Biogeosciences

Open Access

EGU

Supplement of

Estimating marine carbon uptake in the northeast Pacific using a neural network approach

Patrick J. Duke et al.

*Correspondence to:* Patrick J. Duke (pjduke@ucalgary.ca)

Contents of this file

Text S1

The reported fCO$_2$ estimates were converted to pCO$_2$ using the equation S1 (Körtzinger, 1999):

15 $$p\text{CO}_2 = f\text{CO}_2 \times \exp[\text{P}_{\text{atm}}^{\text{surf}} \frac{B+(2\delta)}{RT}],\tag{S1}$$

where $\text{P}_{\text{atm}}^{\text{surf}}$ is the total atmospheric surface pressure, B and δ are viral coefficients (Weiss, 1974), R is the gas constant and T is the absolute temperature. National Centers for Environmental Prediction (NCEP) monthly mean sea level pressure was used for $\text{P}_{\text{atm}}^{\text{surf}}$ (Kalnay et al., 1996).
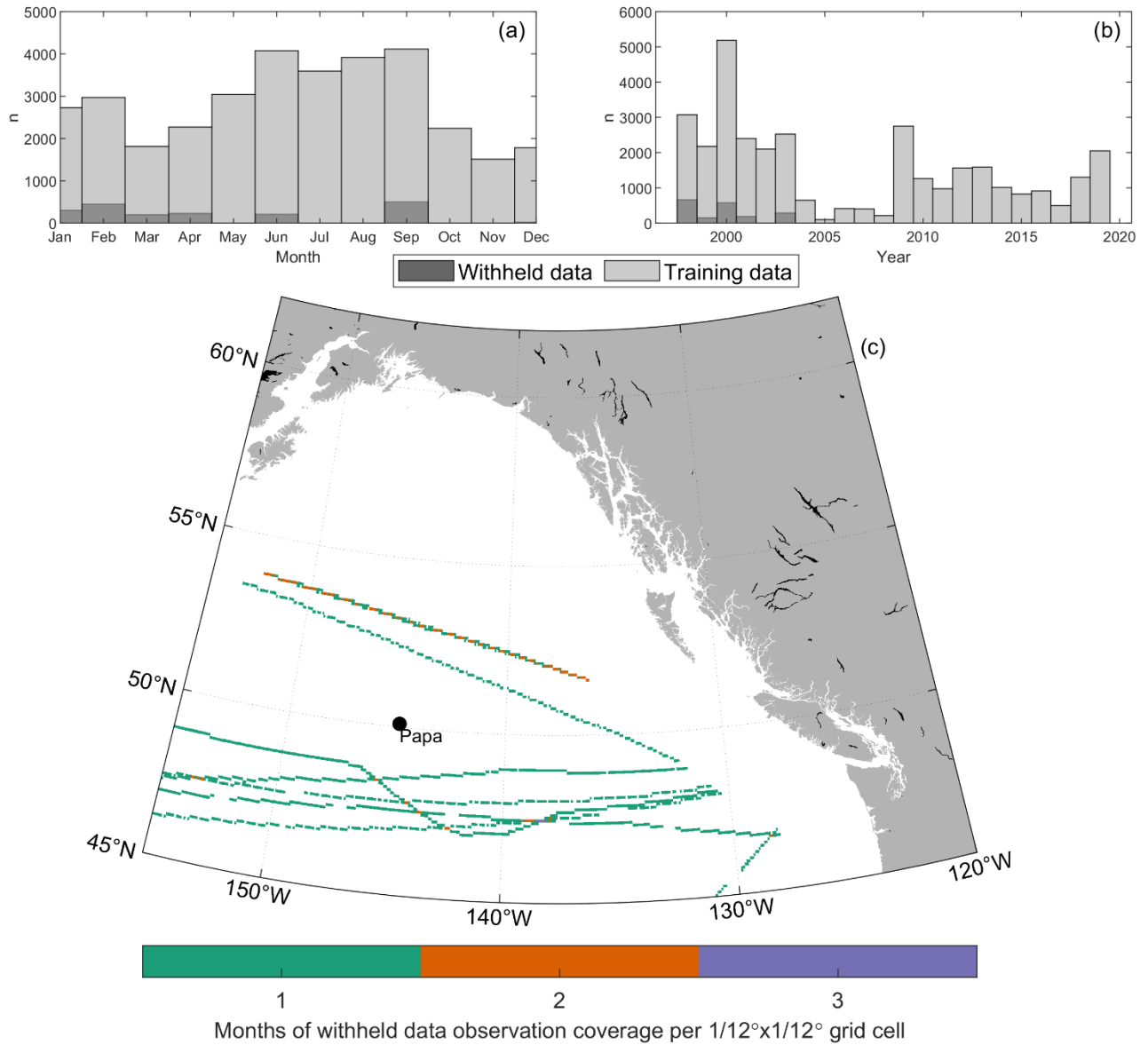
Figure S1 Withheld data (dark grey) distribution in (a) months, and (b) years relative to training data (light grey), and (c) geographically as the number of months of observational coverage per 1/12°x1/12° grid cell. Ocean Station Papa is shown for reference.
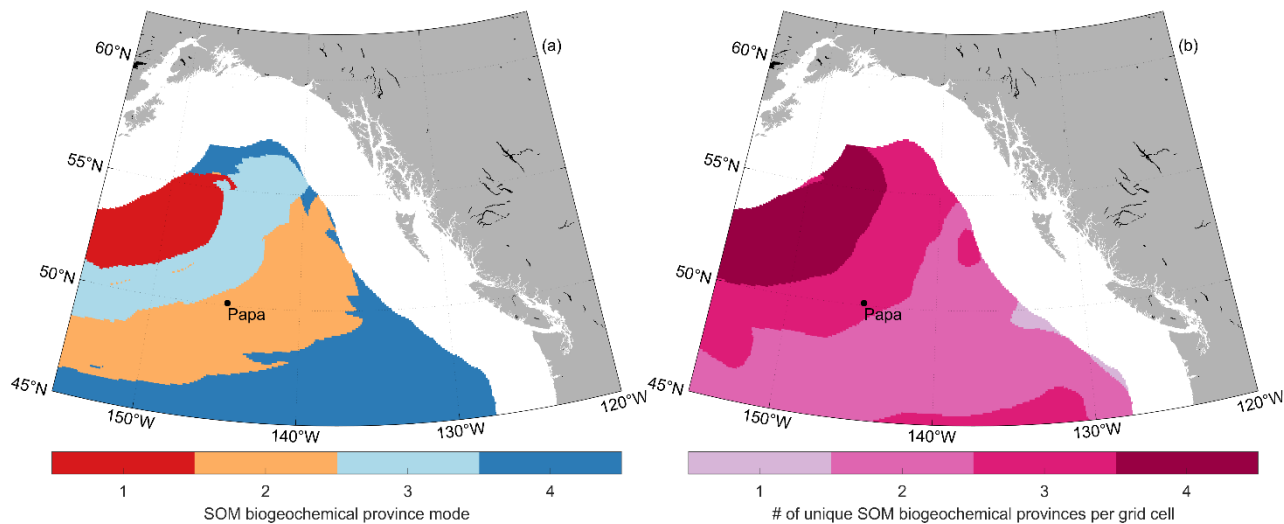
Figure S2 Mapped (a) mode of SOM biogeochemical provinces (i.e., most frequent occurrence), and (b) the number of unique SOM biogeochemical provinces each grid cell belongs to for each month from January 1998 to December 2019.
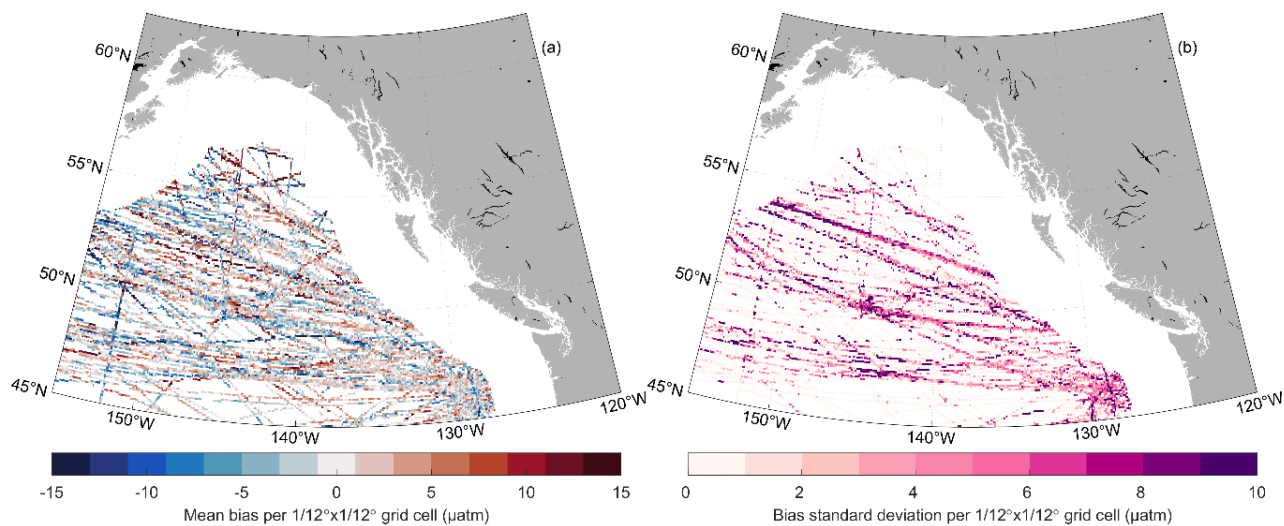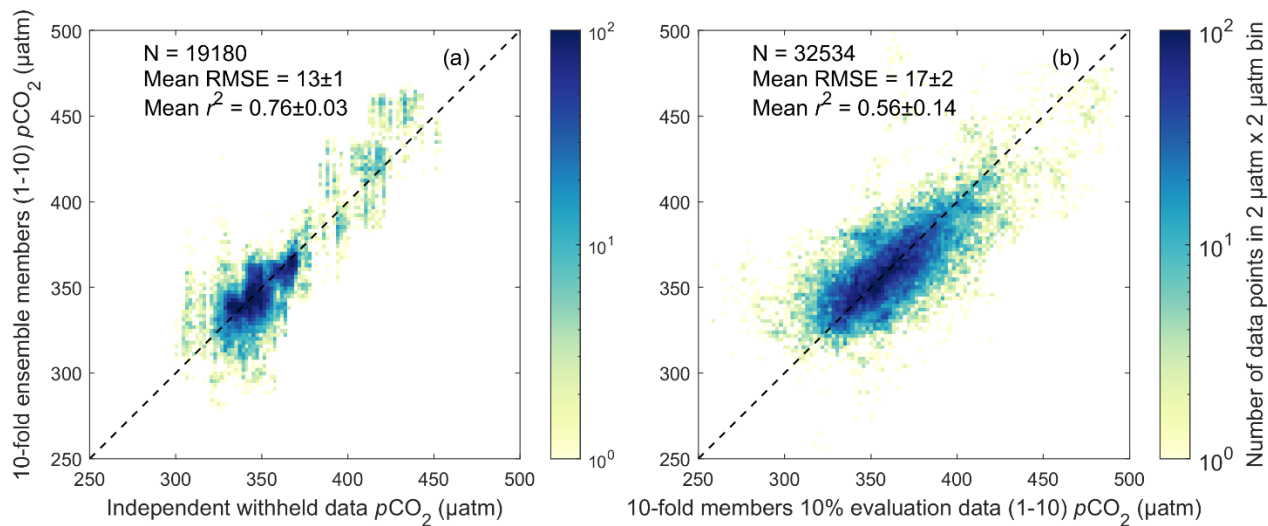

