Global Biogeochemical Cycles

Supporting Information for

Trends in Sea-Air CO_2 Fluxes and Sensitivities to Atmospheric Forcing Using an Extremely Randomized Trees Machine Learning Approach

Rik Wanninkhof[1], Joaquin Triñanes[1,2,3], Denis Pierrot[1], David R. Munro[4,5], Colm Sweeney[4], and Amanda R. Fay[6]

[1]Atlantic Oceanographic and Meteorological Laboratory, National Oceanic and Atmospheric Administration, 4301 Rickenbacker Causeway, Miami, FL, USA.

[2]Cooperative Institute for Marine and Atmospheric Studies, Rosenstiel School of Marine, Atmospheric and Earth Science, University of Miami, 4600 Rickenbacker Causeway, Miami, FL, USA.

[3]Dept. of Electronics and Computer Science, Universidade de Santiago de Compostela, Santiago, Spain

[4]Global Monitoring Laboratory, National Oceanic and Atmospheric Administration, Boulder, CO, USA

[5]Cooperative Institute for Research in Environmental Sciences (CIRES), University of Colorado, Boulder, CO, USA

[6]Columbia University and Lamont-Doherty Earth Observatory, Palisades NY, USA

Contents of this file

Introduction

The supporting information provides further detail on AOML_ET machine learning and its sensitivity to changing predictor variables. The text S1 details the extreme randomized trees approach and other machine learning methods. Text S2 describes the mapping step of AOML-ET and compares AOML-ETv2020 with that of AOML-ETv2023 to illustrate the tail effects. Text S3 shows the influence of predictor variables on AOML-ET. Text S4 gives the references cited in Text S1 through S4. Figures S1 through S6 provide graphics as referenced in the text S1 through S3. Table S1 shows the predictor and target variables used in ET along with units and references. The table also shows the differences between the variables and nomenclature used in RECCAP2 and in this work. Table S2 shows the average global flux values and trends for different permutations of AOML-ET and the consensus values provided in the Global Carbon Budget, GCB2023.

Text S1.

The Extremely Randomized Trees (ET) and other machine learning methods for mapping fCO_{2w}

The Extremely Randomized Trees or Extra Trees (ET) Method is a machine learning (ML) ensemble approach that combines many independent trees trained over random and equally distributed subsets of the data (Geurts et al., 2006). Its learning stage consists of creating many independent decision trees from slightly different input data. All of the input data, or bagging ensemble, is randomly subsampled. All trees are evaluated independently and averaged to compute the forest estimate. The probability that a given input belongs to a given class is interpreted as the proportion of trees that classify that input as a member of that class. The ET differs from the more common Random Forest (RF) method in that the RF subsamples the input data with replacement, whereas the ET uses the whole original sample. The RF chooses the optimum split points in the decision process, while ET chooses it randomly (Aznar, 2020). The ET approach is computationally efficient and therefore significantly faster than RF and other ML methods, making it amenable to testing different permutations to create monthly fCO_{2w} maps.

Different ML mapping products have been compared for their ability to determine fCO_{2w} fields, notably under the aegis of the Surface Ocean CO_2 Mapping (SOCOM) intercomparison effort (Rödenbeck et al., 2015). The products have been used in several assessments, including the global ocean carbon assessment in RECCAP2 (deVries et al., 2023). Detailed regional and global comparisons of different mapping products and ensemble approaches have been undertaken (e.g., Fay et al., 2021; Gregor et al., 2019; Rödenbeck et al., 2022; Chau et al., 2022). The analysis by Gregor et al. (2019) includes several different ML approaches and suggests that overall skill of the methods at the global scale is similar and that the skill for any given approach is mainly limited by fCO_{2w} data availability in undersampled regions and seasons. Gregor et al. (2019) also show broad similarity in overall patterns of the fCO_{2w} fields and the magnitude of interannual variability of fCO_{2w} for the various ML approaches. In particular, the Northern Hemisphere oceans show agreement in fCO_{2w} between methods, while areas with fewer fCO_{2w} observations, such as the mid- and high-latitude Southern Hemisphere oceans, and regions with large interannual variability, such as the Equatorial Pacific, show greater differences between approaches. Aside from differences in the ML-derived fCO_{2w}, inconsistencies in modeled surface areas, wind speed products, and the method of calculation of fluxes contribute to differences in sea-air CO_2 fluxes. To account for these differences, area normalization and ensembles (or multi-product averages) are increasingly common in analyses and improve consistency (Fay et al., 2021; Roobaert et al., 2018). A summary of the annual global sea-air CO_2 fluxes using different ML approaches used in the Global Carbon Budget (Friedlingstein et al., 2023) is shown in

Figure 1. It shows correspondence at the global scale in sea-air CO_2 fluxes over time between the observation-based methods using ML and interpolation approaches.

The fCO_{2w} and corresponding sea-air CO_2 flux maps created with the ET approach, referred to as AOML-ET, that are submitted to RECCAP2 cover the period from 1998 through 2018 (Müller, 2023). It uses the gridded SOCAT v2020 data for training and validation. Gridded sea surface temperature (SST), sea surface salinity (SSS), mixed layer depth (MLD), chlorophyll-a, position, and time are the predictor variables. The default configuration of the AOML-ET sea-air CO_2 flux product was produced to meet the RECCAP2 requirements and nomenclature. The output is pCO_{2w} values and sea-air CO_2 fluxes on a monthly 1˚ by 1˚ grid. The details of the predictor and target variables, units, and nomenclature as prescribed by RECCAP2, and those used in this paper are provided in Table S1. When the AOML-ET parameters or nomenclature used in this manuscript are different than specified in RECCAP2, they are added in italics.

Text S2.

