High-resolution Neural Network Demonstrates Strong CO2 Source-Sink Juxtaposition in the Coastal Zone

Patrick J. Duke\(^1\), Roberta C Hamme\(^1\), Debby Ianson\(^2\), Peter Landschützer\(^3\), Neil C. Swart\(^4\), and Paul Covert\(^5\)

\(^1\)University of Victoria
\(^2\)Institute of Ocean Sciences
\(^3\)Flanders Marine Institute
\(^4\)Environment Canada
\(^5\)Institute of Ocean Sciences, Fisheries and Oceans Canada

March 21, 2024

Abstract

Coastal oceans may play an important role in regulating the concentration of carbon dioxide in the atmosphere. Quantification of carbon fluxes at this highly dynamic land-ocean interface will aid in monitoring, reporting, and verification for marine carbon dioxide removal. Here, we use a two-step neural network approach to generate basin-wide estimates from sparse observational data in the coastal Northeast Pacific Ocean at an unprecedented spatial resolution of 1/12\(^\circ\) with coverage in the nearshore (0 - 25 km offshore). We compiled partial pressure of carbon dioxide (pCO2) observations as well as a range of predictor variables including satellite-based and physical oceanographic reanalysis products. With the predictor variables representing processes affecting pCO2, we created non-linear relationships to interpolate observations from 1998-2019. Compared to in situ shipboard and mooring observations, our coastal pCO2 product captures broad spatial patterns and seasonal cycle variability well. A sensitivity analysis identifies that the parameters responsible for the neural network’s ability to capture regional pCO2 variability agrees with mechanistic processes. Using wind speed and atmospheric CO2, we calculated air-sea CO2 fluxes. We report an anticorrelation between net annual air-sea CO2 flux and air-sea CO2 flux seasonal amplitude and suggest the relationship is driven by regional processes. We show the inclusion of nearshore net outgassing fluxes lowers the overall regional net flux. Overall, our results suggest that the region is a net sink (-0.7 mol m\(^{-2}\) yr\(^{-1}\)) for atmospheric CO2 with trends indicating increasing oceanic uptake due to strong connectivity to subsurface waters.

Hosted file


Hosted file

High-resolution Neural Network Demonstrates Strong CO₂ Source-Sink Juxtaposition in the Coastal Zone

P. J. Duke¹, R. C. Hamme¹, D. Ianson¹,², P. Landschützer³, N. C. Swart¹,⁴, P. A. Covert²

¹School of Earth and Ocean Sciences, University of Victoria, Victoria, BC, Canada
²Institute of Ocean Sciences, Fisheries and Oceans Canada, Sidney, BC, Canada
³Flanders Marine Institute (VLIZ), Ostend, Belgium
⁴Canadian Centre for Climate Modelling and Analysis, Environment and Climate Change Canada, Victoria, BC, Canada

Corresponding author: P. J. Duke (pjduke@ucalgary.ca)

Key Points:

• The coastal Northeast Pacific is a net sink for atmospheric CO₂ with increasing air-sea $p$CO₂ disequilibrium trends in most of the region.
• Regional processes drive net annual air-sea CO₂ flux to be anticorrelated with air-sea CO₂ flux seasonal amplitude.
• Estimated $p$CO₂ reproduces observed seasonal cycle phase and amplitude well along with broad spatial patterns of variability.
Abstract
Coastal oceans may play an important role in regulating the concentration of carbon dioxide in the atmosphere. Quantification of carbon fluxes at this highly dynamic land-ocean interface will aid in monitoring, reporting, and verification for marine carbon dioxide removal. Here, we use a two-step neural network approach to generate basin-wide estimates from sparse observational data in the coastal Northeast Pacific Ocean at an unprecedented spatial resolution of 1/12° with coverage in the nearshore (0 - 25 km offshore). We compiled partial pressure of carbon dioxide ($pCO_2$) observations as well as a range of predictor variables including satellite-based and physical oceanographic reanalysis products. With the predictor variables representing processes affecting $pCO_2$, we created non-linear relationships to interpolate observations from 1998-2019. Compared to in situ shipboard and mooring observations, our coastal $pCO_2$ product captures broad spatial patterns and seasonal cycle variability well. A sensitivity analysis identifies that the parameters responsible for the neural network’s ability to capture regional $pCO_2$ variability agrees with mechanistic processes. Using wind speed and atmospheric CO$_2$, we calculated air-sea CO$_2$ fluxes. We report an anticorrelation between net annual air-sea CO$_2$ flux and air-sea CO$_2$ flux seasonal amplitude and suggest the relationship is driven by regional processes. We show the inclusion of nearshore net outgassing fluxes lowers the overall regional net flux. Overall, our results suggest that the region is a net sink (-0.7 mol m$^{-2}$ yr$^{-1}$) for atmospheric CO$_2$ with trends indicating increasing oceanic uptake due to strong connectivity to subsurface waters.

Plain Language Summary
The importance of the coastal ocean as a hub of exchange for carbon between terrestrial ecosystems, the open ocean, and the atmosphere is still unclear. In this study, we investigate how much carbon dioxide moves between the ocean and the atmosphere in the coastal Northeast Pacific. We use a mathematical technique (i.e., machine learning) to transform limited observational data to a high-resolution estimate of this exchange across the entire region. We found this method effectively captured the big picture patterns and seasonal changes in ocean carbon dioxide levels. We report that the coastal Northeast Pacific absorbs slightly more carbon dioxide than it releases, helping regulate atmospheric levels of this greenhouse gas. However, there are large differences regionally with some coastal zones absorbing substantial amounts of carbon dioxide and others releasing the gas, such as the nearshore. We report a trend of increasing ocean uptake over time, suggesting the region may play an increasingly important role...
in reducing atmospheric carbon dioxide levels. This study provides valuable baseline information for efforts to reduce carbon dioxide in the atmosphere through artificially enhancing ocean uptake in the region.

1 Introduction

The global ocean takes up nearly a quarter of anthropogenic carbon dioxide (CO$_2$) emissions annually (Friedlingstein et al., 2023). It has been suggested coastal oceans contribute disproportionately to oceanic CO$_2$ uptake relative to global ocean by surface area (Bourgeois et al., 2016; Chau et al., 2022; Laruelle et al., 2014; Resplandy et al., 2024; Roobaert et al., 2019, 2024), but exhibit far greater heterogeneity in air-sea CO$_2$ fluxes (Liu et al., 2010) and may be changing at a different rate compared to the open ocean (Laruelle et al., 2018; Resplandy et al., 2024). Coastal oceans serve as an important hub of exchange, outgassing carbon delivered by terrestrial ecosystems to the ocean (Regnier et al., 2022), while facilitating transport between the coast and open ocean, and directly absorbing CO$_2$ from the atmosphere (Bauer et al., 2013; C.-T. A. Chen & Borges, 2009; Mackenzie et al., 1998; Ward et al., 2020). However, the role of the coastal ocean in the global carbon budget is not well-constrained due to lack of observations relative to the complexity of highly localized variability (Chavez et al., 2007; Dai, 2021; Dai et al., 2022).

Gap filling approaches (i.e., methods to interpolate sparse observations) used to inform coastal ocean air-sea CO$_2$ flux estimates are often at coarse resolution and often operate as a “black box.” Interpolation techniques have been widely used to inform air-sea CO$_2$ flux estimates in the coastal ocean both regionally and globally (e.g., S. Chen et al., 2016; Hales et al., 2012; Laruelle et al., 2017; G. Parard et al., 2015; Gaëlle Parard et al., 2016; Roobaert et al., 2019, 2024; Sharp et al., 2022; Xu et al., 2019). These approaches extend the temporal and spatial coverage of partial pressure of CO$_2$ in seawater (pCO$_2$) observations from community synthesis efforts (e.g., through the Surface Ocean CO$_2$ Atlas (SOCAT); Bakker et al., 2016) and can be used to calculate air-sea CO$_2$ fluxes using wind speed and atmospheric CO$_2$ (Wanninkhof, 2014). Historically, coastal ocean approaches have been adopted from their open ocean counterparts (Chau et al., 2022; Landschützer, Laruelle, et al., 2020), and thus most of these estimates have at best a monthly, 1/4°x1/4° latitude by longitude resolution, which is incapable of resolving smaller scale processes in coastal regions, especially nearshore, that experience high
variability and short autocorrelation length scales (Jones et al., 2012). Interpolation techniques, which lack transparency, also rarely probe internal relationship dependency between variables.