Figure S3 Mapped mean bias and standard deviation in residuals between ANN-NEP pCO$_2$ estimate and 1/12°x1/12° gridded SOCAT data.

30   **Figure S4 10-fold cross-evaluation (Section 2.4) individual ensemble member estimated $p\text{CO}_2$ against the (a) independent withheld data, and (b) 10% 10-fold evaluation data specific to that ensemble member. Mean root mean squared error (RMSE) and coefficient of determination (r^2) are across all individual ensemble members. Data is binned into 2 μatm by 2 μatm bins. The dashed black line represents a perfect fit of slope (c_1) = 1 and intercept = 0.**
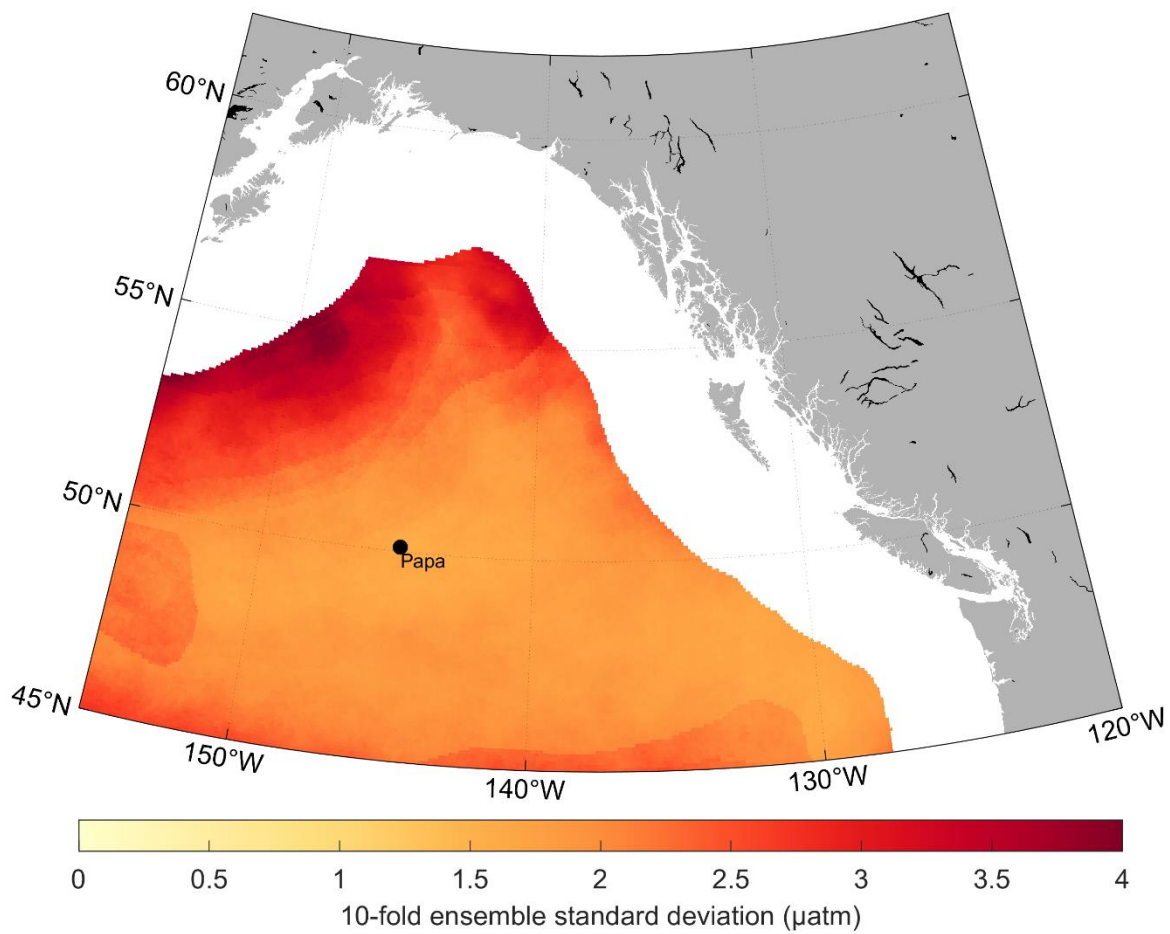
35 **Figure S5 Mean pCO$_2$ standard deviation between ANN-NEP 10-fold ensemble members. Ocean Station Papa is shown for reference.**
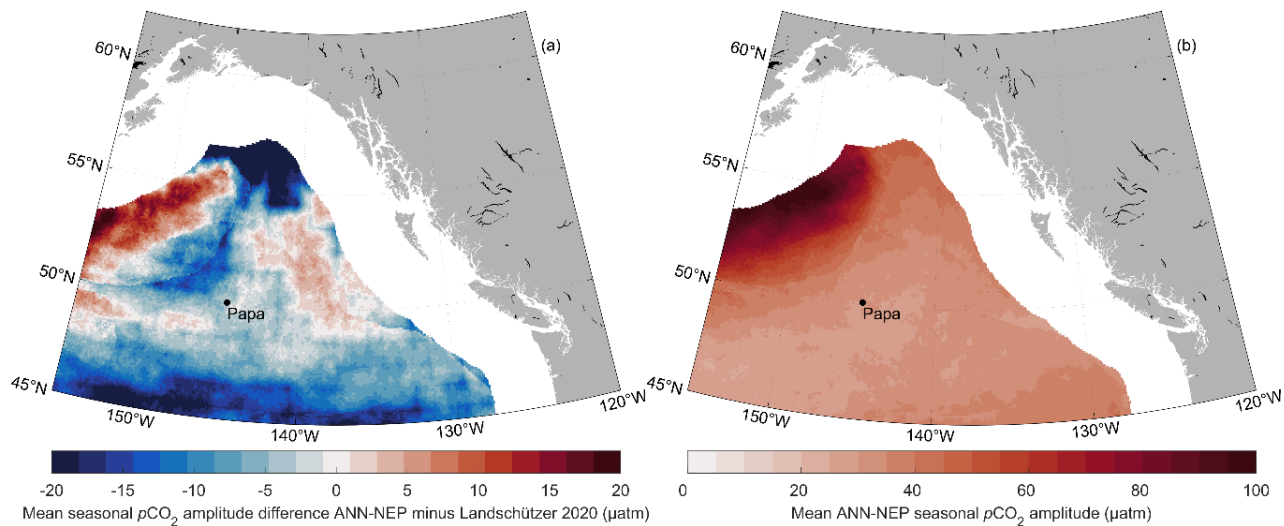
