Estimating temporal and spatial variation of ocean surface $p$CO$_2$ in the North Pacific using a self-organizing map neural network technique

S. Nakaoka$^1$, M. Telszewski$^{1,*}$, Y. Nojiri$^1$, S. Yasunaka$^1$, C. Miyazaki$^{1,**}$, H. Mukai$^1$, and N. Usui$^2$

$^1$National Institute for Environmental Studies, 1-2 Onogawa, Tsukuba, Ibaraki 305-8506, Japan
$^2$Meteorological Research Institute, 1-1 Nagamine, Tsukuba, Ibaraki 305-0052, Japan
* now at: International Ocean Carbon Coordination Project, Institute of Oceanology of Polish Academy of Sciences, Ul. Powstańców Warszawy 55, 81-712 Sopot, Poland
** now at: Hokkaido University, Kita 8 Nishi 5, Kita-ku, Sapporo, Hokkaido 060-0810, Japan

Correspondence to: S. Nakaoka (nakaoka.shinichiro@nies.go.jp)

Received: 4 February 2013 – Published in Biogeosciences Discuss.: 8 March 2013
Revised: 25 June 2013 – Accepted: 9 August 2013 – Published: 26 September 2013

Abstract. This study uses a neural network technique to produce maps of the partial pressure of oceanic carbon dioxide ($p$CO$_2^{sea}$) in the North Pacific on a 0.25° latitude × 0.25° longitude grid from 2002 to 2008. The $p$CO$_2^{sea}$ distribution was computed using a self-organizing map (SOM) originally utilized to map the $p$CO$_2^{sea}$ in the North Atlantic. Four proxy parameters – sea surface temperature (SST), mixed layer depth, chlorophyll $a$ concentration, and sea surface salinity (SSS) – are used during the training phase to enable the network to resolve the nonlinear relationships between the $p$CO$_2^{sea}$ distribution and biogeochemistry of the basin. The observed $p$CO$_2^{sea}$ data were obtained from an extensive dataset generated by the volunteer observation ship program operated by the National Institute for Environmental Studies (NIES). The reconstructed $p$CO$_2^{sea}$ values agreed well with the $p$CO$_2^{sea}$ measurements, with the root-mean-square error ranging from 17.6 µatm (for the NIES dataset used in the SOM) to 20.2 µatm (for independent dataset). We confirmed that the $p$CO$_2^{sea}$ estimates could be improved by including SSS as one of the training parameters and by taking into account secular increases of $p$CO$_2^{sea}$ that have tracked increases in atmospheric CO$_2$. Estimated $p$CO$_2^{sea}$ values accurately reproduced $p$CO$_2^{sea}$ data at several time series locations in the North Pacific. The distributions of $p$CO$_2^{sea}$ revealed by 7 yr averaged monthly $p$CO$_2^{sea}$ maps were similar to Lamont-Doherty Earth Observatory $p$CO$_2^{sea}$ climatology, allowing, however, for a more detailed analysis of biogeochemical conditions. The distributions of $p$CO$_2^{sea}$ anomalies over the North Pacific during the winter clearly showed regional contrasts between El Niño and La Niña years related to changes of SST and vertical mixing.

1 Introduction

The ocean plays an important role as a major carbon reservoir for CO$_2$ emitted to the atmosphere from fossil fuel burning, cement production, and biomass burning. The ocean has absorbed about 48% of the CO$_2$ emitted to the atmosphere by fossil fuel combustion since the Industrial Revolution (Sabine et al., 2004). To evaluate the global budget of oceanic CO$_2$ uptake, measurements of the partial pressure of oceanic CO$_2$ ($p$CO$_2^{sea}$) in surface seawater have been carried out over the global ocean, with the highest intensity in the equatorial Pacific (Feely et al., 1987, 2006; Ishii et al., 2009), the North Atlantic (Cooper et al., 1998; Olsen et al., 2003; Schuster et al., 2009), and the North Pacific (Inoue et al., 1995; Murphy et al., 2001a; Zeng et al., 2002; Chierici et al., 2006). A compilation of worldwide efforts to measure $p$CO$_2^{sea}$ on a global scale can be found in Takahashi et al. (2009). The authors, led by a team at the Lamont-Doherty Earth Observatory (LDEO), computed a 35 yr $p$CO$_2^{sea}$ climatology (for a reference year 2000) on 4° latitude × 5° longitude resolution
and estimated the annual global air–sea CO$_2$ exchange at $-1.6 \pm 0.9$ Pg C yr$^{-1}$.

Neural network (NN) techniques can be generally described as empirical statistical tools that resolve, to a certain degree, the nonlinear and often discontinuous relationships among proxy parameters without any a priori assumptions. In the past decade a handful of authors have reported the application of an NN technique to basin-scale $p$CO$_2$ analysis (e.g., LeFebvre et al., 2005; Jamet et al., 2007; Friedrich and Oschlies, 2009a, b; Telszewski et al., 2009), concentrating mainly on the North Atlantic Ocean. Most recently, Telszewski et al. (2009) successfully applied a self-organizing-map (SOM) based NN technique to reconstruct $p$CO$_2$ distribution in the North Atlantic (10.5 to 75.5° N, 9.5° E to 75.5° W) for three years (2004 to 2006) by examining nonlinear/discontinuous relationship between $p$CO$_2$ and ocean parameters of sea surface temperature (SST), mixed layer depth (MLD), and chlorophyll $a$ concentration (CHL). One of the main benefits of this approach over the more traditional techniques, such as multiple linear regression (MLR), is that there are numerous empirical relationships established (e.g., 2220 in Telszewski et al., 2009) between examined parameters, allowing for more accurate representation of the highly variable system of interconnected water properties.