Extrapolation using AOML-ET and comparison of AOML-ETv2023 with AOML-ETv2020

While the ML approaches are commonly used for mapping of fCO_{2w} and determination of flux densities, they can be used to extrapolate fCO_{2w} data with time. That is, once the approach is trained, the results can be extrapolated if predictor variables are available for the appropriate period. For AOML-ETv2020 that has training data through 2019, the monthly fCO_{2w} are extended through 2023 and corresponding annual sea-air CO_2 fluxes are determined. These are compared with AOML-ET trained with the updated SOCAT database SOCATv2023, AOML-ETv2023 that has gridded data through 2022. The influence on training of AOML-ET to create the fCO_{2w} fields can be gleaned from comparing the last years of the time series using AOML-ETv2020 and AOML-ETv2023, where the former extrapolates findings for 2019 through 2022 without training data in those years. As shown in Figure S1, extrapolating the results of AOML-ETv2020 beyond 2020 shows a large increase in influx of -0.7 Pg C between 2020 and 2022. This trend is not observed in AOML-ETv2023, which is trained with predictors through 2022, suggesting that training data over the entire time interval is critical. A summary of the statistics of the linear regressions of the annual sea-air CO_2 fluxes shown in Figure 9 is provided in Table S2.

The SOCATv2020 has 277,462 gridded fCO_{2w} values, while the SOCATv2023 has 316,963 values, with the tally per year shown in Figure S2, and the number of observations in each cell provided in Figure S3a. There are more observations in data-sparse regions in SOCATv2023 compared to SOCATv2020 (Figure S3b), notably in the Davis Straight and Hudson Bay, the Southern Ocean, and Subtropical South Pacific and South Atlantic. The extra data of AOML-ETv2023 from 2000-2022 does not only affect the results at the end of the record. AOML-ETv2023 shows very similar global fluxes up to 2016 to AOML-ETv2020 but a strong divergence thereafter. This is attributed to

changes in predictor variables and because there is no training data for AOML-ETv2020 beyond 2019. Thus, the ET method as applied is not suitable for extrapolation (Figure S1). Examples of the spatial differences in flux densities between extrapolated AOML-ETv2020 values and AOML-ETv2023, for January and July 2020, are shown in the maps of Figure S4. For January 2020, significantly greater effluxes are observed in the Western Subpolar and Polar North Pacific for AOML-ETv2023 and less uptake in the Subpolar and Polar regions of the Western Atlantic. Less uptake is also evident in AOML-ETv2023 in the Southern Ocean Antarctic regions around and south of the tip of South America (Figure S4a). For July 2020, there is less influx with AOML-ETv2023 in the Northern Subpolar regions and around the tip of the coast of South America. A significantly greater efflux is in the subtropics of the South Pacific. Of note are the negative values in the difference along the coast of Somalia and the Arabian Peninsula in July, indicating that AOML-ETv2023 has greater outgassing during the monsoon season (Figure S4d), which is poorly represented in AOML-ETv2020. While the differences described for many regions can be attributed to new data (Fig S3a), the changes observed in the Arabian Sea and in the Western Subpolar and Polar North Pacific are the result of changed algorithms using the expanded training dataset, as there is no new data in SOCAT v2023 in these regions (Figure S3b).

Text S3.

The influence of predictor variables on AOML-ET

Different sources are used for predictor variables (Table S1) that sometimes change with time. Changes in satellite remote sensors and platforms can make obtaining consistent time series challenging, which can affect the trends in the predictors. For AOML-ET, widely available and commonly used gridded products of predictors are used (Table S1) that are adjusted to the monthly 1˚ by 1˚ grid. The ET method and other ML approaches preclude determining direct causal relationships with individual predictors, but qualitative insights can be gained from the global annual trends of predictors and corresponding global sea-air CO_2 fluxes. Annual area-weighted values versus year are shown for SST and MLD (Figure S5a) and for SSS and Chl-a (Figure S5b). The SST, SSS, and MLD show a positive trend with time, while MLD decreases with time up to 2018. A large step change occurs in 2018 with decreases in SST, SSS, and, in particular, MLD, while Chl-a shows a significant increase. The increase in global MLD is likely spurious due to a change in the Hybrid Coordinate Ocean Model (HYCOM) model setup going from version GOFS3.0 to GOFS3.1 used to estimate the MLDs (https://www.hycom.org/faqs). This step change is not apparent in the global fluxes for AOML-ET. Interestingly, omitting either MLD or Chl-a has the same effect on the global sea-air CO_2 fluxes derived from AOML-ETv2020, with a large decrease in ocean sink from 2018-2020 and an accelerated uptake thereafter.

The impact of omitting and adding predictor variables is used as a means to determine the robustness of the predicted fCO_{2w} values using AOML-ET. The MLD and Chl-a were selected for omission as their quality and resolution are of lower fidelity than the other predictors, particularly at the start of the record. MLD are model-derived, and

Chl-a is a satellite ocean color product interpolated for regions and times with cloud cover and darkness. The AOML-ET method is relatively insensitive to withholding Chl-a and MLD predictor variables. That is, withholding the predictors results in no appreciable differences in flux estimates on a global scale up to 2018, but differences show up in regional patterns. All runs show approximately the same magnitude, variability, and trends in global sea-air CO_2 fluxes within their monthly variability of 0.3 PgC yr^{-1} (Figure S1). Omitting Chl-a increases the annual global uptake by about 0.2 PgC yr^{-1} up to 2018. Omitting MLD has a smaller global effect with differences < 0.1 PgC yr^{-1} for up to 2018. The second moment of the wind, $<u^2>$, is added as a predictor as it can both directly and indirectly affect fCO_{2w}. The direct effect on fCO_{2w} is through the impact of the CO_2 flux adding CO_2 to the ocean. This has a small influence; for example, a flux density of 0.5 mol m^{-2} yr^{-1} and a 50 m mixed layer will yield a change in fCO_{2w} of ≈ 0.06 µatm day^{-1}. Indirect effects of $<u^2>$ of enhanced mixing, erosion of the mixed layer, and upwelling are bigger factors for fCO_{2w}. However, adding wind $<u^2>$ as a predictor variable does not show any large differences in annual global uptake with the default AOML-ET up to 2018 (Figure S1). The omission of Chl-a and MLD degrades the skill of AOML-ET slightly and in similar amounts compared to the default configuration as summarized in a Taylor diagram (Figure 6). Figure 6 includes AOML-ET output created without MLD, without Chl-a, and adding $<u^2>$ as predictors. The RMSE of fCO_{2w} for both permutations is 23 µatm; variability (as expressed as a standard deviation of all data over the 23-year timespan) is 35 and the correlation coefficient, r^2, is 0.83. The addition of $<u^2>$ shows similar skill as the default configuration with a RMSE of 21 µatm, a r^2 of 0.88, and a standard deviation between 36 µatm compared to the standard deviation of the data withheld for testing of 42 µatm.

