https://doi.org/10.5194/essd-2026-375
Preprint. Discussion started: 11 June 2026

Open Access Earth System Science Data Discussions

FDU-BTR: a physics-guided ensemble learning reconstruction of global surface-ocean pCO_2 (1982–2024) with uncertainty diagnostics

Zhenguo Wang[1], Weiwei Fu[1,2]

[1]Department of Atmospheric and Oceanic Sciences, Fudan University, Shanghai, 200438, China

[2]Institute of Eco-Chongming (IEC), 1050 Baozhen, Lühua Town, Chongming District, Shanghai 202151, China

*Correspondence to: Weiwei Fu (*wwfu@fudan.edu.cn*)*

Abstract. The ocean takes up roughly 25% of anthropogenic CO_2 emissions, yet quantifying the magnitude and variability of this sink is limited by the uneven, sparse sampling of surface-ocean partial pressure of CO_2 (pCO_2). Here we present FDU-BTR, a global monthly 1° × 1° reconstruction of surface-ocean pCO_2 for 1982–2024, produced with a background–thermal residual (BTR) ensemble learning framework that embeds first-order physical structure in a machine-learning workflow (Wang and Fu, 2026, https://doi.org/10.5281/zenodo.20152530). Observed pCO_2 is decomposed into a multi-product background climatology, an explicit thermal-anomaly term, and a residual field; region-specific CatBoost ensembles then reconstruct the residual, with boundary blending ensuring spatial continuity. This decomposition simplifies the learning target while preserving physically meaningful constraints. Validated against the independent Hawaii Ocean Time-series (HOT) and Bermuda Atlantic Time-series Study (BATS) observations, FDU-BTR achieves a correlation of 0.93 and a root-mean-square error of 8.34 µatm, comparable to leading products, with a mean total uncertainty of 12.90 µatm. Cross-product comparisons and coverage–entropy diagnostics localize structural disagreement to coastal, marginal, and high-latitude regions where observations are sparse and processes are complex. Controlled thinning experiments further reveal a strong asymmetry in the observational error budget: reducing spatial coverage degrades reconstruction skill approximately twice as much as equivalent reductions in temporal coverage. FDU-BTR therefore provides a physically constrained, uncertainty-quantified pCO_2 product for air–sea CO_2 flux assessment and identifies spatial observational sparsity — not algorithm choice — as the dominant remaining limit on reconstructing the global ocean carbon sink, with direct implications for the design of future ocean carbon observing systems.

Keywords: surface-ocean pCO_2; machine learning; residual learning; observational coverage; uncertainty

1 Introduction

The ocean is a dominant dynamic carbon reservoir that has absorbed approximately 25% of anthropogenic CO_2 emissions, substantially slowing atmospheric CO_2 growth and moderating climate change (Sabine et al., 2004; Gruber et al., 2019; Friedlingstein et al., 2025). However, this sink is neither spatially uniform nor temporally steady. It responds to rising

https://doi.org/10.5194/essd-2026-375
Preprint. Discussion started: 11 June 2026

atmospheric CO_2, sea surface temperature variability, circulation-driven transport, vertical mixing, the biological pump, and carbonate buffering, producing pronounced variability on seasonal to decadal timescales (Takahashi et al., 2002; Landschützer et al., 2016; Crisp et al., 2022). Because the sea surface partial pressure of carbon dioxide (pCO_2) directly sets the air–sea pCO_2 gradient, it is the central variable for computing air–sea CO_2 fluxes and tracking changes in the ocean carbon sink. Therefore, a continuous, long-term, regularly gridded, and physically interpretable global surface-ocean pCO_2 product is fundamental for understanding sink evolution and supporting global carbon budget assessments (Peters et al., 2017; Chau et al., 2024; Wanninkhof et al., 2025).

Current global surface-ocean pCO_2 reconstructions rely mainly on high-quality in situ observations, primarily from SOCAT, together with multi-source environmental predictors. SOCAT compiles millions of ship-based and underway observations and provides the core observational constraint for observation-based products (Bakker et al., 2016). However, these observations remain highly uneven in both time and space. Coverage is denser along repeat cruise tracks, in Northern Hemisphere mid-latitude regions, and in recent decades, whereas the Southern Ocean, high-latitude winters, the South Pacific, the South Atlantic, marginal seas, and some coastal regions remain persistently undersampled (Gregor et al., 2019; Chau et al., 2022; Heimdal et al., 2024). Therefore, constructing global pCO_2 products is fundamentally a gap-filling problem: models must fit pCO_2 where observations exist and extrapolate to regions and periods with few or no data.

