Accepted Manuscript

Satellite-based prediction of pCO$_2$ in coastal waters of the eastern North Pacific

Burke Hales, Peter Strutton, Martin Saraceno, Ricardo Letelier, Taro Takahashi, Richard Feely, Christopher Sabine, Francisco Chavez
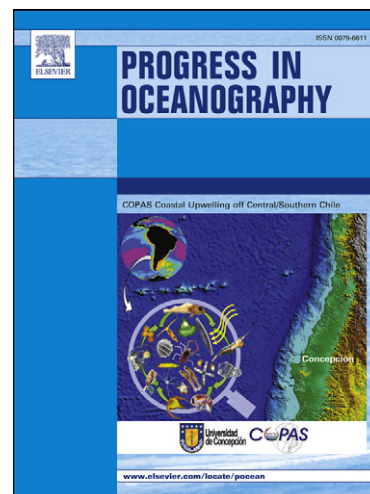
Please cite this article as: Hales, B., Strutton, P., Saraceno, M., Letelier, R., Takahashi, T., Feely, R., Sabine, C., Chavez, F., Satellite-based prediction of pCO$_2$ in coastal waters of the eastern North Pacific, *Progress in Oceanography* (2012), doi: 10.1016/j.pocean.2012.03.001

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

Burke Hales[1,*], Peter Strutton[1,**], Martin Saraceno[2,3], Ricardo Letelier[1], Taro Takahashi[4], Richard Feely[5], Christopher Sabine[5], and Francisco Chavez[6]

1. Oregon State University College of Oceanic and Atmospheric Sciences, Corvallis, OR 97331
2. Centro de Investigaciones del Mar y la Atmosfera, Consejo Nacional de Investigaciones Científicas y Técnicas, Argentina
3. Departamento de Ciencias de la Atmósfera y de los Océanos, Facultad de Ciencias Exactas y Naturales, Universidad de Buenos Aires, Argentina
4. Lamont-Doherty Earth Observatory of Columbia University, Palisades, NY 10964
5. NOAA Pacific Marine Environmental Laboratory, Seattle, WA 98115
6. Monterey Bay Aquarium Research Institute, Moss Landing, CA 95039

* corresponding author: 541 737-8121; bhales@coas.oregonstate.edu

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

1 **Abstract**

2 Continental margin carbon cycling is complex, highly variable over a range of space and
3 time scales, and forced by multiple physical and biogeochemical drivers. Predictions of
4 globally significant air-sea CO_2 fluxes in these regions have been extrapolated based on
5 very sparse data sets. We present here a method for predicting coastal surface-water pCO_2
6 from remote-sensing data, based on self organizing maps (SOMs) and a non-linear semi-
7 empirical model of surface water carbonate chemistry. The model used simple empirical
8 relationships between carbonate chemistry (total dissolved carbon dioxide (T_{CO2}) and
9 alkalinity (T_{Alk})) and satellite data (sea surface temperature (SST) and chlorophyll (Chl)).
10 Surface-water CO_2 partial pressure (pCO_2) was calculated from the empirically-predicted
11 T_{CO2} and T_{Alk}. This directly incorporated the inherent nonlinearities of the carbonate
12 system, in a completely mechanistic manner. The model's empirical coefficients were
13 determined for a target study area of the central North American Pacific continental margin
14 (22-50°N, within 370 km of the coastline), by optimally reproducing a set of historical
15 observations paired with satellite data. The model-predicted pCO_2 agreed with the highly
16 variable observations with a root mean squared (RMS) deviation of < 20 µatm, and with a
17 correlation coefficient of > 0.8 (r = 0.81; r^2 = 0.66). This level of accuracy is a significant
18 improvement relative to that of simpler models that did not resolve the biogeochemical
19 sub-regions or that relied on linear dependences on input parameters. Air-sea fluxes based
20 on these pCO_2 predictions and satellite-based wind speed measurements suggest that the
21 region is a ~14 Tg C yr^{-1} sink for atmospheric CO_2 over the 1997-2005 period, with an
22 approximately equivalent uncertainty, compared with a ~0.5 Tg C yr^{-1} source predicted by
23 a recent bin-averaging and interpolation-based estimate for the same area.
24

1

2. 1. Introduction and Background

3 Coastal waters have been alternately reported as globally important sources or sinks of
4 atmospheric CO_2 (Bianchi et al. 2005; Borges 2005; Cai et al. 2006; Ducklow and McAllister,
5 2005; Chen et al. 2004; DeGrandpre et al. 2002; Frankignoulle and Borges, 2001; Hales et al.
6 2005; Smith and Hollibaugh, 1993; Thomas et al. 2004; Tsunogai et al. 1999, as
7 summarized by Hales et al., 2008). The uncertainty stems from several factors. First, the
8 dynamic range of surface pCO_2 in coastal waters spans hundreds of μatm over a variety of
9 time and space scales (Friederich et al., 2002; Cai et al. 2003; Cai, 2003; Hales et al., 2005;
10 Bates et al., 2005; Chavez et al., 2007). This makes adequate observational constraint
11 difficult to achieve, and global or regional flux estimations have always been the result of
12 extrapolating a few spatially and temporally limited observations far beyond their scope
13 (Hales et al., 2005; Thomas et al., 2005; Cai et al., 2006; Borges, 2005). However, the vast
14 majority of flux estimates based on direct observation of pCO_2 in coastal waters suggest
15 that coastal waters are net sinks of atmospheric CO_2. Complicating matters is the fact that
16 terrestrial inputs of carbon via rivers supply nearly a petagram per year (Pg C y^{-1}; 1 Pg =
17 10^{15} g) of terrestrial carbon to coastal waters, almost none of which can be accounted for in
18 coastal sediment or water column reservoirs (Degens et al. 1991; Ittekkot and Laane, 1991;
19 Spitzy and Leenheer, 1991; Hedges et al. 1997; Meybeck and Vorosmarty, 1999;
20 Aitkenhead and McDowell, 2000; Schlunz and Schneider, 2000 as summarized by Bauer et
21 al., 2008). This imbalance suggests a large efflux of CO_2, as argued by Smith and Hollibaugh
22 (1993) that has not been observed in coastal waters. Borges et al. (2005, 2006) offered the
23 possibility that estuaries may be the location of efflux of terrestrial carbon, while Cai
24 (2011) suggested that marine-source carbon degradation supported estuarine efflux while
25 terrestrial material is degraded offshore. Again, sampling coverage is limited.

26

27 Chavez et al. (2007) compiled all the North American coastal observations of pCO_2 in the
28 Lamont-Doherty Earth Observatory (LDEO) database (now hosted by the Carbon Dioxide
29 Information Analysis Center (CDIAC) data server), to assess the contribution of coastal
30 waters to the continental carbon budget. They found that North American coastal waters
31 contributed a small net source of CO_2 to the atmosphere (~2 Tg C yr^{-1}; 1 Tg = 0.001 Pg), but
32 this near-neutral flux was the result of large high-latitude sinks (~35 Tg C yr^{-1} into the
33 Pacific and Atlantic north of 50°N and the Bering and Chukchi seas) that were balanced by
34 large low-latitude sources (~40 Tg C yr^{-1} out of the Pacific and Atlantic south of 25°S and of
35 the Gulf of Mexico and Caribbean Sea). Although the data set included nearly 1 million
36 observations, the ship of opportunity-based sampling did not provide uniform spatial or
37 temporal coverage. In the bin-averaging approach employed by Chavez et al. (2007) it is
38 clear that the opposing large magnitude high and low latitude sink and source terms were
39 based on pixels with inadequate observations throughout the calendar year (Hales et al.,
40 2008).

41

42 With these concerns in mind, we have attempted to produce synthetic approaches for
43 estimating surface-water pCO_2 distributions in coastal waters. True mechanistic modeling
44 including accurate physical circulation fields and biogeochemistry, driven by actual
45 physical forcing is ultimately the ideal approach. However, modeling coastal environments
46 is challenging even for only the physical circulation, and to date the few truly coupled

3

1   biogeochemical/ physical models that predict surface pCO_2 distributions are limited to
2   narrow geographical regimes such as the southern California Current system (e.g. Gruber
3   et al., 2006) or the shelves of the northwest North Atlantic (Fennel et al., 2008; Fennel and
4   Wilkin, 2009).
5
6   We present here a semi-empirical approach to developing algorithms that link surface
7   water pCO_2 to remotely sensed data, in addition to position and time. We found that two
8   factors were critical: 1) an objective means for identifying biogeochemical sub-regions
9   within coastal waters, since there was very low success for single predictive algorithms
10   applied across wide geographical areas such as the continental waters of North America;
11   and 2) a justified application of high-order dependences on input parameters, since simple
12   linear dependences were incapable of generating the large dynamic ranges seen in the
13   observed distributions. Given these factors, we present a method that uses (1) self-
14   organizing maps (SOMs) based on satellite observations for distinguishing distinct
15   provinces, and (2) a semi-mechanistic representation of the relationships between
16   seawater carbonate chemistry (total dissolved CO_2 (T_{CO2}) and alkalinity (T_{alk})) and the
17   input parameters of sea surface temperature (SST) and chlorophyll (Chl). pCO_2 was then
18   calculated from T_{CO2} and T_{alk}, thus incorporating the inherent non-linearities of the
19   seawater carbonate system. We test this approach for the Pacific coastal waters of the
20   central North American continent, from 22°N to 50°N, within 370 km (~200 nautical miles)
21   of the shore.
22
23   Although this is not the first attempt to link surface-ocean pCO_2 to more extensively-
24   observable parameters, earlier approaches have been substantively different.   Most were
25   in the open ocean where dynamic ranges were smaller and pCO_2 was not impacted by the
26   continental margins (Cosca et al., 2003; Lefevre et al., 2005; Feely et al., 2006; Friedrich and
27   Oschlies 2009a,b; Park et al., 2010), or in narrowly defined margin settings where
28   individual drivers such as river plumes (Lohrenz and Cai, 2006) or thermal forcing
29   (Wanninkhof et al., 2006) dominate the pCO_2 distributions. Lefevre et al. (2005), Friedrich
30   and Oschlies (2009a,b), and Telszewski et al., (2009) did use the neuronal network
31   approach, but applied it only over the parameter space of their observations, as opposed to
32   the regional province identification that we applied here. None of the previous efforts have
33   attempted synthesis of a region so dynamic and diverse as the North American Pacific coast,
34   nor have any included mechanistically-justified nonlinearities as we do here.
35
36   **2. Materials and Methods**
37
38   *2.1 pCO_2 Observations*
39   We used the compilation of pCO_2 data prepared by Chavez et al., (2007), for waters
40   surrounding the North American continent to train the algorithm. These data were
41   originally stored in the LDEO pCO_2 data repository, and are now available from CDIAC
42   (http://cdiac.esd.ornl.gov/). They consist of nearly 800,000 observations made in the
43   interval 1978-2005 (heavily weighted for more recent observations), all based on analysis
44   of equilibrated gas headspace over flowing seawater streams or discrete samples. A
45   significant portion of these data, particularly those closest to shore were excluded from the

