NEAR-REAL-TIME MONITORING OF GLOBAL OCEAN CARBON SINK

Piyu Ke[1,2,3]*, **Xiaofan Gui**[3]*, **Wei Cao**[3], **Dezhi Wang**[4], **Ce Hou**[5,6], **Lixing Wang**[1],
Xuanren Song[1], Yun Li[7], Biqing Zhu[8], Jiang Bian[3], Stephen Sitch[2], Philippe Ciais[9], 10,
Pierre Friedlingstein[2], Zhu Liu[1], 11
[1] Department of Earth System Science, Tsinghua University, Beijing, China
[2] Department of Mathematics and Statistics, Faculty of Environment, Science and Economy, University of Exeter, Exeter, UK
[3] Microsoft research
[4] School of Mathematics and Statistics, Lanzhou University, Lanzhou, China
[5] Department of Civil and Environmental Engineering, The Hong Kong University of Science and Technology, Hong Kong SAR, China
[6] Institute of Remote Sensing and Geographical Information System, School of Earth and Space Sciences, Peking University, Beijing, China
[7] Department of architecture, faculty of engineering science, KU Leuven, Leuven, Belgium
[8] Integrated Assessment and Climate Change Research Group and Exploratory Modeling of Human-natural Systems Research Group, International Institute for Applied Systems Analysis (IIASA), 2361 Laxenburg, Austria
[9] Laboratoire des Sciences du Climate et de l'Environnement LSCE, Orme de Merisiers 91191 Gif-sur-Yvette, France
[1]0 Climate and Atmosphere Research Center (CARE-C) The Cyprus Institute 20 Konstantinou Kavafi Street, 2121, Nicosia, Cyprus
[1]1 Institute of Climate and Carbon Neutrality, Department of Geography, The University of Hong Kong, Hong Kong SAR, China

ABSTRACT

The ocean, absorbing about 25% of anthropogenic CO_2 emissions, plays a crucial role in mitigating climate change. However, the delayed (by one year) traditional estimates of ocean-atmosphere CO_2 flux hinder timely understanding and response to the global carbon cycle's dynamics. Addressing this challenge, we introduce Carbon Monitor Ocean (CMO-NRT), a pioneering dataset providing near-real-time, monthly gridded estimates of global surface ocean fugacity of CO_2 (fCO_2) and ocean-atmosphere CO_2 flux from January 2022 to July 2023. This dataset marks a significant advancement by updating the global carbon budget's estimates through a fusion of data from 10 Global Ocean Biogeochemical Models (GOBMs) and 8 data products into a near-real-time analysis framework. By harnessing the power of Convolutional Neural Networks (CNNs) and semi-supervised learning techniques, we decode the complex nonlinear relationships between model or product estimates and observed environmental predictors. The predictive models, both for GOBM and data products, exhibit exceptional accuracy, with root mean square errors (RMSEs) maintaining below the 5% threshold. This advancement supports more effective climate change mitigation efforts by providing scientists and policymakers with timely and accurate data.

1 INTRODUCTION

The ocean is a pivotal component in the Earth's climate system, acting as a major sink for anthropogenic heat and carbon dioxide (CO2). This crucial role underscores the necessity for timely and accurate estimates of the global ocean carbon sink to inform climate change mitigation efforts and support the global stocktake process under the Paris Climate Agreement. Traditionally, the annual

*Equal contribution.

Global Carbon Budget report has provided estimates of the global ocean carbon sink, yet these figures are historically delayed by approximately one year Friedlingstein et al. (2022) due to computational and data gathering constraints. This latency hinders the timely assessment and response to the changing state of global carbon sinks, emphasizing the need for more immediate data solutions.

In response to this need, numerous methodologies have been developed, utilizing both in situ measurements and global ocean biogeochemical models (GOBMs) to estimate surface ocean fugacity of CO_2 (fCO_2), surface ocean partial pressure of CO_2 (pCO_2) and air-sea CO_2 flux. The Surface Ocean CO_2 Atlas (`www.socat.info`) (`SOCAT`), a community-led database, has played a central role by providing a comprehensive repository of quality-controlled surface ocean fCO_2 measurements. These data, spanning from 1957 to 2022 in its latest update Bakker et al. (2016), form the backbone of observation-based products that estimate pCO_2 across the globe. These products leverage sparse pCO_2 observations from `SOCAT`, applying multivariate linear regression or machine learning algorithms alongside observations of related variables to estimate pCO_2 at any given location and times (Landschützer et al., 2016; Rödenbeck et al., 2022; Chau et al., 2022; Gloege et al., 2022; Watson et al., 2020; Zeng et al., 2014; Iida et al., 2021; Gregor & Gruber, 2021). Parallelly, GOBMs offer a holistic simulation of the ocean's carbonate system by integrating its physical, biological, and chemical dynamicsy (Wright et al., 2021; Schwinger et al., 2016; Lacroix et al., 2021; Berthet et al., 2019; Hauck et al., 2020; Liao et al., 2020; Doney et al., 2009; Aumont et al., 2015; Nakano et al., 2011; Urakawa et al., 2020; Long et al., 2021). Though these models deliver extensive insights into the ocean's carbon cycle, their utility is constrained by computational demands and the inherent delay in `SOCAT` data updates, culminating in a significant latency in current global ocean carbon sink estimates.

To bridge this gap, we introduce Carbon Monitor Ocean (`CMO-NRT`), a dataset providing near-real-time, monthly gridded global surface ocean fCO_2 and ocean-atmosphere CO_2 flux data from January 2022 to July 2023. CMO-NRT represents a paradigm shift in the monitoring of oceanic carbon by employing a deep learning approach that amalgamates temporal, spatial, and environmental variables, offering a timely alternative to the delayed updates characteristic of existing methods. This dataset not only addresses the critical need for near-real-time data but also showcases the potential of advanced computational techniques to enhance our understanding of global biogeochemical cycles. This paper details the development and validation of CMO-NRT, illustrating its methodology, data integration processes, and the implications of its findings for global carbon monitoring efforts.

2 Data and methods

Carbon Monitor Ocean (`CMO-NRT`) utilizes advanced deep learning methods to build models that connects the output of GOBMs or ocean data products with relevant environmental variable observations. We then use `CMO-NRT` to extend the monthly gridded global surface ocean fCO_2 and ocean-atmosphere CO_2 flux estimates from each of the 10 GOBMs and 8 data products used in the Global Carbon Budget 2022, from January 2022 to July 2023. A pipeline of `CMO-NRT` is shown in Figure 1. Below we describe the calculation process in detail.