Large heterogeneity in air-sea CO₂ fluxes exist in the coastal Northeast Pacific, with substantial expanses of the coast completely absent of observations (Benway et al., 2016). Large discrepancies exist between previous air-sea CO₂ flux estimates within this region, with disagreement over the net annual flux magnitude and direction (i.e., as a net sink or source for atmospheric CO₂; Duke, Richaud, et al., 2023; Fennel et al., 2019). Air-sea CO₂ flux variability in the region is heavily impacted by coastal processes such as upwelling, river plumes, tidal mixing, and coastal currents (Evans et al., 2012, 2019; Evans & Mathis, 2013; Hales et al., 2005; Ianson et al., 2003; Nemcek et al., 2008). Upwelling along the Pacific eastern boundary shelf has contrasting impacts on the oceanic CO₂ sink reflected in complex interactions between biological carbon drawdown fueled by upwelled nutrient and carbon-rich waters (Hales et al., 2005; Messié & Chavez, 2015; Ribalet et al., 2010) and outgassing associated with the same subsurface waters brought to the surface (Chan et al., 2017; Christensen, 1994; Evans et al., 2011; Feely et al., 2008; Hales et al., 2005; Ianson & Allen, 2002). Closer to shore, within the Salish Sea, and along Alaska’s Inside Passage, air-sea CO₂ fluxes into and out of the ocean are highly episodic and spatially heterogeneous (Evans et al., 2022; Jarníková, Ianson, et al., 2022). Binning regional pCO₂ observations in three dimensions into monthly, 1/12°x1/12° grid cells over the period 1998–2019, reveals the data scarcity (Figure 1). Of the 6,030,816 spatial and temporal grid cells just 0.6% have an associated gridded pCO₂ value. Observations are concentrated along shipping lanes, have a summer bias, and increase in frequency during later years (Figure 1). No observations exist in vast areas of the coastal Gulf of Alaska and along extensive stretches of shoreline (Figure 1c).
Figure 1. Number of grid cells (of 54782 total spatial grid cells) with coastal $p$CO$_2$ observation data (Section 2.1) in (a) months reveals a summer bias, and (b) years showing increased sampling closer to present. (c) Total number of months of observational coverage per grid cell displays better coverage along shipping routes. 300 km offshore line shown for coastal/open oceanic boundary used in this study (solid blue line labelled ‘300’).

Here we investigate how well a high-resolution regional artificial neural network (ANN) approach can determine air-sea CO$_2$ fluxes in the coastal Northeast Pacific (NEPc). We build on an existing global setup (Landschützer et al., 2013) adopted previously in stepping to a higher spatial resolution in the open Northeast Pacific (Duke, Hamme, et al., 2023b). In Section 2, we describe the creation of a gridded $p$CO$_2$ data product for the coastal Northeast Pacific monthly from January 1998 to December 2019 at an unprecedented $1/12^\circ \times 1/12^\circ$ resolution to resolve coastal ocean processes. In Section 3, we demonstrate that our product robustly recreates gridded
observation data, comparable to a less variable open ocean product. In Section 4, we directly
compare our $p$CO$_2$ product with *in situ* shipboard and mooring observations and detail potential
capabilities and limitations in the continuous, gridded product. In Section 5, we examine the
regional patterns of variability in the net annual air-sea CO$_2$ flux relative to the seasonal cycle
and describe potential drivers through a spatial sensitivity analysis. We conclude by calculating
surface ocean $p$CO$_2$ trends in the last decades.

2 Data and methods

We created a coastal $p$CO$_2$ data product spanning a geographic area between 45°-62°N
and 120°-155°W, and within 6 to 300 km of shore building on the methods of Duke, Hamme, et
al. (2023b). (ANN-NEPc; Duke et al., 2024). Briefly, our first step identified grid cells with
similar environmental characteristics, provinces, using a self-organizing map approach (SOM)
(Landschützer et al., 2013). In the second step, within each province, we used a feed-forward
neural network (FFN) to create non-linear functional relationships between $p$CO$_2$ observations
and independent predictor variables (Landschützer et al., 2013). Third, we applied these
relationships to the predictor data to generate continuous monthly sea surface $p$CO$_2$ maps from
1998-2019 in the coastal Northeast Pacific (NEPc). ANN-NEPc fills the gap between open ocean
(> 300 km offshore) $p$CO$_2$ (Duke, Hamme, et al., 2023b) to as close to the shoreline as reanalysis
and satellite-based products reach. In stepping to 1/12° spatial resolution (approximately 9 km by
5 km, latitude by longitude), this work represents a three times increase in spatial resolution over
previous 1/4° global and regional coastal ocean products with an overlapping domain
(Landschützer, Laruelle, et al., 2020; Laruelle et al., 2017; Roobaert et al., 2024; Sharp et al.,
2022), with extended coverage into the nearshore (defined here as 0 - 25 km offshore).

2.1 $p$CO$_2$ observations

ANN target $p$CO$_2$ data came from the Surface Ocean CO$_2$ Atlas (SOCAT) v2021 (Bakker
et al., 2016), the Fisheries and Oceans Canada February 2019 Line P cruise
(https://www.waterproperties.ca/linep/), a West Coast Ocean Acidification cruise from July and
August 2010 (Evans et al., 2012), and La Perouse cruises from May 2007 and May 2010 (Tortell
et al., 2012). Sea surface CO$_2$ fugacity ($f$CO$_2$) was converted to sea surface $p$CO$_2$ (supplementary
Text S1; Körtzinger, 1999). We did not correct *in situ* $p$CO$_2$ observations to sea surface mass
boundary layer temperature, because following previous techniques introduced significant
additional uncertainty in our coastal study area (supplementary Text S2). $p$CO$_2$ observations were bin-averaged (monthly from 1998–2019, at 1/12°x1/12°), computing the mean and standard deviation within each grid cell.

2.2 Predictor data

Predictor data were chosen based on accessibility and ability to represent processes that mechanistically impact surface ocean $p$CO$_2$ (Table 1). Selected predictor variables primarily originate from satellite observations or reanalysis models (Table 1; supplementary Text S3). Predictors differ slightly from a regional open ocean estimate (Duke, Hamme, et al., 2023b). Here, we used a high-resolution regional wind speed product and not reanalysis model derived mixed layer depth. Capturing greater variability in the coastal ocean required a high-resolution regional wind speed product over a low-resolution global product (supplementary Figure S2). Latitude, longitude, and time were not used as predictor variables.
Table 1. Northeast Pacific Coastal Ocean artificial neural network predictor variables, and their corresponding source, original temporal and spatial resolutions, and processing steps used for this study.

<table>
<thead>
<tr>
<th>Predictor variable</th>
<th>Source</th>
<th>Original resolution</th>
<th>Processing</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Temporal</td>
<td>Spatial</td>
</tr>
<tr>
<td>Satellite-based product</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sea surface temperature  (SST)</td>
<td>SST_cci: Level 4 Analysis Climate Data Record, version 2.1</td>
<td>Daily</td>
<td>1/20°x1/20°</td>
</tr>
<tr>
<td>Chlorophyll-a (Chl)</td>
<td>OceanColour_cci: Version 5.0</td>
<td>Daily</td>
<td>1/24°x1/24°</td>
</tr>
<tr>
<td>Satellite and in-situ observation data assimilated reanalysis product</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sea surface salinity (SSS)</td>
<td>Copernicus Marine Service GLOBAL_REANALYSIS_PHY_001_030</td>
<td>Monthly</td>
<td>1/12°x1/12°</td>
</tr>
<tr>
<td>Sea surface height (SSH)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Atmospheric-measurement-based interpolation product</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Atmospheric pCO₂</td>
<td>Landschützer et al. (2020b) - NCEI Accession 0160558</td>
<td>Monthly</td>
<td>1°x1°</td>
</tr>
<tr>
<td>High-resolution regional forecast model</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Wind speed</td>
<td>Regional Deterministic Reforecast System (RDRS-v2.1)</td>
<td>Hourly</td>
<td>1/11°x1/11°</td>
</tr>
</tbody>
</table>

