



Satellite-based prediction of pCO₂ in coastal waters of the eastern North Pacific

Burke Hales^{a,*}, Peter G. Strutton^g, Martin Saraceno^{b,c}, Ricardo Letelier^a, Taro Takahashi^d,
Richard Feely^e, Christopher Sabine^e, Francisco Chavez^f

^a Oregon State University College of Oceanic and Atmospheric Sciences, Corvallis, OR 97331, United States

^b Centro de Investigaciones del Mar y la Atmosfera, Consejo Nacional de Investigaciones Científicas y Técnicas, Argentina

^c Departamento de Ciencias de la Atmósfera y de los Océanos, Facultad de Ciencias Exactas y Naturales, Universidad de Buenos Aires, Argentina

^d Lamont-Doherty Earth Observatory of Columbia University, Palisades, NY 10964, United States

^e NOAA Pacific Marine Environmental Laboratory, Seattle, WA 98115, United States

^f Monterey Bay Aquarium Research Institute, Moss Landing, CA 95039, United States

^g Institute for Marine and Antarctic Studies, University of Tasmania, Hobart, Australia and the Australian Research Council Centre of Excellence for Climate System Science, Australia

ARTICLE INFO

Article history:

Received 25 February 2011

Received in revised form 13 February 2012

Accepted 4 March 2012

Available online 16 March 2012

ABSTRACT

Continental margin carbon cycling is complex, highly variable over a range of space and time scales, and forced by multiple physical and biogeochemical drivers. Predictions of globally significant air–sea CO₂ fluxes in these regions have been extrapolated based on very sparse data sets. We present here a method for predicting coastal surface-water pCO₂ from remote-sensing data, based on self organizing maps (SOMs) and a nonlinear semi-empirical model of surface water carbonate chemistry. The model used simple empirical relationships between carbonate chemistry (total dissolved carbon dioxide (T_{CO_2}) and alkalinity (T_{Alk})) and satellite data (sea surface temperature (SST) and chlorophyll (Chl)). Surface-water CO₂ partial pressure (pCO₂) was calculated from the empirically-predicted T_{CO_2} and T_{Alk} . This directly incorporated the inherent nonlinearities of the carbonate system, in a completely mechanistic manner. The model's empirical coefficients were determined for a target study area of the central North American Pacific continental margin (22–50°N, within 370 km of the coastline), by optimally reproducing a set of historical observations paired with satellite data. The model-predicted pCO₂ agreed with the highly variable observations with a root mean squared (RMS) deviation of <20 μatm, and with a correlation coefficient of >0.8 ($r = 0.81$; $r^2 = 0.66$). This level of accuracy is a significant improvement relative to that of simpler models that did not resolve the biogeochemical sub-regions or that relied on linear dependences on input parameters. Air–sea fluxes based on these pCO₂ predictions and satellite-based wind speed measurements suggest that the region is a ~14 Tg C yr⁻¹ sink for atmospheric CO₂ over the 1997–2005 period, with an approximately equivalent uncertainty, compared with a ~0.5 Tg C yr⁻¹ source predicted by a recent bin-averaging and interpolation-based estimate for the same area.

© 2012 Elsevier Ltd. All rights reserved.

1. Introduction and background

Coastal waters have been alternately reported as globally important sources or sinks of atmospheric CO₂ (Bianchi et al., 2005; Borges, 2005; Cai et al., 2006; Ducklow and McAllister, 2005; Chen et al., 2004; DeGrandpre et al., 2002; Frankignoulle and Borges, 2001; Hales et al., 2005; Smith and Hollibaugh, 1993; Thomas et al., 2004; Tsunogai et al., 1999, as summarized by Hales et al. (2008)). The uncertainty stems from several factors. First, the dynamic range of surface pCO₂ in coastal waters spans hundreds of μatm over a variety of time and space scales (Friederich et al., 2002; Cai et al., 2003; Cai, 2003; Hales et al., 2005; Bates et al., 2005; Chavez et al., 2007). This makes adequate observational constraint difficult to achieve, and global or regional

flux estimations have always been the result of extrapolating a few spatially and temporally limited observations far beyond their scope (Hales et al., 2005; Thomas et al., 2004; Cai et al., 2006; Borges, 2005). However, the vast majority of flux estimates based on direct observation of pCO₂ in coastal waters suggests that coastal waters are net sinks of atmospheric CO₂. The fact that terrestrial inputs of carbon via rivers supply nearly a petagram per year (Pg C yr⁻¹; 1 Pg = 10¹⁵ g) of terrestrial carbon to coastal waters, almost none of which can be accounted for in coastal sediment or water column reservoirs (Degens et al., 1991; Ittekkot and Laane, 1991; Spitzky and Leenheer, 1991; Hedges et al., 1997; Meybeck and Vorosmarty, 1999; Aitkenhead and McDowell, 2000; Schlunz and Schneider, 2000 as summarized by Bauer et al., 2008) further complicates matters. This imbalance suggests a large efflux of CO₂, as argued by Smith and Hollibaugh (1993) that has not been observed in coastal waters. Borges et al. (2005, 2006) offered the possibility that estuaries may be the location of efflux of terrestrial

* Corresponding author. Tel.: +1 541 737 8121.

E-mail address: bhales@coas.oregonstate.edu (B. Hales).

carbon, while Cai (2011) suggested that marine-source carbon degradation supported estuarine efflux while terrestrial material is degraded offshore. Again, sampling coverage is limited.

Chavez et al. (2007) compiled all the North American coastal observations of $p\text{CO}_2$ in the Lamont-Doherty Earth Observatory (LDEO) database (now hosted by the Carbon Dioxide Information Analysis Center (CDIAC) data server), to assess the contribution of coastal waters to the continental carbon budget. They found that North American coastal waters contributed a small net source of CO_2 to the atmosphere ($\sim 2 \text{ Tg C yr}^{-1}$; $1 \text{ Tg} = 0.001 \text{ Pg}$), but this near-neutral flux was the result of large high-latitude sinks ($\sim 35 \text{ Tg C yr}^{-1}$ into the Pacific and Atlantic north of 50°N and the Bering and Chukchi seas) that were balanced by large low-latitude sources ($\sim 40 \text{ Tg C yr}^{-1}$ out of the Pacific and Atlantic south of 25°S and of the Gulf of Mexico and Caribbean Sea). Although the data set included nearly 1 million observations, the ship of opportunity-based sampling did not provide uniform spatial or temporal coverage. In the bin-averaging approach employed by Chavez et al. (2007) it is clear that the opposing large magnitude high and low latitude sink and source terms were based on pixels with inadequate observations throughout the calendar year (Hales et al., 2008).

With these concerns in mind, we have attempted to produce synthetic approaches for estimating surface-water $p\text{CO}_2$ distributions in coastal waters. True mechanistic modeling including accurate physical circulation fields and biogeochemistry, driven by actual physical forcing is ultimately the ideal approach. However, modeling coastal environments is challenging even for only the physical circulation, and to date the few truly coupled biogeochemical/physical models that predict surface $p\text{CO}_2$ distributions are limited to narrow geographical regimes such as the southern California Current system (e.g. Gruber et al., 2006) or the shelves of the northwest North Atlantic (Fennel et al., 2008; Fennel and Wilkin, 2009).

We present here a semi-empirical approach to developing algorithms that link surface water $p\text{CO}_2$ to remotely sensed data, in addition to position and time. We found that two factors were critical: (1) an objective means for identifying biogeochemical sub-regions within coastal waters, since there was very low success for single predictive algorithms applied across wide geographical areas such as the continental waters of North America; and (2) a justified application of high-order dependences on input parameters, since simple linear dependences were incapable of generating the large dynamic ranges seen in the observed distributions. Given these factors, we present a method that uses (1) self-organizing maps (SOMs) based on satellite observations for distinguishing distinct provinces, and (2) a semi-mechanistic representation of the relationships between seawater carbonate chemistry (total dissolved CO_2 (T_{CO_2}) and alkalinity (T_{Alk})) and the input parameters of sea surface temperature (SST) and chlorophyll (Chl). $p\text{CO}_2$ was then calculated from T_{CO_2} and T_{Alk} , thus incorporating the inherent nonlinearities of the seawater carbonate system. We test this approach for the Pacific coastal waters of the central North American continent, from 22°N to 50°N , within 370 km (~ 200 nautical miles) of the shore.

Although this is not the first attempt to link surface-ocean $p\text{CO}_2$ to more extensively-observable parameters, earlier approaches have been substantively different. Most were in the open ocean where dynamic ranges were smaller and $p\text{CO}_2$ was not impacted by the continental margins (Cosca et al., 2003; Lefevre et al., 2005; Feely et al., 2006; Friedrich and Oschlies, 2009a,b; Park et al., 2010), or in narrowly defined margin settings where individual drivers such as river plumes (Lohrenz and Cai, 2006) or thermal forcing (Wanninkhof et al., 2006) dominate the $p\text{CO}_2$ distributions. Lefevre et al. (2005), Friedrich and Oschlies (2009a,b), and Telszewski et al. (2009) did use the neuronal network approach, but applied it only over the parameter space of their observations,

as opposed to the regional province identification that we applied here. None of the previous efforts have attempted synthesis of a region so dynamic and diverse as the North American Pacific coast, nor have any included mechanistically-justified nonlinearities as we do here.

2. Materials and methods

2.1. $p\text{CO}_2$ observations

We used the compilation of $p\text{CO}_2$ data prepared by Chavez et al. (2007), for waters surrounding the North American continent to train the algorithm. These data were originally stored in the LDEO $p\text{CO}_2$ data repository, and are now available from CDIAC (<http://cdiac.esd.ornl.gov/>). They consist of nearly 800,000 observations made in the interval 1978–2005 (heavily weighted for more recent observations), all based on analysis of equilibrated gas headspace over flowing seawater streams or discrete samples. A significant portion of these data, particularly those closest to shore were excluded from the global analyses of Takahashi et al. (1995, Takahashi et al., 2002, Takahashi et al., 2009), and thus these observations from continental margins have not been included in broader estimates of air–sea CO_2 fluxes.

We limited our analyses to observations made within 370 km of the nearest major coastline, roughly consistent with the position of the 200 nautical mile limit of national Exclusive Economic Zones (EEZs). This limit is functional as well as operational. Essentially all signals of coastal influences are gone by this distance from shore, although exceptions do occur (Chavez et al., 2007). This extends the spatial limit of the analysis far enough to meet nearly all of the landward-most boundaries of the global ocean syntheses of Takahashi et al. (e.g. Takahashi et al., 2009), which were bin-averaged at 4° latitude by 5° longitude bins. Finally, the EEZ boundary represents the furthest distance where local governments may restrict access to local waters and the data collected therein. Thus, this represents a distance from shore within which data may have been excluded from regional or global compilations. Limiting the observations to those made with coincident SeaWiFS satellite observations, which started in September 1997, reduced the sample size to about 300,000. The North American data set was then divided into four regions—Atlantic, Gulf of Mexico/Caribbean, Pacific, and Bering/Chukchi coastal waters. For the high resolution regional analyses described below, we focused on the North American Pacific coast between 22°N and 50°N , where about 96,000 observations exist within 370 km of the coastline and with coincident SeaWiFS data (see Section 2.2).

We did not attempt to correct the surface $p\text{CO}_2$ data to a reference year, as done for the Takahashi global syntheses (e.g. Takahashi, 2009). The dynamics of the coastal ocean are complicated and unique within regions, with some being dominated by terrestrial inputs, others by upwelling, others by local heating and cooling. Water residence times within coastal regions can be short relative to the age of the water masses themselves, and thus it is questionable whether to treat freshly upwelled waters that might have last been at the surface some decades ago (Feely et al., 2008) with a modern adjustment to the in-water $p\text{CO}_2$. Limiting the analysis to the SeaWiFS era suggests a very small correction for the 1995–2005 interval.

For the generation of the predictive algorithms, we used the in situ SST recorded along with the $p\text{CO}_2$ and satellite chlorophyll measurements. We viewed this as important for the closest coupling of the SST and $p\text{CO}_2$ data. We examined the results using the satellite-based SST (not shown) and found comparable results, albeit with slightly inferior model-data deviation statistics. This is not too surprising, given the reasonably good agreement between

in situ and remotely-sensed SST. Salinity has promise as a predictive input, especially in coastal areas where many water masses may be present. However, we did not see significant improvement in our predictive algorithms using reported in situ salinity, and while remote-sensing salinity products are now available from Aquarius (Lagerloef et al., 2008, and SMOS, Mecklenburg et al., 2008), they will have the spatial, temporal, and salinity resolution to distinguish only the largest features, such as river plumes.