To investigate how the time-dependent predictor contributes to the trend, JDN expressed as month since October 1997 was substituted by annual global averaged XCO_{2a} in AOML-ETv2023, and omitted. As described in the main text, using XCO_{2a} as a predictor shows very little difference compared to using JDN in global fluxes up to 2019, after which the run with XCO_{2a} shows ≈ 0.4 Pg C yr^{-1} greater uptake attributed to rapidly increasing XCO_{2a} (Figure S6). Omitting time and XCO_{2a} as predictors shows a near constant global flux of 1.6 ± 0.1 PgC yr^{-1} for 1998-2022, indicating that the atmospheric CO_2 increases are the dominant driver of increasing sea-air CO_2 fluxes over time. The slight decrease in uptake in the scenario without JDN and XCO_{2a} as predictors indicates the expected decrease due to changing ocean conditions with surface warming and decreasing MLD.

Text S4
References for Supplemental Material

Aznar, P. What is the difference between Extra Trees and Random Forest? *17/06/2020* https://quantdare.com/what-is-the-difference-between-extra-trees-and-random-forest/

Chau, T. T. T., Gehlen, M., & Chevallier, F. (2022). A seamless ensemble-based reconstruction of surface ocean pCO_2 and air–sea CO_2 fluxes over the global coastal and open oceans. *Biogeosciences, 19*(4), 1087-1109. https://bg.copernicus.org/articles/19/1087/2022/

DeVries, T., Yamamoto, K., Wanninkhof, R., Gruber, N., Hauck, J., Müller, J. D., et al. (2023). Magnitude, trends, and variability of the global ocean carbon sink from 1985-2018. *Global Biogeochemical Cycles,* https://doi.org/10.1029/2023GB007780

Fay, A. R., Gregor, L., Landschützer, P., McKinley, G. A., Gruber, N., Gehlen, M., et al. (2021). SeaFlux: harmonization of air–sea CO_2 fluxes from surface pCO_2 data products using a standardized approach. *Earth Syst. Sci. Data, 13*(10), 4693-4710. https://essd.copernicus.org/articles/13/4693/2021/

Friedlingstein, P., O'Sullivan, M., Jones, M. W., Andrew, R. M., Bakker, D. C. E., Hauck, J., et al. (2023). Global Carbon Budget 2023. *Earth Syst. Sci. Data, 15*(12), 5301-5369. https://essd.copernicus.org/articles/15/5301/2023/

Gregor, L., Lebehot, A. D., Kok, S., & Scheel Monteiro, P. M. (2019). A comparative assessment of the uncertainties of global surface ocean CO_2 estimates using a machine-learning ensemble (CSIR-ML6 version 2019a) – have we hit the wall? *Geosci. Model Dev., 12*(12), 5113-5136. https://gmd.copernicus.org/articles/12/5113/2019/

Geurts, P., Ernst, D., & Wehenkel, L. (2006). Extremely randomized trees. *Machine Learning, 63*(1), 3-42. https://doi.org/10.1007/s10994-006-6226-1

Müller, J. D. (2023). RECCAP2-ocean data collection [Version V1]. https://doi.org/10.5281/zenodo.7990823, Zenodo

Pierrot, D., Neil, C., Sullivan, K., Castle, R., Wanninkhof, R., Lueger, H., et al. (2009). Recommendations for autonomous underway pCO_2 measuring systems and data reduction routines. *Deep -Sea Res II, 56*, 512-522.

Rödenbeck, C., D. C. E. Bakker, N. Gruber, Y. Iida, A. R. Jacobson, S. Jones, et al. (2015). Data-based estimates of the ocean carbon sink variability –first results of the Surface Ocean pCO_2 Mapping intercomparison (SOCOM). *Biogeosciences, 12*, 7251-7278. https://doi.org/10.5194/bg-12-7251-2015

Rödenbeck, C., DeVries, T., Hauck, J., Le Quéré, C., & Keeling, R. F. (2022). Data-based estimates of interannual sea–air CO_2 flux variations 1957–2020 and their relation to environmental drivers. *Biogeosciences, 19*(10), 2627-2652. https://bg.copernicus.org/articles/19/2627/2022/

Roobaert, A., Laruelle, G. G., Landschützer, P., & Regnier, P. (2018). Uncertainty in the global oceanic CO_2 uptake induced by wind forcing: quantification and spatial analysis. *Biogeosciences, 15*(6), 1701-1720. https://www.biogeosciences.net/15/1701/2018/

Takahashi, T., Sutherland, S. C., Wanninkhof, R., Sweeney, C., Feely, R. A., Chipman, D. W., 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 II, 2009*, 554-577.