2.3 Neural network construction

To reach the optimal ANN-NEPc architecture, we performed a series of tuning tests using the MATLAB Neural Network Toolbox, with sequential improvements impacting future tests (Duke, Hamme, et al., 2023b). The choice of three dynamic (i.e., changing shape at every timestep) self-organizing map (SOM) based clusters represented the lowest number for a typical clustering structure to emerge (supplementary Figure S3a). All spatial grid cells within the study area belong to more than one SOM cluster at some point over 1998-2019 (supplementary Figure S3b). SOM predictor variables (SST, SSS, SSH only; Table 1) were normalized to a mean of 0 and standard deviation of 1. The second FFN step used all six predictor variables in Table 1, in addition to each predictor variable anomaly (i.e., deseasonalized; calculated by subtracting the climatological monthly mean), bringing the total number of predictors to 12. Anomaly values were used to highlight interannual to decadal variability within our predictor data sets. The
number of neurons within the first hidden layer varied by province with the optimal number of neurons determined in a pre-training run (Landschützer et al., 2013, 2014). The second hidden layer used seven static neurons, which slightly improved performance. To further decrease the risk of overfitting, we used a 10-fold cross-evaluation approach to create an ANN ensemble (Duke, Hamme, et al., 2023b; Li et al., 2019, 2020) and a bootstrapping method (Landschützer et al., 2013). Observation cruises were randomly divided into 10 equal subsamples (10% each) using expocodes (i.e., unique identifiers corresponding to complete underway cruise tracks or mooring deployments) prior to gridding, leaving some data splits with more (or less) gridded $p$CO$_2$ targets (Section 2.1). We repeated the FFN training step 10 times, using each of the 10 subsamples once as the internally withheld evaluation dataset and the rest as the training dataset (with a separate independent data always withheld; Section 2.4). In each iteration, we trained the ANN for 10 rounds. The robustness and reliability of an ANN estimate has been shown to be significantly improved by combining a ANN ensemble (Duke, Hamme, et al., 2023b; Fourrier et al., 2020; Linares-Rodriguez et al., 2013; Sharkey, 1999). Here, we take the mean of the 10-fold estimates.

2.4 Evaluation

Comparisons of ANN output to training and independent withheld data were made throughout tuning tests. ANN-NEPc performance for each tuning test was evaluated using five statistical metrics: root mean squared error (RMSE), coefficient of determination ($r^2$), mean absolute error (MAE), mean bias (calculated as the mean residual), and the slope of the linear regression ($c_1$) between the ANN and the corresponding gridded $p$CO$_2$ observations. One subset of data was selected from the observation data using associated expocodes to be entirely withheld from the FFN training step. We tested 100 random independent withheld data splits and selected the one with the best observational coverage over a wide range of seasons, years, and locations (supplementary Figure S4). These independent withheld data represented approximately 4.5% of the total study area gridded $p$CO$_2$ data.

2.5 Sensitivity analysis

We used a perturbation approach to quantitatively assess the impact of each predictor variable on estimated $p$CO$_2$ (e.g., Broullón et al., 2018; Li et al., 2020; Sun et al., 2021). To diagnose how important different predictor variables were across the study area, a single set of
non-linear relationships was used inside a single uniform SOM cluster. We then applied this single FFN to our continuous, gridded predictor dataset and to perturbed versions of that dataset. For each predictor variable separately, we introduced a perturbation increasing the value within each grid cell by 50% of the standard deviation within that grid cell ($X' = X + 0.5(\text{std}(X)); N = 264$ months per grid cell; de Oña & Garrido, 2014) and calculated the resulting predicted $p$CO$_2$. We then took the difference between the perturbed run and a baseline run using unperturbed predictor variables.

2.6 Computation of air-sea fluxes

Using our $p$CO$_2$ product, we calculated the air-sea CO$_2$ flux ($F$CO$_2$; mol m$^{-2}$ yr$^{-1}$):

$$F_{CO_2} = K_0 k \Delta p_{CO_2},$$

from the Henry’s Law solubility constant ($K_0$; mmol m$^{-3}$ μatm$^{-1}$) as a function of temperature and salinity (Table 1; Weiss, 1974), gas transfer velocity ($k$; m day$^{-1}$), and the gradient between $p$CO$_2$ in the surface ocean and the atmosphere ($\Delta p_{CO_2}$; μatm). Here, the gas transfer velocity is derived from Wanninkhof (2014), a function of wind-speed at 10 meter elevation (Table 1) and the temperature dependent Schmidt number specific to CO$_2$ (Wanninkhof, 2014). Negative flux values indicate CO$_2$ uptake by the ocean. We assume that the uncertainty in our air-sea CO$_2$ flux estimate results from a 20% uncertainty in $k$ (Wanninkhof, 2014) and the overall product uncertainty in estimated $p$CO$_2$ ($\theta p_{CO_2}$; Section 3.3 below). As the uncertainty of $\Delta p_{CO_2}$ is dominated by the uncertainty in estimated surface ocean $p$CO$_2$, we neglect the small contribution from atmospheric CO$_2$ (< 1 μatm; Landschützer et al., 2014).

3 Network performance

3.1 Evaluation with respect to observational data

Comparing our estimated $p$CO$_2$ product with the gridded observations across both the training data (Figure 2a) and independent withheld data (Figure 2b) demonstrates fits with an MAE less than 30 μatm and RMSE of around 40 μatm. The mean bias is negligible over the full range (< 0.2 μatm, smaller than observational uncertainty; Section 3.3). 70% of the calculated residuals fall within the -20 to 20 μatm range, while 47% of the grid cells have absolute residuals < 10 μatm especially further offshore (supplementary Figure S5). Despite seasonal and annual biases in observations (Figure 1; Section 2.1), our product performs similarly across different
months and years (supplementary Table S1). The ANN ensemble model mean demonstrated improved performance compared to each individual ensemble member (supplementary Figure S6; supplementary Text S4).

Larger bias exists at the upper and lower limits of the gridded $pCO_2$ observational range. Our product underestimates $pCO_2$ observations greater than the 90\textsuperscript{th} percentile (> 412 \mu atm; mean bias = -28 \mu atm), and overestimates values less than the 10\textsuperscript{th} percentile (< 306 \mu atm; mean bias = 13 \mu atm). The spatial structure of the residuals reflects this bias distribution (supplementary Figure S5), with negative residuals in the strong mixing regions of the Salish Sea commonly characterized by high $pCO_2$ (Evans et al., 2012, 2019; Jarníková, Ianson, et al., 2022), and positive residuals along the upwelling zone off the west coast of Oregon and Washington States characterized by low $pCO_2$ (Evans et al., 2011). Observation-based $pCO_2$ products commonly overestimate $pCO_2$ in highly biologically productive coastal upwelling regions (Chau et al., 2022; Hales et al., 2012; Roobaert et al., 2024; Sharp et al., 2022). Chlorophyll (Table 1) as a proxy for biological productivity in training may not fully represent biological control on $pCO_2$. Ford et al. (2022) showed that in regions with high biological activity and nutrients supplied from depth (i.e., South Atlantic upwelling mesoscale eddies) regional, algorithm-derived net community production estimates (Ford et al., 2021) improved ANN $pCO_2$ estimates. Creation of coastal, regionally specific net community production algorithms, and inclusion as a predictor variable, may help reduce bias of low $pCO_2$ values in our study area.
Figure 2. Our ensemble mean $pCO_2$ estimate (ANN-NEPC) against (a) observed $pCO_2$ training data, (b) observed $pCO_2$ independently withheld data, and (c) individual ensemble member estimates. Data are binned into 5 μatm by 5 μatm bins with data density shown in the colorbar on a log scale (note order of magnitude difference between panels). Dashed black line is the 1:1. Dotted blue line is the least squares best fit. Also shown are number of observations (N), root mean squared error (RMSE), coefficient of determination ($r^2$), mean absolute error (MAE), mean bias (calculated as the mean residual), and the slope of the linear regression ($c_1$).