We used recent data collected on a 2007 cruise off the North American Pacific coast as a qualitative validation of the predictive algorithm (Feely et al., 2008). The pCO₂ data presented here were analyzed using a flow-through membrane contactor interfaced with a tangential-flow filter. This system has been described elsewhere (Evans et al., 2011), and is a modification of earlier membrane-contactor based systems (Hales et al., 2004) adapted to facilitate surface-underway mapping with reduced user maintenance while retaining the fast response times of the membrane contactor.

2.2. Remote sensing products

We used a variety of satellite products. For the SOMs (see below) we used annual climatologies at 0.25° spatial resolution for MODIS SST and Chl (July 2002 to December 2006), and QuikSCAT wind stress (July 1999 to December 2006). Coincident match-up with the pCO₂ observational data used in the development of the predictive algorithms were derived by averaging the 9 (3 × 3) nearest 9 km pixels from 8-day averaged SeaWiFS Chl bracketing each observation. For creation of the predicted maps of pCO₂ distributions and air–sea CO₂ flux, we used the monthly average SeaWiFS Chl, MODIS SST, and QuikSCAT wind data at 0.25° spatial resolution. It warrants repeating that the Chl–pCO₂ data match-up is based on remote-sensing data, while the SST–pCO₂ data match-up is based on in situ observations.

3. Calculations

3.1. Self organizing maps

We followed the approach of Saraceno et al. (2006). Briefly, this method relies on an artificial neuronal network to identify biogeochemical regions within the target study area. The method comprises a probabilistic version of the Kohonen (1990, 1995) self-organizing map (SOM) and Hierarchical Ascending Clustering (HAC) algorithms; for brevity we will refer to these as SOM only. The approach is described in detail by Telszewski et al. (2009). Each input parameter was transformed before being input to the SOM. The Chl-a values were initially log-transformed and all three parameters (SST, Chl-a, and wind stress) were mapped to the common range of –1 to 1, with the two extrema corresponding to the raw extrema of each parameter. In the simplest terms, the SOM approach clusters pixels with similar properties and separates them from dissimilar clusters. The scoring function is defined by the interclass inertia, which can be thought of as the ratio of the dissimilarity between clusters to the dissimilarity within clusters. The SOM approach is given a maximum number of regions as a stopping point, but operationally selects the number of regions found when the interclass inertia has dropped to <10% of that seen with the fewest number of regions as the best representation of sub-regional distinctions.

3.2. Predictive models

3.2.1. Multiple linear regression model

We examined two kinds of predictive algorithms. The first was a multiple linear regression (MLR), i.e.

$$p\text{CO}_{2,m\text{lr}} = C_0 + \sum_{i=1}^n C_i p_i \quad (1)$$

where pCO_{2,m\text{lr}}} is the predicted pCO₂. C₀ and C_i are the coefficients, and the p_i terms are the independent variables, in the MLR. In our case we considered the deviation in latitude, longitude from the region's center (Δ*lat* and Δ*lon*, respectively), time of year, SST, and chlorophyll as independent variables. Time of year was sine-transformed, and allowed for an optimization of the seasonal phasing, *e*:

$$t \equiv \sin\left(\frac{2\pi\text{day}}{365}\right) + e \quad (2)$$

where 'day' was the decimal day of year.

3.2.2. Mechanistic nonlinear model

As shown in Section 4.2, we quickly found the MLR approach to be inadequate for the high dynamic range of the observed pCO₂, even after applying the SOM to the study area, and built a new meta-model that included the inherent nonlinearities of the aqueous carbonate system. We expect this model to be applicable only to the SOM-defined sub-regions, and as a result did not attempt to apply it uniformly to the entire study area. Knowing that the pCO₂ is a quantitative nonlinear function of the alkalinity (*T*_{Alk}) and total CO₂ (*T*_{CO₂}), we chose to develop a model that retained this functionality. Conceptually, we assumed that the parameters *T*_{Alk} and *T*_{CO₂} were approximated by some initial values plus perturbation terms

$$T_{\text{CO}_2} = T_{\text{CO}_2}^0 + \Delta T_{\text{CO}_2} \quad (3)$$

$$T_{\text{Alk}} = T_{\text{Alk}}^0 + \Delta T_{\text{Alk}} \quad (4)$$

where the perturbations were simply related to the processes of mixing and biological productivity. These processes were subsequently linked to observed chlorophyll and temperature distributions. Specifically, we assumed first-order Taylor-series approximations:

$$\Delta T_{\text{CO}_2} = \left. \frac{\partial T_{\text{CO}_2}}{\partial T} \Delta T \right]_{\text{mixing}} + \left. \frac{\partial T_{\text{CO}_2}}{\partial \text{Chl}} \Delta \text{Chl} \right]_{\text{biology}} \quad (5)$$

where mixing is reflected in a temperature change term Δ*T*, and biological processes in a chlorophyll change term Δ*Chl*. These two terms are each defined by:

$$\Delta T]_{\text{mixing}} = \phi(T - T^0) \quad (6)$$

$$\Delta \text{Chl}]_{\text{biology}} = (\text{Chl} - \text{Chl}^0) \quad (7)$$

where *T*⁰ and Chl⁰ are empirical initial temperature and chlorophyll values associated with the initial *T*_{CO₂} defined above, and φ is an empirical term that quantifies the proportion of the temperature change due to mixing as opposed to thermal forcing. The *T* and Chl variables are the observed (either in situ or remotely sensed) values of these parameters.

If we assume that the local mixing gradient and stoichiometric relationship between chlorophyll and *T*_{CO₂} can be treated as empirical terms as well (γ and β, respectively), and that the initial chlorophyll is zero (reflecting recently upwelled water), the *T*_{CO₂} approximation becomes:

$$T_{\text{CO}_2} \cong T_{\text{CO}_2}^0 + \gamma\phi(T - T^0) + \beta\text{Chl}. \quad (8)$$

The terms *T*_{CO₂}⁰, *T*_{Alk}⁰, *T*⁰, γ, φ, and β in the above equation are each empirically-determined coefficients. The term β is given additional functionality in that it is set to zero for temperatures above a certain value. This is shaped by the concept of upwelling, in which water newly introduced to the surface is cold and nutrient-rich, and changes in chlorophyll abundance are due primarily to net

photosynthetic production. Waters that have been at the surface for some time are more likely to be in a state of stationary, highly recycled growth, and the dependence of T_{CO_2} on chlorophyll would be much weaker in those cases. We chose this ‘cut-off’ temperature based on the predicted values of the T^0 parameter within each region; specifically, chlorophyll dependence was allowed for all temperatures $\leq T^0 + 10$ °C. This is a somewhat arbitrary choice, but as discussed later, other complications in understanding the chlorophyll dependences are such that more detail is not warranted.

If we further allow each coefficient $T_{\text{CO}_2}^0$, T_{Alk}^0 , T^0 , γ , ϕ , β to have dependence on space and time, e.g.,

$$T_{\text{CO}_2}^0 = c_0 + c_1 \Delta lat + c_2 \Delta lon + c_3 t \quad (9)$$

where c_i are empirical coefficients and Δlat , Δlon , and t are as defined previously. We assume that the seasonal phasing in the t term (e in Eq. (2)) is fixed for all terms in a given region. Finally, since alkalinity is affected by change in NO_3^- , which is, in turn, related to T_{CO_2} via the Redfield N/C ratio of 16/106 (=0.15), we assume that perturbations to alkalinity are directly proportional to the changes in T_{CO_2} , i.e.,

$$\Delta T_{\text{Alk}} = -0.15 \Delta T_{\text{CO}_2}. \quad (10)$$

We have not included factors that may change the $T_{\text{CO}_2}:T_{\text{Alk}}$ stoichiometry from that implied by Redfield C:N stoichiometry. The study site is not an area with large populations of calcifying organisms (Hauri et al., 2009), and we expect that variations in alkalinity production with respect to subtleties in the C:N stoichiometry will be small relative to the other factors. Other factors that will influence the $T_{\text{CO}_2}:T_{\text{Alk}}$ ratio such as terrestrial inputs and variations in source-water characteristics are only captured empirically in the formulation of Eq. (9).

The model thus predicts intermediate products for T_{Alk} and T_{CO_2} , which are then used to calculate pCO_2 at the in situ temperature. For calculating pCO_2 from T_{Alk} and T_{CO_2} , we used an inorganic equilibrium chemistry model for T_{Alk} that included carbonic and boric acid but ignored minor species, with the solubility of CO_2 in seawater by Weiss (1974), the apparent dissociation constants by Mehrbach et al. (1973) for carbonic acid as adjusted to the seawater pH scale by Dickson and Millero (1987) and by Dickson (1990) for boric acid. Thus, while T_{Alk} and T_{CO_2} are predicted by the model, they are not used directly in the evaluation procedure. The calculated pCO_2 is the ultimate model output.

The model has a couple of deficiencies. First is the possibility that there are as many as 25 optimizeable coefficients in the two equations for T_{CO_2} and T_{Alk} that derive from Eqs. (2)–(10). We needed to make some *a priori* limitation of this number, described in Section 4.3, to limit the maximum possible number of coefficients. The second is that the model contains products of some of the terms, which implies the possibility of interdependence and a reduced set of truly optimizeable coefficients. This is addressed in the Appendix.

3.3. Optimization

We chose to use Powell’s method for optimization in multiple dimensions, as described in Press et al. (1989). This method has the advantage in that it does not depend on the form of the equation being optimized, which is essential given the potentially variable forms as empirical terms are included or excluded, and because of the non-linearity of the step in which pCO_2 is calculated from the intermediate Alk and T_{CO_2} products. In every optimization exercise, all 25 empirical constants were explicitly included, even if set to zero, but the optimization approach allowed selection of only reduced sets of coefficients to optimize. For example, we could specify that the chlorophyll dependence β is non-zero but remains

fixed at the nominal value of -7 (e.g. reflecting a 1:1 N:Chl ratio and Redfield C:N stoichiometry).

Our diagnostic of algorithm performance was the root-mean-squared (RMS) deviation between the observed and predicted pCO_2 , over the complete set of the observations within each region. We examined the sensitivity of the optimized results in two ways. First, we tried optimizations with varied values of the initial guesses of the optimized parameters. This included using the optimum set of parameters as initial guesses and restarting the optimization procedure. This verified the stability of the optimum solution. Second, we calculated coefficient-specific sensitivity factors (SF_i) defined by:

$$\text{SF}_i = \left(\frac{1}{\text{RMS}} \frac{\partial \text{RMS}}{\partial C_i} \right) / \left(\frac{1}{C_i} \frac{\partial C_i}{\partial C_i} \right) \quad (11)$$

which effectively describes the relative change in RMS for a relative change in coefficient c_i . Coefficients with $\text{SF} \leq 0.1$ – i.e., the optimum RMS value changed by no more than 1% for a 10% change in coefficient—were deemed insignificant.

3.4. Sensitivity analysis

Training an algorithm to reproduce a set of observations requires additional verification and analysis of the sensitivity of the model parameters to the dataset. The complexities of the model obviated some of the more standard tools that could be easily applied to smaller data sets and linear systems, so we chose to take a more rudimentary approach. After finding the optimum set of model parameters for a given region, we perform ten additional simulations to examine the robustness of the result. Each of these simulations consisted of the same analysis as described above, but with 10% of a region’s data randomly extracted from the total. The remaining 90% were used as training data to repeat the optimization procedure. The optimal model constrained by the training data was then applied to the extracted 10% that had not been used in constraining the optima.

3.5. Air–sea flux calculations

Once we mapped the best set of predicted pCO_2 distributions, we calculated air–sea CO_2 fluxes. We did this by calculating a gas exchange coefficient based on the quadratic wind-speed dependence of Ho et al. (2006) and the QuikSCAT monthly wind speed-squared data from the Scatterometer Climatology of Ocean Winds (SCOWs) available at <http://cioss.coas.oregonstate.edu/scow/>, based on the methods in Risien and Chelton (2008). We multiplied this coefficient by the model-predicted air–sea pCO_2 difference and the gas solubility of Weiss (1974) for satellite SST and a fixed salinity.