2.1 Data sources and pre-processing

2.1.1 Global ocean biogeochemical models and observation-based data products

We utilize monthly data until the end of 2021 from ten GOBMs and eight data products contributing to the Global Carbon Budget 2022 (Friedlingstein et al., 2022; Hauck et al., 2022). Each GOBM, detailed in Table 1, is driven by meteorological reanalysis and atmospheric CO2 levels, encapsulating physical, chemical, and biological influences on surface ocean pCO2 through a system of interconnected differential equations. Each `data product` is based on multivariate linear regression or machine learning techniques correlating `SOCAT` observational data with relevant variable observations. The fCO_2 output from each GOBM or data product is provided at a monthly resolution of `1° × 1°`.

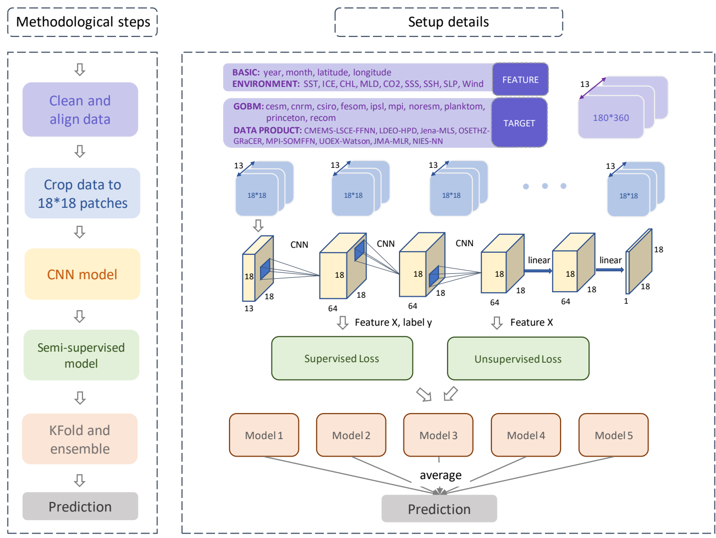

Figure 1: Schematic overview of the methodology and data sources for Carbon Monitor Ocean (CMO-NRT).

2.1.2 Observed predictors

Our predictive variables include biological, chemical, and physical factors linked to f CO2 fluctuations. These factors are SST, ICE, SSS, xCO2, MLD, SSH, chl a, SLP, and wind speed. These variables, detailed in Table 2, are bilinearly interpolated from their original grid to a 1° × 1° monthly resolution to align with our fCO_2 targets. Given that the xCO_2 data is only available until the end of 2022, we use a LightGBM model (Ke et al., 2017) to correlate year, month, latitude, longitude, mean atmospheric CO_2 data and xCO_2 , enabling near-real-time xCO_2 data. Data from 1979-2021 is split into training and validation datasets at an 8:2 ratio. Early stopping with LightGBM is implemented, tested on 2022 data, yielding a test RMSE of 1.74, approximately a 0.5% prediction error.

2.2 Deep learning method

Our study devised a deep learning approach tailored for near-real-time estimation of monthly gridded oceanic carbon fluxes, depicted in Figure 1. Integrating inputs of temporal, spatial, and environmental factors, we used GOBM or ocean data product outputs as prediction targets, transforming each dataset into a 180x360 grid format. For computational efficiency, all environmental factors were subdivided into 18x18 patches.

Recognizing the pivotal influence of surrounding conditions on oceanic carbon absorption, and the proficiency of Convolutional Neural Networks (CNNs) in integrating peripheral data, our model comprises multiple stacked CNN and Linear layers. This allows the capture of both linear and non-linear data relationships. To enhance the model's stability, classic semi-supervised approach , Pseudo-labeling, is employed. For data with labels, the Root Mean Square Error(RMSE) between labels and model predictions was calculated, serving as the supervised loss, L_s. For unlabeled data points, the RMSE between these pseudo-labels and model predictions was calculated as unsupervised loss L_u. The final loss of the model is determined by the weighted average of both supervised and unsupervised losses $wL_u + L_s$.

To bolster model robustness, we adopted a KFold strategy, partitioning the training data into five subsets. We iteratively used four subsets for training and one for validation, generating five models. Each model yielded a set of predictions for the test data, with the final prediction being the average of all five models' outputs. Detailed model information is provided in the Appendix B.

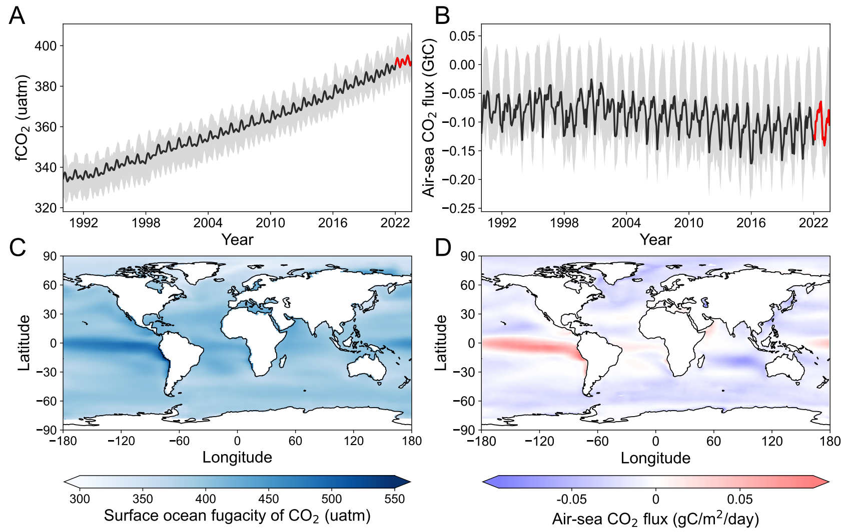

Figure 2: Monthly fCO_2 (A) and air-sea CO_2 flux (positive upward) (B) average of 10 GOBMs and 8 data products over 1990-2021 from Global Carbon Budget 2022 (black lines) and 2022-July 2023 (near-real-time predictions, red lines) from Carbon Monitor Ocean (CMO-NRT). Grey shaded areas represent the range of estimations from 10 GOBMs and 8 data products. Mean fCO_2 (C) and air-sea CO_2 flux (D) over August 2022-July 2023 estimated from CMO-NRT.

Figure 2 displays the monthly oceanic fCO_2 and air-sea CO_2 flux from 1990 through July 2023, alongside their corresponding gridded data for the period of August 2022 to July 2023 as estimated by `CMO-NRT`. More details of output data can be found at Appendix D

2.3 Calculations of pCO2 and air-sea CO2 flux

While our deep learning model enables the near-real-time prediction of fCO_2, a series of further calculations are necessary to determine the amount of CO_2 absorbed by the ocean, resulting in the CO_2 flux. Due to space constraints, the detailed calculation process for this part can be found in Appendix C.

3 Technical validation

To evaluate CMO-NRT's near-real-time forecasts, we trained models using data from 2000-2019, excluding the most recent two years' data from 10 GOBMs and 8 data products, which served as the test set. Our models achieved an RMSE of less than 5% across all targets, with most data products registering around 2% RMSE loss. For more detailed information, please refer to Table 3. We then analyzed the results from three perspectives: correlation, global quantity, and spatial distribution.