Wanninkhof, R. (1992). Relationship between gas exchange and wind speed over the ocean. *J. Geophys. Res., 97*, 7373-7381. https://doi.org/10.1029/92JC00188

Wanninkhof, R. (2014). Relationship between wind speed and gas exchange over the ocean revisited. *Limnol and Oceanogr: Methods, 12*, 351-362. https://doi.org/10.4319/lom.2014.12.351

Weiss, R. F., & Price, B. A. (1980). Nitrous oxide solubility in water and seawater. *Mar. Chem., 8*, 347-359. https://doi.org/10.1016/0304-4203(80)90024-9

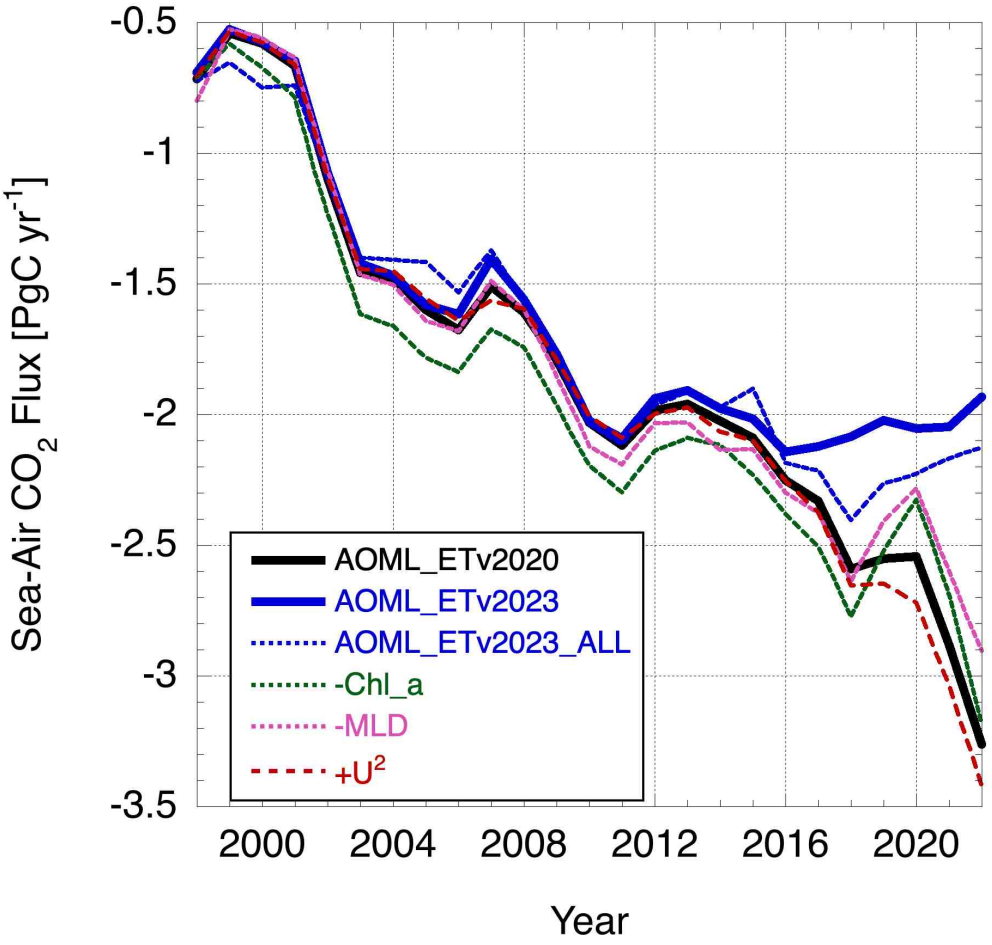


Figure S1. Sea-air CO_2 fluxes using AOML-ETv2020 and AOML-ETv2023. AOML-ETv2020 (default) (black solid line); AOML-ETv2023 (solid blue line); AOML-ETv2023 using all available data for training without withholding data for testing "ALL" (dashed blue line); AOML-ETv2020 omitting Chl-a as predictor (green dashed line); omitting MLD as predictor (pink dashed line); including $<u^2>$ as predictor (red dashed line).