4. Results

4.1. Linear models

We began our analysis with an attempt to fit the pCO_2 observations in the Chavez et al. (2007) synthesis data set, limited as described above to within 370 km of the coasts and with coincident SeaWiFS Chl data, with a multiple linear dependence in the form of Eq. (1). This was without success (Fig. 1), and we recognized the need to distinguish the major coastal regions (Fig. 2). The MLR approach applied to these regions yielded better results, particularly for the Gulf of Mexico/Caribbean waters (Fig. 2, lower right) where our regression yielded statistics similar to the prediction presented by Wanninkhof et al. (2006). The pCO_2 distributions in Pacific coastal waters, however, were nearly as

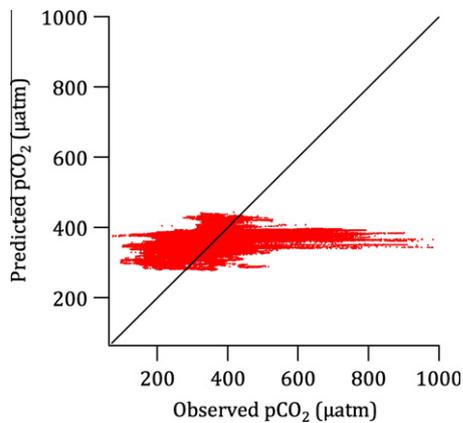


Fig. 1. Multiple linear regression representation of all of the $p\text{CO}_2$ observations in the Chavez et al. (2007) North American Continental Margins dataset. Solid diagonal line represents the perfect-agreement 1:1 relationship.

poorly reproduced by this attempt as were those from the entire North American coastline. The essence of the failure was first that the large dynamic range of the observed $p\text{CO}_2$ was not in any way reproduced by the predictions, and that there seemed to be little meaningful dependence on any of the independent input variables. This result suggested that the Pacific coast contained sub-regions that were not well described by a single set of empirical parameters, and that the inherently linear aspect of a MLR was insufficient for reproducing $p\text{CO}_2$ distributions with large dynamic ranges. Because the Pacific coastal waters presented the greatest challenge, the rest of our efforts were directed at developing predictive algorithms for this region.

4.2. Determination of regions

Fig. 3 shows the climatological distributions of satellite SST, Chl, and wind stress over Pacific coastal waters as observed by these three sensors. These data were used as primary inputs to the self-organizing map determination of Saraceno et al. (2006) to define the regions shown in Fig. 4. The approach found 13 sub-regions before no significant improvement (as defined earlier and in Saraceno et al., 2006) was gained by adding more regions. These

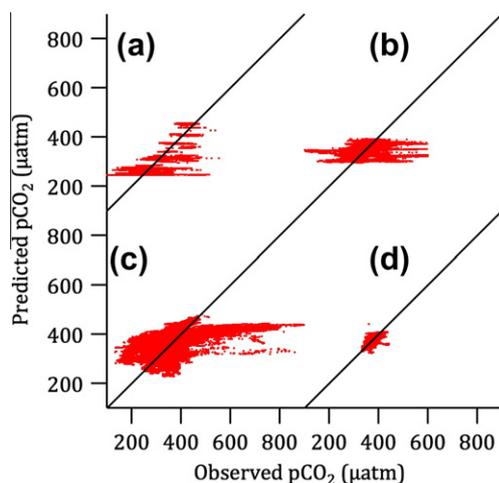


Fig. 2. Comparison of multiple linear regression prediction of sea-surface $p\text{CO}_2$ in the four major coastal sub-regions of North America: (a) Bering Sea and Arctic coastal waters; (b) the Atlantic coast; (c) the Pacific coast; (d) Gulf of Mexico and Caribbean coastal waters. Solid diagonal lines represent the perfect-agreement 1:1 relationship.

objectively-determined sub-regions confirmed our thinking about the distinctions of coastal areas of the Pacific coast and precisely define important biophysical regions within this study area (Table 1). Such an objective sub-division has not been done before in this region, as far as we know. At the global scale, Longhurst (2006) defined bio-geographical regions based on in situ and remote sensing data. In the East Pacific, he defined only two regions that overlap with our results: the California Current and Central America Coastal provinces. Spalding et al. (2007) defined marine ecoregions of the world along coastal and shelf areas. In the area considered in this work they found six regions. In latitude, their divisions approximately coincide with ours. However they did not find cross-shore distinction of regions as we found. These divisions separate coastal upwelled waters from offshore waters remarkably well.

Once the sub-regions were defined and observations assigned to each sub-region (Table 1), we developed distinct algorithm parameters for each region. Our first attempt was to simply apply the multiple linear regression of Eq. (1) to the subsets of data; the results are summarized in Fig. 5 and Table 2.

This result is significantly improved over that applied to the Pacific coastal data as a whole. On a numerically-averaged basis, the RMS deviation between model and observations dropped to 57.8 from the initial 67 μatm . Weighting by the areas of each region, the RMS deviation dropped to 24.5 μatm . Regions 1, 2, 4 and 13 were particularly well-described by this approach, with RMS deviations $<15 \mu\text{atm}$ and correlation coefficients of ~ 0.9 . However improved the results, there were still significant shortcomings. Particularly in regions with large data densities and large dynamic ranges, such as the coastal upwelling areas from central California northward and the regions immediately offshore of them, the familiar pattern of the model inadequately reproducing the dynamic range of the data was evident. This suggests that the large dynamic ranges of the data do not correspond with large ranges in the input parameters, and argues for the application of a higher-order model. In addition, the coefficients of the simple linear model demonstrate odd behavior (Table 2). Region 7, for example, shows a large negative constant term that is compensated for by an exceedingly high temperature dependence term. Furthermore, chlorophyll dependences are insignificant in 9 of the 12 cases. In the remaining three (Regions 1, 10, 13), sensitivity to chlorophyll is barely significant by our $SF < 0.1$ criterion ($SF \leq 0.15$), and the coefficients show large variability in the sign and magnitude of the dependences.

4.3. Mechanistic model

The ocean carbonate system has well known nonlinearities, especially with regard to $p\text{CO}_2$. The dependence of $p\text{CO}_2$ on water temperature is exponential, given established thermodynamics of gas solubility (Takahashi et al., 1993). In-water processes that affect T_{CO_2} and T_{Alk} impact $p\text{CO}_2$ disproportionately; this effect is known as the Revelle Factor (Revelle and Suess, 1957) and suggests relative changes in $p\text{CO}_2$ are ~ 10 th power dependent on relative T_{CO_2} changes for normal oceanic conditions. We experimented with functionality that included a Revelle-factor dependence on an empirically determined T_{CO_2} and incorporated an exponential temperature dependence following Takahashi et al. (1993). The fundamental nature of these relationships, however, is to predict relative changes, and thus the empirical coefficients were not well constrained by the optimization procedure. Empirical nonlinearities are easy to assign—we could simply employ polynomial dependences on the input parameters of the MLR representation, but doing this without mechanistic justification is questionable. As a result, we chose not to pursue any of these approaches and to build

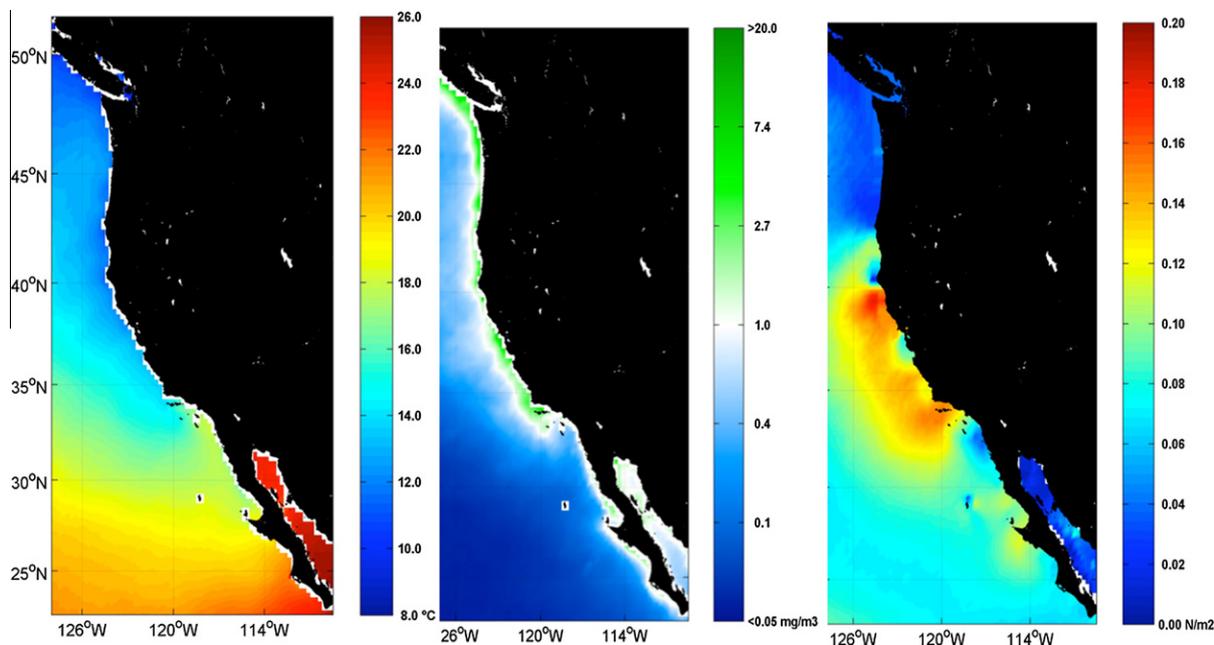


Fig. 3. Input data used for generating the self-organizing map (SOM) of Pacific coastal waters. Left, Sea-Surface Temperature (SST); middle, sea surface chlorophyll (Chl); right, net southward wind stress.

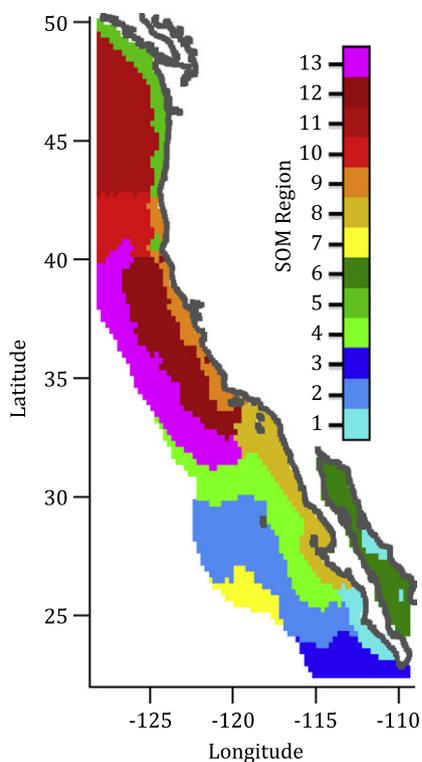


Fig. 4. Self-organizing map of biogeochemical regions in the study area.

the quasi-mechanistic model as described in Section 3.2.2, Eqs. (3)–(10).

As stated previously, the model could potentially include a very large number of optimizable coefficients. We felt that it was reasonable to restrict this number in the following ways. The ratio of T_{CO_2} to T_{Alk} is the leading factor in determining the pCO_2 distributions in the aqueous carbonate system. We opted to allow the

$T_{\text{CO}_2}^0$ term in Eq. (3) to retain the full time and space variability as shown in Eq. (8). Assuming this would allow sufficient variability in the ratio of $T_{\text{CO}_2}^0 : T_{\text{Alk}}^0$, we chose to make T_{Alk}^0 in each region be determined only by the corresponding constant term in Eq. (8). This may have the effect of amplifying the suggested spatio-temporal T_{CO_2} variability, but T_{Alk} is generally the more conservative parameter and this was viewed as a reasonable trade-off for the reduction in empirical coefficients. We further opted to allow the terms ϕ , γ , β , and T^0 to have only temporal variability. As stated earlier, the temporal phasing for all parameters within a given SOM region was assumed to be the same. The empirical model thus depended on a maximum of 14 optimizable coefficients.

This is still a large number of coefficients, and there are two different approaches that could be taken to identify the most important terms. First is a stepwise approach in which terms are sequentially added and model improvement or lack thereof is assessed at each step. Second is the approach we followed, in which we allowed all 14 coefficients to vary initially, and then assessed the sensitivity of the result to each one.