- The correlation between CMO-NRT predictions and original outputs for the 10 GOBMs and 8 data products during 2020-2021, was generally strong with most R2 values above 0.9. Scatter plots indicated more stable model performance with GOBMs due to less deviation from the fit line.
- Analysis of global fCO2 monthly variations revealed a high agreement in seasonality across GOBMs and data products (Figures 4 and 5). Our estimates were slightly higher, but most differences were under 3 µatm. Global aggregate comparisons demonstrated superior performance by GOBMs with most R2 values exceeding 0.85.
- Evaluation of the spatial fCO2 distribution patterns revealed consistency between CMO-NRT predictions and original outputs across latitudes (Figures 6 and 7), with most discrepancies under 20 µatm. Larger errors were observed in extreme fCO2 regions, such as the equatorial Pacific, Arctic Ocean, and areas with missing historical data. Monthly

mean fCO2 values, calculated separately across the 10 GOBMs and 8 data products during 2020-2021, showed high consistency with original results, with generally under 10 µatm discrepancies. Seasonal variations revealed higher original values from GOBMs in the Arctic region during summer, while data product performance was more uniform across months.

4 Conclusion

Our study, CMO-NRT, introduces a deep learning approach tailored for near-real-time estimation of oceanic carbon sinks, with a minimal delay of 1-2 months. This represents the first global effort for near-real-time oceanic carbon sink estimations. Notably, our method achieves an approximate 2% RMSE prediction for most GOBMs and data products, underscoring its robustness and accuracy. The development of the CMO-NRT dataset effectively bridges the gap between the immediacy of data and the requirement for swift analysis, thereby becoming an invaluable tool for climate change mitigation. Despite significant strides in oceanic carbon sink estimations, the complex evaluation of terrestrial carbon sinks, closely linked with human activities, persists as a lagging issue requiring urgent attention. Drawing on our experience in oceanic explorations, our work could serve as a reference for terrestrial carbon sink predictions, thereby advancing climate change solutions. As the urgency for climate change solutions escalates, our work offers a crucial tool for timely, informed decision-making, playing a central role in global climate change management.

References

O. Aumont, C. Ethé, A. Tagliabue, L. Bopp, and M. Gehlen. Pisces-v2: an ocean biogeochemical model for carbon and ecosystem studies. *Geosci. Model Dev.*, 8(8):2465–2513, 2015. ISSN 1991-9603. doi: 10.5194/gmd-8-2465-2015. URL https://gmd.copernicus.org/articles/8/2465/2015/https://gmd.copernicus.org/articles/8/2465/2015/gmd-8-2465-2015.pdf. GMD.

D. C. E. Bakker, B. Pfeil, C. S. Landa, N. Metzl, K. M. O'Brien, A. Olsen, K. Smith, C. Cosca, S. Harasawa, S. D. Jones, S. Nakaoka, Y. Nojiri, U. Schuster, T. Steinhoff, C. Sweeney, T. Takahashi, B. Tilbrook, C. Wada, R. Wanninkhof, S. R. Alin, C. F. Balestrini, L. Barbero, N. R. Bates, A. A. Bianchi, F. Bonou, J. Boutin, Y. Bozec, E. F. Burger, W. J. Cai, R. D. Castle, L. Chen, M. Chierici, K. Currie, W. Evans, C. Featherstone, R. A. Feely, A. Fransson, C. Goyet, N. Greenwood, L. Gregor, S. Hankin, N. J. Hardman-Mountford, J. Harlay, J. Hauck, M. Hoppema, M. P. Humphreys, C. W. Hunt, B. Huss, J. S. P. Ibánhez, T. Johannessen, R. Keeling, V. Kitidis, A. Körtzinger, A. Kozyr, E. Krasakopoulou, A. Kuwata, P. Landschützer, S. K. Lauvset, N. Lefèvre, C. Lo Monaco, A. Manke, J. T. Mathis, L. Merlivat, F. J. Millero, P. M. S. Monteiro, D. R. Munro, A. Murata, T. Newberger, A. M. Omar, T. Ono, K. Paterson, D. Pearce, D. Pierrot, L. L. Robbins, S. Saito, J. Salisbury, R. Schlitzer, B. Schneider, R. Schweitzer, R. Sieger, I. Skjelvan, K. F. Sullivan, S. C. Sutherland, A. J. Sutton, K. Tadokoro, M. Telszewski, M. Tuma, S. M. A. C. van Heuven, D. Vandemark, B. Ward, A. J. Watson, and S. Xu. A multi-decade record of high-quality fco2 data in version 3 of the surface ocean co2 atlas (socat). *Earth Syst. Sci. Data*, 8(2):383–413, 2016. ISSN 1866-3516. doi: 10.5194/essd-8-383-2016. URL https://essd.copernicus.org/articles/8/383/2016/https://essd.copernicus.org/articles/8/383/2016/essd-8-383-2016.pdf. ESSD.

Sarah Berthet, Roland Séférian, Clément Bricaud, Matthieu Chevallier, Aurore Voldoire, and Christian Ethé. Evaluation of an online grid-coarsening algorithm in a global eddy-admitting ocean biogeochemical model. *Journal of Advances in Modeling Earth Systems*, 11(6):1759–1783, 2019. ISSN 1942-2466. doi: https://doi.org/10.1029/2019MS001644. URL https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/2019MS001644.

T. T. T. Chau, M. Gehlen, and F. Chevallier. A seamless ensemble-based reconstruction of surface ocean pco2 and air–sea co2 fluxes over the global coastal and open oceans. *Biogeosciences*, 19(4):1087–1109, 2022. ISSN 1726-4189. doi: 10.5194/bg-19-1087-2022. URL https://bg.copernicus.org/articles/19/1087/2022/https:

//bg.copernicus.org/articles/19/1087/2022/bg-19-1087-2022.pdf. BG.

Climate Data Store Copernicus Climate Change Service. Oras5 global ocean reanalysis monthly data from 1958 to present, 2021.

Andrew Gilmore Dickson, Christopher L Sabine, and James Robert Christian. *Guide to best practices for ocean CO2 measurements*. North Pacific Marine Science Organization, 2007. ISBN 1897176074.

Scott C. Doney, Ivan Lima, Richard A. Feely, David M. Glover, Keith Lindsay, Natalie Mahowald, J. Keith Moore, and Rik Wanninkhof. Mechanisms governing interannual variability in upper-ocean inorganic carbon system and air–sea co2 fluxes: Physical climate and atmospheric dust. *Deep Sea Research Part II: Topical Studies in Oceanography*, 56(8):640–655, 2009. ISSN 0967-0645. doi: https://doi.org/10.1016/j.dsr2.2008.12.006. URL https://www.sciencedirect.com/science/article/pii/S096706450800427X.

Amanda R Fay and Galen A McKinley. Observed regional fluxes to constrain modeled estimates of the ocean carbon sink. *Geophysical Research Letters*, 48(20):e2021GL095325, 2021. ISSN 0094-8276.