To address this problem, several classes of reconstruction methods have been developed (Jiang et al., 2026). Empirical regression, mixed-layer approaches, and statistical interpolation methods first established relationships between pCO_2 and variables such as sea surface temperature, salinity, chlorophyll, mixed-layer depth, wind speed, and atmospheric CO_2, providing the first globally continuous reconstructions (Rödenbeck et al., 2014; Iida et al., 2021). Two-step approaches using self-organizing maps (SOMs) and feed-forward neural networks, such as MPI-SOMFFN, were subsequently widely adopted. These methods classify the ocean into biogeochemical regimes and then build nonlinear regressions within each cluster, improving monthly pCO_2 fields (Zeng et al., 2014; Landschützer et al., 2014; Denvil-Sommer et al., 2019). Ensemble techniques further reduce single-model instability. Multi-member products such as CSIR-ML6 and OceanSODA-ETHZ have shown that combining diverse machine-learning members enhances robustness and provides a basis for assessing inter-product uncertainty (Gregor et al., 2019; Gregor and Gruber, 2021). More recent studies emphasize physical priors and complex spatiotemporal structures. For example, the LDEO-HPD product uses a global ocean biogeochemical model as a prior and corrects model–observation biases with machine learning (Gloege et al., 2022). The pCO_2-residual approach demonstrates that explicitly removing the temperature signal before training reduces target complexity and improves the physical interpretability of the model (Bennington et al., 2022a). Random forests, extremely randomized trees, optimized ensemble models, and Vision Transformers have also been used to capture more complex nonlinearity and spatiotemporal dependencies (Chen et al., 2025; Wanninkhof et al., 2025; Zhang et al., 2025).

These studies have considerably improved global pCO_2 products, but three critical issues persist. First, many products can now

https://doi.org/10.5194/essd-2026-375
Preprint. Discussion started: 11 June 2026

represent the mean climatological state and seasonal cycle reasonably well, but they diverge substantially in their representation of anomalies, interannual variability, and extrapolation behavior in data-sparse regions. Multi-product comparisons such as SOCOM reveal large spreads over the Southern Ocean and other under-observed areas (Rödenbeck et al., 2015; Gregor et al., 2019). Second, a single global model struggles to simultaneously capture the distinct process regimes of tropical upwelling zones, high-latitude mixing regions, marginal seas, and the open ocean. Regionalized modeling can mitigate this heterogeneity, but discrete regional divisions may introduce non-physical discontinuities at boundaries and degrade the spatial continuity of trends and variability (Wang et al., 2026). Third, product evaluation commonly relies on root-mean-square error and bias computed at observation locations. Because validation data are concentrated in relatively well-observed regions, these metrics tend to underestimate uncertainty in undersampled areas and cannot readily separate errors arising from model structure, insufficient temporal coverage, or limited spatial coverage.

Advancing global pCO_2 products requires more than complex machine-learning algorithms. It also demands improved physical interpretability of the target variable, better representation of regional processes, spatially continuous reconstruction fields, and explicit diagnosis of the constraints imposed by observational coverage. Important open questions include how to reduce the burden of large-scale background and thermal signals in direct pCO_2 fitting, how to represent basin-scale differences while avoiding discontinuities at regional boundaries, how to characterize the dominant sources of product uncertainty, and what additional observations would most improve the robustness of historical reconstructions.

To address these issues, we propose a Background–Thermal Residual (BTR) ensemble learning framework and generate a new global monthly surface-ocean pCO_2 product at 1° × 1° resolution for 1982–2024. Instead of learning raw pCO_2, BTR decomposes observed pCO_2 into three components: a multi-product background climatology, an explicit thermal-anomaly term, and a residual term. A regionalized CatBoost ensemble model then learns the residual field. This design retains the physical constraints of the background state and thermal response while allowing the ensemble to focus on non-thermal residual signals associated with upwelling, mixing, horizontal transport, biological processes, freshwater input, and air–sea disequilibrium. Boundary blending between adjacent regional models is further introduced to suppress artificial discontinuities. We also construct an uncertainty characterization framework that combines measurement, sampling, and mapping errors, and we validate the product using independent time-series records from HOT and BATS. In addition, we diagnose the influence of observational coverage on reconstruction uncertainty using multi-product climatological dispersion and time–space coverage entropy. Targeted thinning experiments are further designed to quantify the relative effects of temporal and spatial coverage loss on out-of-sample reconstruction skill.

Overall, this study makes three principal contributions: it delivers a long-term global monthly pCO_2 product, characterizes the spatial patterns and dominant sources of product uncertainty, and quantifies how the structure of observational coverage constrains extrapolation skill. By integrating product construction, methodological validation, and coverage diagnostics, this work provides a foundation for air–sea CO_2 flux estimation, observing-system optimization, and the next generation of global

https://doi.org/10.5194/essd-2026-375
Preprint. Discussion started: 11 June 2026

Open Access Earth System Science Data Discussions

pCO_2 reconstructions.

2 Data

2.1 Training data

This study uses monthly gridded global surface ocean carbon dioxide fugacity (fCO_2) data from SOCATv2025 at a spatial resolution of 1° × 1° to calculate surface ocean pCO_2 (Bakker et al., 2016). This dataset integrates global shipboard observations from 1982 to 2024, with gridded fCO_2 values generated by cruise-weighted averaging. Sea surface pCO_2 is calculated as follows:

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

where P_{atm} is the sea-level atmospheric pressure from ERA5, T is the sea surface temperature from NOAA Optimum Interpolation Sea Surface Temperature (OISST v02r01) (Reynolds et al., 2007), B and δ are the virial coefficients given by Weiss (1974), and R is the gas constant (Dickson et al., 2007). The dataset retains the in situ intake-temperature information from individual cruises, which provides the basis for subsequent temperature normalization and correction.