Results are shown in Fig. 6 and summarized in Tables 3 and 4. The immediate first impression is of the improvement in the predictability of the pCO_2 in the problem regions identified in the MLR analysis. Region 9, in particular, the persistent upwelling region off central California that includes the large dataset from the MBARI time-series, has significantly improved predictability (R increased to 0.75 from 0.15, a 25-fold increase in the percentage of variance explained) as a result of including the nonlinear mechanistic representation. Other regions (5, 8, 10, and 11) that include large dynamic-range pCO_2 observations are also better predicted (R values 0.82 vs 0.57, 0.80 vs 0.60, 0.92 vs 0.79, 0.61 vs 0.07, respectively). Regions can be broadly categorized into three groups—those with high (≥ 0.75) correlation coefficient and low ($\leq 20 \mu\text{atm}$) RMS (Regions 1–4, 8, 10, 13), those with high correlation coefficient and high RMS (Regions 5, 9) and those with high RMS and low correlation coefficient (7, 11, 12). The second group includes the most intense temperate coastal upwelling conditions, where the dynamics of source waters and terrestrial inputs may not be

Table 1
Description of sub-regions.

Region	Area (km ²)	Number of pCO ₂ observations	Description
1	43,924	684	Upwelling-influenced southern Baja and Sea of Cortez
2	261,489	1093	3rd offshore from Baja
3	94,406	959	Southern extreme
4	212,661	2764	2nd offshore from Baja
5	46,037	2139	Cascadia coastal waters; seasonal upwelling
6	106,421	0	Sea of Cortez, non-upwelling-influenced
7	182,506	105	Baja, most offshore
8	113,618	3106	Northern Baja coastal, upwelling
9	55,240	48,514	California coastal, persistent upwelling
10	75,105	861	Northern California offshore
11	169,190	3632	Cascadia offshore
12	175,591	22,542	Central California, 1st offshore
13	224,407	9801	Central California, 2nd offshore

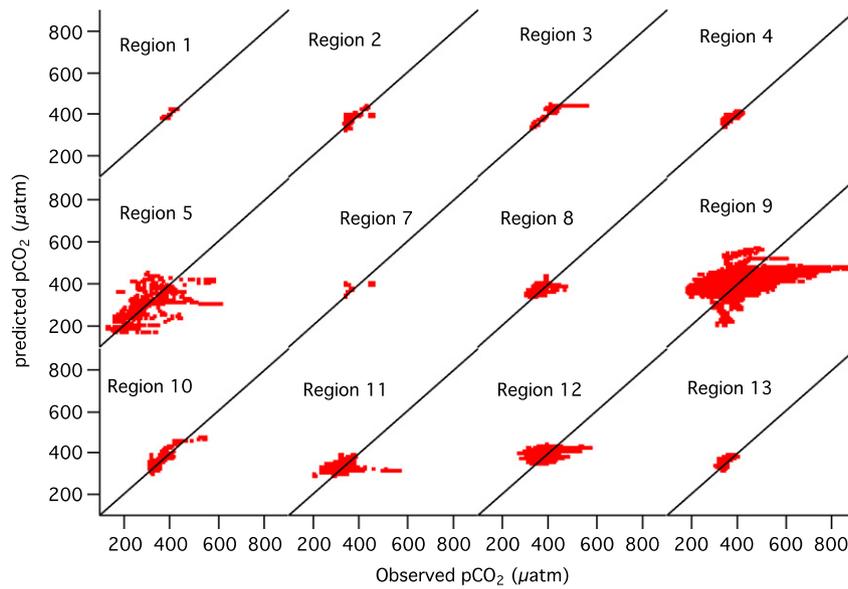


Fig. 5. Region-by-region application of the MLR to the pCO₂ data in each SOM region. There were no observations in Region 6. Results are summarized in Table 2. Solid diagonal lines represent the perfect-agreement 1:1 relationship.

Table 2
Results of multiple-linear regression^a approach applied to data in each SOM region.

Region	C ₀	C ₁	C ₂	C ₃	C ₄	C ₅	e	RMS	R
1	350.5	-14.73	2.80	-75.7	2.92	16.5	0.11	6.1	0.94
2	309.9	3.32	2.97	-30.1	3.38	0.8	1.26	14.0	0.87
3	332.8	14.14	8.88	-24.7	2.95	19.2	1.60	19.8	0.79
4	360.9	-4.71	-0.90	-22.6	0.48	0.6	1.05	8.2	0.93
5	386.5	-25.19	-28.24	-68.3	1.29	-4.8	-0.73	63.4	0.57
6	NSD								
7	-200.2	-4.53	16.05	-90.8	26.85	-53.8	-2.40	25.9	0.58
8	324.4	1.55	14.33	8.6	2.30	-11.2	-0.05	22.3	0.60
9	511.5	-14.7	5.48	37.9	-7.24	0.4	-0.52	88.6	0.15
10	395.0	13.87	13.45	-24.1	-2.13	82.9	2.64	31.8	0.79
11	367.0	-9.47	-16.20	27.2	-5.72	-0.9	3.18	44.8	0.07
12	560.9	8.49	12.77	29.9	-8.79	-14.8	3.74	31.9	0.24
13	377.4	-3.82	-7.96	31.7	-3.64	-66.2	3.65	10.5	0.88
Number-weighted composite statistics:								57.8	0.32
Area-weighted composite statistics:								24.5	0.51

^a Predictive function followed the form of Eq. (1), specifically, $pCO_{2,p} = C_0 + C_1\Delta lat + C_2\Delta lon + C_3t + C_4T + C_5Chl$ where t is as defined in Eq. (2), and Δlat and Δlon are the deviations of the latitude and longitude from each respective mid-point of the region. Coefficients to which the prediction was significantly sensitive ($SF > 0.1$) are highlighted in red text.

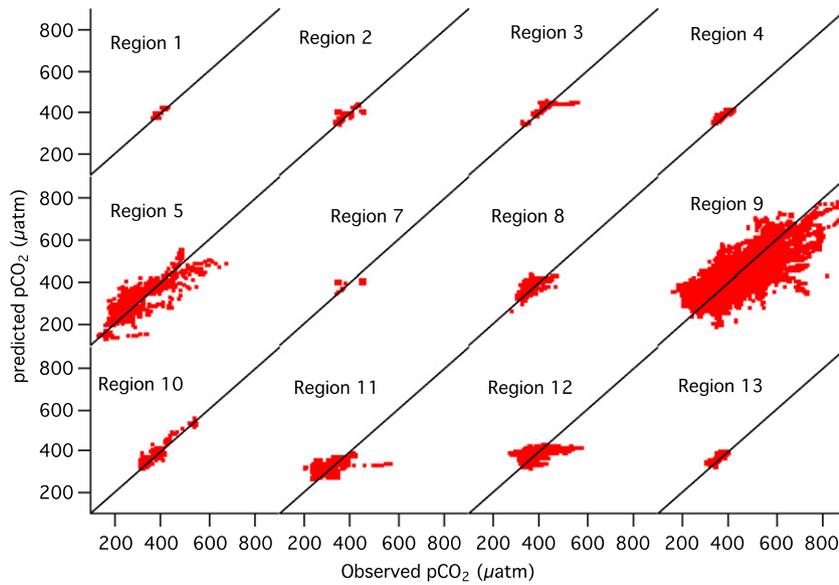


Fig. 6. Region by region application of the nonlinear mechanistic model to the pCO₂ observations in each SOM region. Results are summarized in Table 3. Solid diagonal lines represent the perfect-agreement 1:1 relationsh.

Table 3
Results of the mechanistic optimization applied to the regions identified by the self-organizing map. Coefficients highlighted in red text are those with SF > 0.1. Italicized rows show the combinations of coefficients applied to the regions to generate the pCO₂ and air-sea CO₂ flux maps in Figures.

Region	Alk ⁰ c ₀	c ₀	T _{CO₂⁰}			T ⁰		φ		γ		β		e	RMS	R
			c ₁	c ₂	c ₄	c ₀	c ₄									
1. Full	2309	2202	-5.44	3.37	-43.1	9.86	-1.52	0.82	-0.03	-28.9	-2.2	62.3	-83.3	-0.05	6.3	0.93
Spec.	2270	2164	-6.79	0	44.3	8.75	0.14	0.59	-0.29	-31.5	-2.94	62.9	71.5	0.29	5.8	0.94
Min.	2332	2188	0	0	26.9	10.02	1.01	0.85	-0.10	-34.5	-3.8	0	0	0.99	6.6	0.93
2. Full	2306	2286	-3.17	0.84	57.4	2.12	-6.08	0.70	0.05	-32.7	-9.1	-66.2	242	0.38	12.4	0.90
Spec.	2190	2181	-4.35	0	69.6	2.04	-6.25	0.62	0.06	-24.2	-8.1	-55.3	183	0.52	11.3	0.92
Min.	2298	2191	0	0	-88.9	4.35	12.34	0.64	-0.02	-16.1	0.1	0	0	0.95	12.9	0.90
3. Full	2190	2184	4.53	6.63	-15.5	3.73	-5.13	0.67	0.09	-28.1	-1.5	26.7	-127	0.32	14.6	0.88
Spec.	2340	2238	0	0	-119	7.90	-1.30	0.72	0.23	-26.8	-0.9	0	-0.9	0.99	15.3	0.86
Min.	2228	2148	0	0	-92.5	6.41	-2.88	0.72	0.17	-24.8	1.9	0	0	0.76	15.3	0.86
4. Full	2266	2204	-2.30	-0.22	7.4	1.90	-2.99	0.65	-0.03	-20.4	2.0	-8.6	89.4	0.40	6.8	0.95
Spec.	2291	2224	-2.48	-0.39	-17.2	1.44	-2.92	0.74	0.05	-27.2	-0.2	0	167	0.65	7.1	0.94
Min.	2363	2227	0	0	-8.1	7.80	-0.74	0.79	-0.21	-28.0	1.2	0	0	1.03	9.9	0.88
5. Full	2270	2203	-2.96	-1.93	10.0	8.39	-1.23	0.32	-0.16	-39.8	-8.5	-4.7	2.9	0.23	47.6	0.82
Spec.	2265	2202	0	0	0	8.06	-0.81	0.05	0	-29.8	-11.2	-4.7	0	0.31	49.3	0.81
Min.	2274	2191	0	0	8.5	8.23	-0.82	0.23	-0.19	-33.6	-7.4	0	0	0.20	55.5	0.74
7. Full	2383	2187	1.84	0.54	41.9	12.8	-3.56	0.63	-0.20	-10.9	-19.5	-11.0	532	1.27	25.9	0.63
Spec.	2381	2202	0	0	35.3	9.87	-5.64	0.85	0	-27.7	-30.3	0	393	1.25	26.0	0.62
Min.	2439	2248	0	0	-52.7	6.57	22.0	0.90	-0.16	-31.7	28.0	0	0	1.49	26.3	0.61
8. Full	2312	2188	-3.57	-1.07	-35.3	8.35	-4.40	0.71	-0.02	-23.1	20.6	-6.8	14.6	0.88	16.8	0.80
Spec.	2313	2191	-1.86	0.52	-36.8	6.27	2.85	0.60	0	-13.7	2.9	0	0	1.75	20.4	0.68
Min.	2313	2192	0	0	-38.6	6.08	-3.87	0.77	-0.01	-28.5	25.7	0	0	0.90	22.7	0.58
9. Full	2231	2214	-2.91	5.66	71.7	7.50	-1.28	0.47	-0.08	-37.9	-11.9	-0.8	1.1	-0.15	62.9	0.75
Spec.	2215	2206	0	0	-121	6.89	1.73	0.04	0	-19.2	13.3	0	0	2.73	65.0	0.73
Min.	2173	2174	0	0	-125	6.57	1.79	0.22	-0.01	-23.4	16.5	0	0	2.75	65.0	0.73
10. Full	2246	2192	5.94	1.98	1.7	9.56	2.29	0.47	-0.13	-38.7	-6.9	-132	-144	0.22	20.3	0.92
Spec.	2179	2209	0	0	0	6.30	2.93	0.45	-0.23	-43.3	-2.6	-92.6	-92.0	0.31	22.4	0.90
Min.	2287	2197	0	0	-13.4	7.49	-0.69	0.60	-0.20	-21.3	-0.7	0	0	0.75	25.2	0.87
11. Full	2315	2214	-2.81	-6.61	-17.2	6.83	-0.14	0.59	-0.20	-35.6	13.1	7.9	17.8	0.41	33.4	0.61
Spec.	2287	2229	0	0	0	4.73	0	0.62	-0.18	-40.2	8.6	6.4	20.3	0.3	35.2	0.56
Min.	2347	2189	0	0	44.8	9.43	-0.95	0.67	-0.11	-30.5	-12.1	0	0	-0.60	35.8	0.53
12. Full	2305	2176	2.80	4.79	36.7	11.55	-3.80	0.68	-0.20	-41.1	18.2	-3.5	3.8	0.48	24.1	0.67
Min.	2283	2181	0	0	-0.22	8.33	0.52	0.67	-0.09	-31.2	-0.17	0	0	0.68	26.3	0.59
13. Full	2365	2211	-0.95	-1.32	-14.3	7.50	3.42	0.78	-0.03	-19.3	-2.6	-15.5	14.9	0.92	7.8	0.94
Spec.	2369	2206	0	0	5.6	8.15	2.39	0.88	-0.04	-26.7	-9.7	-13.3	0	1.57	8.2	0.93
Min.	2362	2198	0	0	4.9	7.53	2.42	0.87	-0.05	-25.8	-7.5	0	0	1.16	8.8	0.91

Table 4
Composite statistics for the cases described in Table 3.