P. Friedlingstein, M. O'Sullivan, M. W. Jones, R. M. Andrew, L. Gregor, J. Hauck, C. Le Quéré, I. T. Luijkx, A. Olsen, G. P. Peters, W. Peters, J. Pongratz, C. Schwingshackl, S. Sitch, J. G. Canadell, P. Ciais, R. B. Jackson, S. R. Alin, R. Alkama, A. Arneth, V. K. Arora, N. R. Bates, M. Becker, N. Bellouin, H. C. Bittig, L. Bopp, F. Chevallier, L. P. Chini, M. Cronin, W. Evans, S. Falk, R. A. Feely, T. Gasser, M. Gehlen, T. Gkritzalis, L. Gloege, G. Grassi, N. Gruber, Ö Gürses, I. Harris, M. Hefner, R. A. Houghton, G. C. Hurtt, Y. Iida, T. Ilyina, A. K. Jain, A. Jersild, K. Kadono, E. Kato, D. Kennedy, K. Klein Goldewijk, J. Knauer, J. I. Korsbakken, P. Landschützer, N. Lefèvre, K. Lindsay, J. Liu, Z. Liu, G. Marland, N. Mayot, M. J. McGrath, N. Metzl, N. M. Monacci, D. R. Munro, S. I. Nakaoka, Y. Niwa, K. O'Brien, T. Ono, P. I. Palmer, N. Pan, D. Pierrot, K. Pocock, B. Poulter, L. Resplandy, E. Robertson, C. Rödenbeck, C. Rodriguez, T. M. Rosan, J. Schwinger, R. Séférian, J. D. Shutler, I. Skjelvan, T. Steinhoff, Q. Sun, A. J. Sutton, C. Sweeney, S. Takao, T. Tanhua, P. P. Tans, X. Tian, H. Tian, B. Tilbrook, H. Tsujino, F. Tubiello, G. R. van der Werf, A. P. Walker, R. Wanninkhof, C. Whitehead, A. Willstrand Wranne, R. Wright, et al. Global carbon budget 2022. *Earth Syst. Sci. Data*, 14(11):4811–4900, 2022. ISSN 1866-3516. doi: 10.5194/essd-14-4811-2022. URL https://essd.copernicus.org/articles/14/4811/2022/https://essd.copernicus.org/articles/14/4811/2022/essd-14-4811-2022.pdf. ESSD.

L. Gloege, M. Yan, T. Zheng, and G. A. McKinley. Improved quantification of ocean carbon uptake by using machine learning to merge global models and pco2 data. *Journal of Advances in Modeling Earth Systems*, 14(2):e2021MS002620, 2022. ISSN 1942-2466. doi: https://doi.org/10.1029/2021MS002620. URL https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/2021MS002620.

Simon A. Good, Matthew J. Martin, and Nick A. Rayner. En4: Quality controlled ocean temperature and salinity profiles and monthly objective analyses with uncertainty estimates. *Journal of Geophysical Research: Oceans*, 118(12):6704–6716, 2013. ISSN 2169-9275. doi: https://doi.org/10.1002/2013JC009067. URL https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1002/2013JC009067.

L. Gregor and N. Gruber. Oceansoda-ethz: a global gridded data set of the surface ocean carbonate system for seasonal to decadal studies of ocean acidification. *Earth Syst. Sci. Data*, 13(2):777–808, 2021. ISSN 1866-3516. doi: 10.5194/essd-13-777-2021. URL https://essd.copernicus.org/articles/13/777/2021/https://essd.copernicus.org/articles/13/777/2021/essd-13-777-2021.pdf. ESSD.

Judith Hauck, Moritz Zeising, Corinne Le Quéré, Nicolas Gruber, Dorothee C. E. Bakker, Laurent Bopp, Thi Tuyet Trang Chau, Özgür Gürses, Tatiana Ilyina, Peter Landschützer, Andrew Lenton, Laure Resplandy, Christian Rödenbeck, Jörg Schwinger, and Roland Séférian. Consistency and challenges in the ocean carbon sink estimate for the global carbon budget. *Frontiers in Marine*

Science, 7, 2020. ISSN 2296-7745. doi: 10.3389/fmars.2020.571720. URL https://www.frontiersin.org/articles/10.3389/fmars.2020.571720.

Judith Hauck, Nicolas Mayot, Luke Gregor, Laurent Bopp, Marion Gehlen, Lukas Gloege, Nicolas Gruber, Ozgur Gurses, Yosuke Iida, Tatiana Ilyina, Annika Jersild, Keith Lindsay, Laure Resplandy, Christian Rodenbeck, Jorg Schwinger, Roland Seferian, Jamie Shutler, Hiroyuki Tsujino, Rebecca Wright, and Jiye Zeng. Global carbon budget 2022, surface ocean fugactiy of co2 (fco2) and air-sea co2 flux of individual global ocean biogechemical models and surface ocean fco2-based data-products, 2022. URL <GotoISI>://DRCI:DATA2023016025512889. Data.

Bell B. Hersbach H., Biavati G. Berrisford P., Muñoz Sabater J. Horányi A., Peubey C. Nicolas J., Rozum I. Radu R., Simmons A. Schepers D., Dee D. Soci C., and Thépaut J-N. Era5 monthly averaged data on single levels from 1940 to present, 2023.

B. Y. Huang, C. Y. Liu, V. Banzon, E. Freeman, G. Graham, B. Hankins, T. Smith, and H. M. Zhang. Improvements of the daily optimum interpolation sea surface temperature (doisst) version 2.1. *Journal of Climate*, 34(8):2923–2939, 2021. ISSN 0894-8755. doi: 10.1175/jcli-d-20-0166.1. URL <GotoISI>://WOS:000644147200007. Huang, Boyin Liu, Chunying Banzon, Viva Freeman, Eric Graham, Garrett Hankins, Bill Smith, Tom Zhang, Huai-Min Freeman, Eric/0000-0001-9654-5109 1520-0442.

Yosuke Iida, Yusuke Takatani, Atsushi Kojima, and Masao Ishii. Global trends of ocean co2 sink and ocean acidification: an observation-based reconstruction of surface ocean inorganic carbon variables. *Journal of Oceanography*, 77(2):323–358, 2021. ISSN 1573-868X. doi: 10.1007/s10872-020-00571-5. URL https://doi.org/10.1007/s10872-020-00571-5.

Fabrice Lacroix, Tatiana Ilyina, Moritz Mathis, Goulven G. Laruelle, and Pierre Regnier. Historical increases in land-derived nutrient inputs may alleviate effects of a changing physical climate on the oceanic carbon cycle. *Global Change Biology*, 27(21):5491–5513, 2021. ISSN 1354-1013. doi: https://doi.org/10.1111/gcb.15822. URL https://onlinelibrary.wiley.com/doi/abs/10.1111/gcb.15822.