The North Pacific is dominated by two major current regimes: the subarctic and subtropical gyres (Fig. 1). The cold Oyashio Current and the warm Kuroshio Current are the western boundary currents of the North Pacific subarctic and subtropical gyres, respectively. The two currents meet at midlatitudes in the western North Pacific and turn toward the east as the North Pacific Current. The North Pacific has been typically characterized as a high-nutrient, low-chlorophyll region of the ocean at most of high latitudes because of the low influx of iron to the ocean surface (Dugdale and Wilkerson, 1991), and as a low-nutrient, low-chlorophyll region at the western and central low latitudes (Karl and Letelier, 2008; Lin et al., 2011). The Bering Sea, which is a marginal sea of the North Pacific, and coastal regions are upwelling areas within which the transport of nutrient- and CO$_2$-rich subsurface water to the surface assures high biological productivity (Chierici et al., 2006). In the North Pacific, there are expected to be thus quite large temporal and spatial variations of $p$CO$_2$. Zeng et al. (2002) reported that large temporal amplitude of $\Delta p$CO$_2$ ($p$CO$_2$sea$-p$CO$_2$air) over 60 µatm was apparent in the western-central subarctic and the eastern subtropics based on their measurements between 1995 and 1999.

For analysis of temporal variability of $p$CO$_2$ or $\Delta p$CO$_2$ in the North Pacific, Stephens et al. (1995) estimated basin-scale monthly $\Delta p$CO$_2$ distributions using simple linear regression analysis between $p$CO$_2$ and SST in 1985. Recently, Sarma et al. (2006) used MLR analysis to estimate $p$CO$_2$sea from SST and satellite-based CHL observations in high-latitude regions of the eastern and western North Pacific, but the applicability of the MLR equations was limited to spring and summer. Takamura et al. (2010) also used MLR analysis to reconstruct $p$CO$_2$sea distributions as a function of SST and sea surface salinity (SSS) from 1999 to 2006 in mid-latitudes (25 to 40° N, 120 to 150° W, 140 to 170° E).

The precise time-series analyses of pelagic ocean $p$CO$_2$ sea variability are limited to time-series stations (Bates, 2007; Dore et al., 2009; González-Dávila et al., 2010) where monthly $p$CO$_2$sea observations are available over extended time periods. Two areas of frequent shipboard observations of $p$CO$_2$sea other than time-series stations are the eastern and western equatorial Pacific (e.g., Feely et al., 2006; Ishii et al., 2009), where the observed interannual $p$CO$_2$sea variations are associated with the El Niño–Southern Oscillation (ENSO). Another place where there have been frequent shipboard $p$CO$_2$sea observations in the North Pacific is the 137° E repeat line (Midorikawa et al., 2006), where a weak but significant relationship between $p$CO$_2$sea and ENSO has been observed. A basin-wide analysis of observed $p$CO$_2$sea variability (including the analysis of the interannual signal) has not yet been successfully performed. An atmospheric CO$_2$ inverse model (Patra et al., 2005) and an ocean biogeochemical model (Valsala et al., 2012), however, suggest the possible correlation of the $p$CO$_2$sea variability with Pacific Decadal Oscillation (PDO).

Our goal in this study was to reconstruct temporal and spatial variability of the $p$CO$_2$sea distribution in the North Pacific for seven years from 2002 to 2008 using the SOM technique applied to the observational $p$CO$_2$sea dataset obtained by the NIES Volunteer Observing Ship (VOS) program. We then compared the estimated $p$CO$_2$sea values with measured $p$CO$_2$sea values obtained from the NIES VOS program and independent validation datasets in various areas of the North
Pacific (Fig. 1). We also presented the change of the $p$CO$_2$
sea distribution in response to the ENSO events.

2 Method and datasets

2.1 Method of $p$CO$_2$
sea estimation

The study area includes the North Pacific from 10 to 60° N
and from 120° E to 90° W, and is hereafter called the North
Pacific, although we have excluded coastal (bathymetric
depth $<$ 500 m) and ice-covered (SST $<$ −1.8 °C) areas from
the analysis. In this study, we hypothesized that $p$CO$_2$
sea could be estimated by a linear function of time and an
SOM function ($f_{\text{SOM}}$) of four independent variables: SST,
MLD, CHL, and SSS. The equation for $p$CO$_2$
sea then takes the following form:

$$p\text{CO}_2\text{sea} = a \times (t - t_{\text{ref}}) + f_{\text{SOM}}(\text{SST}, \text{MLD}, \text{CHL}, \text{SSS}). \quad (1)$$

In Eq. (1) $a$ is the secular rate of change of atmospheric
CO$_2$ in µatm day$^{-1}$, $t$ denotes the date, and the reference
date $t_{\text{ref}}$ is set to 30 June 2005. In addition, we assumed $p$CO$_2$
sea to be a linear function of time in order to take into account
the influence of anthropogenic CO$_2$ emissions on $p$CO$_2$
sea, an effect that could not be accounted for by SST, MLD, CHL,
or SSS. The anthropogenic influence on $p$CO$_2$
sea is considered
negligible for relatively short analyses, e.g., three years (cf.,
Lefèvre et al., 2005; Telszewski et al., 2009), but it builds up
to around 10 µatm after seven years. Midorikawa et al. (2006)
reported that the secular trend of $p$CO$_2$
sea varied from 1.3 to 1.8 µatm yr$^{-1}$ (close to the rate of increase
of atmospheric CO$_2$) in the western subtropical North Pacific
based on their measurements over 20 yr along 137° E. Wong
et al. (2010) also reported that their 30 yr time series of meas-
urements along Line P, the line connecting ocean station
P (50° N, 145° W) to the coast, showed that the long-term
trend of $p$CO$_2$
sea tracked the increase of atmospheric CO$_2$
in the eastern subarctic region. Takahashi et al. (2006) con-
ducted that, for the most part, the increase of oceanic CO$_2$
in the North Pacific followed the increase of atmospheric CO$_2$
for the last 35 yr with the increase rate varying geographi-
cally, reflecting differences in local oceanographic geo-
physical processes. We assumed in this study that the secular
trend of $p$CO$_2$
sea was approximately a constant fraction of
the rate of change of atmospheric CO$_2$ over the North Pa-
cific. Specifically, we assumed the value of the coefficient
$a$ in Eq. (1) to be $4.82 \times 10^{-3}$ (= 1.76/365.285) µatm day$^{-1}$,
which is the rate of increase of atmospheric CO$_2$ concentration
converted from the CO$_2$ mole fractions (µCO$_2$) in the
GLOBALVIEW-CO$_2$ dataset (GLOBALVIEW-CO$_2$, 2011)
for the North Pacific region during the period of analysis.