Optimized coefficient set	Number-weighted		Area-weighted	
	RMS	R	RMS	R
Full	41.7	0.76	19.0	0.82
Region-specific	43.6	0.72	20.1	0.79
Minimum	53.7	0.59	21.8	0.75
Applied	43.0	0.75	19.8	0.81

well characterized by the simple model. The third group includes two regions (11, 12) that sit immediately off the upwelling regions 5 and 9. Regions 11 and 12 are primarily offshore but include data-containing pixels that impinge on the coastline, suggesting that the resolution of the SOM may be a factor. The remaining region, 7, sits far offshore in the SW portion of the study area, and data there consists of two clusters that are separated by $\sim 200 \mu\text{atm}$. Other regions in this area (1–4) do not show the same character and we are uncertain of the reason for the poor model performance there.

Sensitivity analysis also shows some interesting results. First, none of the regions required all 14 coefficients to be optimized. Region 1 showed sensitivity to a maximum of 13 coefficients, and Region 9 showed strong sensitivity to only 9 of the tested coefficients. Second, the spatial dependence terms were only rarely significant in the optimizations. Only Region 8 showed significant sensitivity to both latitude and longitude dependence terms, and Regions 1, 2, and 4 were sensitive only to the latitude term. The remaining 8 regions showed no significant sensitivity to the spatial dependence terms. This is satisfying, because one of the objectives of the SOM approach is to cluster data within regions where biogeochemical and hydrographic relationships are consistent, and thus lessen non-mechanistic empirical spatial dependences.

Finally, the sensitivities to chlorophyll, while improved over those seen in the MLR approach, are still puzzling. Regions 8, 9, and 12 show no sensitivity to chlorophyll at all (neither c_0 nor c_4 in the β term were significant; Table 3), while Regions 3–6 and 13 show sensitivity to only one of the two coefficients. Further, the values of the coefficients themselves are puzzling. The magnitudes of β implied by the optimized coefficients are often far from the nominal 'Redfield' expectation of ~ -7 . Equally puzzling in the conceptual context of the model is the variation in the sign of the first-order chlorophyll dependence. Region 7, for example, has coefficients that suggest that β changes from < -500 to > 500 ($\mu\text{mol kg}^{-1}$)/($\mu\text{g l}^{-1}$) over the course of the season. As discussed later, this dependence is hard to justify within the context of the model, and probably indicates the need for a better representation.

We examined two additional subsets of optimizable parameters. The first was the set identified as having sensitivity by the full analysis. In 9 of 12 cases, the optimization statistics were essentially equivalent between these reduced sets of coefficients and the initial full set, with Regions 8, 11, and 12 being the exceptions. The second reduced set of coefficients was one applied to all regions that included no spatial dependence in the $T_{\text{CO}_2}^0$ term, and no chlorophyll dependence. We justified this second choice because so few regions showed spatial dependence, and because the chlorophyll dependences were only significant in half of the regions. This showed only slightly worse performance than the sets of coefficients selected specifically for each region, with Regions 4, 5, and 8 showing significantly worse predictive power with these subsets.

We chose sets of coefficients for use in subsequent analyses based on a combination of the statistics of the optimization and the most conservative combinations of coefficients. When optimization statistics were comparable, for a given region with different sets of coefficients, we selected the result with the lowest number of optimized coefficients. These choices are italicized in Table 3,

and are the values used in Fig. 6 and subsequent pCO_2 and air-sea flux reconstructions (Figs. 8–10).

Composite statistics for the entire study region are presented in Table 4. While the RMS deviation statistics are improved over the SOM/MLR case (43 vs $58 \mu\text{atm}$ on a number-weighted basis and 20 vs $25 \mu\text{atm}$ on an area-weighted basis), the improvement is shown most strongly for the correlation coefficient statistic (0.76 vs 0.32 and 0.81 vs 0.51 on number and area-weighted bases, respectively). As R^2 is often thought of as the fraction of true variance explained by a predictive relationship, the mechanistic model thus explains 2.5 or 5.6 times more of the natural variability than the MLR, based on areal or number weighting, respectively. The areally-weighted mechanistic model applied to the SOM region describes about 66% of the observed variability, and reproduces the observations to within about $20 \mu\text{atm}$. The robustness of this result is addressed in the following discussion.

Although we performed sensitivity analyses for each region, we present only the results of this exercise for Region 9. Region 9 contained the most observations, and difficulty in representing that data had presented us with one of the greatest motivations for developing the detailed model. Results are detailed in Table 5, where we examine the sensitivities of the model coefficients for ten randomly selected subsets of 90% of the total observations in this region. Little difference is seen in the values of the model coefficients either within these ten simulations of between any

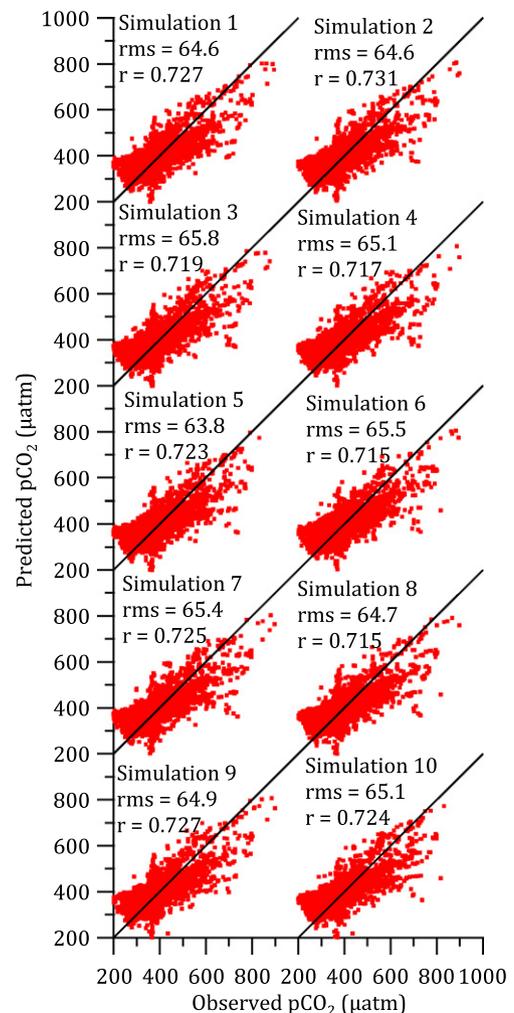


Fig. 7. Demonstration of predictive capability of the algorithm, for ten randomly-selected groups of Region 9 data not included in algorithm training. See Table 5 for details.

Table 5
Sensitivity analysis for Region 9.

Results with training data												Verification data			
<i>n</i>	Alk ⁰		T _{CO₂} ⁰	T ⁰		φ		γ		<i>e</i>	RMS	<i>R</i>	<i>n</i>	RMS	<i>R</i>
	<i>c</i> ₀	<i>c</i> ₀	<i>c</i> ₄	<i>c</i> ₀	<i>c</i> ₄	<i>c</i> ₀	<i>c</i> ₄	<i>c</i> ₀	<i>c</i> ₄						
43,599	2174.4	2174.1	-123.5	6.570	1.782	0.228	-0.008	-23.45	16.31	2.731	65.09	0.727	4915	64.64	0.727
43,655	2173.7	2174.1	-123.8	6.571	1.799	0.226	-0.010	-23.49	16.37	2.735	65.10	0.727	4879	64.56	0.731
43,616	2173.8	2174.2	-123.7	6.568	1.775	0.225	-0.009	-23.46	16.37	2.733	64.96	0.728	4898	65.83	0.719
43,691	2174.4	2174.1	-123.3	6.567	1.788	0.225	-0.010	-23.48	16.27	2.732	65.04	0.729	4823	65.11	0.717
43,794	2173.7	2174.1	-123.7	6.563	1.778	0.226	-0.008	-23.45	16.32	2.738	65.18	0.728	4720	63.77	0.723
43,711	2173.8	2174.2	-123.3	6.570	1.797	0.226	-0.010	-23.52	16.29	2.732	65.00	0.729	4803	65.45	0.715
43,689	2173.5	2174.1	-123.9	6.568	1.813	0.227	-0.009	-23.52	16.24	2.743	65.01	0.727	4825	65.38	0.725
43,503	2173.9	2174.1	-124.1	6.569	1.780	0.225	-0.007	-23.35	16.28	2.730	65.08	0.728	5011	64.74	0.715
43,648	2173.7	2174.1	-122.9	6.570	1.798	0.224	-0.010	-23.48	16.15	2.741	65.07	0.727	4866	64.89	0.727
43,785	2173.5	2174.1	-123.8	6.576	1.789	0.226	-0.010	-23.47	16.43	2.736	65.03	0.728	4729	65.14	0.724

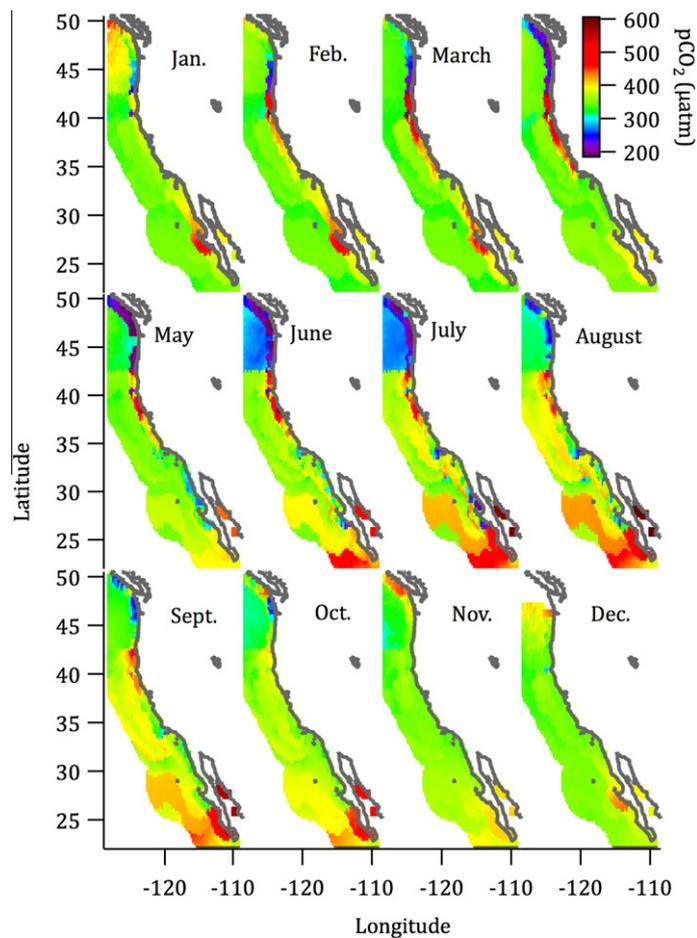


Fig. 8. Monthly maps of surface pCO₂ distributions generated using the coefficients in Table 3 and monthly climatologies of SeaWiFSchl and SST.

of them and the result based on the full data set. This strongly suggests that the model coefficients are not being fortuitously driven by a few anomalous observations. The greater proof of the model's robustness comes in the analysis of the predictive power of a set of coefficients with one set of training data for the 10% of the observations that were not included in the optimizations. These results are shown in Fig. 7, and listed in Table 5. In each case, the model based on the training data reproduces the verification data about as well as the model based on either the reduced training subsets, or the data-set as a whole. These results show relatively low sensitivity to changes in the training and validation data, and suggest that the predictions based on the full data set (Fig. 6; Tables 3 and 4) are broadly applicable.