4

1   global analyses of Takahashi et al., (1995; 2002; 2009), and thus these observations from
2   continental margins have not been included in broader estimates of air-sea CO_2 fluxes.
3
4   We limited our analyses to observations made within 370 km of the nearest major
5   coastline, roughly consistent with the position of the 200 nautical mile limit of national
6   Exclusive Economic Zones (EEZ). This limit is functional as well as operational. Essentially
7   all signals of coastal influences are gone by this distance from shore, although exceptions
8   do occur (Chavez et al., 2007). This extends the spatial limit of the analysis far enough to
9   meet nearly all of the landward-most boundaries of the global ocean syntheses of
10   Takahashi et al. (e.g. Takahashi et al. 2009), which were bin-averaged at 4° latitude by 5°
11   longitude bins. Finally, the EEZ boundary represents the furthest distance where local
12   governments may restrict access to local waters and the data collected therein. Thus, this
13   represents a distance from shore within which data may have been excluded from regional
14   or global compilations. Limiting the observations to those made with coincident SeaWiFS
15   satellite observations, which started in September 1997, reduced the sample size to about
16   300,000. The North American data set was then divided into four regions—Atlantic, Gulf of
17   Mexico/Caribbean, Pacific, and Bering/Chukchi coastal waters. For the high resolution
18   regional analyses described below, we focused on the North American Pacific coast
19   between 22°N-50°N, where about 96,000 observations exist within 370 km of the coastline
20   and with coincident SeaWiFS data (see section 2.2, below).
21
22   We did not attempt to correct the surface pCO_2 data to a reference year, as done for the
23   Takahashi global syntheses (e.g. Takahashi, 2009). The dynamics of the coastal ocean are
24   complicated and unique within regions, with some being dominated by terrestrial inputs,
25   others by upwelling, others by local heating and cooling. Water residence times within
26   coastal regions can be short relative to the age of the water masses themselves, and thus it
27   is questionable whether to treat freshly upwelled waters that might have last been at the
28   surface some decades ago (Feely et al., 2008) with a modern adjustment to the in-water
29   pCO_2. Limiting the analysis to the SeaWiFS era suggests a very small correction for the
30   1995-2005 interval.
31
32   For the generation of the predictive algorithms, we used the in situ SST recorded along
33   with the pCO_2 and satellite chlorophyll measurements. We viewed this as important for the
34   closest coupling of the SST and pCO_2 data. We examined the results using the satellite-
35   based SST (not shown) and found comparable results, albeit with slightly inferior model-
36   data deviation statistics. This is not too surprising, given the reasonably good agreement
37   between in situ and remotely-sensed SST. Salinity has promise as a predictive input,
38   especially in coastal areas where many water masses may be present. However, we did not
39   see significant improvement in our predictive algorithms using reported in situ salinity,
40   and while remote-sensing salinity products are now available from Aquarius (Lagerloeff et
41   al., 2008, and SMOS, Mecklenburg et al., 2008), they will have the spatial, temporal, and
42   salinity resolution to distinguish only the largest features, such as river plumes.
43
44   We used recent data collected on a 2007 cruise off the North American Pacific coast as a
45   qualitative validation of the predictive algorithm (Feely et al., 2008). The pCO_2 data
46   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 pCO_2 observational data used in the development of the predictive algorithms were
derived by averaging the 9 (3×3) nearest 9km pixels from 8-day averaged SeaWiFS Chl
bracketing each observation. For creation of the predicted maps of pCO_2 distributions and
air-sea CO_2 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-pCO_2 data match-
up is based on remote-sensing data, while the SST-pCO_2 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.

$$pCO_{2,mlr} = C_0 + \sum_{i=1}^{n} C_i \, p_i \qquad\qquad 1$$

where $pCO_{2,mlr}$ is the predicted pCO_2. C_0 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 (Δlat and Δlon, 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:

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

where 'day' was the decimal day of year.

3.2.2 Mechanistic non-linear model: As shown in section 4.2, below, we quickly found the MLR approach to be inadequate for the high dynamic range of the observed pCO_2, even after applying the SOM to the study area, and built a new meta-model that included the inherent non-linearities 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 pCO_2 is a quantitative nonlinear function of the alkalinity (T_{ALK}) and total CO_2 (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:

$$T_{CO2} = T_{CO2}^0 + \Delta T_{CO2} \qquad\qquad 3$$

$$T_{Alk} = T_{Alk}^0 + \Delta T_{Alk} \qquad\qquad 4$$

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:

$$\Delta T_{CO2} = \left.\frac{\partial T_{CO2}}{\partial T}\Delta T\right]_{mixing} + \left.\frac{\partial T_{CO2}}{\partial Chl}\Delta Chl\right]_{biology} \qquad\qquad 5$$

where mixing is reflected in a temperature change term ΔT, and biological processes in a chlorophyll change term ΔChl. These two terms are each defined by:

$$\Delta T]_{mixing} = \varnothing(T - T^0) \qquad\qquad 6$$

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

where T^0 and Chl^0 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:

$$T_{CO2} \cong T_{CO2}^0 + \gamma\varnothing(T - T^0) + \beta Chl. \qquad 8$$

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 ß 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_{CO2} 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_{CO2}^0, T_{Alk}^0, T^0, γ, ϕ, β to have dependence on space and time, e.g.,

$$T_{CO2}^0 = c_0 + c_1\Delta lat + c_2\Delta lon + c_3 t \qquad 9$$

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_{CO2} 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_{CO2}, i.e.,

$$\Delta T_{Alk} = -0.15\Delta T_{CO2}. \qquad 10$$

We have not included factors that may change the $T_{CO2}:T_{Alk}$ 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 $T_{CO2}:T_{ALK}$ 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_{CO2}, which are then used to calculate pCO_2 at the in situ temperature. For calculating pCO_2 from T_{Alk} and T_{CO2}, 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_{CO2} are predicted by the model, they are not used directly in the evaluation procedure. The calculated pCO_2 is the ultimate model output.

1   The model has a couple of deficiencies. First is the possibility that there are as many as 25
2   optimizeable coefficients in the two equations for T_{CO2} and T_{Alk} that derive from Eqs. 2-10.
3   We needed to make some *a priori* limitation of this number, described in section 4.3, to
4   limit the maximum possible number of coefficients. The second is that the model contains
5   products of some of the terms, which implies the possibility of interdependence and a
6   reduced set of truly optimizeable coefficients. This is addressed in the Appendix.
7
8   *3.3 Optimization*
9   We chose to use Powell's method for optimization in multiple dimensions, as described in
10 Press et al. (1989). This method has the advantage in that it does not depend on the form
11 of the equation being optimized, which is essential given the potentially variable forms as
12 empirical terms are included or excluded, and because of the non-linearity of the step in
13 which pCO_2 is calculated from the intermediate Alk and T_{CO2} products. In every
14 optimization exercise, all 25 empirical constants were explicitly included, even if set to zero,
15 but the optimization approach allowed selection of only reduced sets of coefficients to
16 optimize. For example, we could specify that the chlorophyll dependence ß is non-zero but
17 remains fixed at the nominal value of -7 (e.g. reflecting a 1:1 N:Chl ratio and Redfield C:N
18 stoichiometry).
19
20 Our diagnostic of algorithm performance was the root-mean-squared (RMS) deviation
21 between the observed and predicted pCO_2, over the complete set of the observations within
22 each region. We examined the sensitivity of the optimized results in two ways. First, we
23 tried optimizations with varied values of the initial guesses of the optimized parameters.
24 This included using the optimum set of parameters as initial guesses and restarting the
25 optimization procedure. This verified the stability of the optimum solution. Second, we
26 calculated coefficient-specific sensitivity factors (SF_i) defined by:
27

$$SF_i = \left. \left(\frac{1}{RMS} \partial RMS \right) \middle/ \left(\frac{1}{c_i} \partial c_i \right) \right. \qquad 14$$

29
30 which effectively describes the relative change in RMS for a relative change in coefficient c_i.
31 Coefficients with SF ≤ 0.1 – i.e., the optimum RMS value changed by no more than 1% for a
32 10% change in coefficient—were deemed insignificant.
33
34 *3.4 Sensitivity Analysis*
35 Training an algorithm to reproduce a set of observations requires additional verification
36 and analysis of the sensitivity of the model parameters to the dataset. The complexities of
37 the model obviated some of the more standard tools that could be easily applied to smaller
38 data sets and linear systems, so we chose to take a more rudimentary approach. After
39 finding the optimum set of model parameters for a given region, we perform ten additional
40 simulations to examine the robustness of the result. Each of these simulations consisted of
41 the same analysis as described above, but with 10% of a region's data randomly extracted
42 from the total. The remaining 90% were used as training data to repeat the optimization

1   procedure. The optimal model constrained by the training data was then applied to the
2   extracted 10% that had not been used in constraining the optima.
3
4   *3.5 Air-sea flux calculations*
5   Once we mapped the best set of predicted pCO_2 distributions, we calculated air-sea CO_2
6   fluxes. We did this by calculating a gas exchange coefficient based on the quadratic wind-
7   speed dependence of Ho et al (2006) and the QuikSCAT monthly wind speed-squared data
8   from the Scatterometer Climatology of Ocean Winds (SCOW) available at
9   http://cioss.coas.oregonstate.edu/scow/, based on the methods in Risien and Chelton
10   (2008). We multiplied this coefficient by the model-predicted air-sea pCO_2 difference and
11   the gas solubility of Weiss (1974) for satellite SST and a fixed salinity.
12

13 **4. Results**
14   *4.1 Linear models.*
15   We began our analysis with an attempt to fit the pCO_2 observations in the Chavez et al
16   (2007) synthesis data set, limited as described above to within 370 km of the coasts and
17   with coincident SeaWiFS Chl data, with a multiple linear dependence in the form of Eq. 1.
18   This was without success (Figure 1), and we recognized the need to distinguish the major
19   coastal regions (Figure 2). The MLR approach applied to these regions yielded better
20   results, particularly for the Gulf of Mexico/Caribbean waters (Fig. 2, lower right) where our
21   regression yielded statistics similar to the prediction presented by Wanninkhof et al (2006).
22   The pCO_2 distributions in Pacific coastal waters, however, were nearly as poorly
23   reproduced by this attempt as were those from the entire North American coastline. The
24   essence of the failure was first that the large dynamic range of the observed pCO_2 was not
25   in any way reproduced by the predictions, and that there seemed to be little meaningful
26   dependence on any of the independent input variables. This result suggested that the
27   Pacific coast contained sub-regions that were not well described by a single set of empirical
28   parameters, and that the inherently linear aspect of a MLR was insufficient for reproducing
29   pCO_2 distributions with large dynamic ranges. Because the Pacific coastal waters
30   presented the greatest challenge, the rest of our efforts were directed at developing
31   predictive algorithms for this region.
32

33   *Figure 1. Multiple linear regression prediction of pCO_2 using the Chavez et al compilation for*
34   *observations in North American coastal waters.*
35

36   *Figure 2. Multiple linear regression prediction of pCO_2 for coastal regions of North America.*
37

38   *Figure 3. Climatological maps of MODIS SST, Chl, and QuikSCAT wind stress.*
39

40   *4.2 Determination of regions*
41   Figure 3 shows the climatological distributions of satellite SST, Chl, and wind stress over
42   Pacific coastal waters as observed by these three sensors. These data were used as
43   primary inputs to the self-organizing map determination of Saraceno et al. (2006) to define
44   the regions shown in Figure 4. The approach found 13 sub-regions before no significant
45   improvement (as defined earlier and in Saraceno et al., 2006) was gained by adding more
46   regions. These objectively-determined sub-regions confirmed our thinking about the

1 distinctions of coastal areas of the Pacific coast and precisely define important biophysical
2 regions within this study area (Table 1). Such an objective sub-division has not been done
3 before in this region, as far as we know. At the global scale, Longhurst (2006) defined bio-
4 geographical regions based on in-situ and remote sensing data. In the East Pacific, he
5 defined only two regions that overlap with our results: the California Current and Central
6 America Coastal provinces. Spalding et al. (2007) defined marine ecoregions of the world
7 along coastal and shelf areas. In the area considered in this work they found six regions. In
8 latitude, their divisions approximately coincide with ours. However they did not find cross-
9 shore distinction of regions as we found. These divisions separate coastal upwelled waters
10 from offshore waters remarkably well.
11
12 *Figure 4. SOM-based sub-regions of central North American Pacific coastal waters.*
13
14 *Table 1. Description of sub-regions*
15
16 Once the sub-regions were defined and observations assigned to each sub-region (Table 1),
17 we developed distinct algorithm parameters for each region. Our first attempt was to
18 simply apply the multiple linear regression of Eq. 1 to the subsets of data; the results are
19 summarized in Figure 5 and Table 2.
20
21 *Figure 5. Plots of data/model agreement for linear model.*
22
23 *Table 2. MLR statistics*
24
25 This result is significantly improved over that applied to the Pacific coastal data as a whole.
26 On a numerically-averaged basis, the RMS deviation between model and observations
27 dropped to 57.8 from the initial 67 µatm. Weighting by the areas of each region, the RMS
28 deviation dropped to 24.5 µatm. Regions 1, 2, 4 and 13 were particularly well-described
29 by this approach, with RMS deviations < 15 µatm and correlation coefficients of ~0.9.
30 However improved the results, there were still significant shortcomings. Particularly in
31 regions with large data densities and large dynamic ranges, such as the costal upwelling
32 areas from central California northward and the regions immediately offshore of them, the
33 familiar pattern of the model inadequately reproducing the dynamic range of the data was
34 evident. This suggests that the large dynamic ranges of the data do not correspond with
35 large ranges in the input parameters, and argues for the application of a higher-order
36 model. In addition, the coefficients of the simple linear model demonstrate odd behavior
37 (Table 2). Region 7, for example, shows a large negative constant term that is compensated
38 for by an exceedingly high temperature dependence term. Furthermore, chlorophyll
39 dependences are insignificant in 9 of the 12 cases. In the remaining three (Regions 1, 10,
40 13), sensitivity to chlorophyll is barely significant by our SF<0.1 criterion (SF≤0.15), and
41 the coefficients show large variability in the sign and magnitude of the dependences.
42
43 *4.3 Mechanistic Model*
44 The ocean carbonate system has well known nonlinearities, especially with regard to pCO_2.
45 The dependence of pCO_2 on water temperature is exponential, given established
46 thermodynamics of gas solubility (Takahashi et al., 1993). In-water processes that impact

1   T_{CO2} and T_{Alk} impact pCO_2 disproportionately; this effect is known as the Revelle Factor

2   (Revelle and Suess, 1957) and suggests relative changes in pCO_2 are ~10th power

3   dependent on relative T_{CO2} changes for normal oceanic conditions. We experimented with

4   functionality that included a Revelle-factor dependence on an empirically determined T_{CO2}

5   and incorporated an exponential temperature dependence following Takahashi et al.