The data were subjected to a two-step quality-control procedure. First, extreme values were removed, including records with pCO_2 greater than 700 µatm and observations with sea surface temperature below −2 °C. To reduce systematic biases caused by differences between in situ intake temperature and gridded sea surface temperature, a temperature correction was applied to the SOCATv2025 observations. Specifically, the original pCO_2 values were corrected as follows:

$$pCO_2^* = pCO_2 \cdot \exp[0.0423 \cdot (SST - T_{\text{obs}})] \quad (2)$$

to obtain pCO_2^*, where SST is the gridded sea surface temperature from OISST v02r01, T_{obs} is the in situ intake temperature measured during each cruise, and 0.0423 K^{-1} is the temperature sensitivity coefficient (Takahashi et al., 1993). To ensure data quality, records were removed if the absolute difference after correction, $|pCO_2^* - pCO_2|$, exceeded 40 µatm. Only observations from 1982 to 2024 were retained.

In addition, this study uses a set of environmental and physical–biogeochemical predictors, which were aligned with the SOCAT data and interpolated onto a 1° grid. Sea surface temperature (SST) was obtained from NOAA OISST v02r01 (Reynolds et al., 2007). Sea surface salinity (SSS) was taken from EN4.2.2 (Good et al., 2013). Chlorophyll-a concentration (Chl-a) was derived from the Copernicus GlobColour monthly 4 km composite product and was $\log_{10}$-transformed (OCTAC, 2023). Mixed-layer depth (MLD) was obtained from CLS data. Atmospheric CO_2 mole fraction (xCO_2) was taken from the NOAA ESRL MBL product and mapped onto the 1° × 1° grid through latitudinal interpolation and longitudinal replication (Lan et al., 2024). Surface wind components (U10/V10) and sea-ice concentration (Siconc) were obtained from ERA5 (Hersbach, 2023), while sea-level anomaly (SLA) was obtained from the Copernicus Climate Data Store (Copernicus Climate

https://doi.org/10.5194/essd-2026-375
Preprint. Discussion started: 11 June 2026

Change Service, 2018). The tropical Pacific ENSO index, represented by the NINO3.4 anomaly, was used to provide information on remote climate coupling. pH and dissolved inorganic carbon data were obtained from the CMEMS Surface Ocean Carbon Fields product (Chau et al., 2022; Chau et al., 2024). Detailed information is provided in Table 1.

Table 1. Observational and environmental variables

Variable name	Abbreviation (Unit)	Period	Temporal resolution	Spatial resolution	Source	Reference
Surface-ocean fugacity of CO_2	fCO_2 (µatm)	1982/01–2024/12	Monthly	1° × 1°	SOCAT v2025	(Bakker et al., 2016)
Atmospheric CO_2 mole fraction	xCO_2 (ppm)	1982/01–2024/12	Monthly	Meridional bands	NOAA ESRL Marine Boundary Layer	(Lan et al., 2024)
Sea surface temperature	SST (°C)	1982/01–2024/12	Daily	0.25° × 0.25°	NOAA OISST v02r01	(Reynolds et al., 2007)
Sea surface salinity	SSS (psu)	1982/01–2024/12	Monthly	0.25° × 0.25°	EN4.2.2	(Good et al., 2013)
Chlorophyll-a concentration	Chl-a (mg m^{-3})	1997/09–2024/12	Monthly	4 km	Copernicus GlobColour	(OCTAC, 2023)
Sea-level anomaly	SLA (m)	1993/01–2023/12	Monthly	0.25° × 0.25°	Copernicus Climate Data Store	(Copernicus Climate Change Service, 2018)
Zonal wind speed at 10 m	U10 (m s^{-1})	1982/01–2024/12	Monthly	0.25° × 0.25°	ERA5	(Hersbach, 2023)
Meridional wind speed at 10 m	V10 (m s^{-1})	1982/01–2024/12	Monthly	0.25° × 0.25°	ERA5	(Hersbach, 2023)
Sea-ice concentration	Siconc (dimensionless)	1982/01–2024/12	Monthly	0.25° × 0.25°	ERA5	(Hersbach, 2023)
Mixed layer depth	MLD (m)	1993/01–2022/12	Monthly	0.125° × 0.125°	CLS	(Guinehut et al., 2012)
Surface-ocean pH	pH (dimensionless)	1985/01–2024/12	Monthly	0.25° × 0.25°	LSCE	(Chau et al., 2022; Chau et al., 2024)

https://doi.org/10.5194/essd-2026-375
Preprint. Discussion started: 11 June 2026

Dissolved inorganic carbon	DIC (μmol kg⁻¹)	1985/01–2024/12	Monthly	0.25° × 0.25°	LSCE	(Chau et al., 2022; Chau et al., 2024)
Niño 3.4 anom (ERSSTv5)	NINO3.4_anom (°C)	1982/01–2024/12	Monthly	Regional index	NOAA PSL/CPC	(Huang et al., 2017)

2.2 Multi-product climatology

To separate the broad, large-scale background structure of sea surface pCO_2 from the local anomalies superimposed on it, we first constructed a monthly climatological prior field constrained by multiple products. Estimating the background state directly from sparse in situ observations is susceptible to strong sampling bias, whereas relying on any single product risks introducing method-dependent systematic biases. By merging information from multiple global surface-ocean carbon products, this climatology is designed to retain robust large-scale features that are consistently supported across products, suppress the influence of individual product errors, and serve as the stable background term within the residual learning framework.