We offer some further validation of the predictive approach via comparison of the climatological May surface pCO₂ predictions with direct observations collected on a cruise (described in the methods section) in May of 2007. The two representations are shown in Fig. 8. The agreement is not perfect, but the predictive algorithm for the climatological May captures many of the features of the observed May 2007 data: The bands of low pCO₂ near the coasts of Oregon, Washington, and Vancouver Island, and regions of strongly elevated pCO₂ off the northern and central California coast between Cape Blanco and Monterey are reproduced. Near-shore pCO₂ off southern California and Baja is lower than the waters further offshore in both representations. Offshore pCO₂ in northern regions is lower than in the south, as seen in the

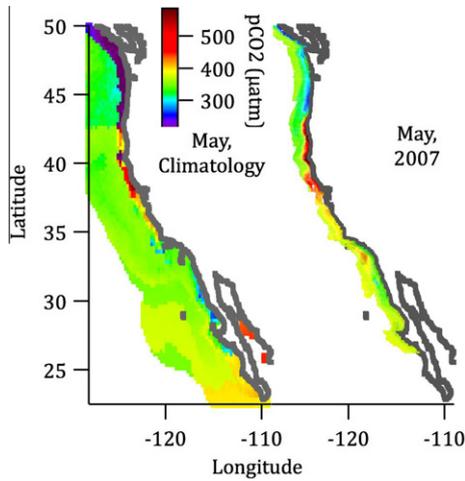


Fig. 9. Comparison of algorithm-predicted surface pCO₂ distributions for a climatological May (left panel) with observed pCO₂ from a cruise from Vancouver, BC, to Baja, Mexico in May, 2007.

predicted distribution. This comparison, while admittedly qualitative, represents a completely independent method of evaluating the empirical approach developed here. It strongly suggests that the features of the climatologically-defined pCO₂ maps are persistent. Thus, while the system is highly variable in space and time, the relationships to remotely observable parameters are persistent

and regional variability is not randomly spatially and temporally ephemeral.

Given the apparently robust nature of the predictive algorithm, we feel justified in producing monthly-resolved climatological maps. Using the italicized sets of coefficients identified in Table 3, we then generated maps of pCO₂ distributions (Fig. 9) based on monthly climatologies of MODIS SST and SeaWiFS Chlorophyll. It is important to note that while the SOM region locations were seasonally static and defined based on climatological fields, the values of SST and Chl-a within those regions was time-variant following the monthly-resolved climatology.

The monthly distributions of air–sea flux are shown in Fig. 10. These distributions demonstrate the spatial variability expected from recent publications (Hales et al., 2005; Chavez et al., 2007; Ianson et al., 2010; Evans et al., 2011), but also reveal seasonal variability that was previously less obvious. Every region, with the possible exception of a small patch furthest offshore at 40°N, experiences a change in sign over the course of the year. Even the high source Region 9 becomes a weak sink from October to January, and the strong-sink Region 5 becomes a weak source from November to January. Overall, the notion that higher latitude regions are stronger sinks than those to the south holds true for most of the year, but even this has seasonal dependence; net out-gassing occurs in northern regions during December–January while southern regions actually take up CO₂ during that time.

In addition to the features discussed above, the monthly maps point out some deficiencies in the discrete SOM approach. In particular, sharp N–S gradients seen in June and July at the boundary

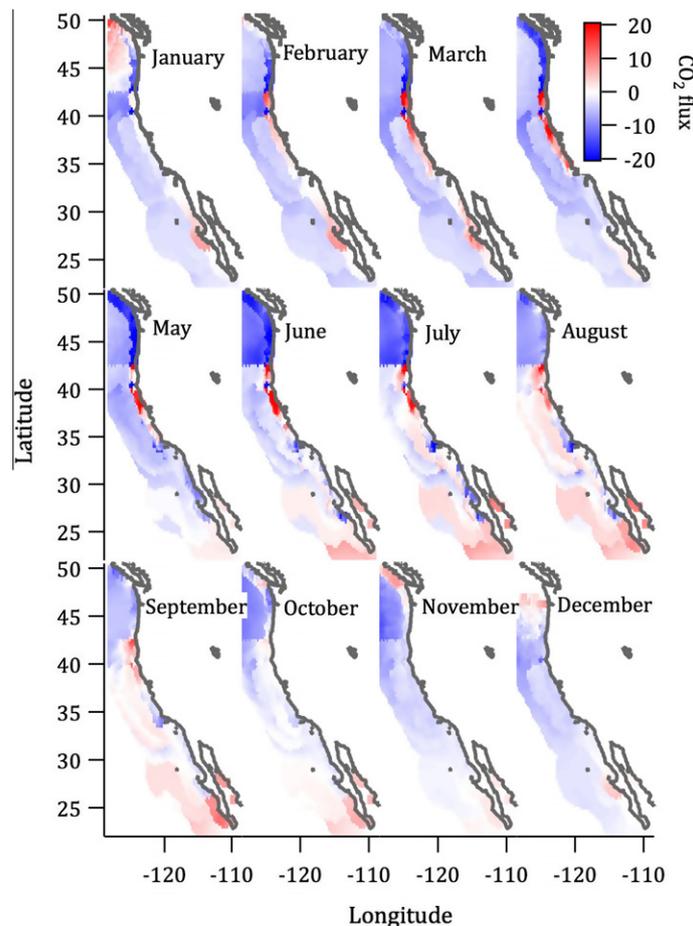


Fig. 10. Monthly climatologies of air–sea CO₂ flux (mmol m⁻² d⁻¹; negative numbers are into the ocean) generated from the pCO₂ maps of Fig. 9 and the monthly climatologies of wind-speed squared taken from the Risien and Chelton database (<http://cioss.coas.oregonstate.edu/scow/>; compiled following Risien and Chelton, 2008).

of regions 10 and 11, and the E–W banding across Regions 8, 4, 2, and 1 in July–September are probably not accurate representations of the true distributions. The SOM regions are distinct, and the changes in the model coefficients at the region boundaries are discontinuous, while the satellite input fields are continuous. We have applied no smoothing to the region transitions, and such artifactual discontinuities are a likely result of this approach. While a smoothing of the region boundaries would be straightforward and result in a more visually pleasing result, we felt that no scientifically-relevant information was gained by that and opted to show an unsmoothed result that revealed this issue.

Air sea flux for North American Pacific coastal waters from 22°N to 50°N and within 370 km of shore amounted to an annual-average sink of $1.8 \text{ mmol CO}_2 \text{ m}^{-2} \text{ d}^{-1}$. The most comparable flux estimate from the Chavez et al. (2007) analysis, based on the three $1^\circ \times 1^\circ$ bins nearest the shore, predicts a source of about $0.07 \text{ mmol m}^{-2} \text{ d}^{-1}$. Although not regionally specific, Borges et al. (2005) and Cai et al. (2006) estimate mid-latitude upwelling systems to be air–sea sinks of 0.3 and $2.7 \text{ mmol CO}_2 \text{ m}^{-2} \text{ d}^{-1}$, respectively. Scaling our new areal average to the whole study region (less the Sea of Cortez) gives a net annual air–sea CO_2 uptake of about 14 Tg C yr^{-1} , compared with the Chavez et al. net out-gassing of 0.5 Tg C yr^{-1} . Our new air–sea transfer estimate is comparable to that estimated by Hales et al. (2005), but only fortuitously. Hales et al. (2005) limited their extrapolation to a smaller cross-shelf extent (water depths <200 m) and a shorter seasonal duration, but also extrapolated a significantly more negative mean air–sea pCO_2 difference. Uncertainty in this estimate is difficult to completely assess, and depends on short-term covariance of wind fields and surface pCO_2 distributions that we could never resolve with this climatological estimate. If we assume that the large-scale wind patterns and surface pCO_2 distributions are adequately represented by this method, and that the smaller-scale variability is uncorrelated, we can crudely estimate the uncertainty from the areally-weighted RMS deviation between predicted and observed pCO_2 ($\sim 20 \mu\text{atm}$) and the areally-averaged air–sea pCO_2 difference ($\sim 20 \mu\text{atm}$), we must place uncertainty on the air–sea transfer estimation that is similar in magnitude to the best estimate (i.e. $\pm 14 \text{ Tg C yr}^{-1}$).

5. Discussion

The combination of SOM-defined regions and a mechanistic, nonlinear predictive algorithm for pCO_2 shows promise for superior spatial and temporal estimation of air–sea CO_2 fluxes. However, the results thus far need to be evaluated in comparison with previous large-scale syntheses, and issues related to the broader applicability need to be elaborated upon. In the discussion, we first investigate the difference between the flux estimate presented above and that resulting from the Chavez et al. (2007) analysis. Second, we attempt to understand the still-unsatisfying predictive capability of chlorophyll. Third, we address the limitation of the climatologically-defined SOM. Finally, we discuss limitations on broader application of the approach.

5.1. Comparison to previous regional flux estimations

The net annual flux for the study area is a net annual sink for atmospheric CO_2 of much larger magnitude than the small source based on the Chavez et al. (2007) synthesis. While uncertainties are large in both estimates, the difference in best estimates amounts to a regionally important air–sea transport difference, and should be discussed. There are a number of reasons for the difference. First is the fact that this analysis was limited to the years 1997–2005, when satellite chlorophyll products were available,

and the ambient atmospheric pCO_2 was higher than over much of the time period of the Chavez et al. (2007) data synthesis, which included observations as far back as the late 1970s. We used an average atmospheric pCO_2 of $375 \mu\text{atm}$ in our flux calculations, appropriate for the location of the study and the mid-point of that time frame. The Chavez et al analysis used constructions of the air–sea pCO_2 difference (ΔpCO_2) in their synthesis, and this suggests that atmospheric increases in pCO_2 were accounted for if the coastal waters were tracking the atmosphere, as is the case with many open ocean regions (Takahashi et al., 2009). If the coastal oceans do not, in fact, follow atmospheric CO_2 increases, then this more recent atmospheric CO_2 value could account for a large part of the estimated flux difference.

Second is the use of satellite scatterometer wind speed estimates in place of NCEP reanalysis products. The former are more spatially and temporally variable than the latter (Risien and Chelton, 2008), and the use of the 2nd-order wind-speed dependence of the gas exchange coefficient is expected to yield a greater mean gas transfer rate. The wind product we used was the monthly average of the square of the wind stress, so this potential enhancement of the gas transfer would be retained. However, the more significant factor is in the correlation or anti-correlation of the gas transfer velocity with the air–sea pCO_2 difference. If our analysis produces regions with undersaturated surface waters coinciding with times when winds are strong, the ocean sink strength will be correspondingly amplified. Some hint of this is present in the waters of the Pacific Northwest, which show strong undersaturation in nearshore waters in early spring when winter storms have not yet subsided, and in offshore waters, which retain undersaturated conditions into the fall when winter winds increase. These regions show up as strong sinks and play a role in setting the net annual flux for the region.

Last is the algorithm itself. The new algorithm includes explicit dependences on temperature and chlorophyll that were not included in the bin-averaging and spatio-temporal interpolation approach of Chavez et al. (2007). These dependences give the possibility of hydrographically-driven variability in the surface pCO_2 distributions that is certain to be different than that in the Chavez et al. (2007) analysis. The reasons for any systematic bias are not clear, and are almost impossible to isolate given the multitude of other differences between the two approaches.

5.2. Chlorophyll as a predictor for pCO_2

Another key result is the ambiguous nature of the Chl dependences in the algorithm. The empirical result is greatly improved over the MLR approach, with twice as many regions requiring Chl dependence in their final representation and 5 of these 6 having coefficients that are at least within the same order as the canonical $T_{\text{CO}_2}:\text{Chl}$ ratio of -7 . The results are not as mechanistically satisfying as hoped, however, with half of the regions still requiring no significant Chl dependence. Some of this may be due to the fact that the Chl data in the predictive algorithm is a remote-sensing product that includes some spatial and temporal averaging not incorporated in the in situ pCO_2 measurements; however, we believe that there are important de-couplings in the T_{CO_2} –chlorophyll relationship that should be discussed.

There is some region-specific justification for these results: Regions 2, 3, and 7, which require no Chl dependence, and Region 4, which requires a surprisingly large-amplitude seasonal variation around a mean value of zero, are in the southernmost part of the study area and not in direct contact with the coast. These temperate, low-biomass, low-productivity regions are likely to have highly recycled phytoplankton production, in which only a small fraction of their primary production actually results in net export. The geochemical result of a this scenario would be a weak linkage

between changes in chlorophyll abundance and changes in T_{CO_2} and thus direct connection of pCO_2 to Chl standing stock may be expected to be unclear. This would be particularly acute if variations in the degree of recycling were decoupled from variations in biomass.