6   (1993). The fundamental nature of these relationships, however, is to predict relative

7   changes, and thus the empirical coefficients were not well constrained by the optimization

8   procedure. Empirical nonlinearities are easy to assign—we could simply employ

9   polynomial dependences on the input parameters of the MLR representation, but doing this

10   without mechanistic justification is questionable. As a result, we chose not to pursue any of

11   these approaches and to build the quasi-mechanistic model as described in section 3.3.2,

12   Equations 3-10.

13

14   As stated previously, the model could potentially include a very large number of

15   optimizable coefficients. We felt that it was reasonable to restrict this number in the

16   following ways. The ratio of T_{CO2} to T_{Alk} is the leading factor in determining the pCO_2

17   distributions in the aqueous carbonate system. We opted to allow the T_{CO2}^0 term in Eq. 7 to

18   retain the full time and space variability as shown in Eq. 8. Assuming this would allow

19   sufficient variability in the ratio of $T_{CO2}^0 : T_{Alk}^0$, we chose to make T_{Alk}^0 in each region be

20   determined only by the corresponding constant term in Eq. 8. This may have the effect of

21   amplifying the suggested spatio-temporal T_{CO2} variability, but T_{Alk} is generally the more

22   conservative parameter and this was viewed as a reasonable trade-off for the reduction in

23   empirical coefficients. We further opted to allow the terms ϕ, γ, β, and T^0 to have only

24   temporal variability. As stated earlier, the temporal phasing for all parameters within a

25   given SOM region was assumed to be the same. The empirical model thus depended on a

26   maximum of 14 optimizable coefficients.

27

28   This is still a large number of coefficients, and there are two different approaches that

29   could be taken to identify the most important terms. First is a stepwise approach in which

30   terms are sequentially added and model improvement or lack thereof is assessed at each

31   step. Second is the approach we followed, in which we allowed all 14 coefficients to vary

32   initially, and then assessed the sensitivity of the result to each one.

33

34   Results are shown in Figure 6 and summarized in Tables 3 and 4. The immediate first

35   impression is of the improvement in the predictability of the pCO_2 in the problem regions

36   identified in the MLR analysis. Region 9, in particular, the persistent upwelling region off

37   central California that includes the large dataset from the MBARI time-series, has

38   significantly improved predictability (R increased to 0.75 from 0.15, a 25-fold increase in

39   the percentage of variance explained) as a result of including the nonlinear mechanistic

40   representation. Other regions (5, 8, 10, and 11) that include large dynamic-range pCO_2

41   observations are also better predicted (R values 0.82 vs 0.57, 0.80 vs 0.60, 0.92 vs 0.79,

42   0.61 vs 0.07, respectively). Regions can be broadly categorized into three groups—those

43   with high (≥ 0.75) correlation coefficient and low (≤ 20 µatm) RMS (regions 1-4, 8, 10, 13),

44   those with high correlation coefficient and high RMS (5, 9) and those with high RMS and

45   low correlation coefficient (7, 11, 12). The second group includes the most intense

46   temperate coastal upwelling conditions, where the dynamics of source waters and

1   terrestrial inputs may not be well characterized by the simple model. The third group
2   includes two regions (11, 12) that sit immediately of the upwelling regions 5 and 9.
3   Regions 11 and 12 are primarily offshore but include data-containing pixels that impinge
4   on the coastline, suggesting that the resolution of the SOM may be a factor. The remaining
5   region, 7, sits far offshore in the SW portion of the study area, and data there consists of
6   two clusters that are separated by ~200 µatm. Other regions in this area (1-4) do not show
7   the same character and we are uncertain of the reason for the poor model performance
8   there.
9
10 *Figure 6. Mechanistic non-linear algorithm predictions of pCO$_2$.*
11
12 Sensitivity analysis also shows some interesting results. First, none of the regions required
13 all 14 coefficients to be optimized. Region 1 showed sensitivity to a maximum of 13
14 coefficients, and Region 9 showed strong sensitivity to only 9 of the tested coefficients.
15 Second, the spatial dependence terms were only rarely significant in the optimizations.
16 Only Region 8 showed significant sensitivity to both latitude and longitude dependence
17 terms, and Regions 1, 2, and 4 were sensitive only to the latitude term. The remaining 8
18 regions showed no significant sensitivity to the spatial dependence terms. This is satisfying,
19 because one of the objectives of the SOM approach is to cluster data within regions where
20 biogeochemical and hydrographic relationships are consistent, and thus lessen non-
21 mechanistic empirical spatial dependences.
22
23 Finally, the sensitivities to chlorophyll, while improved over those seen in the MLR
24 approach, are still puzzling. Regions 8, 9, and 12 show no sensitivity to chlorophyll at all
25 (neither c_0 nor c_4 in the ß term were significant; Table 3), while regions 3-6 and 13 show
26 sensitivity to only one of the two coefficients. Further, the values of the coefficients
27 themselves are puzzling. The magnitudes of ß implied by the optimized coefficients are
28 often far from the nominal 'Redfield' expectation of ~-7. Equally puzzling in the conceptual
29 context of the model is the variation in the sign of the first-order chlorophyll dependence.
30 Region 7, for example, has coefficients that suggest that ß changes from <-500 to >500
31 (µmol kg^{-1})/(µg l^{-1}) over the course of the season. As discussed later, this dependence is
32 hard to justify within the context of the model, and probably indicates the need for a better
33 representation.
34
35 We examined two additional subsets of optimizable parameters. The first was the set
36 identified as having sensitivity by the full analysis. In 9 of 12 cases, the optimization
37 statistics were essentially equivalent between these reduced sets of coefficients and the
38 initial full set, with Regions 8, 11, and 12 being the exceptions. The second reduced set of
39 coefficients was one applied to all regions that included no spatial dependence in the T_{CO2}^0
40 term, and no chlorophyll dependence. We justified this second choice because so few
41 regions showed spatial dependence, and because the chlorophyll dependences were only
42 significant in half of the regions. This showed only slightly worse performance than the
43 sets of coefficients selected specifically for each region, with Regions 4, 5, and 8 showing
44 significantly worse predictive power with these subsets.
45

1   We chose sets of coefficients for use in subsequent analyses based on a combination of the
2   statistics of the optimization and the most conservative combinations of coefficients. When
3   optimization statistics were comparable, for a given region with different sets of
4   coefficients, we selected the result with the lowest number of optimized coefficients. These
5   choices are italicized in Table 3, and are the values used in Figure 6 and subsequent pCO_2
6   and air-sea flux reconstructions (Figures 8-10, below).

7

8   *Table 3. Model output statistics.*

9

10 Composite statistics for the entire study region are presented in Table 4. While the RMS
11 deviation statistics are improved over the SOM/MLR case (43 vs 58 µatm on a number-
12 weighted basis and 20 vs 25 µatm on an area-weighted basis), the improvement is shown
13 most strongly for the correlation coefficient statistic (0.76 vs 0.32 and 0.81 vs 0.51 on
14 number and area-weighted bases, respectively). As R^2 is often thought of as the fraction of
15 true variance explained by a predictive relationship, the mechanistic model thus explains
16 2.5 or 5.6 times more of the natural variability than the MLR, based on areal or number
17 weighting, respectively. The areally-weighted mechanistic model applied to the SOM
18 region describes about 66% of the observed variability, and reproduces the observations to
19 within about 20 µatm. The robustness of this result is addressed in the following
20 discussion.

21

22 *Table 4. Pacific coast composite model regression statistics.*

23

24 Although we performed sensitivity analyses for each region, we present only the results of
25 this exercise for Region 9. Region 9 contained the most observations, and difficulty in
26 representing that data had presented us with one of the greatest motivations for
27 developing the detailed model. Results are detailed in Table 5, where we examine the
28 sensitivities of the model coefficients for ten randomly selected subsets of 90% of the total
29 observations in this region. Little difference is seen in the values of the model coefficients
30 either within these ten simulations of between any of them and the result based on the full
31 data set. This strongly suggests that the model coefficients are not being fortuitously
32 driven by a few anomalous observations. The greater proof of the model's robustness
33 comes in the analysis of the predictive power of a set of coefficients with one set of training
34 data for the 10% of the observations that were not included in the optimizations. These
35 results are shown in Figure 7, and listed in Table 5. In each case, the model based on the
36 training data reproduces the verification data about as well as the model based on either
37 the reduced training subsets, or the data-set as a whole. These results show relatively low
38 sensitivity to changes in the training and validation data, and suggest that the predictions
39 based on the full data set (Figure 6; Tables 3 and 4) are broadly applicable.

40

41 We offer some further validation of the predictive approach via comparison of the
42 climatological May surface pCO_2 predictions with direct observations collected on a cruise
43 (described in the methods section) in May of 2007. The two representations are shown in
44 Figure 8. The agreement is not perfect, but the predictive algorithm for the climatological
45 May captures many of the features of the observed May 2007 data: The bands of low pCO_2
46 near the coasts of Oregon, Washington, and Vancouver Island, and regions of strongly

1 elevated pCO_2 off the northern and central California coast between Cape Blanco and
2 Monterey are reproduced. Nearshore pCO_2 off southern California and Baja is lower than
3 the waters further offshore in both representations. Offshore pCO_2 in northern regions is
4 lower than in the south, as seen in the predicted distribution. This comparison, while
5 admittedly qualitative, represents a completely independent method of evaluating the
6 empirical approach developed here. It strongly suggests that the features of the
7 climatologically-defined pCO_2 maps are persistent. Thus, while the system is highly
8 variable in space and time, the relationships to remotely observable parameters are
9 persistent and regional variability is not randomly spatially and temporally ephemeral.
10
11 *Figure 8. Simulated pCO_2 distributions for climatological May and comparison to a new set of*
12 *observations from May 2007.*
13
14 Given the apparently robust nature of the predictive algorithm, we feel justified in
15 producing monthly-resolved climatological maps. Using the italicized sets of coefficients
16 identified in Table 3, we then generated maps of pCO_2 distributions (Figure 9) based on
17 monthly climatologies of MODIS SST and SeaWiFS Chlorophyll. It is important to note that
18 while the SOM region locations were seasonally static and defined based on climatological
19 fields, the values of SST and Chl-a within those regions was time-variant following the
20 monthly-resolved climatology.
21
22 *Figure 9. Monthly pCO_2 maps.*
23
24 The monthly distributions of air-sea flux are shown in Figure 10. These distributions
25 demonstrate the spatial variability expected from recent publications (Hales et al., 2005;
26 Chavez et al., 2007; Ianson et al., 2010; Evans et al., 2011), but also reveal seasonal
27 variability that was previously less obvious. Every region, with the possible exception of a
28 small patch furthest offshore at 40°N, experiences a change in sign over the course of the
29 year. Even the high source Region 9 becomes a weak sink from October to January, and the
30 strong-sink region 5 becomes a weak source from November to January. Overall, the
31 notion that higher latitude regions are stronger sinks than those to the south holds true for
32 most of the year, but even this has seasonal dependence; net out-gassing occurs in northern
33 regions during December-January while southern regions actually take up CO_2 during that
34 time.
35
36 In addition to the features discussed above, the monthly maps point out some deficiencies
37 in the discrete SOM approach. In particular, sharp N-S gradients seen in June and July at
38 the boundary of regions 10 and 11, and the E-W banding across regions 8, 4, 2, and 1 in
39 July-September are probably not accurate representations of the true distributions. The
40 SOM regions are distinct, and the changes in the model coefficients at the region
41 boundaries are discontinuous, while the satellite input fields are continuous. We have
42 applied no smoothing to the region transitions, and such artifactual discontinuities are a
43 likely result of this approach. While a smoothing of the region boundaries would be
44 straightforward and result in a more visually pleasing result, we felt that no scientifically-
45 relevant information was gained by that and opted to show an un-smoothed result that
46 revealed this issue.