We compiled 11 global monthly surface-ocean carbon products that provide either sea surface pCO_2 or fCO_2 variables (Table 2). To build a temporally homogeneous background field, we selected 6 products that provide complete monthly coverage over 1990–2020: NIES-ML3, LDEO fCO_2-Residual, LDEO-HPD fCO_2-Extended, CMEMS-LSCEv2, OceanSODA-ETHZv2, and SJTU-AViT. The remaining five products, affected by incomplete temporal coverage or pronounced spatial discontinuities, were used only for independent intercomparison and evaluation to ensure a more robust background field.

Because the products differ in their definition of carbon variables, all product variables were first standardized before use. Products released as sea surface fCO_2 were converted to pCO_2 according to Eq. (1). After variable standardization, all products used for background-field construction were remapped onto a global 1° × 1° monthly grid. For products with temporal resolutions finer than monthly, monthly means were first calculated before spatial regridding.

Next, for each product, grid cell, and calendar month, we estimated a linear trend over 1990–2020 and adjusted every monthly value to the reference year 2005, yielding a detrended monthly climatology. Finally, at each grid cell and calendar month, the detrended climatologies from the 6 products were combined by taking the multi-product median. This median field serves as the monthly climatological prior for sea surface pCO_2, providing a background field that is substantially more robust against systematic biases and outliers than any single-product estimate.

Table 2. Global monthly-scale carbon products

Product	Variable	Period	Resolution	Reference
NIES-ML3	fCO_2	1982/01–2024/12	1° × 1°, monthly	(Zeng et al., 2022)
LDEO fCO_2-Residual	fCO_2	1982/01–2023/12	1° × 1°, monthly	(Bennington et al., 2022a)

https://doi.org/10.5194/essd-2026-375
Preprint. Discussion started: 11 June 2026

LDEO-HPD fCO_2-Extended	fCO_2	1959/01–2023/12	1° × 1°, monthly	(Bennington et al., 2022b)
CMEMS-LSCEv2	pCO_2	1985/01–2024/12	0.25° × 0.25°, monthly	(Chau et al., 2024)
OceanSODA-ETHZv2	fCO_2	1982/01–2022/12	25 km, 8-day	(Gregor et al., 2024)
SJTU-AViT	pCO_2	1982/01–2023/12	1° × 1°, monthly	(Zhang et al., 2025)
CSIR-ML6	pCO_2	1982/01–2016/12	1° × 1°, monthly	(Gregor et al., 2019)
JMA-MLR	pCO_2	1990/01–2023/12	1° × 1°, monthly	(Iida et al., 2021)
Stepwise-FFNN	pCO_2	1992/01–2024/12	1° × 1°, monthly	(Zhong et al., 2022)
AOML-ET	pCO_2	1997/09–2023/12	1° × 1°, monthly	(Wanninkhof et al., 2025)
TSSCXG-17	pCO_2	1993/01–2020/12	1° × 1°, monthly	(Chen et al., 2025)

3 Methods

3.1 Background–thermal residual framework

This study proposes a Background–Thermal Residual (BTR) framework as a prior-guided approach for surface-ocean pCO_2 reconstruction. Instead of directly regressing raw sea surface pCO_2, the framework first decomposes observed pCO_2 into a background term, a thermal term, and a remaining residual term, r_T^*, which is then used as the training target. The core objective of the BTR framework is to reduce the target complexity faced by machine learning by using physically interpretable prior information to account for the large-scale climatological structure and the first-order thermal response before model training. The workflow is shown in Fig. 1. Related studies have also improved pCO_2 reconstruction by estimating the climatological state from model data and then training models on observation–model residuals (Bennington et al., 2022b). Raw pCO_2 contains the background climatological state, temperature-driven variability, and non-thermal anomalies. By separating these components in advance, the BTR framework allows the model to focus on residual signals that are not explained by the dynamic prior.