X. Lan, P. Tans, K. Thoning, and NOAA Global Monitoring Laboratory. Noaa greenhouse gas marine boundary layer reference - co2, 2023.

Peter Landschützer, Nicolas Gruber, and Dorothee C. E. Bakker. Decadal variations and trends of the global ocean carbon sink. *Global Biogeochemical Cycles*, 30(10):1396–1417, 2016. ISSN 0886-6236. doi: https://doi.org/10.1002/2015GB005359. URL https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1002/2015GB005359.

Enhui Liao, Laure Resplandy, Junjie Liu, and Kevin W. Bowman. Amplification of the ocean carbon sink during el niños: Role of poleward ekman transport and influence on atmospheric co2. *Global Biogeochemical Cycles*, 34(9):e2020GB006574, 2020. ISSN 0886-6236. doi: https://doi.org/10.1029/2020GB006574. URL https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/2020GB006574.

Matthew C. Long, J. Keith Moore, Keith Lindsay, Michael Levy, Scott C. Doney, Jessica Y. Luo, Kristen M. Krumhardt, Robert T. Letscher, Maxwell Grover, and Zephyr T. Sylvester. Simulations with the marine biogeochemistry library (marbl). *Journal of Advances in Modeling Earth Systems*, 13(12):e2021MS002647, 2021. ISSN 1942-2466. doi: https://doi.org/10.1029/2021MS002647. URL https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/2021MS002647.

S. Maritorena, O. H. F. d'Andon, A. Mangin, and D. A. Siegel. Merged satellite ocean color data products using a bio-optical model: Characteristics, benefits and issues. *Remote Sensing of Environment*, 114(8):1791–1804, 2010. ISSN 0034-4257. doi: 10.1016/j.rse.2010.04.002. URL <GotoISI>://WOS:000278943900015. Maritorena, Stephane d'Andon, Odile Hembise Fanton Mangin, Antoine Siegel, David A. Peters, Steef w.m./A-7433-2011; Siegel, David A/C-5587-2008 Maritorena, Stephane/0000-0002-5508-8391 1879-0704.

Hideyuki Nakano, Hiroyuki Tsujino, Mikitoshi Hirabara, Tamaki Yasuda, Tatsuo Motoi, Masao Ishii, and Goro Yamanaka. Uptake mechanism of anthropogenic co2 in the kuroshio extension region in an ocean general circulation model. *Journal of Oceanography*, 67(6):765–783, 2011. ISSN 1573-868X. doi: 10.1007/s10872-011-0075-7. URL https://doi.org/10.1007/s10872-011-0075-7.

C. Rödenbeck, T. DeVries, J. Hauck, C. Le Quéré, and R. F. Keeling. Data-based estimates of interannual sea–air co2 flux variations 1957–2020 and their relation to environmental drivers. *Biogeosciences*, 19(10):2627–2652, 2022. ISSN 1726-4189. doi: 10.5194/bg-19-2627-2022. URL https://bg.copernicus.org/articles/19/2627/2022/https://bg.copernicus.org/articles/19/2627/2022/bg-19-2627-2022.pdf. BG.

J. Schwinger, N. Goris, J. F. Tjiputra, I. Kriest, M. Bentsen, I. Bethke, M. Ilicak, K. M. Assmann, and C. Heinze. Evaluation of noresm-oc (versions 1 and 1.2), the ocean carbon-cycle stand-alone configuration of the norwegian earth system model (noresm1). *Geosci. Model Dev.*, 9(8):2589–2622, 2016. ISSN 1991-9603. doi: 10.5194/gmd-9-2589-2016. URL https://gmd.copernicus.org/articles/9/2589/2016/https://gmd.copernicus.org/articles/9/2589/2016/gmd-9-2589-2016.pdf. GMD.

Colm Sweeney, Emanuel Gloor, Andrew R Jacobson, Robert M Key, Galen McKinley, Jorge L Sarmiento, and Rik Wanninkhof. Constraining global air-sea gas exchange for co2 with recent bomb 14c measurements. *Global biogeochemical cycles*, 21(2), 2007. ISSN 0886-6236.

L. Shogo Urakawa, Hiroyuki Tsujino, Hideyuki Nakano, Kei Sakamoto, Goro Yamanaka, and Takahiro Toyoda. The sensitivity of a depth-coordinate model to diapycnal mixing induced by practical implementations of the isopycnal tracer diffusion scheme. *Ocean Modelling*, 154:101693, 2020. ISSN 1463-5003. doi: https://doi.org/10.1016/j.ocemod.2020.101693. URL https://www.sciencedirect.com/science/article/pii/S1463500320301955.

Rik Wanninkhof. Relationship between wind speed and gas exchange over the ocean. *Journal of Geophysical Research: Oceans*, 97(C5):7373–7382, 1992. ISSN 0148-0227.

Andrew J. Watson, Ute Schuster, Jamie D. Shutler, Thomas Holding, Ian G. C. Ashton, Peter Landschützer, David K. Woolf, and Lonneke Goddijn-Murphy. Revised estimates of ocean-atmosphere co2 flux are consistent with ocean carbon inventory. *Nature Communications*, 11(1):4422, 2020. ISSN 2041-1723. doi: 10.1038/s41467-020-18203-3. URL https://doi.org/10.1038/s41467-020-18203-3.

R_F Weiss. Carbon dioxide in water and seawater: the solubility of a non-ideal gas. *Marine chemistry*, 2(3):203–215, 1974. ISSN 0304-4203.

R. M. Wright, C. Le Quéré, E. Buitenhuis, S. Pitois, and M. J. Gibbons. Role of jellyfish in the plankton ecosystem revealed using a global ocean biogeochemical model. *Biogeosciences*, 18(4):1291–1320, 2021. ISSN 1726-4189. doi: 10.5194/bg-18-1291-2021. URL https://bg.copernicus.org/articles/18/1291/2021/https://bg.copernicus.org/articles/18/1291/2021/bg-18-1291-2021.pdf. BG.

J. Zeng, Y. Nojiri, P. Landschützer, M. Telszewski, and S. Nakaoka. A global surface ocean fco2 climatology based on a feed-forward neural network. *Journal of Atmospheric and Oceanic Technology*, 31(8):1838–1849, 2014. ISSN 0739-0572. doi: https://doi.org/10.1175/JTECH-D-13-00137.1. URL https://journals.ametsoc.org/view/journals/atot/31/8/jtech-d-13-00137_1.xml.

Table 2: Sources of input data sets.