Regions 1, 9, and 10 are likely to be strongly influenced by upwelling and high net productivity, but still have no requirement for Chl as a predictor. This observation is more difficult to explain, but we can speculate that coastal diatom populations, which have rapid population dynamics, often dominate upwelling systems. These populations can draw large amounts of nitrate down to background levels in a matter of a few days (Dugdale et al., 1990, 2006). Upon reaching nutrient exhaustion, blooms can terminate with similar rapidity, either by aggregation (Prieto et al., 2002) or by viral attack (Bratbak et al., 1990), resulting in massive export events that leave little biomass in surface waters. The impact on surface water chemistry, however can persist until other, physical factors such as gas exchange can restore pre-bloom conditions. In the case of CO_2 , gas exchange is a slow process, taking months to re-equilibrate surface mixed layers, and so the CO_2 depletion resulting from productivity could be retained in surface waters long after biomass had disappeared.

The conclusion must be that the model applied here is inadequate for capturing the complexity of the relationship between CO_2 and Chl-a as defined by MODIS. This is not necessarily a surprise, despite our conceptual improvements over simple linear dependences of pCO_2 on chlorophyll standing stock, and we know that there are a number of better hypothetical approaches. Others have shown that chlorophyll is a poor predictor for pCO_2 , particularly in coastal waters (e.g. Borges and Frankignoulle, 2001; Zhai et al., 2009). One improved approach might be to discern the dominant types of phytoplankton from ocean color remote sensing data, such as the approach being carried out by d'Ovidio et al. (2010) using algorithms like PHYSAT. Another might be to incorporate estimates of net productivity rather than standing stock and some sort of historical evolution of the chlorophyll abundance within a water mass along its trajectory. However these suggestions require more sophisticated modeling and data assimilation than currently exists in order to realize the true capability of remotely-sensed chlorophyll as a predictor for CO_2 .

5.3. Climatologically-defined SOM regions

The objective of this exercise was not only to devise a means for improved pCO_2 prediction, but also to remove the non-mechanistic proxy dependences on independent space and time variables. The SOM approach, by objectively defining biogeochemical regions within which hydrographic relationships were thought to be consistent, was a key step in that process. This succeeded in that there was essentially no need for spatial dependence within regions in the final applied pCO_2 predictions, as shown by the general lack of sensitivity to latitude and longitude as independent variables. It failed, however, in that all regions required empirical time dependence in at least some of their coefficients. While it is easy to explain how there ought to be seasonal dependence in the biogeochemical functioning of a region, that functioning ought to be captured by the hydrographic dependences in the predictive model.

We believe that some of this could be addressed by defining regions that are based on seasonally- or monthly-resolved climatologies. We know that certain parts of these coastal waters are strongly seasonally influenced. For example, the higher latitudes of the study area shift seasonally from upwelling to downwelling physical forcing, and it is unlikely that these waters are hydrographically homogeneous over the course of the year. In such a temporally-resolved SOM exercise, the boundaries of the regions

would not be geographically fixed, but might move north or south, or shrink or expand over the course of the year. This is a complicated next step, and is not as straightforward as simply defining discrete sets of SOM regions for individual temporal intervals. Kavanaugh et al. (submitted for publication) have examined a temporally continuous approach to this problem, applied in the open North Pacific. Applying this approach to these coastal waters is a logical next step.

5.4. Application to CO_2 flux predictions

Two factors limit extrapolation of these results over wider spatial and temporal extents. The first is the obvious data limitation in key regions. The intent of this exercise was to devise a means for expanding sparse data coverage using a synthetic approach; however, sufficient data for algorithm training is still essential. In extremely data-poor regions like those at the northern and southern extents of the North American continental margin, there is still an insufficient observational basis for application of a method such as this. Conversely, the SOM analysis may help to ease the overall observational burden, and to direct observational programs. Once SOM regions are identified for a given area, and the dynamic ranges of the hydrographic factors within each region defined, a sampling approach need only provide pCO_2 data over a reasonable portion of the hydrographic range to be sufficient for algorithm training. This could be a substantial improvement over attempting to exhaustively cover the space and time scales of interest.

The second is that the analysis here was for a monthly-resolved climatology, and as a result has no ability to account for long-term temporal trends that might be unrelated to hydrographic relationships. The pertinent example is of rising atmospheric CO_2 levels, the effects of which we were unable to identify in the surface-water pCO_2 observations. It is unlikely that the coastal surface waters are not responding to the atmosphere at some level, as this would imply that the sink would keep indefinitely increasing as atmospheric levels continued to rise. It is likely that any trend following the $\sim 2 \mu\text{atm yr}^{-1}$ increase in atmospheric CO_2 is just so small relative to the variability in the observations that it is obscured. It may be possible to use an approach such as this to define a reference pCO_2 based on hydrographic relationships, which thus accounts for the biogeochemical forcing. The deviations between the reference pCO_2 and the actual observations could then be used to interpret the long-term temporal trends.

6. Conclusions

We have presented a method for predicting the pCO_2 of surface waters in a highly variable and diverse coastal region, the Pacific coast of central North America. The method requires objective sub-categorization of the study area into biogeochemically consistent sub-regions, and a predictive model that allows nonlinear dependence on input parameters. We accomplished the first by means of a self-organizing map (SOM) that delineated 13 sub-regions in the study area. We accomplished the second by simple parameterization of surface water T_{CO_2} and T_{Alk} as a function of SST, Chl, and time of year, from which pCO_2 was calculated, thus explicitly including the inherent non-linearity of the carbonate chemistry system. The method was trained using a historical compilation of surface-water pCO_2 observations, and yielded predictions with RMS deviations from the observations of less than $20 \mu\text{atm}$ and correlation coefficients of >0.81 . Predictive power was improved by at least a factor of 2.5, for the entire region, with important individual sub-regions having their predictive power improved by over an order of magnitude. The method was validated by comparison

of predicted pCO₂ with that observed on a recent cruise spanning the study region. The predicted pCO₂ distributions coupled with a wind-speed climatology allowed calculation of a regional uptake of 14 Tg C yr⁻¹ of atmospheric CO₂, in contrast to the weak release of CO₂ from the same waters predicted by a previous bin-averaging and interpolation approach. Future application of the method described here should focus on better incorporation of the mechanistic linkages between chlorophyll standing stocks and T_{CO₂}, better mechanistic descriptions of the predicted alkalinity, and better temporal resolution of the SOM analysis. The approach presented here may have application to assessing long-term trends in settings with highly variable pCO₂, and may aid in efficient planning of future observational efforts.

Acknowledgements

The authors thank R. Venegas for executing MS's SOM codes at OSU, and M. Kavanaugh for constructive discussion of the PrSOM approach. This work was supported by NASA grant NNG05GH11G and NOAA grant NA05OAR4311164 This is PMEL contribution #3832.

Appendix A

If each term in Eq. (8) was simply a space-and time-invariant constant (i.e. if parameters c₁–c₃ were all fixed at 0), then the model would reduce to:

$$T_{CO_2} = A + BT + \beta Chl \quad (A.1)$$

where

$$A = T_{CO_2}^0 - BT^0 \quad (A.2)$$

$$B = \gamma\theta \quad (A.3)$$

In this case, the four empirical parameters T_{CO₂}⁰, T⁰, γ, and φ could be reduced to two. We chose to retain the form of Eq. (8), however, because it maintains the conceptual T_{CO₂} representation, and because once independent time- and space-variability of the parameters T_{CO₂}⁰, T⁰, γ, and φ are allowed, the terms then become expanded mixed-polynomial products of equations with the form of Eq. (9). For example,

$$\gamma\theta = (c_0^i + c_1^i\Delta lat + c_2^i\Delta lon + c_3^i t)(c_0^o + c_1^o\Delta lat + c_2^o\Delta lon + c_3^o t) \quad (A.4)$$

where the superscripts refer to the term to which each coefficient c_i applies. The number of coefficients is thus no longer reduced significantly. The form of Eq. (8) with terms that vary spatially and temporally as in Eq. (9) is a more intuitive representation of the model, and we maintained that representation throughout the exercise.

References

Aitkenhead, J.A., McDowell, W.H., 2000. Soil C/N as a predictor of annual riverine DOC flux at local and global scales. *Global Biogeochemical Cycles* 14, 127–138.

Bates, N.R., Best, M.H.P., Hansell, D.A., 2005. Spatio-temporal distribution of dissolved inorganic carbon and net community production in the Chukchi and Beaufort Seas. *Deep-Sea Research II* 52, 3303–3323.

Bauer, J., Goni, M., McKee, B., 2008. North American rivers and Estuaries. In: Hales, B., Cai, W.-J., Mitchell, B.G., Sabine, C.L., Schofield, O. (Eds.), *North American Continental Margins: A Synthesis and Planning Workshop*. Report of the North American Continental Margins Working Group for the US Carbon Cycle Scientific Steering Group and Interagency Working Group. US Carbon Cycle Science Program, Washington, DC, 110 pp.

Bianchi, A., Bianucci, L., Piola, A., Ruiz-Pino, D., Schloss, I., Poisson, A., Balestrini, C., 2005. Vertical stratification and air–sea CO₂ fluxes in the Patagonian shelf. *Journal of Geophysical Research* 110, C07003. <http://dx.doi.org/10.1029/2004JC002488>.

Borges, A.V., 2005. Do we have enough pieces of the jigsaw to integrate CO₂ fluxes in the coastal ocean? *Estuaries* 28, 3–27.

Borges, A.V., Delille, B., Frankignoulle, M., 2005. Budgeting sinks and sources of CO₂ in the coastal ocean: diversity of ecosystems counts. *Geophysical Research Letters* 32, L14601. <http://dx.doi.org/10.1029/2005GL023053>.

Borges, A.V., Schiettecatte, L.-S., Abril, G., Delille, B., Gazeau, F., 2006. Carbon dioxide in European coastal waters, Estuarine. *Coastal and Shelf Science* 70, 375–387.

Bratbak, G., Haldal, M., Norland, S., Thingstad, T.F., 1990. Viruses as partners in spring bloom microbial trophodynamics. *Applied and Environmental Microbiology* 56, 1400–1405.

Cai, W.-J., 2003. Riverine inorganic carbon flux and rate of biological uptake in the Mississippi River plume. *Geophysical Research Letters* 30, 1032.

Cai, W.-J., 2011. Estuarine and coastal ocean carbon paradox: CO₂ sinks or sites of terrestrial carbon incineration? *Annual Review of Marine Science* 3, 123–145. <http://dx.doi.org/10.1146/annurev-marine-120709-142723>.

Cai, W.-J., Wang, Z., Wang, Y., 2003. The role of marsh-dominated heterotrophic continental margins in transport of CO₂ between the atmosphere, the land–sea interface and the ocean. *Geophysical Research Letters* 30, 1849.

Cai, W.-J., Dai, M., Wang, Y., 2006. Air–sea exchange of carbon dioxide in ocean margins: a province based synthesis. *Geophysical Research Letters* 33, L12603. <http://dx.doi.org/10.1029/2006GL026219>.

Chavez, F., Takahashi, T., Cai, W.-J., Friedrich, G., Hales, B., Wanninkhof, R., Feely, R.A., 2007. Coastal oceans. In: King, A.W.L., Dilling, Zimmerman, G.P., Fairman, D.M., Houghton, R.A., Marland, G.H., Rose, A.Z., Wilbanks, T.J. (Eds.), *The First State of the Carbon Cycle Report (SOCCR): North American Carbon Budget and Implications for the Global Carbon Cycle*. A Report by the US Climate Change Science Program and the Subcommittee on Global Change Research. National Ocean and Atmospheric Administration, Climate Program Office, Silver Spring, MD, USA, pp. 157–166, 193. <<http://www.climatechange.gov/Library/sap/sap2-2/final-report/default.htm>> (Chapter 15).

Chen, C.-T.A., Andreev, A., Kim, K.-R., Yamamoto, M., 2004. Roles of continental shelves and marginal seas in the biogeochemical cycles of the North Pacific Ocean. *Journal of Oceanography* 60, 17–44.

Cosca, C.E., Feely, R.A., Boutin, J., Etcheto, J., McPhaden, M.J., Chavez, F.P., Strutton, P.G., 2003. Seasonal and interannual CO₂ fluxes for the central and eastern equatorial Pacific Ocean as determined from fCO₂–SST relationships.