https://doi.org/10.5194/essd-2026-375
Preprint. Discussion started: 11 June 2026

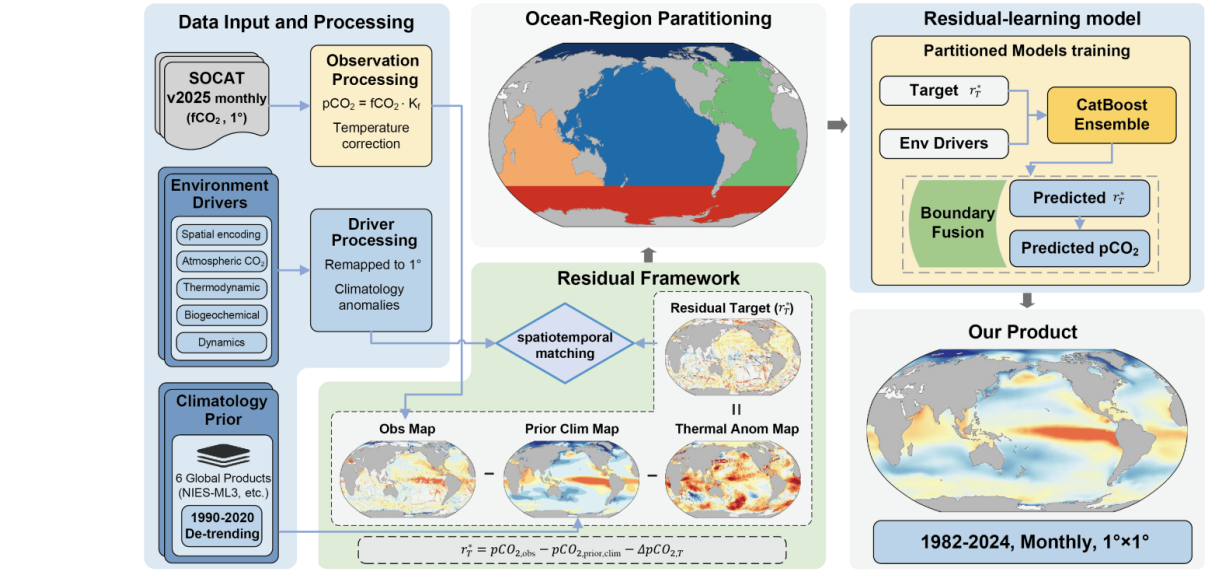

Figure 1. Overall workflow of the Background–Thermal Residual (BTR) framework.

The first component of the BTR framework is the background term, $pCO_{2,\mathrm{prior,clim}}$, which uses the harmonized multi-product climatology as a low-frequency prior background state. For each sample, the corresponding background value is extracted according to month, latitude, and longitude. The multi-product climatology is used only as a detrended prior background field; it provides a robust estimate of the mean spatial–seasonal pCO_2 structure, rather than contemporaneous monthly pCO_2 information, and therefore does not introduce leakage of sample-specific target variability. This term represents the background pCO_2 level at the sample location under the monthly mean state and is used to characterize the large-scale spatial structure and seasonal cycle. The second component is the temperature-driven term, $\Delta pCO_{2,T}$. Sea surface pCO_2 has a rapid and well-defined thermodynamic response to temperature changes; therefore, this component is explicitly represented in this study. The deviation of the sample sea surface temperature from the corresponding monthly climatology is denoted as δSST. The temperature-driven increment is defined as

$$\Delta pCO_{2,T} = pCO_{2,\mathrm{prior,clim}}[\exp(0.0423 \cdot \delta SST) - 1]. \quad (3)$$

Here, 0.0423 K^{-1} is the first-order empirical sensitivity of sea surface pCO_2 to temperature changes. This term represents the thermodynamic change in pCO_2 caused solely by temperature anomalies when other processes remain unchanged. Accordingly, the residual term r_T^* is defined as

$$r_T^* = pCO_{2,\mathrm{obs}} - pCO_{2,\mathrm{prior,clim}} - \Delta pCO_{2,T}, \quad (4)$$

where $pCO_{2,\mathrm{obs}}$ denotes the observed sea surface pCO_2. Thus, r_T^* represents the remaining component after the background and thermal terms are removed. This component mainly corresponds to variability that is not explained by the dynamic prior and is generally associated with non-thermal processes, such as upwelling, mixing, horizontal transport, biological production and respiration, freshwater input, and air–sea disequilibrium exchange.

https://doi.org/10.5194/essd-2026-375
Preprint. Discussion started: 11 June 2026

Open Access Earth System Science Data Discussions

3.2 Residual-learning model training

Our core task is to establish a stable mapping from multi-source environmental predictors to the residual term defined as r_T^*. To this end, we first selected a regression algorithm that can handle nonlinear relationships and missing values, and then constructed an ensemble training framework. We then applied sample weighting to reduce the influence of uneven spatiotemporal observational distributions during training. Finally, because regional processes differ substantially, we adopted a regionalized training strategy and constructed a corresponding feature set for each region.

3.2.1 CatBoost regression algorithm

This study uses a CatBoost regression model to fit r_T^*. CatBoost is a gradient-boosted decision tree machine-learning method that can represent complex nonlinear relationships and interactions among variables, is insensitive to feature scaling, and can directly handle missing values during training (Prokhorenkova et al., 2018). Therefore, this method is well suited for residual learning of sea surface pCO_2 constrained by multiple environmental variables.