In relative terms, our $pCO_2$ product performs nearly as well as an open ocean product, even nearshore (Table 2). Nearshore $pCO_2$ exhibits a much larger range of variability compared to the continental shelf and the offshore marine environment. Table 2 displays relative percent error (RPE) binned by distance offshore ($d$) calculated as:

$$RPE_d = \frac{RMSE_d}{[prctile_5(pCO_2^{obs}_d) - prctile_5(pCO_2^{bs}_d)]} \times 100,$$

(2)
where $\text{RMSE}_d$ is the RMSE from gridded observational data averaged over the distance bin, $\text{prctile}_{95}(pCO_2^{obs})$ is the $95^{\text{th}}$ percentile observed $pCO_2$ in that distance bin and $\text{prctile}_5(pCO_2^{obs})$ is the $5^{\text{th}}$ percentile. Compared to a high-performance, regional open ocean product (Table 2; Duke, Hamme, et al., 2023a), RMSE increases moving toward shore but so does the range in $pCO_2$ such that the RPE is constant within a factor of two.

Table 2. Error statistics for our ensemble mean $pCO_2$ estimate against all gridded observation data binned by distance offshore: number of observations (N) per bin, observed range of variability (range; difference between the $95^{\text{th}}$ and $5^{\text{th}}$ percentile), root mean squared error (RMSE), and relative percent error (RPE; Eq. 2).

<table>
<thead>
<tr>
<th>Distance offshore (km)</th>
<th>N</th>
<th>Range ($\mu$atm)</th>
<th>RMSE ($\mu$atm)</th>
<th>RPE (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>0-25 (nearshore)</td>
<td>8669</td>
<td>481</td>
<td>54</td>
<td>11</td>
</tr>
<tr>
<td>25-50</td>
<td>4763</td>
<td>215</td>
<td>33</td>
<td>16</td>
</tr>
<tr>
<td>50-100</td>
<td>5770</td>
<td>153</td>
<td>24</td>
<td>15</td>
</tr>
<tr>
<td>100-150</td>
<td>3324</td>
<td>114</td>
<td>16</td>
<td>14</td>
</tr>
<tr>
<td>150-200</td>
<td>3317</td>
<td>90</td>
<td>12</td>
<td>13</td>
</tr>
<tr>
<td>200-300</td>
<td>6501</td>
<td>106</td>
<td>10</td>
<td>10</td>
</tr>
<tr>
<td>High-resolution Northeast Pacific open ocean product (Duke, Hamme, et al., 2023a)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>&gt; 300</td>
<td>34096</td>
<td>83</td>
<td>7</td>
<td>8</td>
</tr>
</tbody>
</table>

3.2 Comparison to other products

Our $pCO_2$ estimate agrees well with one other Northeast Pacific coastal ocean estimate but diverges from coarser resolution global products (supplementary Figure S7). The regional Sharp et al. (2022) product within the northern extension of the California current system (45 °N to 59 °N, east of 140 °W) is nearly equivalent to our $pCO_2$ product within reported uncertainties ($r^2 = 0.57$; supplementary Figure S7a). However, our product produces estimates closer to shore (Section 5.2 below). Compared to our product and in situ observations, a global coastal climatology (Landschützer, Laruelle, et al., 2020; Laruelle et al., 2017) and multiyear product (Roobaert et al., 2024) do not capture the same $pCO_2$ range (supplementary Figure S7c&e; supplementary Figure S8). For example, both global products underestimate winter $pCO_2$ values closer to shore in the coastal Gulf of Alaska region (> 52 °N & < 50 km offshore; area-averaged climatological winter $pCO_2$ of 300 $\mu$atm and 290 $\mu$atm respectively compared to 330 $\mu$atm in this study; supplementary Figure S7d&f), highlighting the importance of finer resolution in
coastal systems. This region also has the scarcest $pCO_2$ observations within our study area (0.37% coverage; Figure 1). Global SOM clusters commonly group the California current system with the Northwest European shelf and Sea of Japan (Laruelle et al., 2017; Roobaert et al., 2024). FFN non-linear relationships inside such clusters may not be suitable for regionally specific processes dominated by downwelling (Stabeno et al., 2004), glacial runoff (Pilcher et al., 2018; Siedlecki et al., 2017), significant seasonal biological productivity (Coyle et al., 2012; Fiechter & Moore, 2009; Hermann et al., 2009), and the influence of the upwelling subpolar Alaskan Gyre (Duke, Hamme, et al., 2023b; Hauri et al., 2021). This finding supports the Sharp et al. (2022) recommendation of increasing the number of SOM clusters for observation-based coastal ocean $pCO_2$ estimates to capture more regionally specific non-linear relationships, cognizant of SOCAT observation data density.

3.3 Uncertainty estimate

Uncertainty in the ANN-NEPc estimated $pCO_2$ product was determined following Duke, Hamme, et al. (2023b). The overall $pCO_2$ product uncertainty ($\theta_pCO_2 = 49$ $\mu$atm in our coastal product) is calculated from the square root of the sum of the four squared errors: observational uncertainty based on reported SOCAT QA/QC flags ($\theta_{obs} = 3.7$ $\mu$atm), gridding uncertainty based on the average standard deviation from gridding observations into monthly 1/12°x1/12° bins ($\theta_{grid} = 22.4$ $\mu$atm; with an increasing gradient shoreward), ANN interpolation uncertainty based on the RMSE comparing the ANN-NEPc estimated $pCO_2$ to independent withheld data ($\theta_{map} = 42.9$ $\mu$atm; Section 3.1), and ANN run randomness uncertainty based on the mean standard deviation between 10-fold ensemble members ($\theta_{run} = 6.8$ $\mu$atm; supplementary Figure S9). ANN interpolation uncertainty is the largest contribution overall. Combining the reported uncertainty in the gas transfer velocity (Section 2.6) and the overall $pCO_2$ product uncertainty yields an average uncertainty of $\pm0.18$ mol-C m$^{-2}$ yr$^{-1}$ in the air-sea gas flux across all grid cells, with the largest fraction of the error stemming from the uncertainty in the gas transfer velocity.

Our reported total uncertainty may appear high relative to other coastal $pCO_2$ products, but we include higher variability regions and more stringent error estimates. Other observation-based interpolated $pCO_2$ products in the coastal ocean report lower uncertainty values (RMSE values generally between 10 and 35 $\mu$atm in regional estimates detailed in S. Chen et al., 2016; 29 $\mu$atm globally in Roobaert et al. (2024); approximately 30 $\mu$atm in the California current
However, most other estimates did not use independent withheld data to report total product
uncertainty. Roobaert et al. (2024) point out their largest RMSE values are calculated along the
Cascadia Shelf in our study area (62 µatm). Our $p$CO$_2$ product is also the only estimate that
includes the nearshore, introducing higher variability (Table 2). Excluding the nearshore across
all components of the uncertainty calculation yields an overall uncertainty of 40 µatm, more
compareable to other coastal ocean estimates.

4.0 Comparison to high-resolution observations

Comparison to in situ observations shows that our ANN-NEPc estimated $p$CO$_2$ product
resolves seasonal variability and broad spatial patterns well. Despite high spatial resolution, our
design of a monthly timestep product means the ANN cannot reproduce short temporal (e.g.,
days) events. Predictor variable inaccuracy also contributes to $p$CO$_2$ estimate uncertainty,
particularly in the nearshore where data assimilation into reanalysis models is limited (e.g., SSS
and coastal limitations of Argo float array) and retrieval issues affect satellite estimates (e.g.,
SST and cloud masking, impact of aerosols, diurnal variability, uncertainty estimation, and
validation). In situ measurements show that biogeochemical and hydrographic variability in our
region occurs on spatial scales of less than 20 km (Nemcek et al., 2008), with spatial
autocorrelation lengths increasing offshore (Murphy et al., 2001), and timescales of days to
weeks (Evans et al., 2011, 2012, 2019; Fassbender et al., 2018). Our product is constrained by
initial binning of observations to 1/12°x1/12° (approximately 9 km by 5 km) and a monthly time
step, as well as scarcity of observations used to train (Figure 1). Comparing it directly with in
situ mooring and shipboard underway $p$CO$_2$ system measurements in the coastal zone provides
insight into when and where the ANN is both capable and incapable of resolving variability.