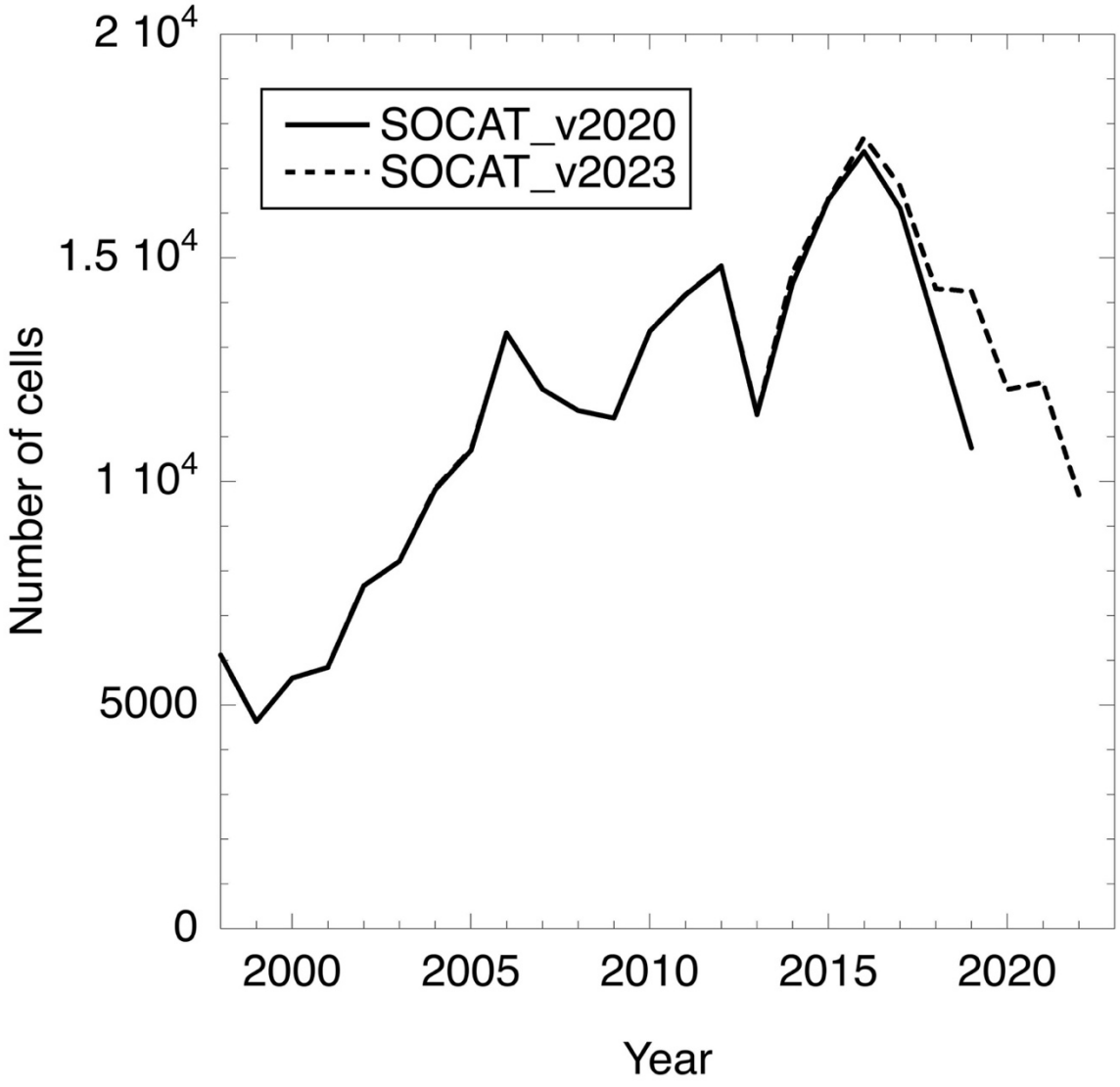
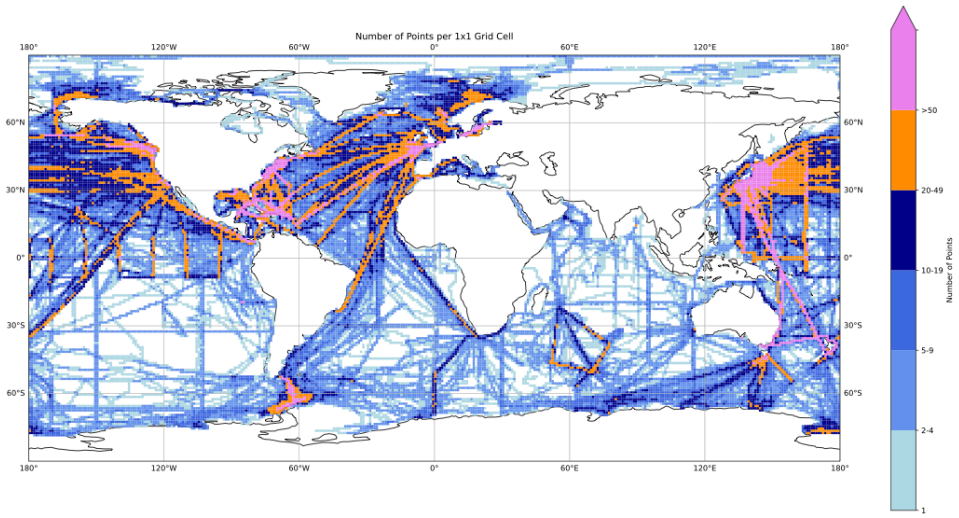
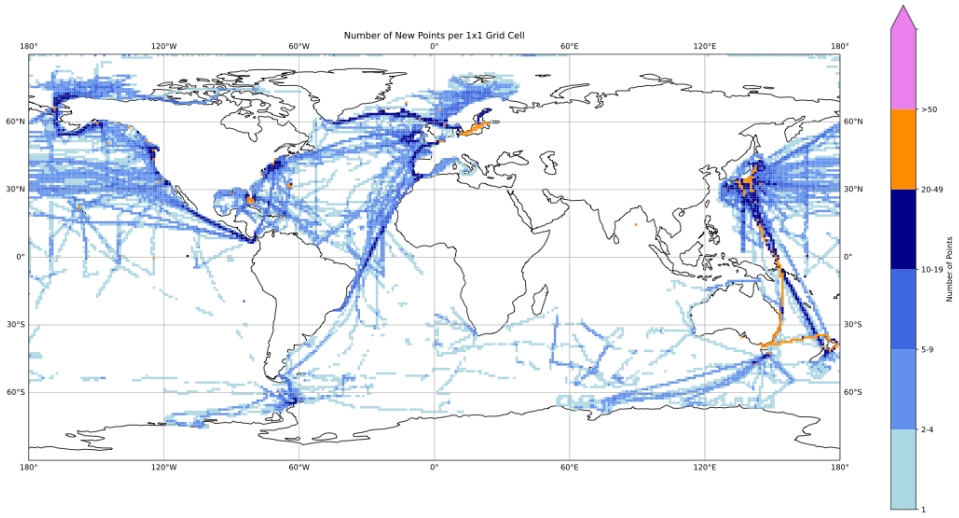


Figure S2. Number of monthly 1˚ by 1˚ cells in SOCATv2020 (solid line) and SOCATv2023 (dashed line). Each annual update of SOCAT contains the data from the previous version and new additions.



a



b

Figure S3. (a) Locations of the grid cells in SOCATv2020 with number of observations in each cell color coded. (b) Locations of the new grid cells in SOCATv2023 compared to SOCATv2020 with number of observations in each cell color coded.