15

1
2   *Figure 10. Flux estimations.*
3
4   Air sea flux for North American Pacific coastal waters from 22°N to 50°N and within 370
5   km of shore amounted to an annual-average sink of 1.8 mmol CO_2 m^{-2} d^{-1}. The most
6   comparable flux estimate from the Chavez et al (2007) analysis, based on the three 1°x1°
7   bins nearest the shore, predicts a source of about 0.07 mmol m^{-2} d^{-1}. Although not
8   regionally specific, Borges et al. (2005) and Cai et al. (2006) estimate mid-latitude
9   upwelling systems to be air-sea sinks of 0.3 and 2.7 mmol CO_2 m^{-2} d^{-1}, respectively. Scaling
10 our new areal average to the whole study region (less the Sea of Cortez) gives a net annual
11 air-sea CO_2 uptake of about 14 Tg C y^{-1}, compared with the Chavez et al. net out-gassing of
12 0.5 Tg C yr^{-1}. Our new air-sea transfer estimate is comparable to that estimated by Hales et
13 al. (2005), but only fortuitously. Hales et al. (2005) limited their extrapolation to a smaller
14 cross-shelf extent (water depths < 200m) and a shorter seasonal duration, but also
15 extrapolated a significantly more negative mean air-sea pCO_2 difference. Uncertainty in this
16 estimate is difficult to completely assess, and depends on short-term covariance of wind
17 fields and surface pCO_2 distributions that we could never resolve with this climatological
18 estimate. If we assume that the large-scale wind patterns and surface pCO_2 distributions
19 are adequately represented by this method, and that the smaller-scale variability is
20 uncorrelated, we can crudely estimate the uncertainty from the areally-weighted RMS
21 deviation between predicted and observed pCO_2 (~20 μatm) and the areally-averaged air-
22 sea pCO_2 difference (~-20 μatm), we must place uncertainty on the air-sea transfer
23 estimation that is similar in magnitude to the best estimate (i.e. ± 14 Tg C y^{-1}).
24
25 **5. Discussion**
26 The combination of SOM-defined regions and a mechanistic, non-linear predictive
27 algorithm for pCO_2 shows promise for superior spatial and temporal estimation of air-sea
28 CO_2 fluxes. However, the results thus far need to be evaluated in comparison with previous
29 large-scale syntheses, and issues related to the broader applicability need to be elaborated
30 upon. In the discussion, we first investigate the difference between the flux estimate
31 presented above and that resulting from the Chavez et al. (2007) analysis. Second, we
32 attempt to understand the still-unsatisfying predictive capability of chlorophyll. Third, we
33 address the limitation of the climatologically-defined SOM. Finally, we discuss limitations
34 on broader application of the approach.
35
36 *5.1 Comparison to previous regional flux estimations*
37 The net annual flux for the study area is a net annual sink for atmospheric CO_2 of much
38 larger magnitude than the small source based on the Chavez et al. (2007) synthesis. While
39 uncertainties are large in both estimates, the difference in best estimates amounts to a
40 regionally important air-sea transport difference, and should be discussed. There are a
41 number of reasons for the difference. First is the fact that this analysis was limited to the
42 years 1997-2005, when satellite chlorophyll products were available, and the ambient
43 atmospheric pCO_2 was higher than over much of the time period of the Chavez et al (2007)
44 data synthesis, which included observations as far back as the late 1970s. We used an
45 average atmospheric pCO_2 of 375 μatm in our flux calculations, appropriate for the location
46 of the study and the midpoint of that time frame. The Chavez et al analysis used

16

1 constructions of the air-sea pCO_2 difference (ΔpCO_2) in their synthesis, and this suggests
2 that atmospheric increases in pCO_2 were accounted for if the coastal waters were tracking
3 the atmosphere, as is the case with many open ocean regions (Takahashi et al., 2009). If
4 the coastal oceans do not, in fact, follow atmospheric CO_2 increases, then this more recent
5 atmospheric CO_2 value could account for a large part of the estimated flux difference.
6
7 Second is the use of satellite scatterometer wind speed estimates in place of NCEP
8 reanalysis products. The former are more spatially and temporally variable than the latter
9 (Risien and Chelton, 2008), and the use of the 2nd-order wind-speed dependence of the gas
10 exchange coefficient is expected to yield a greater mean gas transfer rate. The wind product
11 we used was the monthly average of the square of the wind stress, so this potential
12 enhancement of the gas transfer would be retained. However, the more significant factor is
13 in the correlation or anti-correlation of the gas transfer velocity with the air-sea pCO_2
14 difference. If our analysis produces regions with undersaturated surface waters coinciding
15 with times when winds are strong, the ocean sink strength will be correspondingly
16 amplified. Some hint of this is present in the waters of the Pacific Northwest, which show
17 strong undersaturation in nearshore waters in early spring when winter storms have not
18 yet subsided, and in offshore waters, which retain undersaturated conditions into the fall
19 when winter winds increase. These regions show up as strong sinks and play a role in
20 setting the net annual flux for the region.
21
22 Last is the algorithm itself. The new algorithm includes explicit dependences on
23 temperature and chlorophyll that were not included in the bin-averaging and spatio-
24 temporal interpolation approach of Chavez et al. (2007). These dependences give the
25 possibility of hydrographically-driven variability in the surface pCO_2 distributions that is
26 certain to be different than that in the Chavez et al. (2007) analysis. The reasons for any
27 systematic bias are not clear, and are almost impossible to isolate given the multitude of
28 other differences between the two approaches.
29
30 *5.2 Chlorophyll as a predictor for pCO_2*
31 Another key result is the ambiguous nature of the Chl dependences in the algorithm. The
32 empirical result is greatly improved over the MLR approach, with twice as many regions
33 requiring Chl dependence in their final representation and 5 of these 6 having coefficients
34 that are at least within the same order as the canonical T_{CO2}:Chl ratio of -7. The results are
35 not as mechanistically satisfying as hoped, however, with half of the regions still requiring
36 no significant Chl dependence. Some of this may be due to the fact that the Chl data in the
37 predictive algorithm is a remote-sensing product that includes some spatial and temporal
38 averaging not incorporated in the in situ pCO_2 measurements; however, we believe that
39 there are important de-couplings in the TCO_2-chlorophyll relationship that should be
40 discussed.
41
42 There is some region-specific justification for these results: Regions 2, 3, and 7, which
43 require no Chl dependence, and Region 4, which requires a surprisingly large-amplitude
44 seasonal variation around a mean value of zero, are in the southernmost part of the study
45 area and not in direct contact with the coast. These temperate, low-biomass, low-
46 productivity regions are likely to have highly recycled phytoplankton production, in which

1   only a small fraction of their primary production actually results in net export. The
2   geochemical result of a this scenario would be a weak linkage between changes in
3   chlorophyll abundance and changes in TCO_2 and thus direct connection of pCO_2 to Chl
4   standing stock may be expected to be unclear. This would be particularly acute if
5   variations in the degree of recycling were decoupled from variations in biomass.
6
7   Regions 1, 9, and 10 are likely to be strongly influenced by upwelling and high net
8   productivity, but still have no requirement for Chl as a predictor. This observation is more
9   difficult to explain, but we can speculate that coastal diatom populations, which have rapid
10 population dynamics, often dominate upwelling systems. These populations can draw
11 large amounts of nitrate down to background levels in a matter of a few days (Dugdale et al.,
12 1990; 2006). Upon reaching nutrient exhaustion, blooms can terminate with similar
13 rapidity, either by aggregation (Prieto et al, 2002) or by viral attack (Bratbak et al., 1990),
14 resulting in massive export events that leave little biomass in surface waters. The impact
15 on surface water chemistry, however can persist until other, physical factors such as gas
16 exchange can restore pre-bloom conditions. In the case of CO_2, gas exchange is a slow
17 process, taking months to re-equilibrate surface mixed layers, and so the CO_2 depletion
18 resulting from productivity could be retained in surface waters long after biomass had
19 disappeared.
20
21 The conclusion must be that the model applied here is inadequate for capturing the
22 complexity of the relationship between CO_2 and chl-a as defined by MODIS. This is not
23 necessarily a surprise, despite our conceptual improvements over simple linear
24 dependences of pCO_2 on chlorophyll standing stock, and we know that there are a number
25 of better hypothetical approaches. Others have shown that chlorophyll is a poor predictor
26 for pCO_2, particularly in coastal waters (e.g. Borges and Frankignoulle, 2001; Zhai et al.,
27 2009). One improved approach might be to discern the dominant types of phytoplankton
28 from ocean color remote sensing data, such as the approach being carried out by d'Ovidio
29 et al., (2010) using algorithms like PHYSAT. Another might be to incorporate estimates of
30 net productivity rather than standing stock and some sort of historical evolution of the
31 chlorophyll abundance within a water mass along its trajectory. However these
32 suggestions require more sophisticated modeling and data assimilation than currently
33 exists in order to realize the true capability of remotely-sensed chlorophyll as a predictor
34 for CO_2.
35
36 *5.3 Climatologically-defined SOM regions*
37 The objective of this exercise was not only to devise a means for improved pCO_2 prediction,
38 but also to remove the non-mechanistic proxy dependences on independent space and time
39 variables. The SOM approach, by objectively defining biogeochemical regions within which
40 hydrographic relationships were thought to be consistent, was a key step in that process.
41 This succeeded in that there was essentially no need for spatial dependence within regions
42 in the final applied pCO_2 predictions, as shown by the general lack of sensitivity to latitude
43 and longitude as independent variables. It failed, however, in that all regions required
44 empirical time dependence in at least some of their coefficients. While it is easy to explain
45 how there ought to be seasonal dependence in the biogeochemical functioning of a region,

1   that functioning ought to be captured by the hydrographic dependences in the predictive
2   model.
3
4   We believe that some of this could be addressed by defining regions that are based on
5   seasonally- or monthly-resolved climatologies. We know that certain parts of these coastal
6   waters are strongly seasonally influenced. For example, the higher latitudes of the study
7   area shift seasonally from upwelling to downwelling physical forcing, and it is unlikely that
8   these waters are hydrographically homogeneous over the course of the year. In such a
9   temporally-resolved SOM exercise, the boundaries of the regions would not be
10   geographically fixed, but might move north or south, or shrink or expand over the course of
11   the year. This is a complicated next step, and is not as straightforward as simply defining
12   discrete sets of SOM regions for individual temporal intervals. Kavanaugh et al.
13   (submitted) have examined a temporally continuous approach to this problem, applied in
14   the open North Pacific. Applying this approach to these coastal waters is a logical next step.
15
16   *5.4 Application to CO$_2$ flux predictions.*
17   Two factors limit extrapolation of these results over wider spatial and temporal extents.
18   The first is the obvious data limitation in key regions. The intent of this exercise was to
19   devise a means for expanding sparse data coverage using a synthetic approach; however,
20   sufficient data for algorithm training is still essential. In extremely data-poor regions like
21   those at the northern and southern extents of the North American continental margin,
22   there is still an insufficient observational basis for application of a method such as this.
23   Conversely, the SOM analysis may help to ease the overall observational burden, and to
24   direct observational programs. Once SOM regions are identified for a given area, and the
25   dynamic ranges of the hydrographic factors within each region defined, a sampling
26   approach need only provide pCO$_2$ data over a reasonable portion of the hydrographic range
27   to be sufficient for algorithm training. This could be a substantial improvement over
28   attempting to exhaustively cover the space and time scales of interest.
29
30   The second is that the analysis here was for a monthly-resolved climatology, and as a result
31   has no ability to account for long-term temporal trends that might be unrelated to
32   hydrographic relationships. The pertinent example is of rising atmospheric CO$_2$ levels, the
33   effects of which we were unable to identify in the surface-water pCO$_2$ observations. It is
34   unlikely that the coastal surface waters are not responding to the atmosphere at some level,
35   as this imply that the sink would keep indefinitely increasing as atmospheric levels
36   continued to rise. It is likely that any trend following the $\sim$2 μatm yr^{-1} increase in
37   atmospheric CO$_2$ is just so small relative to the variability in the observations that it is
38   obscured. It may be possible to use an approach such as this to define a reference pCO$_2$
39   based on hydrographic relationships, which thus accounts for the biogeochemical forcing.
40   The deviations between the reference pCO$_2$ and the actual observations could then be used
41   to interpret the long-term temporal trends.
42

6. Conclusions
44   We have presented a method for predicting the pCO$_2$ of surface waters in a highly variable
45   and diverse coastal region, the Pacific coast of central North America. The method requires
46   objective subcategorization of the study area into biogeochemically consistent sub-regions,