Our $p$CO$_2$ estimate captures the observed seasonal cycle (phase and amplitude) at
regional mooring time series sites well (Figure 3; full time series at all five regional mooring
sites in supplementary Figure S8). At NOAA’s Gulf of Alaska Ocean Acidification (GAKOA)
site south of Alaska’s Kenai Peninsula, our product tends to overestimate seasonal summer
minima and winter maxima values. However, it captures seasonal cycle timing well with a
similar average seasonal amplitude even when not all mooring data are included in
SOCATv2021 (this study = 144 µatm; GAKOA = 169 µatm; Figure 3b). At another NOAA Gulf
of Alaska mooring site south of Kodiak Island, our estimate also captures the phase of the seasonal cycle well ($r^2 = 0.89$; $N = 31$ months; supplementary Figure S8a).

![Map and graphs]

**Figure 3.** (a) Map of mean estimated surface ocean $p\text{CO}_2$ seasonal amplitude (1998-2019; range; annual maximum minus minimum) in µatm. Nearshore mooring time series at (b) Gulf of Alaska Ocean Acidification mooring (GAKOA), (c) Quadra, and (d) Cape Elizabeth mooring in situ $p\text{CO}_2$ data (black diamonds; not all included in SOCATv2021) plotted with co-located gridded SOCATv2021 (orange solid line), this study $p\text{CO}_2$ (blue solid line), and atmospheric $p\text{CO}_2$ (light blue dashed line). Kodiak and Chá bá and Roobaert et al. (2024) comparison time series in supplementary Figure S8.

The ANN recreates the seasonal cycle well at Hakai Institute’s Quadra Island Station, but its monthly timestep does not capture higher frequency variability (Figure 3c). In some instances, measured $p\text{CO}_2$ at the Quadra mooring increases over 500 µatm within three days (e.g., June 9-12, 2015), leading to a strong outgassing signal. The ANN monthly estimate does not capture such short events. Monthly binning impacts net annual air-sea $\text{CO}_2$ fluxes within a single grid cell (2015 mean annual flux from daily mooring $p\text{CO}_2$ and wind speed: 0.08 mol m$^{-2}$ yr$^{-1}$; compared to this study: 0.26 mol m$^{-2}$ yr$^{-1}$) but likely has a smaller impact when quantifying the larger regional flux. Near the end of the time series (late 2017 to 2020), the gridded SOCAT data deviates from the *in situ* mooring data due to inclusion of nearby shipboard data, yet our estimated $p\text{CO}_2$ continues to better represent the mooring seasonal cycle. When evaluating ANN
performance (Section 3.1), this difference from the gridded observation data contributes to a higher measure of uncertainty, yet in situ representation is still preserved compared to the mooring data.

The ANN does capture part of the signal from somewhat longer (i.e., weeks) summer high $p$CO$_2$ events at NOAA’s Cape Elizabeth mooring off the west coast of Washington State (Figure 3d). Horizontal advection of freshwater (July 2007) or upwelling events (> 500 μatm; July 2008; Evans et al., 2015) can cause high summer $p$CO$_2$ values. These extreme events impact bin-averaged training data, allowing the ANN to recover some of the short duration signal, albeit at a lower value. Our product reproduces both persistent, weeks long events < 35 km offshore, in line with the monthly averaged observations.
Figure 4. (a) $pCO_2$ along 2010 West Coast Ocean Acidification cruise track from 21 Jul 2010 to 15 Aug 2010 (Evans et al., 2012). Data is gridded into 1/12° by 1/12° bins. Events indicate (1) cruise start, (2) Johnstone Strait, (3) Hecate Strait, (4) intense upwelling plume near Brooks Peninsula, and (5) Juan de Fuca Strait respectively. Subplots against time along cruise track for (b) $pCO_2$ where underway in situ $pCO_2$ data (black diamonds) are plotted with co-located monthly gridded data (orange solid line), this study $pCO_2$ (blue solid line), and atmospheric $pCO_2$ (light blue dashed line). (c) Sea surface salinity (SSS) with underway in situ SSS (light blue dots) and co-located reanalysis SSS (dark blue solid line; used as a predictor variable). SSS values near cruise start as low as 15 in situ and 24 from reanalysis (not shown). (d) Sea surface temperature (SST) with underway in situ SST (red dots) and co-located satellite-based SST (dark red solid line; used as a predictor variable). Gray boxes highlight tidal mixing zones (e.g., Johnstone Strait, Juan de Fuca and Haro Straits and connecting waters).

Direct comparison to a cruise from July/August 2010 provides another example of our $pCO_2$ product’s ability to capture broadscale patterns. The ANN estimate resolves undersaturated $pCO_2$ conditions in the Salish Sea at the start of the cruise well (point 1; Figure 4). Through Johnstone Strait (50.5°N, 126.5°W), a strong tidal mixing zone (Evans et al., 2022), lack of predictor data coverage prevents estimation of $pCO_2$ in those grid cells at all (point 2; Figure 4). The ANN captures the lower variability continental shelf and slope environment in Queen Charlotte Sound and around Haida Gwaii well (between points 2 and 4; Figure 4). Differences
between estimated and observed $p$CO$_2$ exist in Hecate Strait (point 3; Figure 4) likely due to strong underestimation of SSS as a predictor in the reanalysis product (point 3; Figure 4c). Along the west coast of Vancouver Island, shipboard observations captured an upwelling event off Brooks Peninsula (50.14°N, 127.78°W; Asher et al., 2017), visible in decreased temperatures, elevated salinity, and very high in situ $p$CO$_2$ (point 4; Figure 4). The ANN does not replicate this short upwelling event (i.e., days; Asher et al., 2017). High $p$CO$_2$ driven by tidal mixing in the Juan de Fuca and Haro Straits are captured by the ANN (point 5; Figure 4; Jarníková, Olson, et al., 2022). An abundance of consistently high $p$CO$_2$ observations results in a strong reconstruction by the ANN in this region (Evans et al., 2012).

5 Air-sea CO$_2$ flux and $p$CO$_2$ drivers

Long-term (1998–2019) mean air-sea CO$_2$ fluxes display a pronounced juxtaposition between strong uptake and outgassing regions in the coastal Northeast Pacific Ocean (Figure 5c). Overall, air-sea CO$_2$ flux estimates from our product show this coastal zone acts as a net sink for atmospheric CO$_2$, drawing down 0.96±0.25 Tg C yr$^{-1}$ with a mean flux of -0.7 mol m$^{-2}$ yr$^{-1}$ but high variability with a standard deviation of 1.4 mol m$^{-2}$ yr$^{-1}$. Mean $p$CO$_2$ and air-sea CO$_2$ fluxes display similar patterns, with high $p$CO$_2$ nearshore leading to outgassing and low $p$CO$_2$ along the transition zone and continental shelf environments taking up atmospheric CO$_2$ (Figure 5a&c). Canada’s West Coast exclusive economic zone has a CO$_2$ uptake of 0.61±0.11 Tg C yr$^{-1}$.

Compared to the open ocean region of the Northeast Pacific (Duke, Hamme, et al., 2023b), the adjacent coastal ocean is a weaker sink for atmospheric CO$_2$ by area (40% weaker compared to -1.2 mol m$^{-2}$ yr$^{-1}$ in the open ocean), taking up 64% less CO$_2$ total within 40% less area (open ocean uptake = 2.63±0.53 Tg C yr$^{-1}$; open ocean surface area = 1.8x10$^6$ km$^2$; coastal ocean surface area = 1.1x10$^6$ km$^2$). Elevated $p$CO$_2$ and outgassing is also reported in the subpolar Alaskan Gyre system (Figure 5a&c), consistent with Duke, Hamme, et al. (2023b).
Figure 5. (a) Mean $p$CO$_2$ (1998-2019) in µatm. 140 °W meridian divide used in Section 5.2 analysis shown for reference. (b) Ratio of $p$CO$_2$ seasonal amplitude in thermal component (i.e., changes due to temperature; $p$CO$_2$($T$)) and biophysical component (i.e., changes due to circulation, mixing, gas exchange, and biology; $p$CO$_2$($BP$)). (c) Mean air–sea CO$_2$ flux (1998-2019) in mol m$^{-2}$ yr$^{-1}$. Negative flux values indicate CO$_2$ uptake by the ocean. (d) Mean air-sea
CO$_2$ flux seasonal amplitude (range; annual maximum minus minimum) in mol m$^{-2}$ yr$^{-1}$. (e)

Mean air-sea CO$_2$ flux vs. mean air-sea CO$_2$ flux seasonal amplitude (grid cell by grid cell). Dotted blue line is the least squares best fit. Dashed black line separates values of outgassing (positive) from uptake (negative).