The method for reconstructing $p$CO$_2$
sea is based on the
methodology of Telszewski et al. (2009), but we allocated
about three times as many neurons on a flat sheet map
(53 × 115) to improve the estimate. A neuron in this study

$$D (x_i, y_j) = \left[(x_i,_{\text{SST}} - y_j,_{\text{SST}})^2 + (x_i,_{\text{MLD}} - y_j,_{\text{MLD}})^2 + (x_i,_{\text{CHL}} - y_j,_{\text{CHL}})^2 + (x_i,_{\text{SSS}} - y_j,_{\text{SSS}})^2\right]^{0.5}. \quad (2)$$

The neuron closest to the training data point in Euclidean
distance terms, here called the winner, is adjusted towards
its value by a fraction of this distance dictated by the linearly
time-decreasing learning function. At the same time, the neu-
rons in the vicinity of the winner are also adjusted towards
the value of the training data point by a fraction of the win-
ner’s adjustment in accordance with a time-decreasing Gaus-
sian function, as explained by Kohonen (2001). This process
results in clustering of similar neurons and self-organization
of the map. The observed $p$CO$_2$
sea dataset is not required at
this stage of the analysis.

Second, each neuron is labeled with an observed $p$CO$_2$
sea value. Technically, the labeling process follows the same
principles as the training process. The labeling data, which
in this study consist of the observed $p$CO$_2$
sea value assigned to a reference year by adding/subtracting the assumed tem-
poral change of $p$CO$_2$
sea and coincided with normalized SST,
MLD, CHL, and SSS values, is presented to the neural
network, and a winner neuron is found (Labeling Process in
Fig. 2b). Instead of adjusting the winner’s value, it is la-
beled with the $p$CO$_2$
sea value of the labeling data. This pro-
cess is carried out for each of the observed $p$CO$_2$
sea values.

Third, the labeled SOM neurons are used to assign $p$CO$_2$
sea values to the geographical grid points of the North Pacific
(Mapping Process in Fig. 2c). The initial training dataset is
presented to the trained and labeled SOM map. Upon com-
puting the winner neuron, no adjustments are made. Instead,
the training data are assigned a $p$CO$_2$
sea value of the win-
ner neuron. This value becomes a $p$CO$_2$
sea estimate for time and location determined by the spatio/temporal coordinates.
of each training datum after the temporal adjustment is done as expressed in Eq. (1).

Consequently, the \( pCO_2^{sea} \) output produced in this work has originally daily frequency and 0.25° latitude \( \times \) 0.25° longitude resolution. The reconstructed monthly \( pCO_2^{sea} \) distributions obtained as a result of this work will be available for scientific purposes from the NIES’s Ship of Opportunity Program (SOOP) website: http://soop.jp.

2.2 Training dataset (SST, MLD, CHL, SSS)

We used four high-resolution datasets – one each for SST, MLD, CHL, and SSS – to train the SOM. We obtained observed SST datasets from the Merged satellite and in situ data Global Daily Sea Surface Temperatures (MGDSST) project (http://goos.kishou.go.jp/trtdb/database.html) at a daily frequency and 0.25° latitude \( \times \) 0.25° longitude resolution (Kurihara et al., 2006). We obtained daily assimilated MLD estimates from the GLoBal Ocean ReanalYses and Simulations (GLORYS) model by Mercator Ocean (Le Centre National de la Recherche Scientifique, France) with a horizontal resolution of 0.25° latitude \( \times \) 0.25° longitude (Bernard et al., 2006; Ferry et al., 2010). Satellite CHL data were obtained from MODIS-Aqua and SeaWiFS Level 3 Standard products provided by NASA/GFSC/DAAC at a frequency of eight per day and resolution of 9 km (http://oceancolor.gsfc.nasa.gov). We obtained assimilated SSS estimates from the MOVE/MRI.COM-NP model of the Meteorological Research Institute, Japan, at a frequency of 10 per day and horizontal resolution of 0.5° latitude and 0.5° longitude (Usui et al., 2006). For the analysis all parameters were re-gridded onto a frequency of one per day and horizontal resolution of 0.25° latitude \( \times \) 0.25° longitude.

We compared the assimilated datasets of SST and SSS with in situ measurements obtained by the NIES VOS project. The values of their differences were calculated to be about 0.01 \( \pm \) 0.53 °C and 0.03 \( \pm \) 0.18, respectively. O’Reilly et al. (2000) reported that the CHL difference between observed values and satellite-borne data was estimated to be 0.00 \( \pm \) 0.25, while the uncertainty of MLD estimate has not been reported. The above sources of uncertainty compose a fraction of the overhaul uncertainty of the method described in Sect. 2.7.1. In this study we have not attempted to assess the relative significance of various sources of uncertainty in the method.

2.3 \( pCO_2^{sea} \) datasets for labeling

To estimate \( pCO_2^{sea} \) fields in the North Pacific, it was necessary to label the trained SOM neurons with \( pCO_2^{sea} \) values. In the labeling process, observed \( pCO_2^{sea} \) data together with
corresponding SST, MLD, CHL, and SSS values were used. We utilized a subset of the North Pacific dataset collected by the NIES VOS program. The $pCO_2^{sea}$ data are available for public use from NIES’s SOOP website: http://soop.jp. Information related to the four VOS lines is summarized in Table 1, and their composite cruise routes are depicted in Fig. 3. The commercial ships collaborating in the NIES VOS program have taken part in trans-Pacific cruises between Japan and North America (10 to 55° N, 140 to 230° E) since March 1995 and between Japan and Oceania (45° S to 35° N, 140 to 180° E) since July 2006. The ships sail regularly at intervals of about 5–8 weeks between Japan and North America or Oceania. On the North America route the volunteer ship sailed to the northern part of North America in the early part of the NIES VOS program, but since 2003 the route has occasionally shifted to the southeast to pass through the Panama Canal (Supplement Fig. 1). On the Oceania route the volunteer ship has sailed regularly on a biweekly basis, with the shipping route mostly fixed since July 2006.