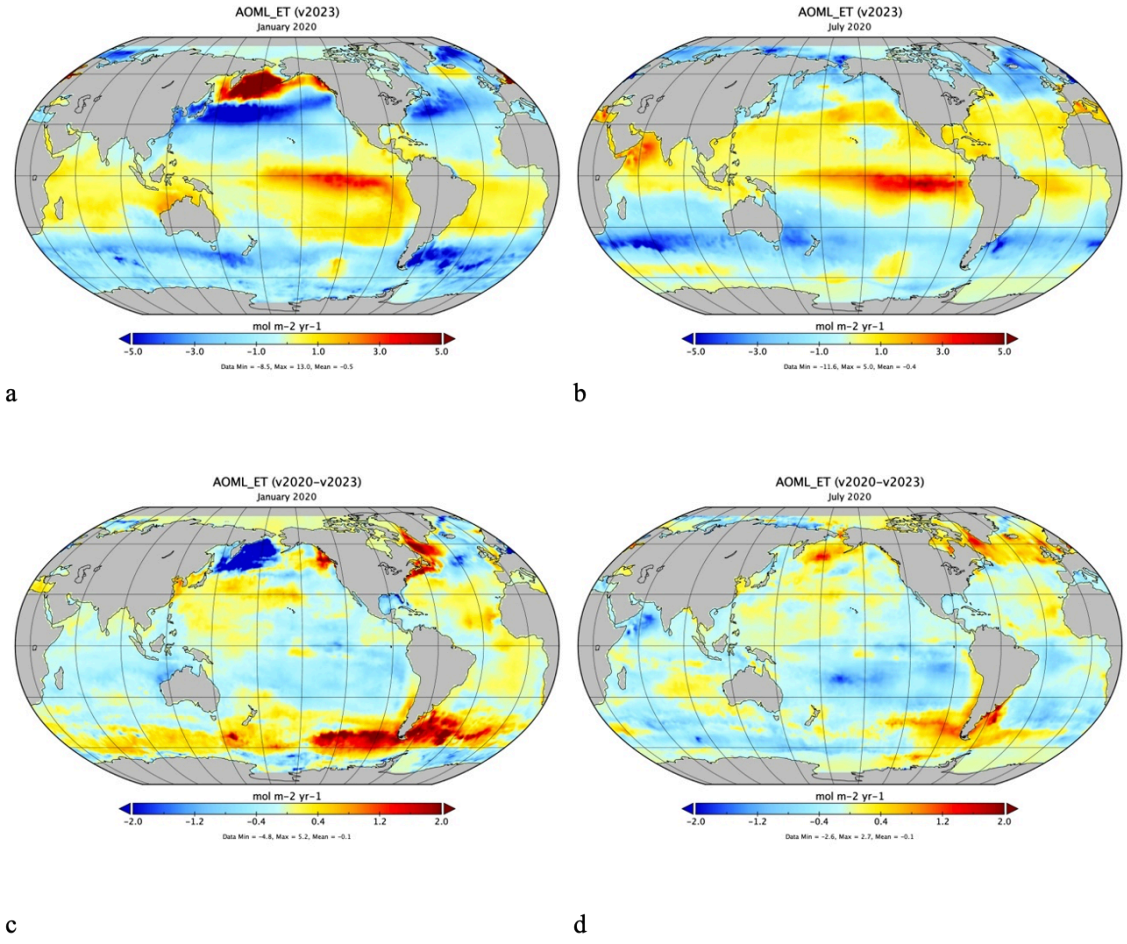
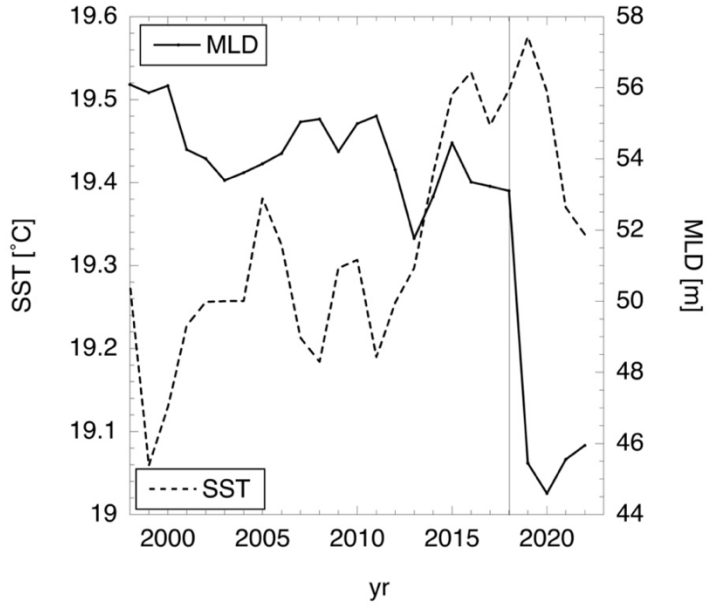
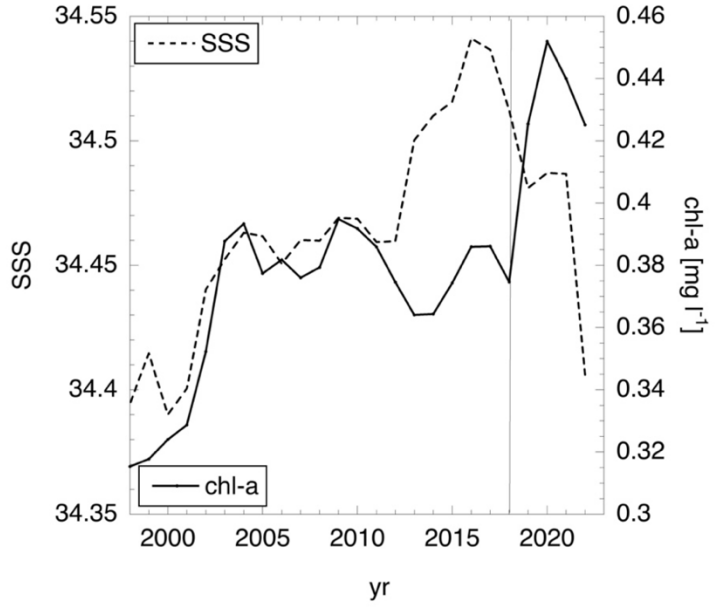


Figure S4. Flux densities of AOML-ETv2023 for January 2020 (a) and July 2020 (b), and difference in fluxes between AOML-ETv2020 and AOML-ETv2023 for January 2020 (c) and July 2020 (d). Note the difference in scales between a, b and c, d. Units are mol m^{-2} yr^{-1}.



a



b

Figure S5. Annual area-weighted global values of predictor variables used in AOML-ET. (a) SST (dashed line, left axis) and MLD (solid line, right axis); (b) SSS (dashed line, left axis) and Chl-a (solid line, right axis).