5.1 Regional patterns

Spatially, the study area can be divided into four distinct regions based on air-sea CO$_2$ flux patterns in our product. The net annual air-sea CO$_2$ flux is anti-correlated with the mean air-sea CO$_2$ flux seasonal amplitude ($r^2 = 0.56$; $p < 0.01$; Figure 5e). We identify four regions that drive this pattern from most offshore to inshore: the transitional zone connecting the open ocean and the coast is a net sink with a small seasonal cycle, the Cascadia Shelf where the net sink is even stronger but the seasonal cycle remains low, nearshore regions with large seasonal cycles, and semi-enclosed estuaries with strong outgassing. To further disentangle driving processes between these four regions we decompose the estimated $p$CO$_2$ into a thermal ($p$CO$_2$ (T)) and biophysical ($p$CO$_2$ (BP)) component (supplementary Text S5; Takahashi et al., 1993, 2002). We then take the ratio ($R_{T/BP}$) of the seasonal amplitude (climatological maximum minus minimum) of the two components ($p$CO$_2$ (T)/$p$CO$_2$ (BP); Figure 5b), where biophysical processes dominate if $R_{T/BP}$ is less than one and vice versa.

Much of the offshore transitional zone (medium blue colours in Figure 5c) acts as a sink for atmospheric CO$_2$ year-round where thermal and biophysical $p$CO$_2$ components are nearly balanced. Low air-sea CO$_2$ flux seasonal amplitudes in the transitional zone (> 50 km offshore; excluding the subpolar Alaska Gyre) correspond to net annual atmospheric CO$_2$ uptake. In the southeast of the study area (Figure 5b), the North Pacific Current region experiences a relative balance of opposing thermal and biophysical $p$CO$_2$ components seasonally ($R_{T/NT-1}$ approximately = 1; Duke, Hamme, et al., 2023b; A. J. Sutton et al., 2017; Takahashi et al., 2006; Wong et al., 2010). Along most of the transitional zone where $R_{T/NT-1}$ is closer to one (Figure 5b), we also report low $p$CO$_2$ seasonal amplitudes (Figure 3a) allowing for continuous $p$CO$_2$ undersaturation with respect to the atmosphere and continuous annual uptake with low air-sea CO$_2$ flux seasonal amplitudes (supplementary Figure S12; Figure 5d). Advection of low $p$CO$_2$ (Duke, Hamme, et al., 2023b; Takahashi et al., 2006) water by the North Pacific Current from the open ocean toward the coast causes overall $p$CO$_2$ undersaturation in this region (Reed & Schumacher, 1986; Thomson, 1981; Weingartner et al., 2002). The low $p$CO$_2$ amplitudes are maintained by the effect of temperature on $p$CO$_2$ (increasing during warming and decreasing
during cooling) dampening changes due to spring phytoplankton blooms (drawing down $pCO_2$) and winter surface mixed layer deepening (increasing $pCO_2$).

The most prominent $CO_2$ sink region is found along the Cascadia Shelf, inshore of the transitional zone, with a mean flux of -1.5 mol m$^{-2}$ yr$^{-1}$ (darkest blue colours in Figure 5c). Along the continental shelf and within much of the nearshore, biophysical processes (e.g., coastal upwelling, seasonal biological drawdown, mixing) dominate the seasonal cycle of $pCO_2$ with $R_T N_T^{-1}$ values < 1. Summer upwelling fuels primary productivity causing surface $pCO_2$ drawdown as waters are advected offshore (Hales et al., 2005; Teeter et al., 2018; Ware & Thomson, 2005). Winter downwelling drives onshore transport of low $pCO_2$ offshore waters and prevents subsurface waters, with elevated respiratory CO$_2$, from mixing to the surface (i.e., coastal nutrient trap; Ianson et al., 2009; F. A. Whitney et al., 2005; Wilkerson & Dugdale, 1987). This general circulation of shelf waters maintains low seasonal flux amplitudes and strong $CO_2$ uptake on the Cascadia Shelf.

Much of the nearshore tends to experience seasonally strong, juxtaposing air-sea $CO_2$ fluxes, leading to near zero net annual $CO_2$ fluxes (nearshore white colours in Figure 5c). For example, closer to shore north of 50°N and south of the Southeast Alaska Archipelago, winter mixed layer deepening brings water rich in nutrients and $CO_2$ from respired organic matter to the surface, increasing $pCO_2$, leading to strong $CO_2$ outgassing to the atmosphere when light is limiting (supplementary Figure S12a; Marchese et al., 2022). In the spring, substantial primary productivity draws down $pCO_2$ (Marchese et al., 2022), reverting the region to a prominent sink for atmospheric $CO_2$ (supplementary Figure S12b). This large seasonal amplitude results in a net neutral flux.

Semi-enclosed, nearshore estuarine environments display strong $CO_2$ outgassing in our product, that is not always observed in regional high-resolution models. High $pCO_2$ values and outgassing fluxes (mean $CO_2$ flux of 0.7 mol m$^{-2}$ yr$^{-1}$) occur in Cook Inlet, the Salish Sea, and the Southeastern Alaska Archipelago (Figure 5c). Globally, the source strength of these integrated estuarine environments is comparable to (or smaller than) other nearshore source regions that decrease averaged coastal ocean $CO_2$ uptake (Section 5.2 below; Duke, Richaud, et al., 2023; Fennel et al., 2019; Laruelle et al., 2018). In high-resolution regional models, the Salish Sea has been reported as a weak net annual source (this study: 1.0 mol m$^{-2}$ yr$^{-1}$;
comparable to Jarníková, Ianson, et al. (2022): 0.69 mol m$^{-2}$ yr$^{-1}$), and Cook Inlet as a net sink (Hauri et al., 2020; Pilcher et al., 2018). Limited observations used to constrain both our observation-based estimate and regional models may create discrepancies between them. Our estimate is based on all available surface ocean $p$CO$_2$ observations along with a suite of predictor variables (Figure 1; Table 1), whereas regional process-based models using data for boundary conditions simplify and parameterise mechanisms (Hauri et al., 2020; Jarníková, Ianson, et al., 2022; Pilcher et al., 2018). Global observation-based estimates and models also disagree, where model fluxes are often more negative (stronger sink) at northern latitudes, attributed to a smaller seasonal $p$CO$_2$ amplitude (Resplandy et al., 2024).

5.2 Nearshore fluxes

The nearshore coastal environment (0 - 25 km offshore) exhibits large air-sea CO$_2$ fluxes, over a relatively small surface area, impacting regional marine carbon budgeting. As our estimate wraps around the coast from primarily E-W to primarily N-S, we split the region along the 140 °W meridian (Figure 5a). Averaging grid cells approximately parallel to the regional coastline along longitudinal bands (155 °W to 140 °W west of 140 °W; Figure 6a&b) and along latitudinal bands (56 °N to 45 °N east of 140 °W; Figure 6c&d), the inclusion or exclusion of the nearshore environment creates large differences in estimated net annual air-sea CO$_2$ fluxes, for example, between 154 °W to 149 °W encompassing Cook Inlet (absolute flux difference of 250%, switching from a net sink to a source; Figure 6b). North to south from 56 °N to the northern extension of the California current system at 45°N (Figure 6d), including the nearshore leads to a slightly weaker net annual sink for atmospheric CO$_2$. The difference is largest within latitudinal bands inclusive of the Salish Sea (49-51 °N; 20% weaker). Differences in zonally averaged $p$CO$_2$ and air-sea CO$_2$ fluxes also exist between products with varying nearshore coverage (Section 3.2; Roobaert et al., 2024; Sharp et al., 2022). Basin-wide, inclusion of the nearshore changes the annual exchange with the atmosphere within the study area by 0.06 Tg C yr$^{-1}$ (6%). These results highlight the importance of including the nearshore in regional marine carbon budgets.
Figure 6. Longitudinally averaged estimates west of 140 °W of mean (a) $p$CO$_2$ and (b) air-sea CO$_2$ flux of: this study (dark blue), this study removing the nearshore (cyan). (c) and (d) are latitudinally averaged estimates east of 140 °W respectively. Additional observation-based estimates with overlapping domains including: Sharp et al. (2022) (dot-dash beige), and Roobaert et al. (2024) (dashed lime green). Sharp et al. (2022) air-sea CO$_2$ fluxes calculated following Section 2.6.