Although we reconstructed $pCO_2^{sea}$ in the North Pacific after 2002, in the analysis we used some in situ data for years 1998–2001 due to the insufficient data coverage especially in the subarctic region for years 2002–2008. The addition of $pCO_2^{sea}$ data from 1998 to 2001 to the labeling dataset improved the coverage of monthly measurements (Supplement, Fig. 2). The improved coverage facilitated reproduction of the rapid drawdown of $pCO_2^{sea}$ due to phytoplankton photosynthesis during the spring bloom in the highly productive western mid-high latitude region.

Murphy et al. (2001b) and Fransson et al. (2006) have both described the technical intricacies of the ocean surface CO$_2$ measurement system used by the NIES VOS program; therefore we only outline the basics here. The nondispersive infrared analyzer used for those measurements was changed from a Licor 6262 to a Licor 7000 for the M/S Pyxis cruises in 2006 (Table 1). The CO$_2$ standard gases were calibrated by the NIES, and are traceable to the World Meteorological Organization scale. The flow-through tandem equilibrator provides a continuous $pCO_2^{sea}$ output with high temporal resolution (Murphy et al., 2001b). The $pCO_2^{sea}$ measurements were made every 10 s, and the $pCO_2^{sea}$ data were 10 min averages of those measurements. The $pCO_2^{sea}$ data were then averaged on a daily basis within 0.25° latitude × 0.25° longitude grid boxes. Consequently, the number of $pCO_2^{sea}$ data by the NIES VOS program amounted to 317 332, and a total of 73 284 $pCO_2^{sea}$ data were binned as the labeling dataset.

### 2.4 Other oceanic CO$_2$ datasets used for the validation of estimated $pCO_2^{sea}$

To validate $pCO_2^{sea}$ values reconstructed by the SOM analysis, we used the fugacity of oceanic CO$_2$ ($fCO_2^{sea}$) dataset from the Surface Ocean CO$_2$ ATlas (SOCAT: http://www.socat.info) version 1.5 database. That dataset has been in the public domain since September 2011, and has been subject to quality control as a part of an international collaboration of more than 10 institutes (including NIES) that work on ocean surface CO$_2$ observations (Pfeil et al., 2013). In the North Pacific, the SOCAT database contains the $fCO_2^{sea}$ values measured mainly by NIES, the Japan Meteorological Agency (JMA), the Japan Agency for Marine-Earth Science and Technology (JAMSTEC), and the United States National Oceanic and Atmospheric Administration (NOAA). For consistency with other datasets used in this study we recalculated $pCO_2^{sea}$ values from the obtained $fCO_2^{sea}$ (Pfeil et al., 2013) wherever necessary.

Underway $pCO_2^{sea}$ data and mooring $pCO_2^{sea}$ data collected by Wong and Johannessen (2010) and Sabine et al. (2010), respectively, were obtained from the Carbon Dioxide Information Analysis Center (CDIAC; http://cdiac.ornl.gov/oceans/). We used those data for the comparisons near ocean station P. In addition, we used $pCO_2^{sea}$ values calculated from measurements of dissolved inorganic carbon (DIC) and total alkalinity (TA) at two stations: station KNOT (44° N, 155° E, Wakita et al., 2010) and station ALOHA (23° N, 202° E, Dore et al., 2009).

### 2.5 Ranges of the training/labeling dataset

As explained by Telszewski et al. (2009), one of the biggest advantages of SOM analysis over the more traditional methods is the fact that the temporal and spatial distribution of proxy parameters in the training and labeling datasets does not influence the analysis. Instead ranges covered by these parameters in each dataset, and more precisely their relative

<table>
<thead>
<tr>
<th>Vessel name</th>
<th>Period</th>
<th>Observed area</th>
<th>NDIR analyzer</th>
</tr>
</thead>
<tbody>
<tr>
<td>M/S Skaugran</td>
<td>Mar 1995–Sep 1999</td>
<td>North Pacific</td>
<td>Rosemount Analytical Model 880A</td>
</tr>
<tr>
<td>M/S Alligator Hope</td>
<td>Nov 1999–Mar 2001</td>
<td>North Pacific</td>
<td>Licor 6262</td>
</tr>
<tr>
<td>M/S Trans Future 5</td>
<td>Jul 2006–present</td>
<td>Western North/ South Pacific</td>
<td>Licor 7000 (Apr 2006–)</td>
</tr>
</tbody>
</table>
overlap, determines whether the SOM will be able to reconstruct the distribution of the predicted parameter. Ranges of the training/labeling datasets and the trained neurons are summarized in Table 2. The training dataset SSTs varied between −1.8 and 32.7 °C; the MLD ranged from 1 m to more than 500 m; CHL varied from 0 to more than 10 mg m−2; and the range of SSS was 30.15–35.69. The values in the labeling datasets and neurons covered most of the range of values in the training dataset. However, the maximum MLDs in the labeling dataset (416 m) and in the neurons (194 m) were substantially lower than the maximum MLD in the training dataset (> 500 m, Table 2). Our results indicate that the correlation between pCO$_2$sea and MLD was not apparent when the MLD was deeper than 200 m (not shown), a result also reported for the North Atlantic by Telszewski et al. (2009). Therefore the MLD dataset is logarithmically normalized, aligning its weight during training (high weight in low values and low weight in high values) with its actual influence on the variability in pCO$_2$sea. Such normalization means that the MLD change from 10 to 100 m is comparable (in terms of change of weight during training) to that from 100 to 1000 m.

### 2.6 Reconstructing pCO$_2$sea distributions in winter at high latitudes

The three products SST, MLD, and SSS provided full basin-wide coverage from 2002 to 2008. However, the CHL data were affected by the lack of satellite coverage from November to January at high latitudes of the North Pacific (north of 45° N) due to the low angle of the sun during that time and enormous atmospheric correction required to retrieve the signal. To reconstruct pCO$_2$sea for this area during those months, we assumed that pCO$_2$sea could be adequately characterized by only three parameters: SST, MLD, and SSS. The rationale for this assumption is that biological activity is relatively low during the winter at high latitudes (e.g., Imai et al., 2002). Therefore, we prepared another SOM trained by the three parameters SST, MLD, and SSS. We generated complete pCO$_2$sea maps in the study area by combining the pCO$_2$sea values obtained with the four-parameter SOM including CHL with the values obtained with the three-parameter SOM excluding CHL in the area north of 45° N (14 % of the study area) during the period from November to January. We calculated the difference between the pCO$_2$sea values estimated with the four-parameter SOM and the three-parameter SOM during the above period in the region between 40 and 45° N and found it to be −2.0 ± 2.2 µatm. We added this difference to the pCO$_2$sea values obtained with the three-parameter SOM in the area north of 45° N.