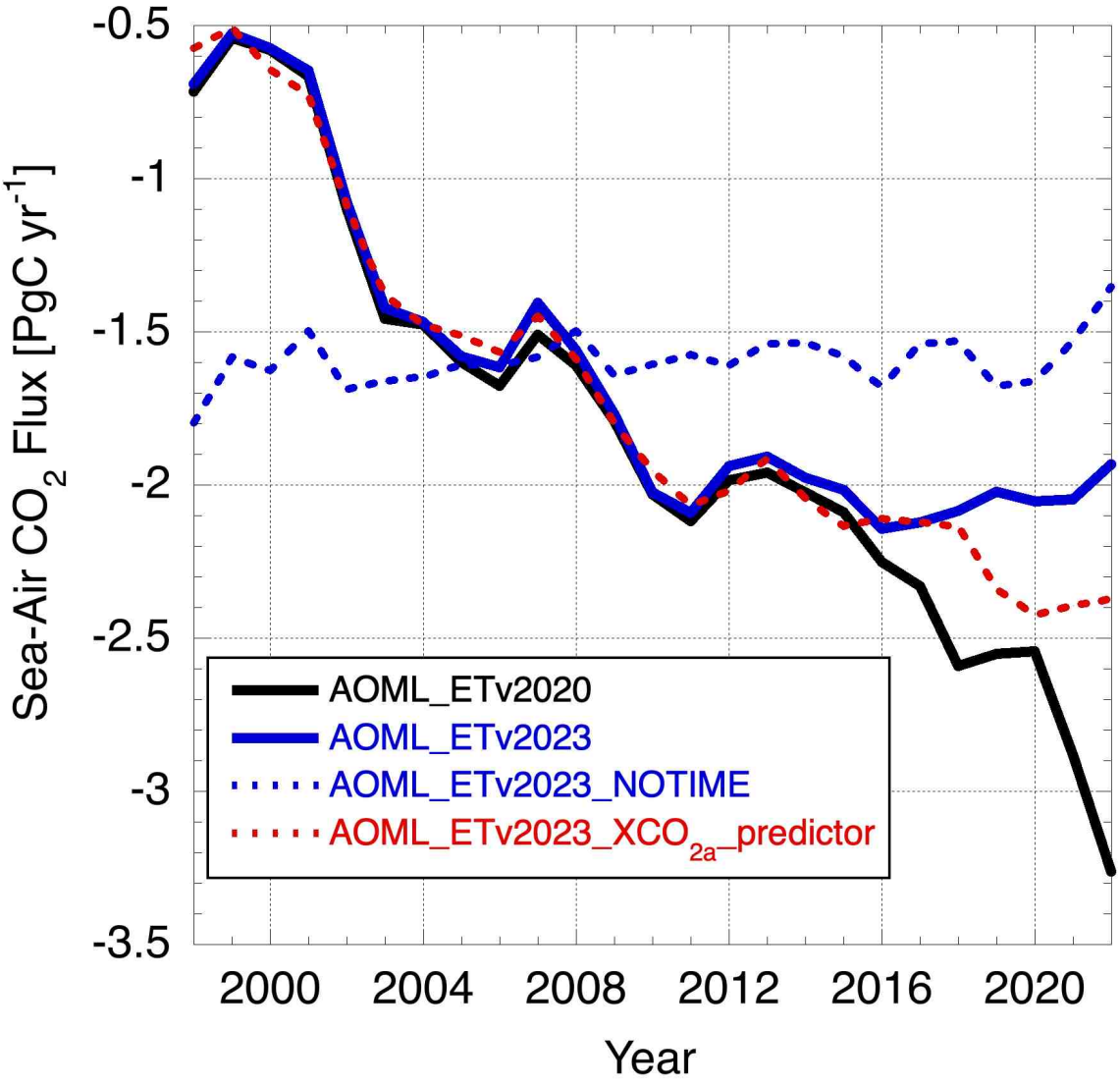


Figure S6. Sea-air CO_2 fluxes using AOML-ETv2020 and AOML-ETv2023. AOML-ETv2020 (default) with fCO_{2w} as target (black solid line); AOML-ETv2023 with fCO_{2w} as target (blue solid line); AOML-ETv2023 without time as predictor (dashed blue line); AOML-ETv2023 XCO_{2a} instead of time as predictor (red dashed line).