1   and a predictive model that allows nonlinear dependence on input parameters. We
2   accomplished the first by means of a self-organizing map (SOM) that delineated 13 sub-
3   regions in the study area. We accomplished the second by simple parameterization of
4   surface water T_{CO2} and T_{Alk} as a function of SST, Chl, and time of year, from which pCO_2 was
5   calculated, thus explicitly including the inherent non-linearity of the carbonate chemistry
6   system. The method was trained using a historical compilation of surface-water pCO_2
7   observations, and yielded predictions with RMS deviations from the observations of less
8   than 20 µatm and correlation coefficients of >0.81. Predictive power was improved by at
9   least a factor of 2.5, for the entire region, with important individual subregions having their
10 predictive power improved by over an order of magnitude. The method was validated by
11 comparison of predicted pCO_2 with that observed on a recent cruise spanning the study
12 region. The predicted pCO_2 distributions coupled with a wind-speed climatology allowed
13 calculation of a regional uptake of 14 Tg C yr^{-1} of atmospheric CO_2, in contrast to the weak
14 release of CO_2 from the same waters predicted by a previous bin-averaging and
15 interpolation approach. Future application of the method described here should focus on
16 better incorporation of the mechanistic linkages between chlorophyll standing stocks and
17 T_{CO2}, better mechanistic descriptions of the predicted alkalinity, and better temporal
18 resolution of the SOM analysis. The approach presented here may have application to
19 assessing long-term trends in settings with highly variable pCO_2, and may aid in efficient
20 planning of future observational efforts.
21
22 **Appendix**
23 If each term in Eq. 8 was simply a space-and time-invariant constant (i.e. if parameters c_1-c_3
24 were all fixed at 0), then the model would reduce to:
25
26 $$T_{CO2} = A + BT + \beta Chl \hspace{5cm} \text{A.1}$$
27
28 where
29
30 $$A = T_{CO2}^0 - BT^0 \hspace{5cm} \text{A.2}$$
31
32 $$B = \gamma\varnothing. \hspace{5cm} \text{A.3}$$
33
34 In this case, the four empirical parameters T_{CO2}^0, T^0, γ, and ϕ could be reduced to two. We
35 chose to retain the form of Eq. 8, however, because it maintains the conceptual T_{CO2}
36 representation, and because once independent time- and space-variability of the
37 parameters T_{CO2}^0, T^0, γ, and ϕ are allowed, the terms then become expanded mixed-
38 polynomial products of equations with the form of Equation 9. For example,

39 $$\gamma\varnothing = \left(c_0^\gamma + c_1^\gamma \Delta lat + c_2^\gamma \Delta lon + c_3^\gamma t\right)\left(c_0^\varnothing + c_1^\varnothing \Delta lat + c_2^\varnothing \Delta lon + c_3^\varnothing t\right) \hspace{1cm} \text{A.4}$$
40
41 where the superscripts refer to the term to which each coefficient c_i applies. The number of
42 coefficients is thus no longer reduced significantly. The form of Eq. 8 with terms that vary
43 spatially and temporally as in Eq. 9 is a more intuitive representation of the model, and we
44 maintained that representation throughout the exercise.

References Cited

Aitkenhead, J.A., and W.H. McDowell, 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., M.H.P. Best, and D.A. Hansell, 2005a. 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., M. Goni., and B. McKee, 2008. North American Rivers and Estuaries. In *North American Continental Margins: A Synthesis and Planning Workshop*. Report of the North American Continental Margins Working Group for the U.S. Carbon Cycle Scientific Steering Group and Interagency Working Group. U.S. Carbon Cycle Science Program, Washington, DC, 110 pp. Hales, B., W.-J. Cai, B. G. Mitchell, C. L. Sabine, and O. Schofield [eds.].

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

Borges, A.V., 2005. Do we have enough pieces of the jigsaw to integrate CO$_2$ fluxes in the coastal ocean? *Estuaries*, 28:3-27.

Borges, A.V., B. Delille, and M. Frankignoulle, 2005. Budgeting sinks and sources of CO$_2$ in the coastal ocean: Diversity of ecosystems counts. *Geophysical Research Letters*, 32:L14601, doi:10.1029/2005GL023053.

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

Bratbak, G., M. Heldal, S. Norland and T. F. Thingstad, 1990. Viruses as partners in spring bloom microbial trophodynamics. Appl. And Env. Microbiol. 56, 1400-1405.

Cai, W-J., 2011. Estuarine and coastal ocean carbon paradox: CO$_2$ sinks or sites of terrestrial carbon incineration? Annu. Rev. Mar. Sci. 3:123-45, doi:10.1146/annurev-marine-120709-142723.

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., Z. Wang, and Y. Wang, 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., M. Dai, and Y. Wang, 2006. Air-sea exchange of carbon dioxide in ocean margins: A province based synthesis. *Geophysical Research Letters*, 33:L12603, doi:10.1029/2006GL026219.

Chavez, F., Takahashi, T., Cai, W-J., Friedrich, G., Hales, B., Wanninkhof, R. and Feely, R. A. (2007). Coastal oceans, Chapter 15, in "The First State of the Carbon Cycle Report (SOCCR): North American Carbon Budget and Implications for the Global Carbon Cycle". A report by the U.S. Climate Change Science Program and the Subcommittee on Global Change Research [King, A.W. L. Dilling. G.P. Zimmerman, D.M. Fairman, R.A.

Houghton, G.H. Marland, A.Z. Rose, and T.J. Wilbanks (eds.)] National Ocean and Atmospheric Administration, Climate Program Office, Silver Spring, MD, USA, 157-166, 193. Available at http://www.climatescience.gov/Library/sap/sap2-2/final-report/default.htm.

Cosca, C. E., R. A. Feely, J. Boutin, J. Etcheto, M. J. McPhaden, F. P. Chavez, and P. G. Strutton, 2003. Seasonal and interannual CO_2 fluxes for the central and eastern equatorial Pacific Ocean as determined from fCO_2-SST relationships

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

DeGrandpre, M.D., T.R. Hammar, G.J. Olbu, and C.M. Beatty, 2002. Air-sea CO_2 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 318.15K. Deep-Sea Res., 37, 755-766.

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

Ducklow, H.W., and S.L. McAllister, 2005. The biogeochemistry of carbon dioxide in the coastal oceans. In: The Global Coastal Ocean: Multiscale Interdisciplinary Processes [K.H. Brink and A.R. Robinson (eds.)]. The Sea, Volume 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., F.P. Wilkerson, V.E. Hogue, A. Marchi, 2006. Nutrient controls on new production in the Bodega Bay, California, coastal upwelling plume. Deep-Sea Research II 53, 3049–3062

Evans, W., B. Hales, and P. Strutton, 2011. The seasonal cycle of surface ocean pCO_2 on the Oregon shelf, *J. Geophys. Res. Oceans*, accepted.

Feely, R. A. T. Takahashi, R. Wanninkhof, M. J. McPhaden, C. E. Cosca, S. C. Sutherland, and M.-E. Carr, 2006. Decadal variability of the air-sea CO_2 fluxes in the equatorial Pacific Ocean. J. Geophys. Res. C—Oceans C08S90, doi:10.1029/2005JC003129.

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

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

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

1 Friedrich, T., and A. Oschlies 2009b. Neural network-based estimates of North Atlantic
2   surface pCO_2 from satellite data: A methodological study. J. Geophys. Res 114, C03020,
3   doi:10.1029/2007JC004646

4 Friedrich, T., and A. Oschlies 2009a. Basin-scale pCO_2 maps estimated from ARGO float
5   data: A model study. J. Geophys. Res 114, C10012, doi:10.1029/2009JC005322

6 Gruber, N., H. Frenzel, S. C. Doney, P. Marchesiello, J. C. McWilliams, J. R. Moisan, J. Oram, G.
7   K. Plattner, and K. D. Stolzenbach, 2006. Eddy-resolving simulation of plankton
8   ecosystem dynamics in the California Current System.   Deep-Sea Res. I, 53,
9   doi:10.1016/j.dsr.2006.06.005

10 Hales, B., D. Chipman and T. Takahashi, 2004. High-frequency measurement of partial
11   pressure and total concentration of carbon dioxide in seawater using microporous
12   hydrophobic membrane contactors. *Limnology and Oceanography: Methods* **2**, 356-
13   364.

14 Hales, B., T. Takahashi, and L. Bandstra, 2005. Atmospheric CO_2 uptake by a coastal
15   upwelling system, *Global Biogeochemical Cycles*, 19, doi:10.1029/2004GB002295.

16 Hales, B., W.-J. Cai, B. G. Mitchell, C. L. Sabine, and O. Schofield [eds.], 2008: *North American
17   Continental Margins: A Synthesis and Planning Workshop*. Report of the North
18   American Continental Margins Working Group for the U.S. Carbon Cycle Scientific
19   Steering Group and Interagency Working Group. U.S. Carbon Cycle Science Program,
20   Washington, DC, 110 pp.

21 Hauri, C., N. Gruber, S. Alin, V. J. Fabry, R. A. Feely, B. Hales, G.-K. Plattner and P. Wheeler.,
22   2009. Ocean acidification in the California Current system. *Oceanography* 22, 60-71.

23 Hedges, J. I., R. G. Keil, and R. Benner, 1997. What happens to terrestrial organic matter in
24   the ocean? *Organic Geochemistry*, 27:195-212.

25 Ho, D. T., C. S. Law, M. J. Smith, P. Schlosser, M. Harvey, and P. Hill, 2006. Measurements of
26   air-sea gas exchange at high wind speeds in the Southern Ocean: Implications for
27   global parameterizations. *Geophys. Res. Lett.* 33, L16611, doi:10.1029/2006GL026817

28 Ianson, D., R. A. Feely, C. L. Sabine, And L. W. Juranek, 2010. Features of coastal upwelling
29   regions that determine net air-sea CO_2 flux. *J. Oceanog.*, 65, 677 - 687

30 Ittekkot, V., and R.W.P.M. Laane, 1991. Fate of riverine particulate organic matter. In:
31   *Biogeochemistry of Major World Rivers* [Degens, E.T., S. Kempe, and J.E. Richey (eds.)].
32   John Wiley & Sons Ltd.

33 Kavanaugh, M. T., B. Hales, M. Saraceno, Y. H. Spitz, A. E. White, and R. M. Letelier, 2012.
34   Towards a quantitative framework for pelagic seascape ecology. *Ecosystems*, in
35   revision.

36 Kohonen, T. The Self-Organizing Map. Proceedings of IEEE 78, 1464–1480, 1990.

37 Kohonen, T., Self-Organizing Maps. Vol. 30. Springer Series in Information Sciences,
38   Springer-Verlag,, Berlin, Heidelberg, New York, 1995.

24

Lagerloef, G., F. R. Colomb, D. Le Vine, F. Wentz, S. Yueh, C. Ruf, J. Lilly, J. Gunn, Y. Chao, A. Decharon, G. Feldman, and C. Swift, 2008. The Aquarius/SAC-D mission: Designed to meet the salinity remote-sensing challenge. Oceanography 21, 68-81.

Lefevre, N., A. J. Watson, and A.R. Watson, 2005. A comparison of multiple regression and neural network techniques for mapping in situ pCO_2 data. Tellus 57B, 375–384

Lewis, E., and D. Wallace, 1998. Program developed for CO_2 system calculations. Oak Ridge National Laboratory Environmental Sciences Division Publication 4735, ORNL/CDIAC-105.

Lohrenz, S.E., and W.-J. Cai, 2006. Satellite ocean color assessment of air-sea fluxes of CO_2 in a river-dominated coastal margin. *Geophysical Research Letters*, 33:L01601, doi:10.1029/2005GL023942.

Longhurst, A. R., Ecological Geography of the Sea (Second Edition), 2006, Academic Press, ISBN: 978-0-12-455521-1, 560pp.

McKee, B.A., R.C. Aller, M.A. Allison, T.S. Bianchi, T.S. and G.C. Kineke (2004) Transport and transformation of dissolved and particulate materials on continental margins by major rivers: benthic boundary layer and seabed processes. *Cont. Shelf Res*. 24, 899-926.

Mecklenburg, S., Y. Kerr, J. Font, and A. Hahne, 2008. The Soil Moisture and Ocean Salinity (SMOS) Mission – An overview. *Geophysical Research Abstracts* 10, EGU2008-A-01922

Mehrbach, C., C. Culberson, J. E. Hawley, and R. M. Pytkowicz, (1973). Measurement of the apparent dissociation constants of carbonic acid in seawater at atmospheric pressure. *Limnol. and Oceanogr*., 18, 897-907.