5.3 Dominant controls on variability

Four distinct tiers of predictor variable importance rankings emerged from a perturbation-based spatial sensitivity analysis in estimated $p$CO$_2$ (Figure 7a). The ANN is purely a set of empirical, not mechanistic, relationships between $p$CO$_2$ observations and predictor variables, though variables were selected with mechanism in mind (Table 1). We used a perturbation-based spatial sensitivity analysis (Section 2.5) to probe the dependency of the ANN relationships on each variable, as they cannot be viewed directly (unlike a multiple linear regression).

Atmospheric $p$CO$_2$ and atmospheric $p$CO$_2$ anomaly (removing the seasonal cycle; Section 2.2) are the most important predictors, followed by SST, and then process-driven controls whose importance varies spatially. Atmospheric $p$CO$_2$ and atmospheric $p$CO$_2$ anomaly are the only two predictor variables that capture a trend in time from 1998 to 2019 (i.e., increase of 2.12 μatm yr$^{-1}$ due to anthropogenic emissions). Due to the trend, these variables also experienced the largest absolute value perturbation (mean basin-wide increase of 7 μatm), at least one order of magnitude greater than other variables. The third most important predictor for estimating $p$CO$_2$ is SST. Basin-wide, the sensitivity test introduced a mean SST increase of 1.5 °C, resulting in a mixed $p$CO$_2$ response where generally there was a decrease, outside of the Gulf of Alaska central
glacial drainage basin where $pCO_2$ increased (supplementary Figure S13a). This result does not follow the mechanistic reduced solubility of CO$_2$ in warmer water. However, it emphasizes the importance of the SST seasonal cycle as a predictor (strong correlation, typically negative, between $pCO_2$ and SST; supplementary Figure S13b).
Figure 7. (a) Predictor variables ordered by absolute mean $p$CO$_2$ change from baseline run during perturbation-based spatial sensitivity analysis (Section 2.5). (b) Most dominant process-based predictor variable mapped by largest absolute mean $p$CO$_2$ change from baseline run during perturbation-based spatial sensitivity analysis (excluding top three variables from (a)). No grid.
cells displayed Chl or Chl anomaly as the largest absolute mean $\rho$CO$_2$ change from baseline over the full study time range (1998-2019). Major river outflows are labelled for reference.

Excluding the three most dominate controls (atmospheric $\rho$CO$_2$, atmospheric $\rho$CO$_2$ anomaly, and SST), the spatial distribution of predictor variable importance rankings can be explained by mechanistic drivers even though the ANN is purely empirical. SSH anomaly is important along the Alaskan Gyre boundary, where the upwelling gyre exerts control over local biogeochemistry (Figure 7b; Duke, Hamme, et al., 2023b; Hauri et al., 2021). Wind speed (as a proxy for mixed layer depth) is important throughout most regions along the continental shelf and the outer coast as winter mixed layer deepening brings CO$_2$-rich subsurface waters to the surface (mean basin-wide increase of 0.4 m s$^{-1}$ resulting in a $\rho$CO$_2$ increase of 1.7%; Figure 7b). SSH and SSH anomaly are additionally important offshore of Sitka, Alaska (57 °N, 143 °W) and Haida Gwaii (52 °N, 133 °W) where mesoscale anticyclonic eddies with enhanced primary productivity and high SSH propagate away from the continental margin (Figure 7b; Batten & Crawford, 2005; Crawford et al., 2007; Crawford & Whitney, 1999; F. A. Whitney et al., 2005; F. Whitney & Robert, 2002). In the North Pacific Current influenced region southeast of the study area, SST anomaly and wind speed anomaly are the most important predictors linked to the relative balance of opposing mechanisms (i.e., thermal and biophysical $\rho$CO$_2$ components; Figure 5b).

Nearshore regions experience a range of predictors with prominent features mostly controlled by salinity (SSS and SSS anomaly) in coastal estuarine areas (Figure 7b), and tidally mixed areas (e.g., Juan de Fuca Strait, Johnstone Strait; Figure 4a). In additional regions where freshwater discharge is important (e.g., supplementary Table S2), SSH and SSH anomaly emerge as important predictors potentially linked to discharge associated changes to nearshore sea level (Figure 7b; Durand et al., 2019). Neither perturbation to Chl nor Chl anomaly resulted in the largest absolute mean $\rho$CO$_2$ change from baseline over 264 months in a single grid cell (Figure 7b). However, seasonally Chl emerges as a prominent predictor in scattered grid cells along nearshore West Coast Vancouver Island and in the Southeast Alaska Archipelago during the spring (i.e., March, April, and May; not shown).
5.4 Air-sea \( pCO_2 \) trends

Trends in the last decades (1998-2019) in \( \Delta pCO_2 \) (sea – air) display spatial heterogeneity in the coastal Northeast Pacific, with a gradient of smaller trends moving offshore. A linear fit was applied to the \( \Delta pCO_2 \) anomaly time series within each grid cell to calculate the trend and standard error (i.e., deseasonalized; Section 2.2). Regions that experience an increase in surface ocean \( pCO_2 \) close to the increase in atmospheric (i.e., resulting in a small \( \Delta pCO_2 \) trend) are spatially distinct from those that have an insignificant trend in \( pCO_2 \) leading to a large divergence with the atmosphere (i.e., large \( \Delta pCO_2 \) trend). Grid cells with a small \( \Delta pCO_2 \) trend are dominantly located in the outer coast (> 50 km offshore) and in the southeast of the study area (Figure 8a). Trends are closer to the atmospheric trend in this region (2.12 \( \mu \text{atm yr}^{-1} \)), meaning any change in the carbon sink due to anthropogenic climate change will require long observation time series to detect, as the signal is small relative to internal variability (Gooya et al., 2023; McKinley et al., 2016; Resplandy et al., 2015; Adrienne J. Sutton et al., 2019). We report trends in \( pCO_2 \) that are similar to those observed at time series sites along Fisheries and Ocean Canada Line P stations (this study: \( P4 = 1.3 \pm 0.1 \mu \text{atm yr}^{-1} \); \( P12 = 1.6 \pm 0.1 \mu \text{atm yr}^{-1} \); comparable to Franco et al. (2021): \( P4 = 1.0 \pm 1.4 \mu \text{atm yr}^{-1} \); \( P12 = 1.5 \pm 0.6 \mu \text{atm yr}^{-1} \)).
Figure 8. 1998-2019 trend in (a) $\Delta$CO$_2$ anomaly (i.e., deseasonalized) where more negative (darker) values indicate an increase in air-sea $p$CO$_2$ disequilibria with time. Black crosshatches show grid cells with an insignificant calculated trend (outside the 95% confidence level; $p \geq 0.05$; 0.4% of total grid cells). (b) Standard error of the estimated slope in the $\Delta$pCO$_2$ trend fit.