Degens, E.T., Kempe, S., Richey, J.E. (Eds.), 1991. *Biogeochemistry of Major World Rivers*. John Wiley & Sons Ltd.

DeGrandpre, M.D., Hammar, T.R., Olbu, G.J., Beatty, C.M., 2002. Air–sea CO₂ fluxes on the US Middle Atlantic Bight. *Deep-Sea Research II* 49, 4355–4367.

Dickson, A.G., 1990. Thermodynamics of the dissociation of boric acid in synthetic seawater from 273.15 to 318.15 K. *Deep-Sea Research* 37, 755–766.

Dickson, A.G., Millero, F.J., 1987. A comparison of the equilibrium constants for the dissociation of carbonic acid in seawater. *Deep-Sea Research* 34, 1733–1743.

d'Ovidio, F., De Monte, S., Alvain, S., Dandonneau, Y., Lévy, M., 2010. Fluid dynamical niches of phytoplankton types, 2010. *Proceedings of the National Academy of Sciences of the United States of America*. <http://dx.doi.org/10.1073/pnas.1004620107>.

Ducklow, H.W., McAllister, S.L., 2005. The biogeochemistry of carbon dioxide in the coastal oceans. In: Brink, K.H., Robinson, A.R. (Eds.), *The Global Coastal Ocean: Multiscale Interdisciplinary Processes*. The Sea, vol. 13. Harvard University Press, Cambridge, MA.

Dugdale, R.C., Wilkerson, F.P., Morel, A., 1990. Realization of new production in coastal upwelling areas: a means to compare relative performance. *Limnology and Oceanography* 35, 822–829.

Dugdale, R.C., Wilkerson, F.P., Hogue, V.E., Marchi, A., 2006. Nutrient controls on new production in the Bodega Bay, California, coastal upwelling plume. *Deep-Sea Research II* 53, 3049–3062.

Evans, W., Hales, B., Strutton, P., 2011. The seasonal cycle of surface ocean pCO₂ on the Oregon shelf. *Journal of Geophysical Research Oceans*, 116, C05012. <http://dx.doi.org/10.1029/2010JC006625>.

Feely, R.A., Takahashi, T., Wanninkhof, R., McPhaden, M.J., Cosca, C.E., Sutherland, S.C., Carr, M.-E., 2006. Decadal variability of the air–sea CO₂ fluxes in the equatorial Pacific Ocean. *Journal of Geophysical Research C—Oceans* C08S90. <http://dx.doi.org/10.1029/2005JC003129>.

Feely, R.A., Sabine, C.L., Hernandez-Ayon, J.M., Ianson, D., Hales, B., 2008. Evidence for upwelling of corrosive 'acidified' water onto the continental shelf. *Science* 320, 1490–1492.

Fennel, K., Wilkin, J., Previdi, M., Najjar, R., 2008. Denitrification effects on air–sea CO₂ flux in the coastal ocean: Simulations for the Northwest North Atlantic. *Geophys. Res. Lett.* 35 L24608. <http://dx.doi.org/10.1029/2008GL036147>.

Fennel, K., Wilkin, J., 2009. Quantifying biological carbon export for the northwest North Atlantic continental shelves. *Geophys. Res. Lett.* L18605. <http://dx.doi.org/10.1029/2009GL039818>.

Frankignoulle, M., Borges, A.V., 2001. European continental shelf as a significant sink for atmospheric carbon dioxide. *Global Biogeochemical Cycles* 15, 569–576.

Friedrich, G.E., Chavez, F.P., Walz, P.M., Burczynski, M.G., 2002. Inorganic carbon in the central California upwelling system during the 1997–1999 El Niño–La Niña event. *Progress in Oceanography* 54, 185–203.

Friedrich, T., Oschlies, A., 2009a. Neural network-based estimates of North Atlantic surface pCO₂ from satellite data: a methodological study. *Journal of Geophysical Research* 114, C03020. <http://dx.doi.org/10.1029/2007JC004646>.

Friedrich, T., Oschlies, A., 2009b. Basin-scale pCO₂ maps estimated from ARGO float data: A model study. *Journal of Geophysical Research* 114, C10012. <http://dx.doi.org/10.1029/2009JC005322>.

Gruber, N., Frenzel, H., Doney, S.C., Marchesiello, P., McWilliams, J.C., Moisan, J.R., Oram, J., Plattner, G.K., Stolzenbach, K.D., 2006. Eddy-resolving simulation of

- plankton ecosystem dynamics in the California Current System. *Deep-Sea Research I*, 53. <http://dx.doi.org/10.1016/j.dsr.2006.06.005>.
- Hales, B., Chipman, D., Takahashi, T., 2004. High-frequency measurement of partial pressure and total concentration of carbon dioxide in seawater using microporous hydrophobic membrane contactors. *Limnology and Oceanography: Methods* 2, 356–364.
- Hales, B., Takahashi, T., Bandstra, L., 2005. Atmospheric CO₂ uptake by a coastal upwelling system. *Global Biogeochemical Cycles* 19. <http://dx.doi.org/10.1029/2004GB002295>.
- Hales, B., Cai, W.-J., Mitchell, B.G., Sabine, C.L., Schofield, O. (Eds.), 2008. North American Continental Margins: A Synthesis and Planning Workshop. Report of the North American Continental Margins Working Group for the US Carbon Cycle Scientific Steering Group and Interagency Working Group. US Carbon Cycle Science Program, Washington, DC, 110 pp.
- Hauri, C., Gruber, N., Alin, S., Fabry, V.J., Feely, R.A., Hales, B., Plattner, G.-K., Wheeler, P., 2009. Ocean acidification in the California Current system. *Oceanography* 22, 60–71.
- Hedges, J.L., Keil, R.G., Benner, R., 1997. What happens to terrestrial organic matter in the ocean? *Organic Geochemistry* 27, 195–212.
- Ho, D.T., Law, C.S., Smith, M.J., Schlosser, P., Harvey, M., Hill, P., 2006. Measurements of air–sea gas exchange at high wind speeds in the Southern Ocean: implications for global parameterizations. *Geophysical Research Letters* 33, L16611. <http://dx.doi.org/10.1029/2006GL026817>.
- Ianson, D., Feely, R.A., Sabine, C.L., Juranek, L.W., 2010. Features of coastal upwelling regions that determine net air–sea CO₂ flux. *Journal of Oceanography* 65, 677–687.
- Ittekkot, V., Laane, R.W.P.M., 1991. Fate of riverine particulate organic matter. In: Degens, E.T., Kempe, S., Richey, J.E. (Eds.), *Biogeochemistry of Major World Rivers*. John Wiley & Sons Ltd.
- Kavanaugh, M.T., Hales, B., Saraceno, M., Spitz, Y.H., White, A.E., Letelier, R.M., submitted for publication. Towards a quantitative framework for pelagic seascape ecology. *Ecosystems*.
- Kohonen, T., 1990. The self-organizing map. *Proceedings of IEEE* 78, 1464–1480.
- Kohonen, T., 1995. *Self-Organizing Maps*. Springer Series in Information Sciences, vol. 30. Springer-Verlag, Berlin, Heidelberg, New York.
- Lagerloef, G., Colomb, F.R., Le Vine, D., Wentz, F., Yueh, S., Ruf, C., Lilly, J., Gunn, J., Chao, Y., Decharon, A., Feldman, G., Swift, C., 2008. The aquarius/SAC-D mission: designed to meet the salinity remote-sensing challenge. *Oceanography* 21, 68–81.
- Lefevre, N., Watson, A.J., Watson, A.R., 2005. A comparison of multiple regression and neural network techniques for mapping in situ pCO₂ data. *Tellus* 57B, 375–384.
- Lohrenz, S.E., Cai, W.-J., 2006. Satellite ocean color assessment of air–sea fluxes of CO₂ in a river-dominated coastal margin. *Geophysical Research Letters* 33, L01601. <http://dx.doi.org/10.1029/2005GL023942>.
- Longhurst, A.R., 2006. *Ecological Geography of the Sea*, second ed. Academic Press, 560pp. (ISBN: 978-0-12-455521-1).
- Mecklenburg, S., Kerr, Y., Font, J., Hahne, A., 2008. The Soil Moisture and Ocean Salinity (SMOS) Mission – an overview. *Geophysical Research (Abstracts)* 10, EGU2008-A-01922).
- Mehrbach, C., Culbertson, C., Hawley, J.E., Pytkowicz, R.M., 1973. Measurement of the apparent dissociation constants of carbonic acid in seawater at atmospheric pressure. *Limnology and Oceanography* 18, 897–907.
- Meybeck, M., Vorosmarty, C., 1999. Global transfer of carbon by rivers. *Global Change Newsletter* 37, 18–19.
- Park, G.H., Wanninkhof, R., Doney, S.C., Takahashi, T., Lee, K., Feely, R.A., Sabine, C.L., Trinanes, J., 2010. Variability of global net air–sea CO₂ fluxes over the last three decades using empirical relationships. *Tellus* 62B, 352–368. <http://dx.doi.org/10.1111/j.1600-0889.2010.00498x>.
- Press, W.H., Flannery, B.P., Teukolsky, S.A., Vetterling, W.T., 1989. *Numerical Recipes—The Art of Scientific Computing*. Cambridge University Press.
- Prieto, L., Ruiz, J., Echevarria, F., Garcia, C.M., Bartual, A., Galvez, J.A., Corzo, A., Macias, D., 2002. Scales and processes in the aggregation of diatom blooms: high time resolution and wide size range records in a mesocosm study. *Deep-Sea Research I* 49, 1233–1253.
- Revelle, R., Suess, H.E., 1957. Carbon dioxide exchange between atmosphere and ocean and the question of an increase of atmospheric CO₂ during the past decades. *Tellus* 9, 18–27.
- Risien, C.M., Chelton, D.B., 2008. A global climatology of surface wind and wind stress fields from eight years of QuikSCAT scatterometer data. *Journal of Physical Oceanography* 38, 2379–2413.
- Saraceno, M., Provost, C., Lebbah, M., 2006. Biophysical regions identification using an artificial neuronal network: a case study in the South Western Atlantic. *Advances in Space Research* 37 (2006), 793–805.
- Schlunz, B., Schneider, R.R., 2000. Transport of terrestrial organic carbon to the oceans by rivers: re-estimating flux and burial rates. *International Journal of Earth Sciences* 88, 599–606.
- Smith, S.V., Hollibaugh, J.T., 1993. Coastal metabolism and the oceanic organic carbon balance. *Reviews of Geophysics* 31, 75–89.
- Spalding, M.D. et al., 2007. Marine ecoregions of the world: a bioregionalization of coastal and shelf areas. *BioScience* 57, 573–583.
- Takahashi, T., Olafsson, J., Goddard, J., Chipman, D.W., Sutherland, S.C., 1993. Seasonal variation of CO₂ and nutrients in the high-latitude surface oceans: a comparative study. *Global Biogeochemical Cycles* 7, 843–878.
- Takahashi, T. et al., 2009. Climatological mean and decadal change in surface ocean pCO₂ and net sea-air CO₂ flux over the global oceans. *Deep-Sea Research*. <http://dx.doi.org/10.1016/j.dsr2.2008.12.009>.
- Telszewski, M., Chazottes, A., Schuster, U., Watson, A.J., Moulin, C., Bakker, D.C.E., González-Dávila, M., Johannessen, T., Kortzinger, A., Luger, H., Olsen, A., Omar, A., Padin, X.A., Rios, A.F., Steinhoff, T., Santana-Casiano, M., Wallace, D.W.R., Wanninkhof, R., 2009. Estimating the monthly pCO₂ distribution in the North Atlantic using a self-organizing neural network. *Biogeosciences* 6, 1405–1421.
- Thomas, H., Bozec, Y., Elkalay, K., de Baar, H.J.W., 2004. Enhanced open ocean storage of CO₂ from shelf sea pumping. *Science* 304, 1005–1008.
- Tsunogai, S., Watanabe, S., Sato, T., 1999. Is there a “continental shelf pump” for the absorption of atmospheric CO₂? *Tellus* B 51, 701–712.
- Wanninkhof, R., Olsen, A., Triñanes, J., 2006. Air–sea CO₂ fluxes in the Caribbean Sea from 2002–2004. *Journal of Marine Systems* 66, 272–284.
- Weiss, R.F., 1974. Carbon dioxide in water and seawater: the solubility of a non-ideal gas. *Marine Chemistry* 2, 203–215.