Figure S6 (a) Mean difference in the surface ocean pCO₂ seasonal amplitude in µatm between the ANN-NEP estimate (this study) and the Landschützer et al. (2020) global product. Positive (negative) differences indicate higher pCO₂ seasonal amplitude for the ANN-NEP (Landschützer et al. (2020)) estimate. The Landschützer et al. (2020) estimates were interpolated to the 1/12°x1/12° grid of this study. (b) Mean ANN-NEP seasonal surface ocean pCO₂ seasonal amplitude in µatm. Ocean Station Papa is shown for reference.
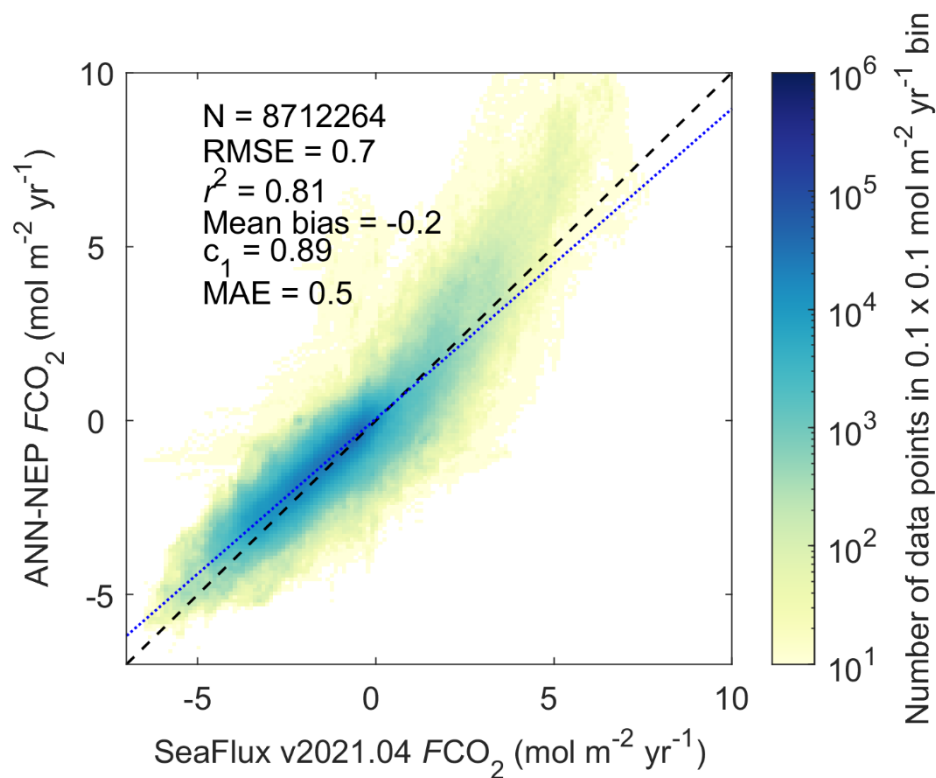
Figure S7 Property to property plot of air-sea CO₂ flux density values calculated from ANN-NEP and from SeaFlux v2021.04 (Fay et al., 2021). The SeaFlux estimates were interpolated to the 1/12°x1/12° grid of this study. Number of