Meybeck, M., and C. Vorosmarty (1999) Global transfer of carbon by rivers. *Global Change Newsletter* 37, 18-19.

d'Ovidio, F., S. De Monte, S. Alvain, Y. Dandonneau, and M. Lévy, 2010. Fluid dynamical niches of phytoplankton types, 2010, *Proceedings of the National Academy of Sciences*, doi:10.1073/pnas.1004620107

Park, G. H., R. Wanninkhof, S. C. Doney, T. Takahashi, K. Lee. R. A. Feely, C. L. Sabine, and J. Trinanes, 2010. Variability of global net air-sea CO_2 fluxes over the last three decades using empirical relationships. *Tellus*, in press.

Press, W. H., B. P. Flannery, S. A. Teukolsky, and W. T. Vetterling, 1989: *Numerical Recipes— The Art of Scientific Computing.* Cambridge University Press.

Prieto, L., J. Ruiz, F. Echevarria, C. M. Garcia, A. Bartual, J. A. Galvez, A. Corzo, and D. Macias, 2002. Scales and processes in the aggregation of diatom blooms: High time resolution and wide size range records in a mesocosm study. *Deep-Sea Res* I 49, 1233-1253.

Revelle, R., and H. E. Suess, 1957. Carbon Dioxide Exchange between Atmosphere and Ocean and the Question of an Increase of Atmospheric CO_2 During the Past Decades. *Tellus* **9**: 18-27.

Risien, C.M., and D.B. Chelton, 2008: A Global Climatology of Surface Wind and Wind Stress Fields from Eight Years of QuikSCAT Scatterometer Data. *J. Phys. Oceanogr.*, 38, 2379-2413.

Saraceno, M., C. Provost, and M. Lebbah, 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. and R. R. Schneider (2000) Transport of terrestrial organic carbon to the oceans by rivers: Re-estimating flux and burial rates. Internat. J. Earth Sci. **88**: 599-606.

Smith, S.V. and J.T. Hollibaugh (1993) Coastal metabolism and the oceanic organic carbon balance. *Rev. Geophys.*, **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. and 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 pCO_2, and net sea-air CO_2 flux over the global oceans. *Deep-Sea Res.*, doi:10.1016/j.dsr2.2008.12.009.

Telszewski, M., A. Chazottes, U. Schuster, A. J. Watson, C. Moulin, D. C. E. Bakker, M. Gonz´alez-D´avila,T. Johannessen, A. Kortzinger, H. Luger, A. Olsen, A. Omar, X. A. Padin, A. F. Rıos , T. Steinhoff, M. Santana-Casiano**,** D. W. R. Wallace, and R. Wanninkhof, 2009. Estimating the monthly pCO_2 distribution in the North Atlantic using a self-organizing neural network. *Biogeosciences*, 6, 1405–1421

Thomas, H., Y. Bozec, K. Elkalay, and H.J.W. de Baar, 2004. Enhanced open ocean storage of CO_2 from shelf sea pumping. *Science*, 304:1005-1008.

Tsunogai, S., S. Watanabe, and T. Sato, 1999. Is there a "continental shelf pump" for the absorption of atmospheric CO_2? *Tellus B*, 51:701-712.

Wanninkhof, R., 1992. Relationship between wind-speed and gas exchange over the ocean. *Journal of Geophysical Research*, 97:7373-7382.

Wanninkhof, R., A Olsen, and J. Triñanes, 2006. Air–sea CO_2 fluxes in the Caribbean Sea from 2002–2004. J. Mar. Syst. 66, 272-284.