Large $\Delta$pCO$_2$ trends (and low or insignificant $p$CO$_2$ trends) occur in regions experiencing strong connectivity to the older subsurface waters of the Northeast Pacific (e.g., subpolar Alaskan Gyre, west coast upwelling zone; Figure 8a). This older water has a lower anthropogenic carbon load (Carter et al., 2019; Clement & Gruber, 2018; Gruber et al., 2019; Sabine et al., 2004), which may be responsible for the lag in the increase in surface ocean $p$CO$_2$ (e.g., Duke, Hamme, et al., 2023b). The $\Delta$pCO$_2$ trend in the Alaska Gyre is dominated by the winter trend, whereas the west coast upwelling zone is dominated by the summer trend (supplementary Figure S14). These seasonal trends coincide with the timing of greatest connectivity to depth in each region. Strongest Alaskan gyre upwelling occurs in winter (Gargett, 1991; Talley, 1985), whereas the coastal upwelling season is spring and summer (Dorman & Winant, 1995; Hsieh et al., 1995) with downwelling occurring in the winter (Section 5.1; Thomson & Ware, 1996). In the nearshore (e.g., Southeast Alaska Archipelago, Salish Sea), subsurface waters exchange through estuarine flow and tidal mixing. In these regions, we report low or insignificant winter $\Delta$pCO$_2$ trends and large negative summer trends in agreement with regional model results (e.g., Jarníková, Ianson, et al., 2022). Increasing summer air-sea $p$CO$_2$ disequilibria enhances ocean CO$_2$ uptake, whereas winter air-sea disequilibria has remained relatively constant, maintaining ocean outgassing. In winter, light limits biological productivity, resulting in higher total CO$_2$ in the surface (Evans et al., 2019; Ianson et al., 2016; Simpson et
al., 2022). This increase in total CO₂ reduces the buffer capacity of the carbonate system (Revelle & Suess, 1957), so that the pCO₂ increase due to anthropogenic carbon uptake is larger than it is in summer in many temperate zones (e.g., Jarníková, Ianson, et al., 2022; Landschützer et al., 2018). Our findings are consistent with global ΔpCO₂ trend estimates where most coastal regions appear to exhibit negative ΔpCO₂ trends (i.e., likely becoming stronger atmospheric CO₂ sinks or weaker sources; Fennel et al., 2019; Laruelle et al., 2018; Resplandy et al., 2024; Roobaert et al., 2024; Wang et al., 2017).

6 Conclusions

Our high-resolution, neural network created pCO₂ product reproduces observed coastal Northeast Pacific Ocean variability well, from the outer transitional zone to the nearshore (0 – 25 km offshore). We interpolated sparse observations using non-linear relationships developed with a neural network based on predictor data from satellite and reanalysis products to create a continuous, gridded monthly pCO₂ estimate at a 1/12° spatial resolution, inclusive of the nearshore. This pCO₂ product provides a baseline environmental context for pCO₂ and air-sea CO₂ flux variability in the study area with an uncertainty of 49 µatm and 0.24 mol-C m⁻² yr⁻¹, respectively. The product resolves seasonal variability (phase and amplitude) and broad spatial patterns well compared to high-resolution in situ observations. The product is not designed to capture daily – weekly variability.

A unique ANN sensitivity analysis shows that variations in pCO₂ results agree with mechanistic drivers even though the ANN itself is purely empirical. ANNs are not based on predefined equations but their ability to capture information inherent to the training data, preventing any explicit explanation of how predictor variables and their output are related. We suggest a new systematic sensitivity analysis introducing perturbations to predictor variables, with a consideration for natural spatial variability, to produce mapped variable importance rankings. This approach offers insight providing greater transparency to ANN “black box” techniques.

We describe the coastal Northeast Pacific as a net sink for atmospheric CO₂ with large spatial heterogeneity between outgassing in the nearshore and uptake on the outer coast. Net annual air-sea CO₂ flux is largely anticorrelated with seasonal air-sea CO₂ flux amplitude. Patterns inherent to specific regions drive this anticorrelation, including circulation and opposing
seasonal upwelling or relaxation vs. downwelling, and may make the relationship regionally specific rather than applicable to the wider global coastal ocean. Our results also emphasize the importance of including nearshore fluxes (often omitted by other coastal products), which are likely to be a source reducing the net coastal sink, when constructing marine carbon budgets (e.g., Legge et al., 2020). These findings could be potentially important considerations for reporting marine carbon dioxide removal approaches in the study area, as interventions impacting source areas are treated differently from those enhancing natural sinks (Verra, 2023).

Trends over the last decades show outer coast $p\text{CO}_2$ may be experiencing the largest increase in air-sea $p\text{CO}_2$ disequilibrium, due to strong connectivity with subsurface waters low in anthropogenic $\text{CO}_2$, while $p\text{CO}_2$ in the North Pacific Current region tracks increasing atmospheric $p\text{CO}_2$ more closely. Trends reported here across the coastal Northeast Pacific indicate most regions are likely to become stronger atmospheric $\text{CO}_2$ sinks or weaker sources.

Improving regional observational coverage and continuity and advancing the ANN approach will improve future air-sea $\text{CO}_2$ flux estimates. Some regions in the coastal Gulf of Alaska display large net annual air-sea $\text{CO}_2$ fluxes (e.g., Cook Inlet) yet are extremely sparsely monitored. A higher temporal resolution, such as daily, could enable the ANN to capture highly episodic air-sea $\text{CO}_2$ flux events common to the nearshore. However, this approach would dramatically reduce the percent coverage of observation training targets. A solution may be creating ANN non-linear relationships to interpolate $p\text{CO}_2$ directly from in situ observations. Using high frequency, collocated sensors and non-uniform “highest available resolution” satellite and reanalysis datasets for predictor variables not collected in situ, a higher temporal and/or spatial resolution coastal product could be developed without substantial loss in ANN training targets.

Acknowledgments

Funding for this project was provided by the Natural Sciences and Engineering Research Council of Canada (NSERC) through the Advancing Climate Change Science in Canada program (grant# ACCPJ 536173-18) to RH. Funding from Fisheries and Oceans Canada’s Aquatic Climate Change Adaptation Service Program to DI supported the analysis of recent underway $p\text{CO}_2$ measurements made by the Line-P program (grant# 96036). PD financial support also provided by a Natural Sciences and Engineering Research Council of Canada (NSERC) Doctoral
Postgraduate Scholarship and the joint Fisheries and Oceans Canada (DFO) and National Oceanic and Atmospheric Administration (NOAA) Ocean Acidification Collaborative Funding Initiative.

The authors declare no conflicts of interest relevant to this study.

Fisheries and Oceans Canada (DFO) runs the Line P monitoring program. National Oceanic and Atmospheric Administration (NOAA) operates the Kodiak, Gulf of Alaska Ocean Acidification, Chá bā, and Cape Elizabeth moorings. Hakai Institute operates Quadra Island Station. We extend our thanks to Jamie Shutler and Dan Ford for their valuable brainstorming efforts regarding the \(\text{pCO}_2\) temperature correction section. We are grateful to Wiley Evans for sharing previously published \(\text{pCO}_2\) data from the 2010 West Coast Ocean Acidification cruise. We also want to thank Philippe Tortell and Robert Izett for providing previously published \(\text{pCO}_2\) data from the May 2007 and May 2010 La Perouse cruises.

**Open Research**

All data used is publicly available. ANN-NEPc \(\text{pCO}_2\) and air-sea \(\text{CO}_2\) flux fields created for this publication are available through the National Center for Environmental Information (NCEI Record ID: BHE12VV0E). \(\text{pCO}_2\) data are from the Surface Ocean \(\text{CO}_2\) Atlas (SOCAT) v2021 (available at https://www.socat.info/) as well as additional data from the Fisheries and Oceans Canada February 2019 Line P cruise, a West Coast Ocean Acidification cruise from July and August 2010 (Evans et al., 2012), and La Perouse cruises from May 2007 and May 2010 (available at https://www.waterproperties.ca/linep/). Sea surface temperature and chlorophyll-a are from the European Space Agency Climate Change Initiative (available at https://climate.esa.int/en/odp/#/dashboard). Sea surface salinity and sea surface height are from Copernicus Marine Environment Monitoring Service (available at https://data.marine.copernicus.eu/product/GLOBAL_MULTIYEAR_PHY_001_030/description). Ocean surface wind data at 10 m height are from Regional Deterministic Reforecast System (available at https://caspar-data.ca/; detailed here https://github.com/julemai/CaSPAr). Mooring data used in analysis are also available through the National Center for Environmental Information (NOAA moorings: NCEI Accession 0173932; and Hakai Institute Quadra Island Field Station: NCEI Accession 0208638).
References


https://doi.org/10.1016/j.csr.2004.02.007