overlapping grid cells withing the study area (N), root mean squared error (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). The observed linear relationship is represented by the dotted blue line. Data is binned into 0.1 by 0.1 mol m^{-2} yr^{-1} bins. The dashed black line represents a perfect fit of slope (c_1) = 1 and intercept = 0. Colorbar shows data density on a log scale.
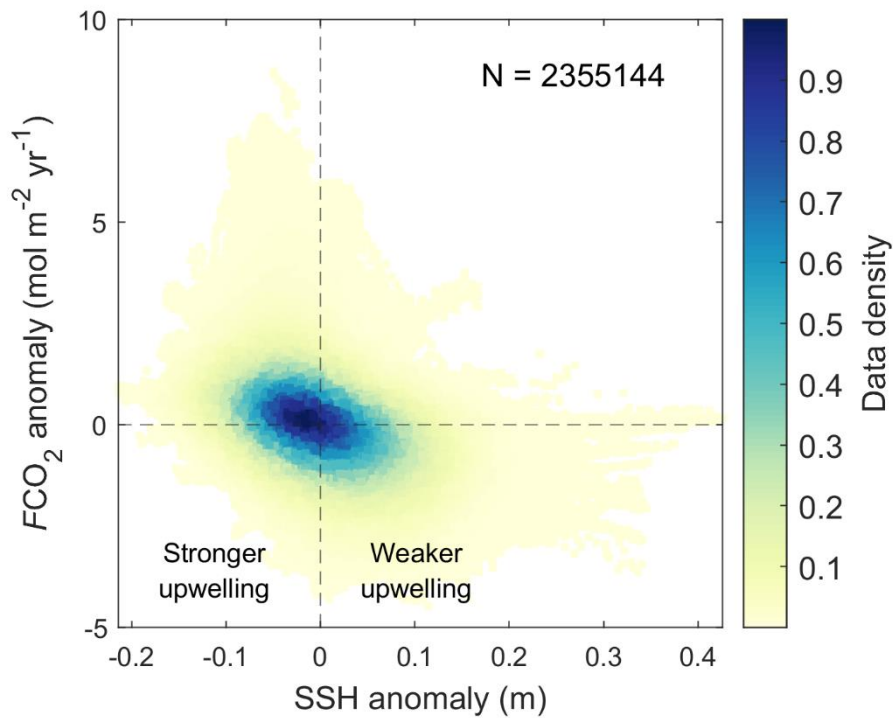
Figure S8 Property to property plot of air-sea CO₂ flux density anomalies and sea surface height (SSH) anomalies (grid cell by grid cell) in the subpolar Alaskan Gyre region of our study area (latitudes north of 52° N). Stronger (weaker) upwelling label relates to gyre upwelling strength driven by winds enhancing (damping) Ekman pumping and depressing (elevating) SSH.