Creation of fCO_2 maps

Variables[1]	abbrev.	unit	Source/notes
Training set			
Partial pressure of CO_2	$spCO_2$	µatm	monthly gridded data SOCATv2020[2]
Fugacity of CO_2	fCO_{2w}	µatm	SOCAT v2020[3]
Sea surface temperature	SST	˚C	gridded data SOCAT v2020
Sea surface salinity	SSS		gridded data SOCAT v2020
Mixed layer depth	MLD		HYCOM model[4]
Julian day	JDN	mo	month since Oct. 1997
Latitude	LAT	degree	
Longitude	SLON	degree	vector longitude (SIN)
Longitude	CLON	degree	vector longitude (COS)
Chlorophyll-a	Chl-a	log (mg/l)	oceancolor.gsfc.nasa.gov
Dependent variable/ Target			
Partial pressure of CO_2	$spCO_2$	µatm	for surface water
Fugacity of CO_2	fCO_{2w}	µatm	for surface water
Predictor/Interpolation variable			
Sea surface temperature	STT	˚C	NOAA OISST
Sea surface salinity	SSS		HYCOM
Mixed layer depth	MLD	m	HYCOM
Chlorophyll-a	Chl-a	log (mg/l)	oceancolor.gsfc.nasa.gov[5]
Time	JDN	mo	month since Oct. 1997
Latitude	Lat		
Longitude	SLON		vector longitude (SIN)
Longitude	CLON		vector longitude (COS)

Creation of flux maps

Dependent variable			
Sea-air CO_2 flux density	F_{sa}	mol m^{-2} s^{-1}	$F_{sa} = k\ K_0\ (1-f_{ice})\ (pCO_{2atm}-spCO_2)$
Sea-air CO_2 flux density	F_{sa}	mol m^{-2} y^{-1}	$F_{sa} = k\ K_0\ (1-f_{ice})\ (fCO_{2w}-fCO2_a)$
Sea-air piston velocity	K_w	m s^{-1}	Wanninkhof (1992, 2014)
Gas transfer velocity	k	cm hr^{-1}	Wanninkhof (2014) .
Schmidt number	Sc		Wanninkhof (2014)
Second moment wind	$<u^2>$	m^2s^{-2}	ERA5 wind[6]
Solubility	alpha	mol $kg^{-1}atm^{-1}$	Weiss and Price (1980)
Solubility	K_0	mol $l^{-1}atm^{-1}$	Weiss and Price (1980)
Ice cover	f_{ice}	fraction	NOAA OISST[7]
Water partial pressure	$spCO_2$	µatm	SOCAT
Fugacity of CO_2 in water	fCO_2	µatm	SOCATv2020
Air partial pressure	pCO_{2atm}[8]	µatm	zonal mo. average xCO_2 MBL-RS
Fugacity of CO_2 in air	fCO_{2a}[9]	µatm	zonal mo. average xCO_2 MBL-RS
Partial pressure difference	ΔpCO_2	µatm	pCO_{2atm}- $spCO_2$

Air-water fugacity difference	ΔfCO_2	µatm	fCO_{2w}-fCO_{2a}
Global Flux	fgco2_glob	Pg C y^{-1}	Efflux negative in RECCAP2
Flux	F	Pg C y^{-1}	Efflux positive

1. Extra Trees (ET) regressors used to estimate the $spCO_2/fCO_{2w}$ values are: time, location, sea surface temperature, sea surface salinity, mixed layer depth, and chlorophyll-a.
2. SOCAT data are converted from fCO_2 to pCO_2 to meet the RECCAP2 submission criteria. These products are based on the monthly 1˚ by 1˚ gridded SOCATv2020 data holdings using datasets with QC flags of A through D and SOCAT data points flagged with WOCE flag values of 2 (= good). See, https://www.ncei.noaa.gov/data/oceans/ncei/ocads/data/0210711/ SOCATv2020_Gridded_Dat/ SOCATv2020_tracks_gridded_monthly.nc. The submission to RECCAP2 is for the period from October 1997 through December 2018.
3. The SOCATv2020 product is used in the analyses.
4. Mixed layer depth is based on a criterium of 0.03 change in density and obtained from http://orca.science.oregonstate.edu/2160.by.4320.monthly.hdf.mld030.hycom.php.
5. Chlorophyll-a values are from the NASA Ocean color monthly fields from SeaWiFS, and from AQUA/TERRA-MODIS. See, https://oceancolor.gsfc.nasa.gov/.
6. From https://www.ecmwf.int/en/forecasts/datasets/reanalysis-datasets/era5 where the 6-hourly 1/4˚ winds are aggregated on a monthly 1˚ by 1˚ grid to produce the second and third moments of the wind, $<u^2>$, and $<u^3>$.
7. From ftp://ftp.cdc.noaa.gov/Datasets/noaa.oisst.v2/icec.mnmean.nc following the approach of Takahashi et al. (2009), where k is scaled by (1-f), where f is the fraction of sea-ice covering the monthly 1˚ by 1˚ grid.
8. pCO_{2atm} = (P-pH_2O) XCO_{2a} , where XCO_{2a} is the interpolated MBL-RS product from NOAA/GML: https://www.esrl.noaa.gov/gmd/ccgg/mbl/mbl.html#ghg_product.
9. fCO_{2a} = Gf(T,S) (P -pH_2O) XCO_{2a}, where Gf(T,S) is the fugacity correction and pH_2O is the water vapor correction as summarized in Pierrot et al. (2009). P is the barometric pressure.

Table S1. Parameters used in AOML-ET

Method	Slope PgC yr^{-1} $decade^{-1}$	Std. error slope[b]	r^2	Average 1998-2020 PgC yr^{-1}	St. dev. PgC yr^{-1}
AOML_ETv2020	-0.89	0.059	0.92	-1.66	0.63
AOML_ETv2020_k_isotope	-0.90	0.066	0.90	-1.05	0.64
AOML_ETv2023	-0.72	0.075	0.82	-1.61	0.56
AOML_ETv2020_ΔfCO_2 target	-0.51	0.034	0.91	-1.65	0.35
AOML_ET v2023_ΔfCO_2 target	-0.45	0.030	0.92	-1.58	0.31
GCB-2023[a]	-0.46	0.034	0.89	-1.85	0.32

a. From Global_Carbon_Budget_2023v1.1; https://globalcarbonbudgetdata.org/latest-data.html
b. Standard error of the slope; the uncertainties in the text listed the uncertainty of the linear regression

Table S2. Linear regressions of the different ML approaches for 1998-2020