### 2.7 Uncertainty and improvement of the pCO$_2$sea estimate

#### 2.7.1 Uncertainty

For each in situ pCO$_2$sea measurement, the corresponding SOM pCO$_2$sea estimate was determined on the basis of the spatial (0.25° longitude × 0.25° latitude grid) and temporal (daily intervals between 1 January 2002 and 31 December 2008) coordinates associated with the measurement. We calculated the root-mean-square error (RMSE) between observed pCO$_2$sea and estimated pCO$_2$sea values as follows:

\[
\text{RMSE} = \sqrt{\frac{\sum (\text{pCO}_2\text{sea} \text{ (estimate)} - \text{pCO}_2\text{sea} \text{ (observed)})^2}{n}},
\]

where \( n \) is the number of points in the labeling dataset. The RMSE provided an estimate of the uncertainty of the method in reproducing the in situ measurements, and equaled 17.6 µatm, or 5.0 % of the average pCO$_2$sea of the in situ dataset. A scatter plot of the estimated pCO$_2$sea against the observed pCO$_2$sea (Fig. 4) shows that the values are clustered around the 1 : 1 line with slightly more scatter at very high pCO$_2$sea. It should be noted that the reported RMSE is fairly large for some applications of small geographical extent such as determining air–sea CO$_2$ flux at local and regional scales.

As an independent validation exercise, we calculated the RMSE between the subset of the SOCAT dataset (all North Pacific data from 10 to 60° N and from 120° E to 90° W for 2002–2008 inclusive) and our SOM estimate. Such a calculated uncertainty estimate turns out to be 20.1 µatm, which makes this study similar to or more accurate than previous reports for the region, despite its largest temporal extent to date. Zeng et al. (2002) estimated the distribution of monthly averaged pCO$_2$sea in the North Pacific based on data from the NIES VOS program from 1995 to 1999, and reported that the estimated pCO$_2$sea agreed with the in situ pCO$_2$sea to within an RMSE of 24.9 µatm. Sarma et al. (2006) used an MLR method to estimate the distribution of monthly average

---

**Table 2.** Ranges of SST, MLD, CHL, and SSS in the training dataset, the labeling dataset, and the trained neurons. Percentages of the training data within the range of the labeling dataset and the neurons are given for each parameter.

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Min.</th>
<th>Max.</th>
<th>Cover (%)</th>
<th>Min.</th>
<th>Max.</th>
<th>Cover (%)</th>
<th>Min.</th>
<th>Max.</th>
<th>Cover (%)</th>
<th>Min.</th>
<th>Max.</th>
<th>Cover (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>SST (°C)</td>
<td>-1.8</td>
<td>32.7</td>
<td></td>
<td>1</td>
<td>&gt; 500</td>
<td></td>
<td>0.00</td>
<td>10</td>
<td></td>
<td>30.15</td>
<td>35.69</td>
<td>99.778</td>
</tr>
<tr>
<td>MLD (m)</td>
<td></td>
<td></td>
<td></td>
<td>1</td>
<td>416</td>
<td>99.995</td>
<td>0.00</td>
<td>10</td>
<td>100</td>
<td>30.15</td>
<td>35.69</td>
<td>100</td>
</tr>
<tr>
<td>CHL (mg m$^{-2}$)</td>
<td></td>
<td></td>
<td></td>
<td>1</td>
<td>194</td>
<td>99.807</td>
<td>0.00</td>
<td>3.2</td>
<td>99.778</td>
<td>31.79</td>
<td>35.58</td>
<td>99.886</td>
</tr>
</tbody>
</table>

---

**Biogeosciences, 10, 6093–6106, 2013**

www.biogeosciences.net/10/6093/2013/
The improvement achieved by using the SSS dataset, we generated another $pCO_{2}^{sea}$ map derived with a three-parameter SOM that excluded SSS and compared the result with the four-parameter SOM result. The RMSE between NIES dataset and the three-parameter SOM estimate was 20.0 µatm. Use of SSS in the training dataset therefore reduced the RMSE by 12%. The $pCO_{2}^{sea}$ distributions were also improved by the use of the SSS data. To visualize the differences, we mapped 7 yr averaged monthly $pCO_{2}^{sea}$ distributions in February and August derived with and without inclusion of SSS in the training dataset (Fig. 5). The estimated $pCO_{2}^{sea}$ derived from the three-parameter SOM in February is characterized by a smaller longitudinal difference in mid-latitudes than the $pCO_{2}^{sea}$ derived from the four-parameter SOM. Furthermore, use of the four-parameter SOM enabled reconstruction of quite high $pCO_{2}^{sea}$ values in August in the eastern low/midlatitude region, where the North Pacific Current flows, whereas use of the three-parameter SOM failed to reproduce this feature. Figure 6 shows the temporal variation of $pCO_{2}^{sea}$ derived with the two SOMs in the North Pacific Current region (36 to 38° N, 138 to 142° W). It clearly shows that the agreement between observed and estimated $pCO_{2}^{sea}$ values was better for the four-parameter SOM than the three-parameter SOM. The RMSE in the region was improved from 15.9 to 10.6 µatm by inclusion of SSS. The improvement was especially apparent during the summer, when high $pCO_{2}^{sea}$ values (about 400 µatm) were observed.