The target variable for model training is r_T^*, as defined in the BTR framework. The model inputs consist of environmental predictors, and the output is the residual value for each corresponding sample. Because some environmental features are missing during the early part of the study period, their missing status is retained without additional interpolation or imputation and is handled automatically by CatBoost during training.

To improve the robustness of model parameter settings, this study uses Optuna for hyperparameter optimization. The optimization is conducted only within the training set and does not use the independent test data, thereby avoiding information leakage from the test set. The searched parameters include learning rate, tree depth, leaf-node L_2 regularization, bagging temperature, and random strength. The optimization objective is the validation root-mean-square error (RMSE) under cross-validation, and the search ranges for each parameter are listed in Table 3.

Table 3. CatBoost hyperparameters and Optuna search priors.

Hyperparameter	Explanation	Search range
iterations	Maximum boosting rounds	500–2000
learning_rate	Learning rate (shrinkage)	0.02–0.08
depth	Tree depth	4–10
l2_leaf_reg	L2 regularization on leaf values	5–20
bagging_temperature	Temperature for Bayesian bootstrap	0.1–0.8
random_strength	Randomness in split scoring	0.1–2.0

Cross-validation is performed using the decade-block K-fold cross-validation. Specifically, samples are grouped by decade,

https://doi.org/10.5194/essd-2026-375
Preprint. Discussion started: 11 June 2026

and five-fold cross-validation is constructed at the decadal scale. This design avoids placing samples from adjacent years simultaneously in the training and validation sets, as can occur under random splitting, thereby reducing overly optimistic validation performance caused by temporal autocorrelation. For long-term reconstruction tasks, this temporally blocked validation strategy better reflects the model's generalization ability to unseen periods. In each training fold, early stopping is applied according to the validation RMSE, and the corresponding optimal number of iterations is recorded.

After obtaining the optimal hyperparameters, this study further constructs an ensemble model to reduce the sensitivity of a single training run to sample perturbations and random initialization. Specifically, 20 CatBoost member models are trained. Each member uses a different random seed and resamples the training samples with bootstrap rows enabled. The bootstrap sampling fraction is set to 1.0, meaning that each member model is trained on a dataset of the same size but with a slightly different sample composition. The final prediction is obtained by ensemble averaging. For any sample, the mean prediction from the 20 member models is taken as the best estimate of r_T^*, while the spread among member predictions is used for subsequent uncertainty characterization. To avoid fluctuations in the number of iterations caused by single-fold training, this study uses the median of the optimal iteration numbers across folds as the final number of boosting iterations and retrains each ensemble member on the full training set. This approach improves training consistency and stability while controlling model complexity.

To evaluate out-of-sample generalization, this study adopts a fixed-year validation design. Six SOCAT observational years—1984, 1988, 1995, 2004, 2013, and 2021—were randomly sampled and designated as independent test years prior to model development. All remaining years were used for model training, whereas the held-out test years were excluded from model fitting, feature selection, and hyperparameter optimization.

3.2.2 Feature engineering

The model inputs consist of four categories of features: spatial location encoding, ocean physical state variables, biogeochemical variables, and external forcing indicators. Because the training target r_T^* explicitly removes the background climatological state and the thermal contribution of temperature, these predictors are mainly used to explain the remaining non-background and non-thermal variability in the residual. Spatial information was represented using a sine-transformed latitude feature, lat_sin, to account for broad meridional structure in surface-ocean pCO_2:

$$lat\_sin = \sin\phi \tag{5}$$

where ϕ denotes latitude. Compared with the direct use of latitude, this transformation provides a bounded and continuous representation of meridional position and is suitable for tree-based models to learn broad latitudinal gradients.

SST_anom and SSS_anom are defined as the differences between SST and SSS and their corresponding monthly climatologies, respectively, and are used to characterize deviations in temperature and salinity from the climatological background. Chl-a is $\log_{10}$-transformed to reduce the influence of its right-skewed distribution on model training. pH_anom

https://doi.org/10.5194/essd-2026-375
Preprint. Discussion started: 11 June 2026

and DIC_anom denote the anomalies of pH and DIC relative to their monthly climatologies, respectively, and are used to characterize deviations in carbonate-system conditions. Surface-ocean pH was included because recent studies have identified it as an important predictor for machine-learning-based pCO_2 reconstruction (Wu et al., 2026).

The candidate predictors entering the feature-selection procedure include lat_sin, SST, SST_anom, SSS, SSS_anom, SLA, MLD, U10, V10, Siconc, $\log_{10}$(Chl-a), atmospheric CO_2 mole fraction xCO_2, pH, pH_anom, DIC, DIC_anom and $NINO3.4\_anom$. These variables constrain spatial location, local oceanic conditions, biogeochemical states, and air–sea forcing, and together constitute the candidate feature set for residual learning.