55 **Table S1 Regional high-resolution artificial neural network Northeast Pacific (ANN-NEP) pCO_2 product performance against all SOCAT pCO_2 observation data grouped by year and month. Number of observations (N), root mean squared error (RMSE), coefficient of determination (r^2), and mean bias (calculated as the mean residual).**

	N	RMSE	r^2	Mean bias
Year				
1998	3085	8.1	0.73	-1.4
1999	2184	7.5	0.91	1.4
2000	5195	8.1	0.93	0.0
2001	2404	5.7	0.96	1.2
2002	2110	7.1	0.91	1.0
2003	2526	5.2	0.84	-1.3
2004	652	4.5	0.94	1.2
2005	110	6.8	0.69	-3.6
2006	413	3.3	0.93	-2.6
2007	405	6.4	0.87	-2.8
2008	217	4.0	0.95	-1.3
2009	2751	5.0	0.94	-0.4
2010	1267	6.5	0.96	1.9
2011	980	9.1	0.87	-1.0
2012	1567	6.1	0.93	1.3
2013	1593	4.3	0.99	0.9
2014	1017	5.1	0.95	-1.5
2015	836	9.1	0.95	1.5
2016	919	6.3	0.91	-0.4
2017	511	4.7	0.89	1.0
2018	1303	5.7	0.94	1.2
2019	2051	8.1	0.93	-0.6
Month				
January	2731	5.2	0.97	0.7
February	2971	5.8	0.98	-0.7