Weiss, R. F. (1974). Carbon dioxide in water and seawater: The solubility of a non-ideal gas, Marine Chem., 2, 203-215.
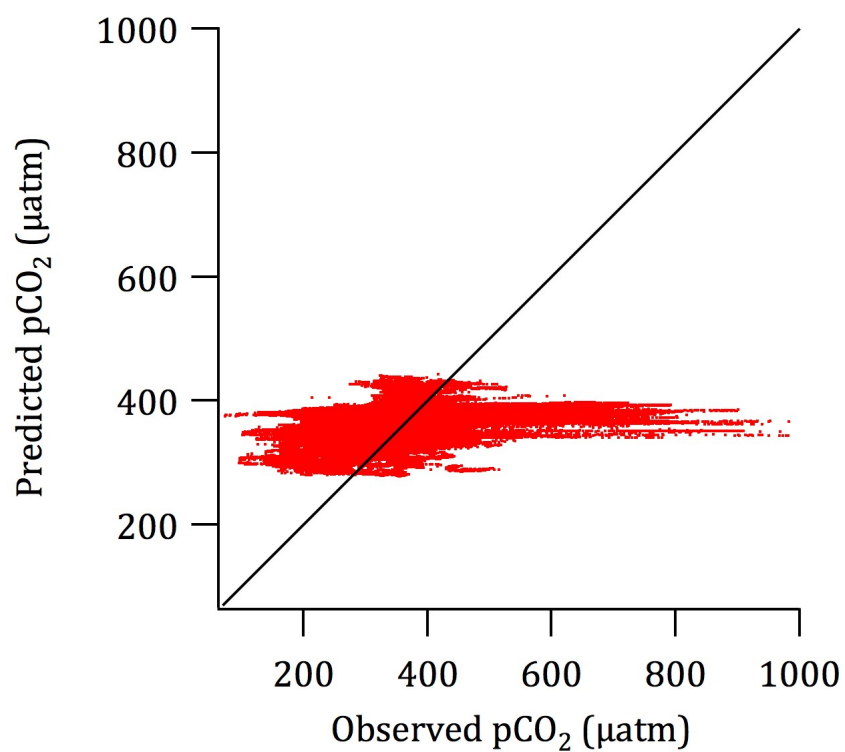
Figure 1. Multiple Linear Regression representation of all of the pCO_2 observations in the Chavez et al (2007) North American Continental Margins dataset. Solid diagonal line represents the perfect-agreement 1:1 relationship.
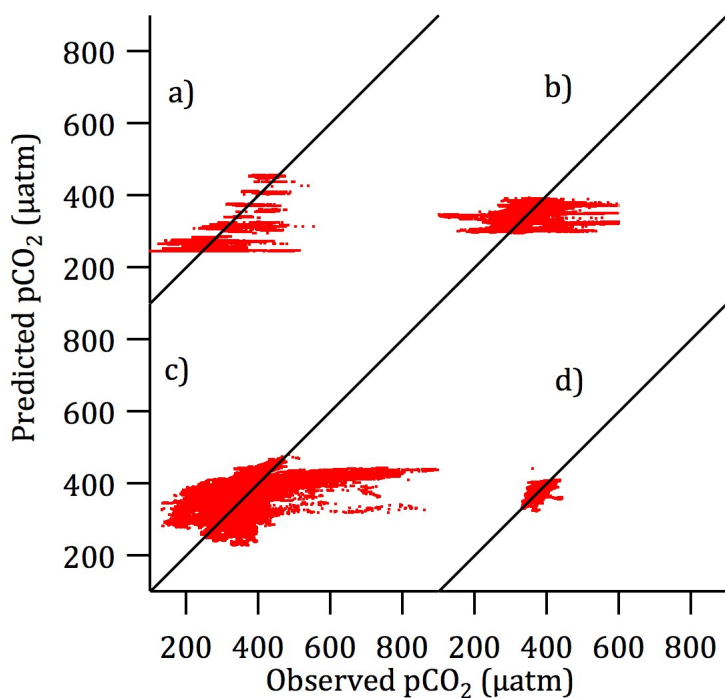
Figure 2. Comparison of Multiple Linear Regression prediction of sea-surface pCO_2 in the four major coastal sub-regions of North America: a) Bering Sea and Arctic coastal waters; b) the Atlantic coast; c) the Pacific coast; d) Gulf of Mexico and Caribbean coastal waters. Solid diagonal lines represent the perfect-agreement 1:1 relationship.
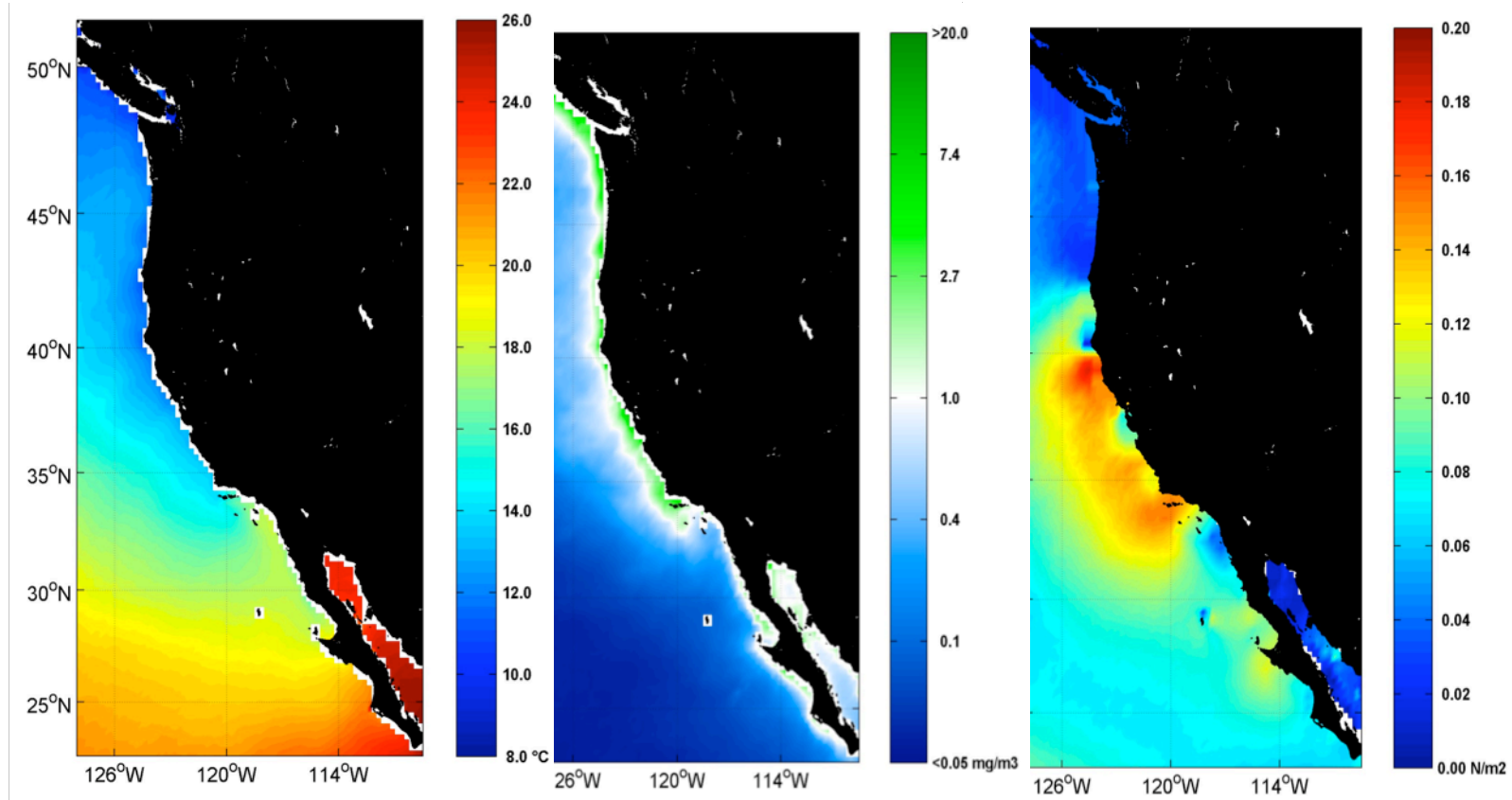
Figure 3. Input data used for generating the Self-Organizing Map (SOM) of Pacific coastal waters. Left, Sea-Surface Temperature (SST); Middle, Sea Surface chlorophyll (Chl); Right, net southward wind stress.
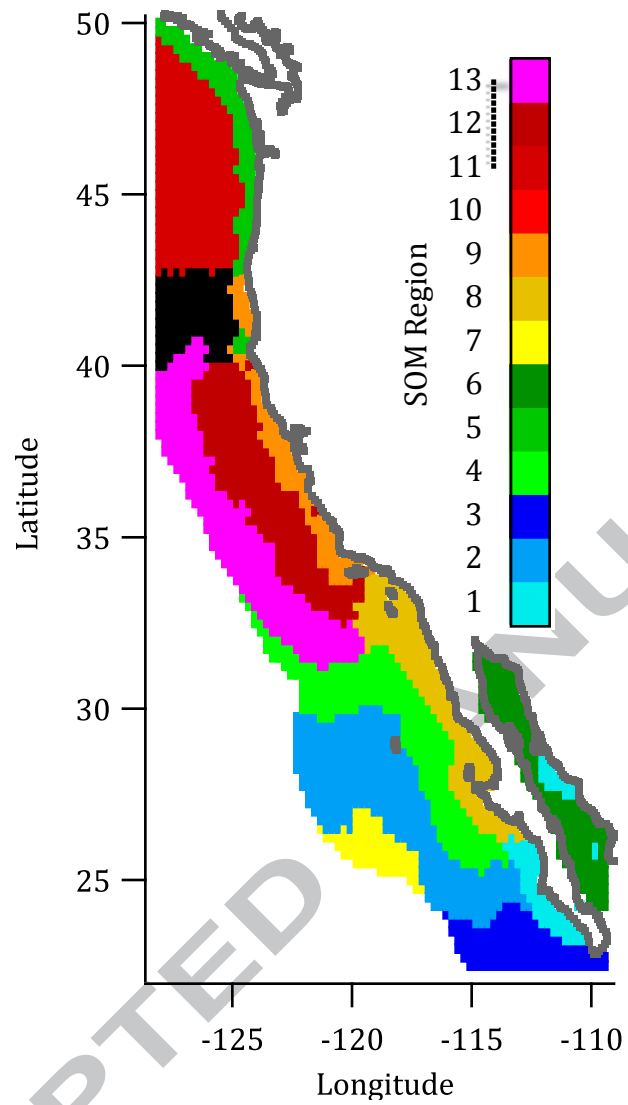
Figure 4. Self-organizing map of biogeochemical regions in the study area.
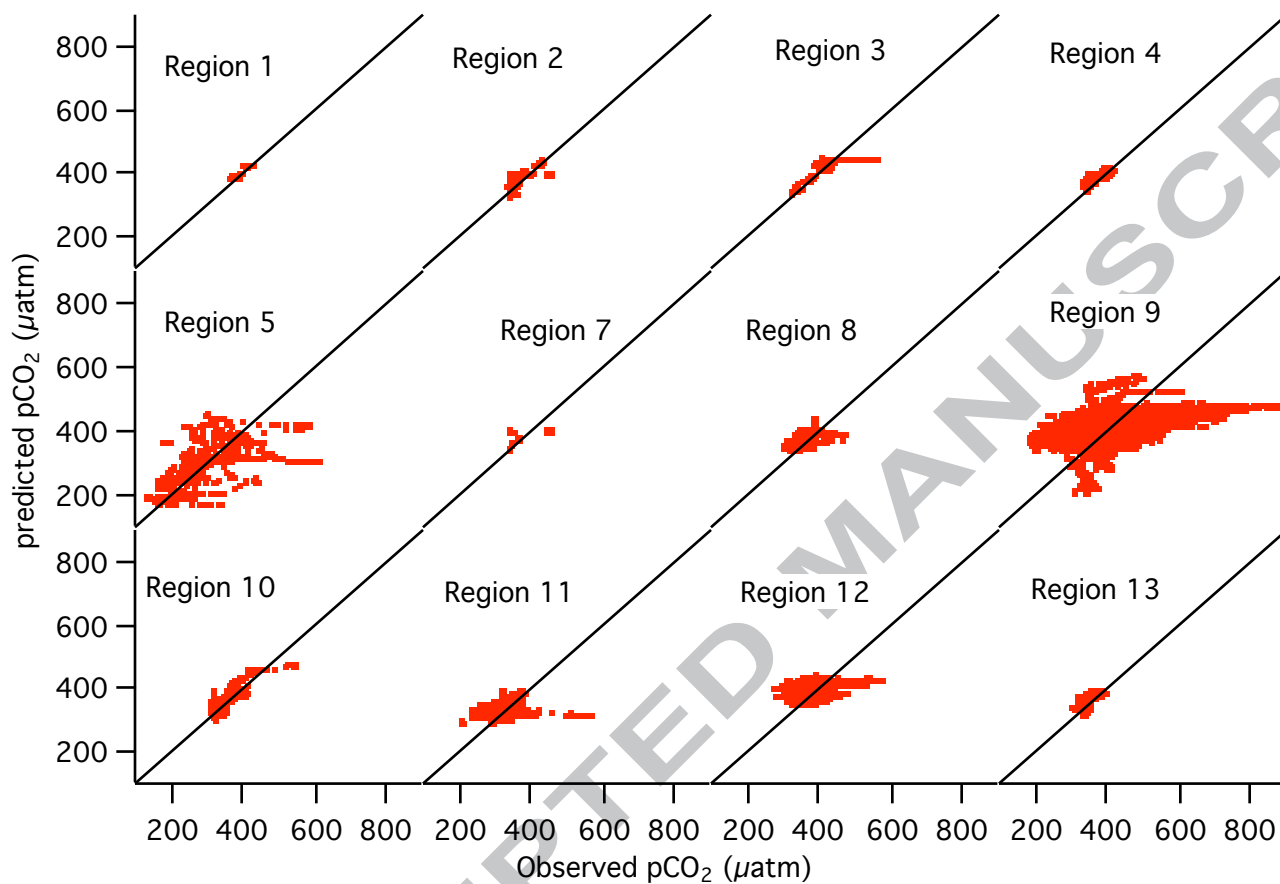
Figure 5. Region-by-region application of the MLR to the pCO_2 data in each SOM region. There were no observations in Region 6. Results are summarized in Table 2. Solid diagonal lines represent the perfect-agreement 1:1 relationship.
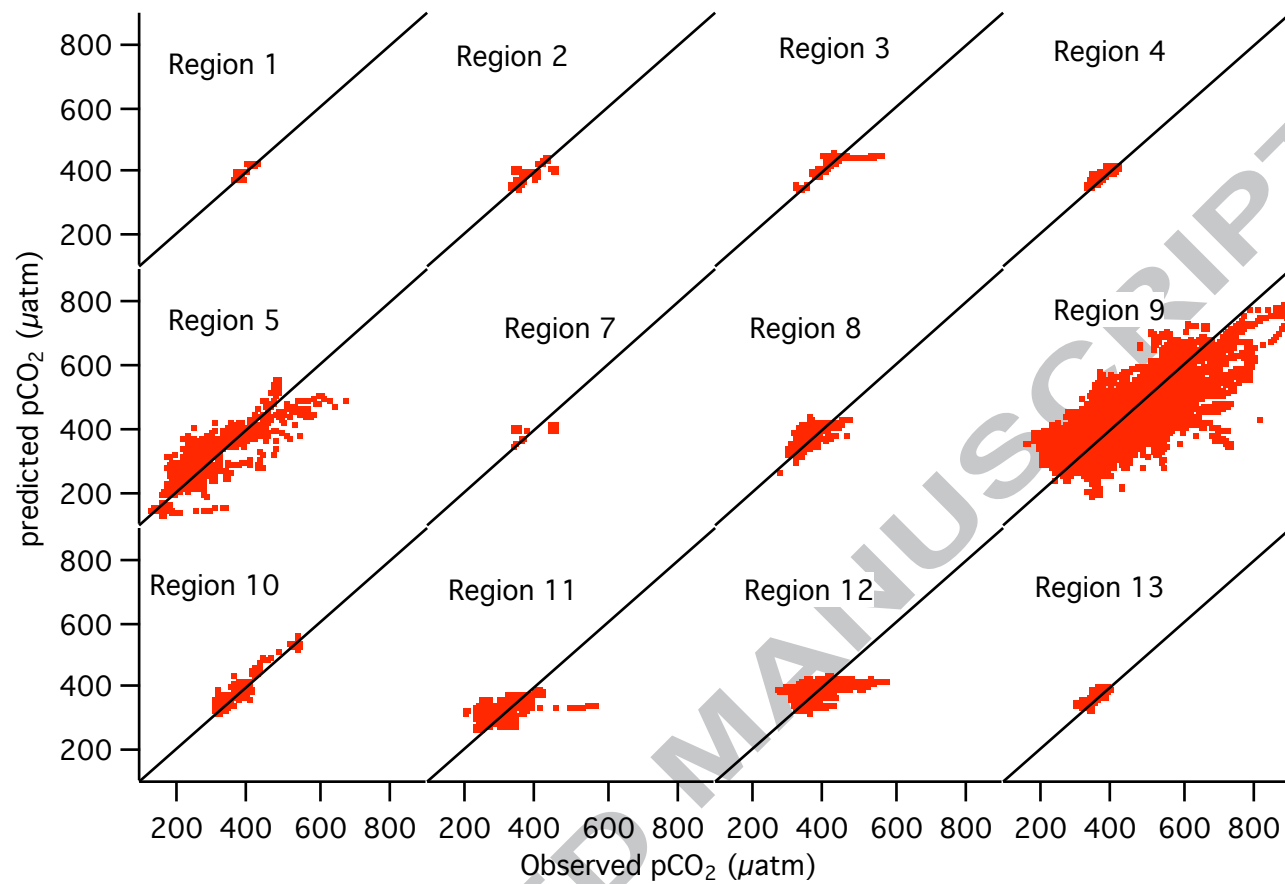
Figure 6. Region by region application of the nonlinear mechanistic model to the pCO_2 observations in each SOM region. Results are summarized in Table 3. Solid diagonal lines represent the perfect-agreement 1:1 relationsh
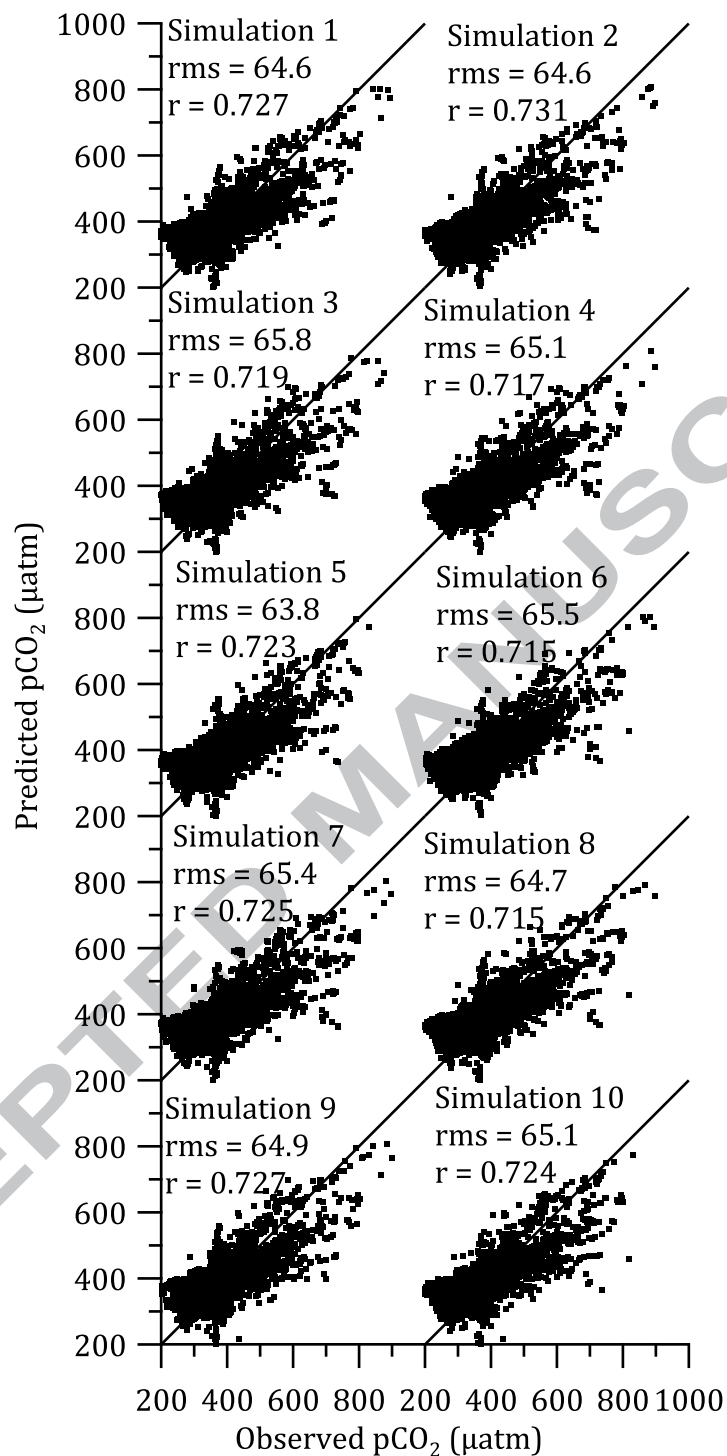
Figure 7. Demonstration of predictive capability of the algorithm, for ten randomly-selected groups of Region 9 data not included in algorithm training. See Table 5 for de
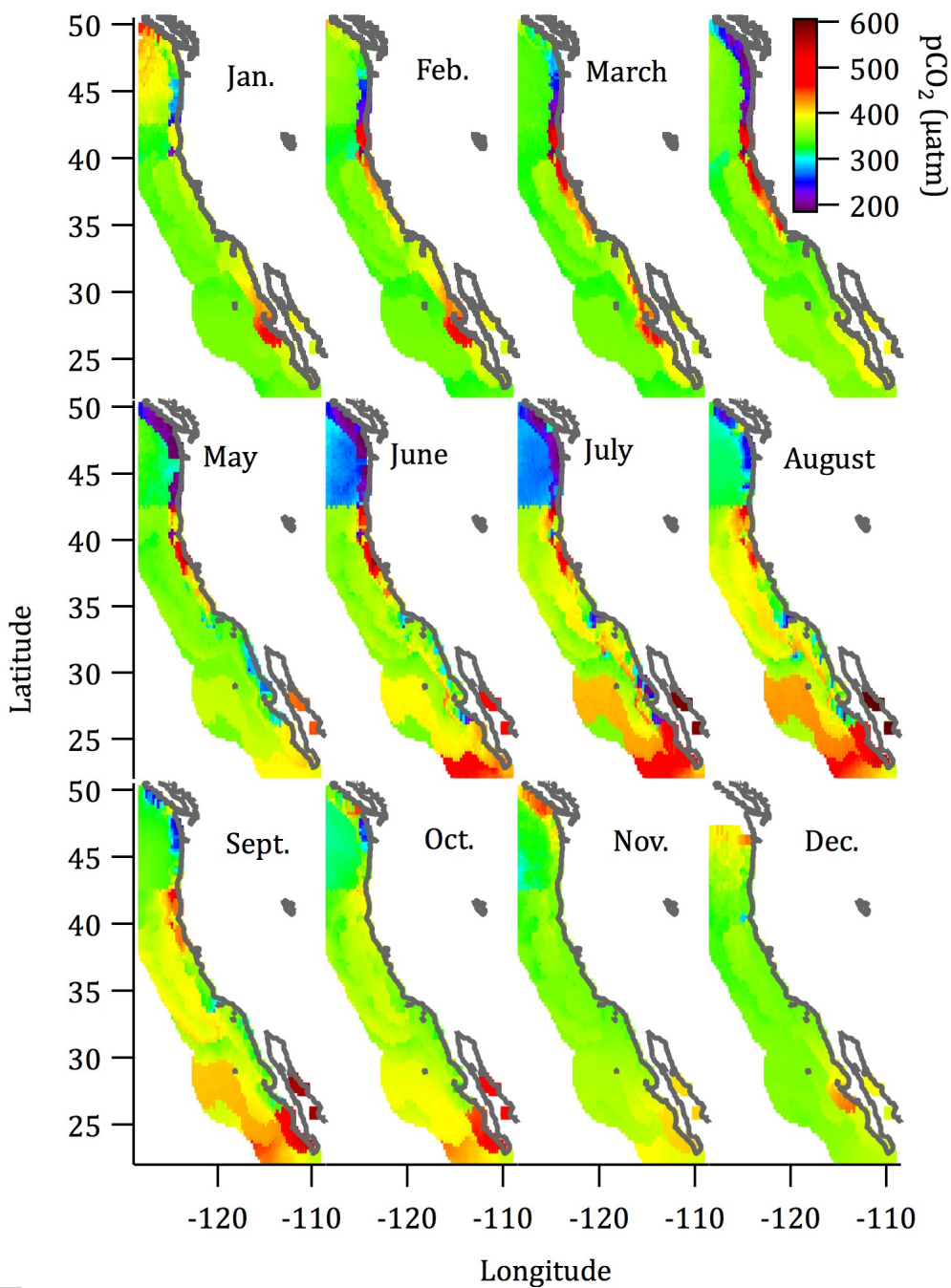
Figure 8. Monthly maps of surface pCO_2 distributions generated using the coefficients in Table 3 and monthly climatologies of SeaWiFS Chl and SST.
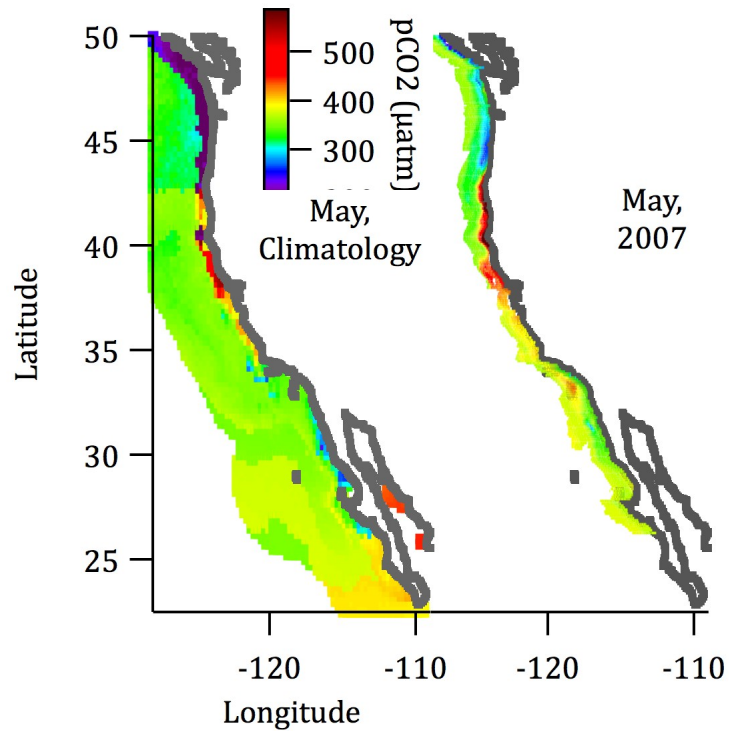
Figure 9. Comparison of algorithm-predicted surface pCO$_2$ distributions for a climatological May (left panel) with observed pCO$_2$ from a cruise from Vancouver, BC, to Baja. Mexico in May, 2007.
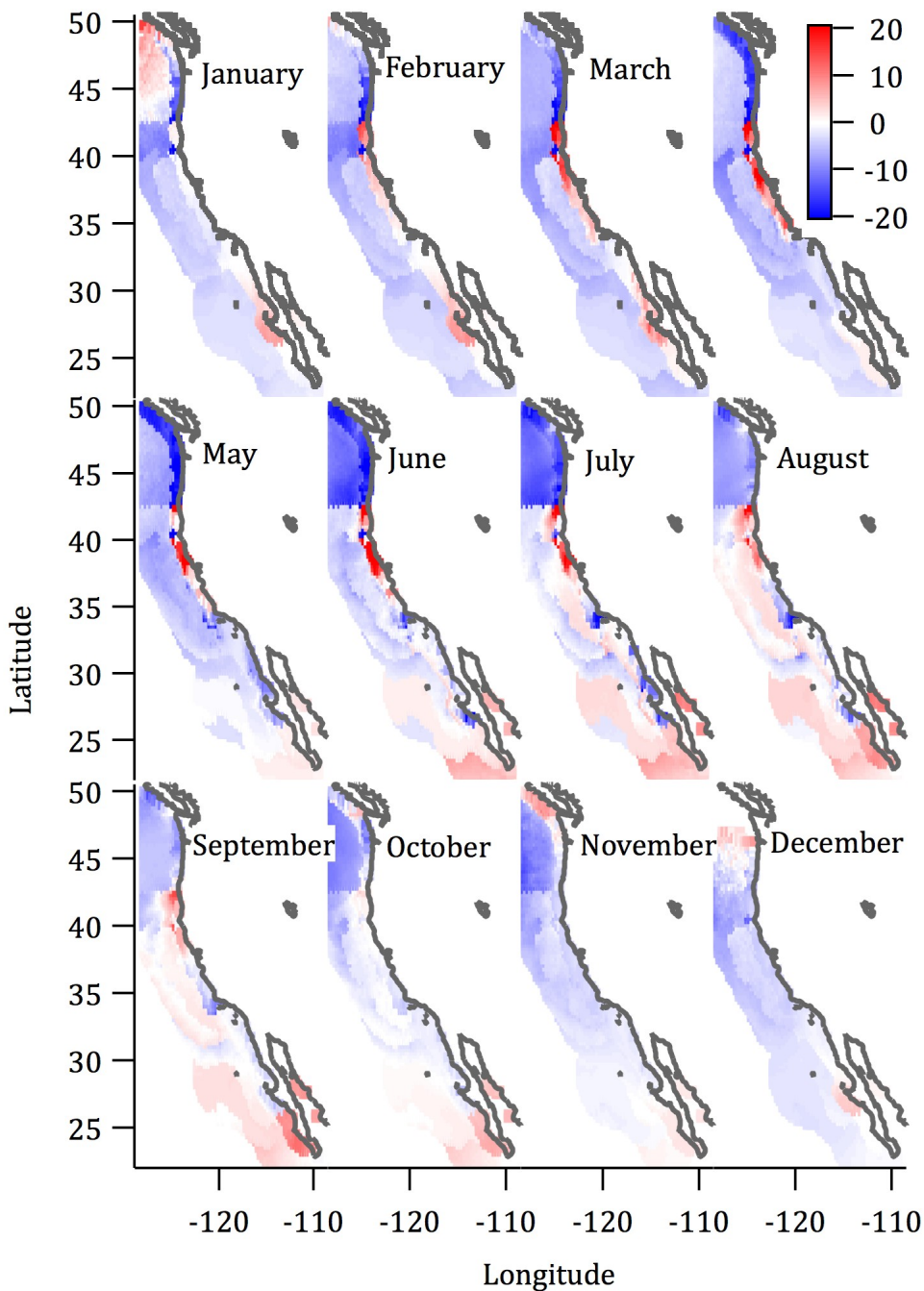
Figure 10. Monthly climatologies of air-sea CO_2 flux (mmol m^{-2} d^{-1}; negative numbers are into the ocean) generated form the pCO_2 maps of Figure 9 and the monthly climatologies of wind-speed squared taken from the Risien and Chelton database (http://cioss.coas.oregonstate.edu/scow/; compiled following Risien and Chelton 2008).