Taking into account the influence of anthropogenic CO₂ emissions on the trend of $pCO_{2}^{sea}$ was the second improvement introduced in this study. As described above it was done by adding or subtracting 1.76 µatm yr⁻¹ (4.82 × 10⁻³ µatm day⁻¹) to project observed $pCO_{2}^{sea}$ values to the $pCO_{2}^{sea}$ values in the reference year of 2005 (Eq. 1). The improvement of the $pCO_{2}^{sea}$ estimate by making this correction was not spatially uniform. For example, the RMSEs were reduced by adding the term from 10.2 to 9.1 µatm in the station P area (48 to 52° N, 142.5 to 147.5° W), from 8.8 to 7.4 µatm in the western subtropics (WST) area (14 to 18° N, 135.5 to 140.5° W), and from 10.8 µatm to 7.9 µatm in the station ALOHA area (21 to 25° N, 155.5 to 160.5° W). In contrast, the improvements at station KNOT area (43.5 to 45.5° N, 142.5 to 147.5° W) was 20.0 µatm. Use of SSS in the training dataset therefore improved the agreement between observed and estimated $pCO_{2}^{sea}$ values (about 400 µatm) were observed.

Fig. 4. Scatter plot of estimated $pCO_{2}^{sea}$ with observed $pCO_{2}^{sea}$. Colors indicate the number of data in a 1 µatm × 1 µatm bin.

Fig. 5. Comparison of 7 yr averaged monthly $pCO_{2}^{sea}$ distributions from 2002 to 2008 in February (upper) and August (bottom). Figures on the left are the $pCO_{2}^{sea}$ distributions estimated from the four-parameter SOM including SSS. Figures on the right were estimated from the three-parameter SOM without SSS.
3 Temporal and spatial variation of $p\text{CO}_2^\text{sea}$

3.1 Mapping of 7 yr averaged monthly $p\text{CO}_2^\text{sea}$ distributions

Figure 7 presents a comparison of 7 yr (2002–2008) averaged monthly $p\text{CO}_2^\text{sea}$ distributions derived from SOM results for February, May, August, and November with LDEO $p\text{CO}_2^\text{sea}$ climatology (Takahashi et al., 2009). The SOM-reconstructed $p\text{CO}_2^\text{sea}$ distributions in this study clearly show a tongue of very low $p\text{CO}_2^\text{sea}$ (about 320 µatm) water distributed (except in August) uniformly between the western and central midlatitude regions of the North Pacific (Fig. 7). Such low $p\text{CO}_2^\text{sea}$ values are attributed to high rates of photosynthesis (Kameda, 2003) and cooling of the seawater that occurred mainly in the subtropics. In addition, a band of relatively high $p\text{CO}_2^\text{sea}$ caused mainly by a seasonal rise in temperature was also apparent during the period from May to September in the western North Pacific between 15° and 30° N. The temperature rise began in April and amounted to about 2–5 °C. Following the temperature dependence of $p\text{CO}_2^\text{sea}$ given by Takahashi et al. (1993), $\delta \ln p\text{CO}_2^\text{sea} / \delta T = 0.0423 \ ^\circ C^{-1}$, the expected $p\text{CO}_2^\text{sea}$ rise due to the temperature effect is about 30–70 µatm. The observed increase in expected $p\text{CO}_2^\text{sea}$ is only about half of the expected $p\text{CO}_2^\text{sea}$ rise due to temperature effects. The increase may have been attenuated by other factors such as photosynthetic uptake of CO$_2$.

The comparison with the LDEO climatology shows that the SOM-reconstructed $p\text{CO}_2^\text{sea}$ maps reveal similar large-scale patterns to these known from the LDEO climatology. However, the SOM results, due to its much higher spatio-temporal resolution, allow for more detailed analysis of local and regional features. Both studies show high $p\text{CO}_2^\text{sea}$
values (over 400 µatm) at high latitudes in the North Pacific in February; however, the SOM-reconstructed $p$CO$_2$sea distribution shows $p$CO$_2$sea-rich water between the Bering Sea and the coast of northern Japan along the axis of the cold, southward-flowing Eastern Kamchatka Current. As described in Sect. 2.7.2, high $p$CO$_2$sea values are apparent from June to October in the eastern low/midlatitude region, where the North Pacific Current and the California Current flow, and the high $p$CO$_2$sea field dominates. With respect to the coastal region, low estimates of $p$CO$_2$sea stretch along the coastline from the Aleutian Islands to the California Peninsula from May to October, when the concentration of phytoplankton is high.

The map of differences between SOM results and LDEO climatology for reference year 2005 is shown in Fig. 8. The difference distribution is positive in the western subarctic and the western subtropics and negative in the central-eastern subtropics, the calculated monthly mean difference is close to zero ($-0.8$ µatm), and its standard deviation is $11.2$ µatm.

3.2 Reproducibility of temporal $p$CO$_2$sea variations in each of six regions

To facilitate a discussion about the temporal variations of $p$CO$_2$sea in the North Pacific, Fig. 9 shows the time series of area-averaged $p$CO$_2$sea estimated in this study for six specific regions of the North Pacific along with observations made during several campaigns at these locations as well as computed estimates from Takamura et al. (2010). The grid size of all the averaged areas except in the station KNOT area is set to $4\degree$ latitude $\times 5\degree$ longitude, whereas the station KNOT area is set to $43.5$–$45.5\degree$ N, $153$–$157\degree$ E. Blue circles and red dots are in situ $p$CO$_2$sea values obtained from NIES measurements and the SOCAT database, respectively. Black dots and crosses on panel (a) and (e) are the $p$CO$_2$sea values calculated from measurements of DIC and TA reported by Wakita et al. (2010) and Dore et al. (2009), respectively. Purple dots on panel (b) are the $p$CO$_2$sea values observed by Wong and Johannessen (2010) and Sabine et al. (2010). In panel (e), the solid green line denotes the $p$CO$_2$sea values during the 2002–2006 period estimated by Takamura et al. (2010). Note that the range of the ordinate in the station KNOT area is larger than those of other station areas.