Variable	Abbreviation	Data Product	Resolution
Sea Surface Temperature	SST	NOAA: OISST Huang et al. (2021)	1981/9-2023/07, daily, 720*1440 (latitude*longitude)
Sea Ice Fraction	ICE		
Sea Surface Salinity	SSS	Met Office: EN4 Good et al. (2013)	1959/01-2023/07, monthly, 173*360
Atmospheric CO_2 mixing ratio	xCO_2	NOAA: GreenhouseGas Marine Boundary Layer Reference Lan et al. (2023)	1979/01-2022/12, weekly, 180*1
Mixed Layer Depth	MLD	ECMWF: ORAS5 Copernicus Climate Change Service (2021)	1959/01-2023/08, monthly, 1021*1442
Sea Surface Height	SSH		
Chlorophyll-a	Chl a	ESA: GlobColour Maritorena et al. (2010)	1997/09-2023/09, monthly, 180*360
Sea Level Pressure	SLP	ECMWF: ERA5 Hersbach H. et al. (2023)	1959/01-2023/08, monthly, 1021*1442
Wind Speed	Wind		
Year, month, longitude and latitude			

A Data

Table 1: Global ocean biogeochemical models and ocean data products used in the Global Carbon Budget 2022 (Friedlingstein et al., 2022; Hauck et al., 2022)

Type	Datasets	Data information
Global ocean biogeochemistry models	NEMO-PlankTOM12 Wright et al. (2021)	Surface ocean fugacity of CO_2, 1959/01-2021/12, monthly, 180*360
	MICOM-HAMOCC (NorESM-OCv1.2) Schwinger et al. (2016)	
	MPIOM-HAMOCC6 Lacroix et al. (2021)	
	NEMO3.6-PISCESv2-gas (CNRM) Berthet et al. (2019)	
	FESOM-2.1-REcoM2 Hauck et al. (2020)	
	MOM6-COBALT (Princeton) Liao et al. (2020)	
	CESM-ETHZ Doney et al. (2009)	
	NEMO-PISCES (IPSL) Aumont et al. (2015)	
	MRI-ESM2-1 (Nakano et al., 2011; Urakawa et al., 2020)	
	CESM2 Long et al. (2021)	
Ocean data products	MPI-SOMFFN Landschützer et al. (2016)	
	Jena-MLS Rödenbeck et al. (2022)	
	CMEMS-LSCE-FFNNv2 Chau et al. (2022)	
	LDEO-HPD Gloege et al. (2022)	
	UOEx-Watson Watson et al. (2020)	
	NIES-NN Zeng et al. (2014)	
	JMA-MLR Iida et al. (2021)	
	OS-ETHZ-GRaCER Gregor & Gruber (2021)	

B Methods

The framework of the model is illustrated in Figure 3. The data points with label values were denoted as D_l, while those with no labels were termed D_u. In the initial phase, the model was trained exclusively on D_l data points. The `RMSE`between labels and predictions was calculated, serving as the supervised loss, L_s . Simultaneously, we aimed to ensure stability in predictions even on data without labels. To this end, we employed a classic semi-supervised approach known as pseudo-labeling. Predictions were made by randomly removing 10% of features, which were then used as pseudo-labels. The results predicted by eliminating 30% of features randomly from the input were used as predictions. The `RMSE` between these pseudo-labels and predictions was calculated as unsupervised loss, L_u. The model was subsequently updated through backward propagation, utilizing the weighted sum of $wL_u + L_s$ as the model's loss. This method has been shown to enhance the model's stability.

Our model's detail architecture, as illustrated in the bottom right corner of Figure 3, is an integration of multi-layered CNN and linear models. The model has been meticulously structured to process our input data efficiently and generate accurate predictions.

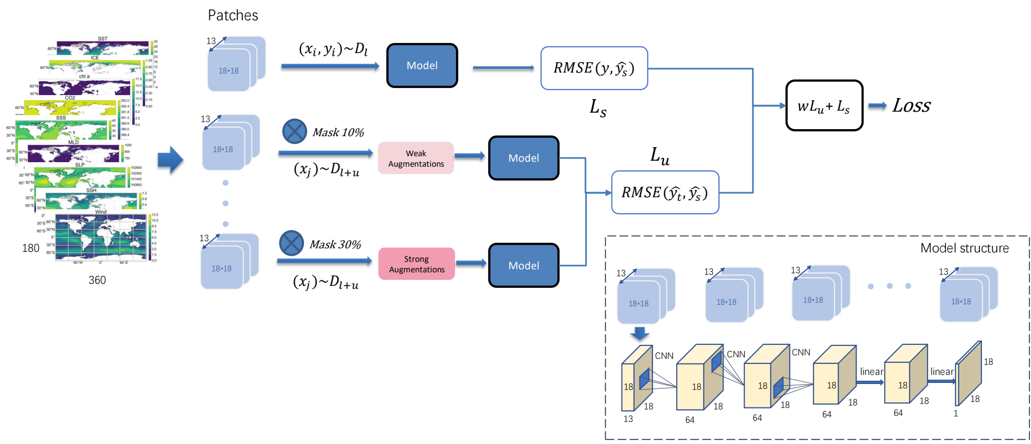

Figure 3: Framework of our methodology.

The input layer is designed to align with the format of our data, preserving a dimension of `18x18x13`. This dimensionality accounts for the width, height, and depth (which corresponds to the thirteen input factors considered in this study) and is consistently upheld across all CNN and linear layers. Within the multi-layered CNN models, the hidden layers are configured with dimensions set at 13, 64 and 64. These layers are intentionally designed to automatically and adaptively identify and learn spatial hierarchies from the data. Following the CNN layers, the linear model layers, structured with dimensions of 64, 64, and 1, perform mathematical operations that linearly transform the feature space, aiding in the generation of effective predictions. The architecture of the model culminates in an output layer with a dimension of 1, indicating that it yields a single output value. This value represents our final predicted value for oceanic carbon fCO_2. The network comprises several layers, including the Convolutional Layer, Rectified Linear Unit (ReLU) Layer, and Fully Connected Layer. These layers work in harmony, each contributing to the processing and transformation of input data, ultimately leading to the output prediction.

C CALCULATIONS OF pCO_2 AND AIR-SEA CO_2 FLUX

The pCO_2 is calculated by the following equation:

$$pCO_2 = fCO_2 \times \exp\left(-P_{\text{atm}}^{\text{surf}} \times \frac{(B + 2\delta)}{R \times T}\right)^{-1} \tag{1}$$

where $P_{\text{atm}}^{\text{surf}}$ is the atmospheric surface pressure from ECMWF Reanalysis version 5 Hersbach H. et al. (2023), T is the sea surface temperature(SST) in Kelvin from National Oceanic and Atmospheric Administration (NOAA) optimally interpolated SST (OISST) Huang et al. (2021), B and δ are virial coefficients from Weiss Weiss (1974), R is the gas constant Dickson et al. (2007).