3.2.3 Sample weighting

SOCAT observations are unevenly distributed in space and time. Samples are more abundant in recent years and along densely sampled cruise tracks, whereas early periods and sparsely sampled ocean regions contain fewer observations. If the raw samples are used directly for training, the model may be dominated by high-density regions, thereby reducing its ability to represent undersampled regions. To mitigate this bias, this study introduces inverse-density weighting during model training. Specifically, the training samples are grouped into 5° × 5° spatial grid cells and non-overlapping 10-year temporal windows, and the number of samples $n_{b(i)}$ in each spatiotemporal bin is counted. For sample i, the initial weight is defined as

$$\widetilde{w}_i = \frac{1}{\sqrt{n_{b(i)}}}, \tag{6}$$

where $b(i)$ denotes the spatiotemporal bin containing sample i. This definition increases the influence of samples from sparsely sampled regions and periods while avoiding excessive weight amplification. The weights are then normalized to have a unit mean:

$$w_i = \frac{\widetilde{w}_i}{\frac{1}{N}\sum_{k=1}^{N}\widetilde{w}_k}, \tag{7}$$

where N is the total number of training samples within the current modeling domain.

The normalized weights were explicitly incorporated into the subsequent training workflow. First, during feature selection, permutation importance was calculated using weighted training and validation errors, and cross-validation recursive feature elimination also incorporated these weights where supported by the algorithm. Second, during hyperparameter optimization, a weighted objective function was used in each training fold, so that the optimal parameters based on validation RMSE were searched after correcting for the spatiotemporal sampling distribution. Finally, during the training of ensemble member models, the sample weights continued to contribute to the loss-function calculation. Together, these steps reduced the dominance of densely sampled regions and periods in parameter estimation and improved the model's ability to represent undersampled regions and periods.

3.2.4 Regionalization strategy

The dominant controls on sea surface pCO_2 vary substantially among ocean basins. In tropical regions, pCO_2 variability is

https://doi.org/10.5194/essd-2026-375
Preprint. Discussion started: 11 June 2026

mainly regulated by upwelling and ENSO events, whereas high-latitude regions are strongly modulated by convective cooling, deep vertical mixing, sea ice, and intense seasonal biological production. In addition, the background carbonate system and large-scale circulation structure differ systematically among the major basins. If a single global model is trained on all data simultaneously, relationships learned from data-rich regions tend to dominate the optimization, weakening the model's ability to generalize to data-sparse regions or to regions governed by fundamentally different mechanisms. For this reason, we adopt a regionally independent modeling strategy.

Many previous studies have relied on complex biogeochemical provinces with highly irregular boundaries. However, such province-based partitioning can introduce pronounced step-like artifacts between adjacent regions after modeling (Iida et al., 2021). In this study, the horizontal partitioning follows the basin definitions of the World Ocean Atlas 2023 (WOA23) (Garcia et al., 2024), with the Southern Ocean treated as an independent region (Fig. 1). The global open ocean is ultimately divided into five regions: the Atlantic Ocean (ATL), Pacific Ocean (PAC), Indian Ocean (IND), Southern Ocean (SO), and Arctic Ocean (ARC). This scheme preserves stable and well-defined basin boundaries, avoids the boundary complexity introduced by fine-scale provinces, and helps reduce the boundary discontinuities that often accompany regionalized modeling. Semi-enclosed seas, such as the Mediterranean Sea and the Red Sea, are excluded from this partitioning.

Within each basin defined above, a dedicated residual-learning model was trained independently. Because both the physical control relationships and observational coverage differ among basins, we do not impose a uniform set of predictors across all regions. Instead, to enhance model interpretability and generalization ability, a two-stage feature-selection procedure was implemented within each basin using only the training data and adopting a decadal-scale cross-validation design consistent with the model evaluation. First, permutation importance was calculated using decade-grouped five-fold cross-validation, and an initial feature subset was selected according to the adaptive rule $K = \max(10, 2\sqrt{p})$, where p denotes the number of candidate features. Cross-validation recursive feature elimination (RFECV) was then applied to iteratively remove the least important features, and the optimal feature combination for each basin was determined by minimizing the validation RMSE. The resulting basin-specific feature sets were used to independently train the residual field in each ocean basin. This design allows the models to operate within a unified physical framework and training workflow while retaining regional adaptability according to process differences among basins, thereby improving the representation of the global sea surface pCO_2 residual field.

3.3 Product generation with boundary blending

To generate a spatially continuous global surface ocean pCO_2 product, this study introduces boundary blending based on the regional predictions. For each prediction point, the prediction is first obtained from the model corresponding to its primary region. Because the regional models are trained independently, numerical discontinuities may occur near regional boundaries. To reduce such discontinuities, regional blending is applied only within transition zones located within 300 km of regional

https://doi.org/10.5194/essd-2026-375
Preprint. Discussion started: 11 June 2026

boundaries; outside these boundary zones, only the prediction from the primary regional model is retained.

