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.

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.

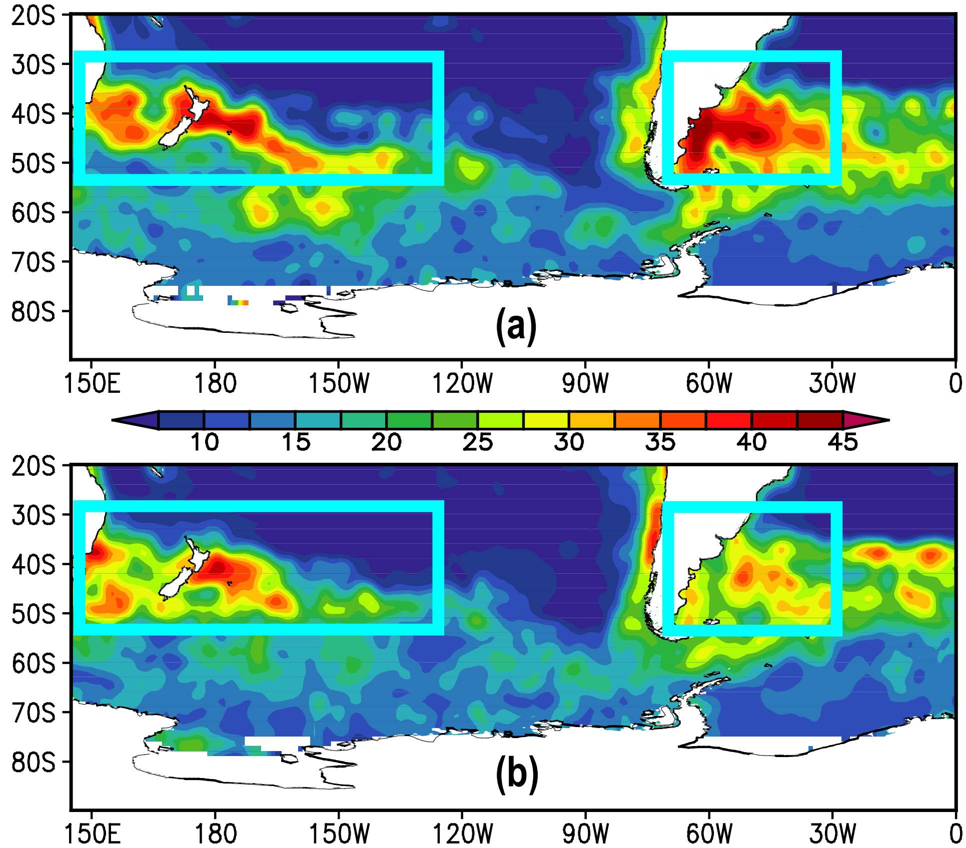

Figure S 1. November NCP (mmol C m^{2}d^{-1}) for (a) 2003, and (b) 2004. The blue squares mark the two regions discussed in the supporting text.

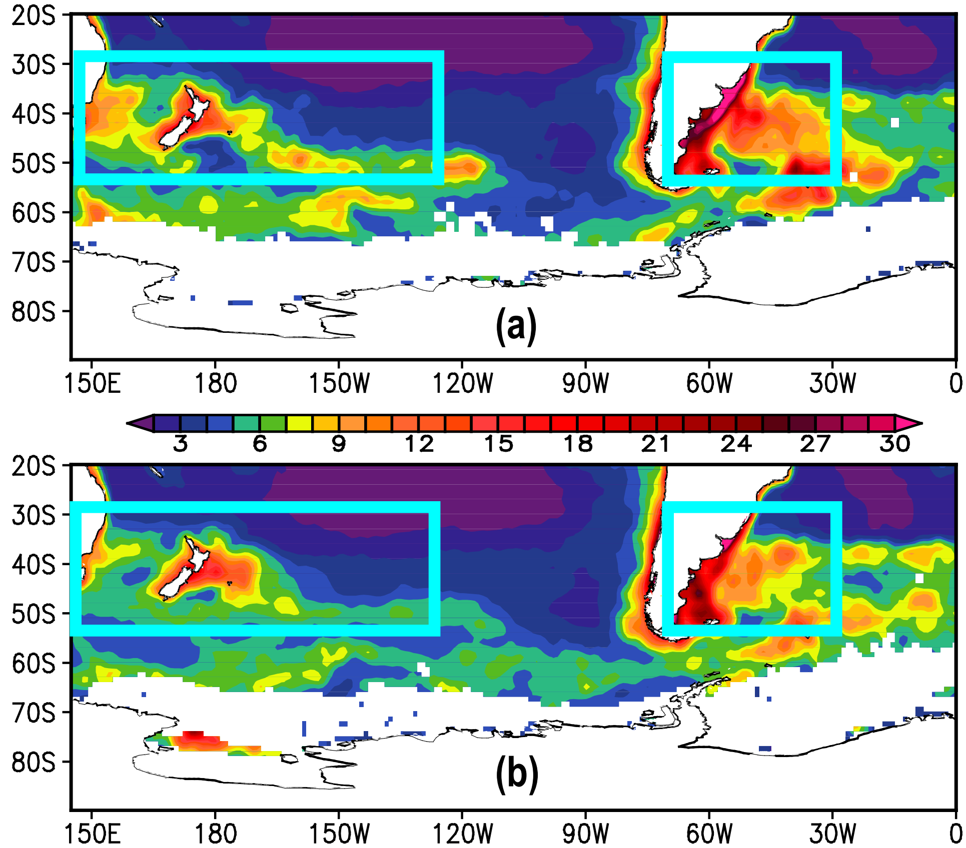

Figure S 2 As in Figure S1 but for POC (mmol C m^{-3}).