The air–sea CO_2 flux is calculated here by the standard bulk equation:

$$F_{CO_2} = k_w \times S_{CO_2} \times (1 - f_{\text{ice}}) \times (pCO_2^{\text{atm-moist}} - pCO_2^{\text{ocean}}) \tag{2}$$

which parameterizes the air-sea CO_2 flux F_{CO_2} as a function of the gas transfer velocity (k_w), CO_2 solubility (S_{CO_2}), ice fraction (f_{ice}), and partial pressure of CO_2 in moist air $(pCO_2^{\text{atm-moist}})$ and surface ocean (pCO_2^{ocean}). Solubility is calculated following Weiss Weiss (1974) and partial pressure of moist air ($pCO_2^{\text{atm-moist}}$) is calculated following Equation,

$$pCO_2^{\text{atm-moist}} = xCO_2 \times (P_{\text{atm}} - pH_2O) \tag{3}$$

where xCO_2 is the dry air mixing ratio of atmospheric CO_2 from NOAA Greenhouse Gas Marine Boundary Layer Reference Lan et al. (2023), P_{atm} is the total atmospheric pressure from ECMWF Reanalysis version 5 Hersbach H. et al. (2023), and pH_2O is the saturation vapor pressure Dickson et al. (2007). We use the Wanninkhof Wanninkhof (1992) formulation for the gas transfer velocity:

$$k_w = k_{w,\text{scaled}} \times u^2 \times \left(\frac{S_c}{660}\right)^{-0.5} \tag{4}$$

which parameterizes k_w as a function of wind speed squared (u^2) and the Schmidt number (S_c). k_w is scaled by a factor of $k_{w,scaled}$ for each wind product to match the invasion of bomb as of 1994) (Sweeney et al., 2007; Fay & McKinley, 2021). The wind product is from ECMWF Reanalysis version 5 Hersbach H. et al. (2023).

D Result

The CMO-NRT dataset, formatted in Network Common Data Form (NetCDF), provides monthly global oceanic surface fCO_2 and air-sea CO_2 flux data. This dataset spans from January 2022 through July 2023 and is structured into a 1° x 1° grid. It includes three dimensions and two variables. The dimensions are as follows:

- **Time:** Monthly data, from January 2022 to July 2023.
- **Latitude (lat):** Ranges from -90° to 90° North.
- **Longitude (lon):** Spans from -180° to 180° East.
- **Product:** Involves 10 GOBMs and 8 data products.

The two variables covered are:

- $sfCO_2$: This represents the surface ocean fCO_2, quantified in units of uatm. It is measured across 18 products, 19 time points, and a grid of 180 latitudes by 360 longitudes.
- $fgCO_2$: This indicates the flux density of total air-sea CO_2 exchange, expressed in gC/m²/day with a positive value indicating an upward direction. Similar to $sfCO_2$, it is measured across the same dimensions. As of the time of this writing, the dataset includes data up to July 2023.

Figure 2 displays the monthly oceanic fCO_2 and air-sea CO_2 flux from 1990 through July 2023, alongside their corresponding gridded data for the period of August 2022 to July 2023 as estimated by CMO-NRT.

E Technical validation

- Overall, the predictions for both GOBMs and data products demonstrated a strong performance, with most having an R2 value above 0.9. Scatter plot comparisons reveal that, in data products, there are occasional points where predictions significantly deviate from the fit line. In contrast, for GOBMs, most points cluster close to the fit line, indicating more stable performance of the model with GOBMs.
- Global quantity: We further analyzed the monthly variations of global fCO2, comparing CMO-NRT predictions with original outputs during 2020-2021 as shown in Figures 4 and 5. The results exhibit a high degree of agreement in terms of seasonality across both GOBMs and data products. Generally, our estimated values were slightly higher than the original values, with most differences being less than 3 µatm. In terms of global aggregate comparisons, GOBMs demonstrated superior performance, with most R2 values exceeding 0.85.

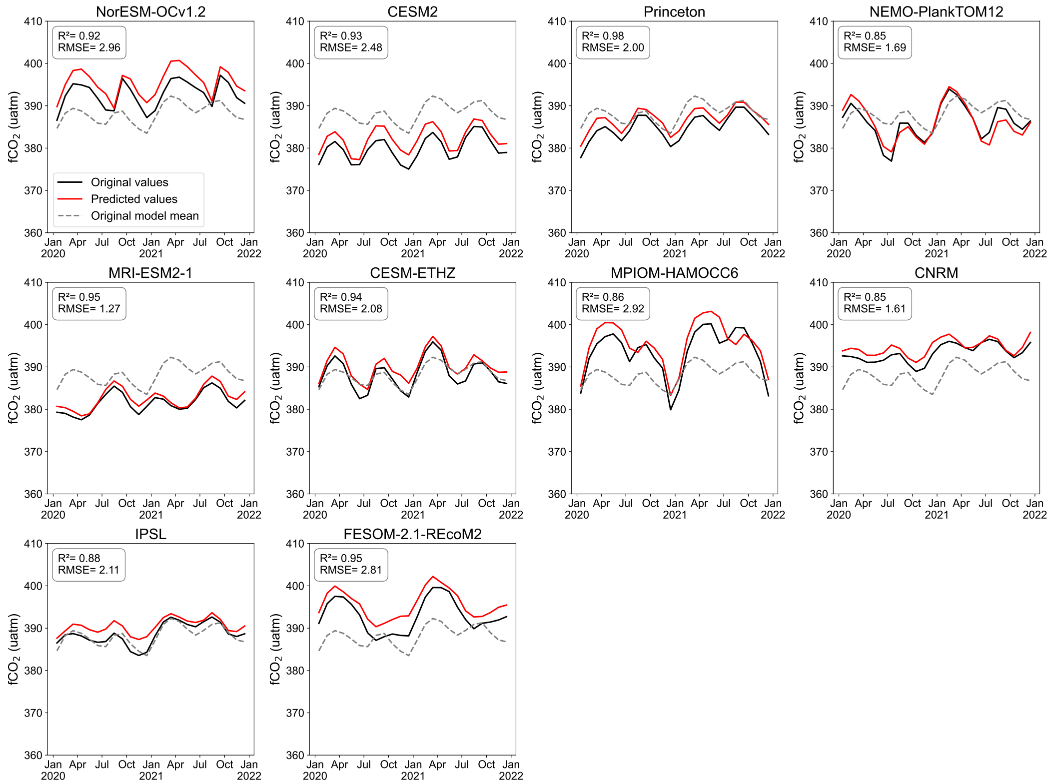

Figure 4: Comparison of global average monthly fCO2 between CMO-NRT predictions and original outputs for each of the 10 GOBMs during 2020-2021. The red lines are the predicted results from CMO-NRT. The black lines are the original results from GOBMs. The dashed grey lines are the average results of 10 GOBMs.