March	1816	6.8	0.96	-1.0
April	2277	4.3	0.97	0.3
May	3041	5.6	0.94	0.1
June	4077	9.0	0.92	-1.1
July	3601	9.6	0.86	-0.1
August	3922	7.5	0.94	1.0
September	4121	6.6	0.93	0.2
October	2240	4.9	0.91	1.2
November	1511	3.9	0.97	-0.4
December	1788	4.4	0.97	0.8

Table S2 Regional artificial neural network Northeast Pacific pCO_2 product performance at varying resolutions against training and independent withheld SOCAT pCO_2 observations. Mean and standard deviation between lower 10th percentile (5 of 50 runs) of overfitting metric values for each resolution with varying internal data division ratios between the pCO_2 training data used by the ANN to train and internally evaluate. 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). Courser resolution product uncertainties are also included where overall pCO_2 product uncertainty (θ_{pCO2}) is calculated from the square root of the sum of the three squared errors: observational uncertainty (θ_{obs}), gridding uncertainty (θ_{grid}), and ANN interpolation uncertainty (θ_{map}). The 10-fold ensemble approach was not run for the courser resolution products, likely leading to a slight underestimate of overall uncertainty as ANN run randomness uncertainty (θ_{run}) was excluded.

Resolution	N	RMSE	r^2	Mean Bias	c_1	MAE
Training data						
1°	2547	11.8±0.6	0.79±0.02	-0.1±0.1	0.73±0.02	8.8±0.5
1/2°	5569	10.9±0.7	0.83±0.02	0.1±0.2	0.77±0.02	7.9±0.6
1/4°	11253	11.3±0.9	0.82±0.03	0.0±0.1	0.77±0.03	8.2±0.7
1/8°	21869	12.2±1.0	0.79±0.03	0.0±0.2	0.74±0.03	8.8±0.8
1/12°	31392	10.5±1.0	0.84±0.03	0.0±0.2	0.79±0.03	7.4±0.8
Independent withheld data						
1°	155	11.7±0.3	0.76±0.01	-0.7±0.5	0.88±0.01	8.4±0.1
1/2°	350	11.5±0.6	0.78±0.02	0.1±0.7	0.93±0.03	8.3±0. 5
1/4°	716	11.5±0.8	0.79±0.03	0.1±0.6	0.98±0.02	8.6±0. 7
1/8°	1387	12.5±0.6	0.76±0.01	-0.1±1.2	0.93±0.04	9.2±0. 5
1/12°	1857	11.4±0.5	0.79±0.01	2.1±0.6	0.92±0.03	8.4±0. 5

	pCO_2 product uncertainty			
	θ_{obs}	θ_{grid}	θ_{map}	θ_{pCO2}
1°	3.1	3.7	11.7	12.6
1/2°	3.1	2.8	11.5	12.3
1/4°	3.1	2.0	11.5	12.2
1/8°	3.1	2.0	12.5	13.0