and observed $p$CO$_2$sea, as exemplified in the area surrounding station KNOT (Fig. 9a), occur occasionally, but there is no systematic overestimate by the SOM in this region. The calculated $p$CO$_2$sea in station P area generally agree well with the data from the NIES VOS program as well as with $p$CO$_2$sea values measured by an underway system from 2002 to 2003 and by a moored buoy system from 2007 to 2008 (Fig. 9b). The largest seasonal amplitudes tend to coincide with the largest disagreements between the estimates (Zeng et al., 2002). The calculated $p$CO$_2$sea values in the KE area of the eastern midlatitude region (Fig. 9c) agree well with the NIES dataset as well as with the $f$CO$_2$sea values from the SOCAT dataset, with all $p$CO$_2$sea values lying within the spatial variability. The results of Takamura et al. (2010) also agree with the $p$CO$_2$sea measurements to within $15$–$20$ µatm, and the temporal pattern of those data is generally consistent with the $p$CO$_2$sea estimates within the spatial variability from this study. The temporal variations of $p$CO$_2$sea in the WST (Fig. 9d) and station ALOHA area (Fig. 9e) agree well with the $p$CO$_2$sea values in the SOCAT dataset, even though the observed $p$CO$_2$sea data used for the labeling process in the SOM analysis rarely existed in these areas. The calculated
$pCO_{2}^{sea}$ values in the eastern subtropics (EST) area (14 to 18° N, 115.5 to 119.5° W) also agree well with the data from the NIES VOS program (Fig. 9f). As shown in Fig. 9d–f, the patterns of variation were similar in the WST, station ALOHA, and EST areas. Keeping in mind that only data obtained by the NIES VOS program were used in the SOM labeling process, these results suggest that the labeling process allows for labeled SOM neurons to effectively learn $pCO_{2}^{sea}$ variations from $pCO_{2}^{sea}$ values observed in other subtropical areas. This confirms the earlier suggestions that the SOM technique, to a larger extent than more traditional mapping techniques, overcomes problems associated with temporal and spatial scarcity of the labeling data (in situ) by putting significant weight on the availability and quality of the training data (satellite and assimilation).

Finally, as an additional independent validation exercise, we calculated the RMSE between all the independent data visualized in Fig. 9 and equivalent SOM estimates. Such a calculated uncertainty estimate turns out to be 20.1 µatm, almost identical to that obtained for SOCAT dataset, giving more confidence in our error estimate.

### 3.3 Difference of $pCO_{2}^{sea}$ distributions during ENSO events

Fig. 10. Anomalies from the monthly climatology for the period box 2002–2008 for detrending $pCO_{2}^{sea}$ (upper), SST (middle), and MLD (bottom) distributions during the winter of 2003 (panels on the left) and 2008 (panels on the right).

The ENSO has a large influence on the climate of the North Pacific (IPCC, 2007), and large fluctuations of $pCO_{2}^{sea}$ coincided with the ENSO cycle have also been observed in the equatorial Pacific (Feely et al., 2006; Ishii et al., 2009). Based on their measurements from 1983 to 2003, Midorikawa et al. (2006) have suggested that the interannual variation of $pCO_{2}^{sea}$ in the western subtropical North Pacific is also related to the ENSO. Although the extent of the ENSO influence on oceanic and atmospheric variables is known to be global (Trenberth and Caron, 2000), the impact of the ENSO on the distribution of $pCO_{2}^{sea}$ over the entire area of the North Pacific is not well understood. Figure 10 depicts the estimated distributions of the detrended $pCO_{2}^{sea}$, SST, and MLD anomalies during the winters of 2003 (i.e., El Niño) and 2008 (i.e., La Niña). Anomalies in Fig. 10 are deviations from the monthly climatology for the period of 2002–2008. El Niño/La Niña periods were chosen in accordance with JMA’s definition based on the 5-month running mean SST deviation for the Niño.3 region (5° S to 5° N, 90 to 150° W).

The patterns of SST anomalies in Fig. 10 are typical of El Niño and La Niña winters (Trenberth and Caron, 2000; Alexander et al., 2002). The $pCO_{2}^{sea}$ anomaly related to ENSO events is easily discernible in the western-central subtropical region, in the eastern subarctic region, and in the eastern midlatitude region south of 30° N. For example, a negative $pCO_{2}^{sea}$ anomaly is apparent in the western-central subtropical region in 2003 (El Niño), when the SST anomaly was negative, whereas a positive $pCO_{2}^{sea}$ anomaly is apparent in 2008 (La Niña), when the SST anomaly is positive. The opposite pattern is observed for the eastern midlatitude region south of 30° N. The amplitudes of the associated $pCO_{2}^{sea}$ anomalies are about 15 µatm, and their SST amplitudes are 1 °C. The $pCO_{2}^{sea}$ change closely tracked the SST change in accordance with the iso-chemical temperature dependency of Takahashi et al. (1993).

A negative relationship between $pCO_{2}^{sea}$ and SST is apparent in the eastern subarctic North Pacific, where the signal of thermodynamic changes on variations of $pCO_{2}^{sea}$ was opposite to that seen in the subtropics. As indicated in Fig. 10, the MLD anomaly clearly showed the typical pattern of ENSO events (Alexander et al., 2002), and the MLD was approximately 10 m deeper in 2008 than in 2003 in the region. CLIVAR Repeat Section Line P data provided by Miller et al. (2010) showed that surface (< 10 m) DIC concentration in station P in February 2003 is about 35 µmol kg$^{-1}$ lower than in February 2008. By using CO2SYS program (Lewis and Wallace, 1998; Robbins et al., 2010), the estimated $pCO_{2}^{sea}$ difference between February 2003 and February 2008 in the region caused by the changes of surface DIC, TA, temperature and salinity, is about 14 µatm. Since the $pCO_{2}^{sea}$ difference between 2002 and 2008 based on the DIC measurements is well consistent with the difference derived by the SOM results, it strongly suggests that more CO$_2$-rich subsurface water was entrained into surface waters during the La Niña period than during the El Niño period.
4 Summary

In this study we used the SOM technique of Telszewski et al. (2009) to examine the temporal and spatial variations of \( p\text{CO}_2 \) in the North Pacific during the period 2002–2008. To improve the \( p\text{CO}_2 \) estimates, we used SSS as an additional training parameter and assumed a trend of increasing \( p\text{CO}_2 \) to take into account the effect of anthropogenic CO\(_2\) emissions on \( p\text{CO}_2 \). The estimated results revealed that the SOM technique could satisfactorily reconstruct variations of \( p\text{CO}_2 \) associated with bio-geophysical processes expressed by the variability in four proxy parameters: SST, MLD, CHL, and SSS. We calculated the uncertainty of the \( p\text{CO}_2 \) estimation to be from 17.8 µatm for the NIES labeling dataset to 20.2 µatm for the SOCAT dataset. The fact that the uncertainty was reduced by about 12% by inclusion of SSS in the training dataset suggests that SSS can be a useful parameter for the estimation of temporal and spatial variation of \( p\text{CO}_2 \). We also found that \( p\text{CO}_2 \) estimates were improved by taking account of the temporal trend associated with anthropogenic CO\(_2\) emissions.