Table 1.

Region	Area (km^2)	Number of pCO$_2$ observations	Description
1	43924	684	Upwelling-influenced southern Baja and Sea of Cortez
2	261489	1093	3rd offshore from Baja
3	94406	959	Southern extreme
4	212661	2764	2nd offshore from Baja
5	46037	2139	Cascadia coastal waters; seasonal upwelling
6	106421	0	Sea of Cortez, non upwelling-influenced
7	182506	105	Baja, most offshore
8	113618	3106	Northern Baja coastal, upwelling
9	55240	48514	California coastal, persistent upwelling
10	75105	861	Northern California offshore
11	169190	3632	Cascadia offshore
12	175591	22542	Central California, 1st offshore
13	224407	9801	Central California, 2nd offshore

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, $pCO_{2,p} = C_0 + C_1 \Delta lat + C_2 \Delta lon + C_3 t + C_4 T + C_5 Chl$ 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.

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 CO_2 flux maps in Figures

Region	Alk^0 c_0	T_{CO2}^0 c_0	c_1	c_2	c_4	T^0 c_0	c_4	ϕ c_0	c_4	γ c_0	c_4	β c_0	c_4	e	RMS	R
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

Table 3, continued.

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

Table 5. Sensitivity analysis for Region 9

													Results With Training Data / Verification Data		
	Alk^0	T_{CO2}^0		T^0		ϕ		γ		e				Verification Data	
n	c_0	c_0	c_4	c_0	c_4	c_0	c_4	c_0	c_4		RMS	R	n	RMS	R
43599	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
43655	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
43616	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
43691	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
43794	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
43711	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
43689	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
43503	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
43648	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
43785	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

Research Highlights

1. In this study we assess multiple linear regression predictions of coastal-ocean pCO_2, and find them inadequate.
2. We apply a satellite-based self-organizing map to the Pacific coastal waters of Central North America and define 13 biogeochemically distinct regions between Baja, Mexico and Vancouver Island, Canada.
3. We develop a semi-mechanistic model of alkalinity and total CO_2 based on dependences on remotely-sensible parameters that captures the inherent non-linearity of the carbonate system in seawater, and this representation yields good predictive capability for coastal water pCO_2.
4. These predictions were combined with satellite wind products to calculate a net annual air-sea CO_2 exchange, and this predicts a larger magnitude sink for CO_2 than a previous synthesis.