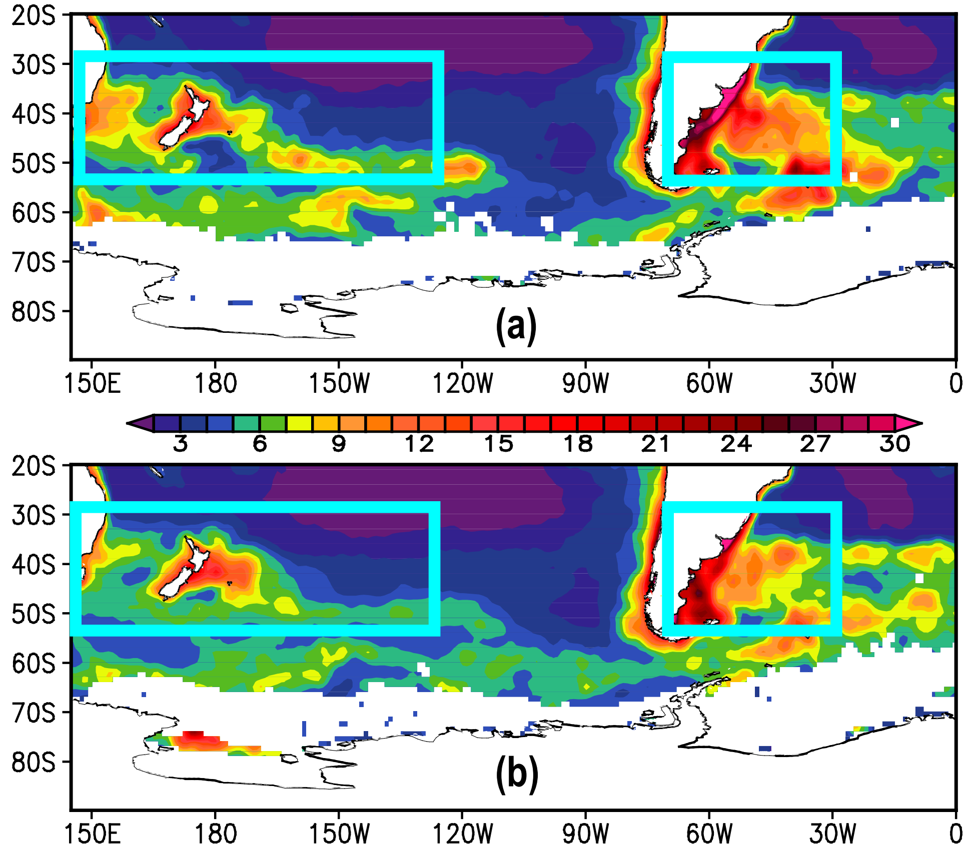

Figure S 2 As in Figure S1 but for POC (mmol C m^{-3}).