The calculated \( p\text{CO}_2 \) variations in six ocean areas generally agreed well not only with the NIES VOS program \( p\text{CO}_2 \) data used for the labeling process but also with other in situ \( p\text{CO}_2 \) datasets. Seven-year (2002–2008) averaged monthly \( p\text{CO}_2 \) distributions were similar to 35 yr climatology \( p\text{CO}_2 \) distributions (Takahashi et al., 2009). However, the SOM-based \( p\text{CO}_2 \) mapping, with its high spatial resolution, reflected oceanic conditions with more detail. The estimated interannual \( p\text{CO}_2 \) variability revealed a difference in the spatial pattern of \( p\text{CO}_2 \) during the winter of the El Niño period in 2003 and the La Niña period in 2008. A negative \( p\text{CO}_2 \) anomaly was apparent in 2003 in the western subtropical North Pacific and in the eastern subarctic North Pacific off the coast of Alaska, whereas a positive anomaly was apparent in 2008 in the same regions. In the western subtropical and eastern midlatitude regions, the correlation of the \( p\text{CO}_2 \) variability with ENSO events seemed to be related mainly to changes in the thermodynamic properties of seawater. In contrast, similar correlation in the subarctic North Pacific seemed to be related to changes in vertical transport of CO\(_2\)-rich subsurface waters.

Further improvement of \( p\text{CO}_2 \) estimates will most certainly require an increase in the number of data points used for labeling. With new datasets becoming available (SOCAT version 2 and LDEO V2012) and offering relatively dense annual data coverage in several oceans regions, we are now in a position to commence a sensitivity study allowing for a meaningful quantitative assessment to be made of the uncertainty related to the amount of labeling data utilized during the mapping process. In this study, 7% of the neurons were not labeled, suggesting that in situ measurements covering a wider range of environmental conditions (as approximated by SST, MLD, CHL, and SSS) are needed to enable the full mapping potential of the method. We plan to undertake a longer-term study covering global ocean using the community quality-controlled (Pfeil et al., 2013) SOCAT collection as the labeling dataset. This work will include a sensitivity study hopefully allowing for quantification of the relationship between the amount of the in situ data and the method’s uncertainty estimate.

The number of neurons is also crucial for accurate \( p\text{CO}_2 \) estimation. In this study we used three times as many neurons as Telszewski et al. (2009) to achieve adequate reproducibility of the \( p\text{CO}_2 \) estimates. However, the number of neurons used in this study was based on the available computing power rather than determined by scientific need. It might also be possible to improve the \( p\text{CO}_2 \) estimate by inclusion of more ocean parameters. Sea surface height is a potential training parameter with basin-wide coverage.

In addition to estimates in the North Pacific, long-term global \( p\text{CO}_2 \) mapping based on such measurements is also important for understanding interannual variations of air–sea CO\(_2\) exchanges. Although \( p\text{CO}_2 \) variations related to climate changes such as the PDO have been reported (Valsala et al., 2012), the overall impact of such changes on global \( p\text{CO}_2 \) variations is not well understood. In the present study, the study area was confined to the North Pacific. However, the SOM technique used in the present study has the potential to estimate \( p\text{CO}_2 \) in regions where there are insufficient numbers of observations, and such regions will be our next target. It is axiomatic to say that further \( p\text{CO}_2 \) measurements are critical, especially in the South Pacific, where few \( p\text{CO}_2 \) measurements have been made (Sabine et al., 2013).

Supplementary material related to this article is available online at http://www.biogeosciences.net/10/6093/2013/bg-10-6093-2013-supplement.pdf.

Acknowledgements. We deeply appreciate the generous cooperation of Seaboard International Shipping Co., Mitsui O. S. K. Lines Co., Toyo Fuji Shipping Co., and Kagoshima Senpaku Co. with the NIES VOS program. We would like to thank the crew of the M/S Skaugran, M/S Alligator Hope, M/S Pyxis, and M/S Trans Future 5. We also thank S. Kariya and T. Yamada of the Global Environmental Forum for their constant assistance with the observations, and appreciate the work of K. Katsumata for calibrating the CO\(_2\) standard gases. We are thankful to contributors of the SOCAT database and the CDIAC database for providing oceanic CO\(_2\) data. We gratefully acknowledge Mercator Ocean for providing the GLORYS model output. The research was financially supported by the Global Environment Research Account for National Institutes by the Ministry of Environment, Japan. Finally, we also express our deep thanks to R. Wannikhof, A. Lenton, and two anonymous reviewers, who provided us with many useful comments.

Edited by: F. Chai
References


S. Nakaoka et al.: Estimating temporal and spatial variation of ocean surface $p$CO$_2$


GLOBALVIEW-CO2: Cooperative Atmospheric Data Integration Project – Carbon Dioxide, CD-ROM, NOAA ESRL, Boulder, Colorado [Also available on Internet via anonymous FTP to ftp.cmdl.noaa.gov, Path: ccg/co2/GLOBALVIEW], 2011.


S. Nakaoka et al.: Estimating temporal and spatial variation of ocean surface pCO₂

6105


www.biogeosciences.net/10/6093/2013/

Biogeosciences, 10, 6093–6106, 2013


Takahashi, T., Olafsson, J., Godard, J. G., Chipman, D. W., and Sutherland, S. C.: Seasonal variation of CO₂ and nutrient in the

Biogeosciences, 10, 6093–6106, 2013