- Spatial distribution:We evaluate the spatial patterns of mean fCO2 of the CMO-NRT prediction against the original output for each of the 10 GOBMs and 8 data products during 2020-2021 Figures 6 and 7 . The spatial variations in fCO2 were largely consistent between the CMO-NRT predictions and the original results, with most discrepancies being under 20 µatm. The predicted values also align well with the original outputs in both trend and magnitude across latitudes. Notably larger errors were observed in the equatorial Pacific, including adjacent coastal areas near Peru and Chile to the west and northeast regions near Indonesia and Papua New Guinea, as well as the Arctic Ocean. These errors are primarily due to these regions typically exhibiting extreme maximum and minimum fCO2 values. Our model tends to perform less accurately in predicting extremes compared to values closer to the mean, resulting in slightly poorer performance in these areas. Additionally, the MPI-SOMFFN shows lower predictive accuracy in the Arctic region, primarily due to the high frequency of missing historical data in this area.

  We also calculated the monthly mean fCO2 values by averaging them separately across the 10 GOBMs and 8 data products during 2020-2021. The spatial variation in fCO2 showed high consistency between the CMO-NRT predictions and the original results, with discrepancies generally under 10 µatm for GOBMs and data products. The predicted values align well with the original outputs in both trend and magnitude across latitudes. Examining different months, we observed that in the summer months, the original values from GOBMs in the Arctic region were significantly higher than our predictions. In contrast, the performance of data products was more uniform across different months. Typically, the original values from data products were lower than the predictions in the Arctic Ocean and higher in the equatorial Pacific region.

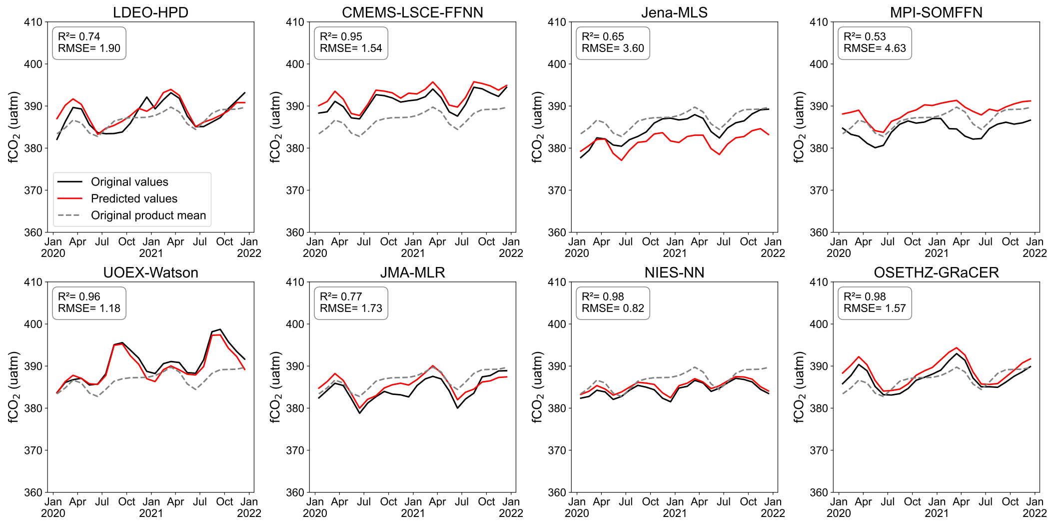

Figure 5: Comparison of global average monthly fCO2 between CMO-NRT predictions and original outputs for each of the 8 data products during 2020-2021. The red lines are the predicted results from CMO-NRT. The black lines are the original results from data products. The dashed grey lines are the average results of 8 data products.

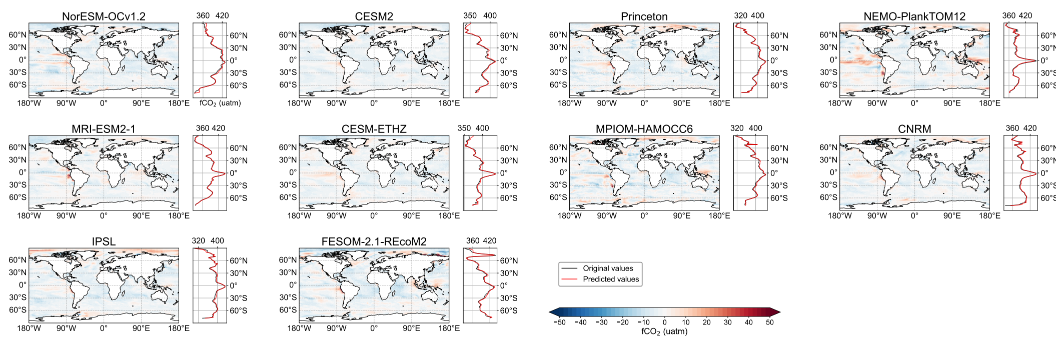

Figure 6: Spatial differences between CMO-NRT predictions and original outputs for each of the 10 GOBMs during 2020-2021. The line graphs represent the values of predictions and original outputs across latitudes.

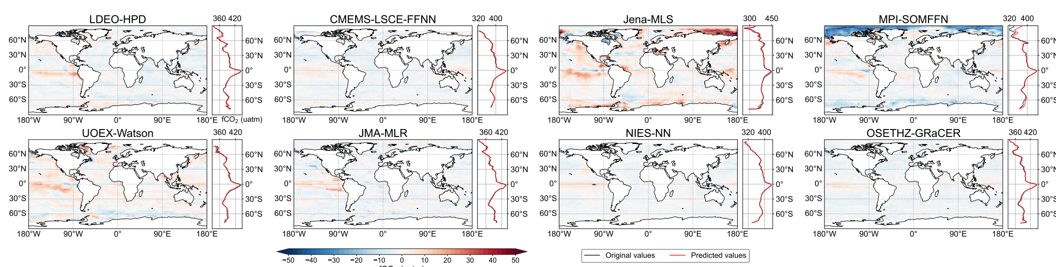

Figure 7: Spatial differences between CMO-NRT predictions and original outputs for each of the 8 data products during 2020-2021. The line graphs represent the values of predictions and original outputs across latitudes.

Table 3: Root Mean Square Error (RMSE) for Different Models

GOBM Models			
NAME	RMSE	MEAN	NRMSE
NorESM-OCv1.2	10.8	392	2.7%
CESM2	6.8	379	1.7%
MRI-ESM2-1	7.7	381	2.0%
CNRM	11.2	393	2.8%
IPSL	10.6	388	2.7%
MPIOM-HAMOCC6	15.6	393	3.9%
NEMO-PlankTOM12	13.3	386	3.4%
FESOM-2.1-REcoM2	15.1	392	3.8%
CESM-ETHZ	8.7	388	2.2%
Princeton	10.6	384	2.7%
Data Product Models			
NAME	RMSE	MEAN	NRMSE
NIES-NN	5.3	384	1.3%
OSETHZ-GRaCER	6.7	387	1.7%
LDEO-HPD	8.4	387	2.1%
MPI-SOMFFN	16.4	384	4.2%
Jena-MLS	15.1	384	3.9%
UOEX-Watson	16.2	390	4.1%
JMA-MLR	9.9	384	2.5%
CMEMS-LSCE-FFNN	8.4	391	2.1%