For a prediction point x located within a transition zone, this study calculates the minimum great-circle distance from the point to the spatial footprint of each candidate region. This distance is then used to weight and combine the predictions from the primary and neighboring regional models. The unnormalized weight of the ith candidate regional model is defined as

$$w_i(x) = \exp\left[-\frac{1}{2}\left(\frac{d_i(x)}{\sigma}\right)^2\right] \quad (8)$$

where $d_i(x)$ is the minimum great-circle distance from point x to the spatial footprint of the ith region, and σ is the smoothing-scale parameter, set to 180 km in this study. The normalized weight is defined as

$$\tilde{w}_i(x) = \frac{w_i(x)}{\sum_{j=1}^{K} w_j(x)} \quad (9)$$

where K is the number of regional models involved in the blending, with a maximum of 3. The blended predictive mean is then given by

$$\hat{y}(x) = \sum_{i=1}^{K} \tilde{w}_i(x)\hat{y}_i(x) \quad (10)$$

where $\hat{y}_i(x)$ is the predictive mean from the ith regional ensemble model at point x.

To retain the uncertainty information from the ensemble predictions, this study further calculates the blended predictive variance as

$$\hat{\sigma}^2(x) = \sum_{i=1}^{K} \tilde{w}_i(x)[\hat{\sigma}_i^2(x) + \hat{y}_i^2(x)] - \hat{y}^2(x) \quad (11)$$

where $\hat{\sigma}_i(x)$ is the predictive standard deviation from the ith regional ensemble model at point x. This treatment reduces boundary discontinuities while preserving the predictive spread after cross-region blending.

After obtaining the blended residual prediction $\hat{r}_T^*$, this study adds it to the dynamic prior to recover sea surface pCO_2:

$$\widehat{pCO_2} = pCO_{2,\text{prior,clim}} + \Delta pCO_{2,T} + \hat{r}_T^* \quad (12)$$

where $pCO_{2,\text{prior,clim}}$ denotes the prior climatological state, and $\Delta pCO_{2,T}$ denotes the temperature-anomaly-driven term.

Based on the workflow described above, we release FDU-BTR, a global monthly surface-ocean pCO_2 dataset (Wang and Fu, 2026, https://doi.org/10.5281/zenodo.20152530). The dataset follows the CF-1.8 and ACDD-1.3 metadata conventions and provides 516 monthly fields from January 1982 to December 2024 on a regular 1° × 1° latitude–longitude grid. The grid contains 180 latitude points from 89.5° S to 89.5° N and 360 longitude points from 179.5° W to 179.5° E. Time is stored using a Gregorian calendar as days since 00:00:00 UTC on 1 January 1950, with each monthly field represented by the mid-month timestamp.

The primary variable is pCO_2, provided in µatm as a single-precision array. During product generation, land grid cells were excluded, and a basin-partition mask consistent with the regionalized reconstruction framework was applied to define the final valid ocean domain. Grid cells outside the predefined basin domains were set to missing values in the final product. Missing or invalid grid cells are stored as NaN. No observation-based spatial interpolation or gap filling was applied after the BTR predictions were aggregated onto the monthly 1° grid.

https://doi.org/10.5194/essd-2026-375
Preprint. Discussion started: 11 June 2026

4 Model performance

4.1 BTR model validation

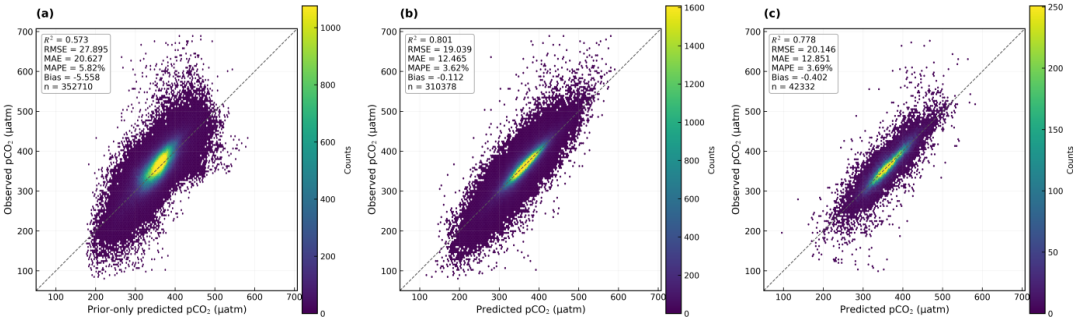

Figure 2. Prediction–observation comparison for the sea surface pCO_2 prior baseline and model reconstruction. (a) Prior baseline obtained by simply adding the background and thermal terms; (b) reconstruction results for the training set based on residual learning; and (c) reconstruction results for the independent test set. The grey dashed line denotes the 1:1 reference line, and the colors indicate the number of samples.

The comparison between the prior baseline and the residual-learning reconstruction provides an initial validation of the BTR framework and establishes the credibility of the reconstructed product. The prior baseline, defined as the sum of the background climatology and the thermal-anomaly term, captures the broad range of surface-ocean pCO_2 variability but remains insufficient to reproduce the observations quantitatively (Fig. 2). Its moderate explanatory power and relatively large error, with an R^2 of 0.57 and an RMSE of 27.89 µatm, indicate that the background and thermal terms provide an effective first-order constraint but cannot fully represent local and extreme pCO_2 variability. This limitation is also reflected by a systematic negative bias of −5.56 µatm and the clear dispersion of points from the 1:1 line, particularly at high and low pCO_2 values.

Residual learning markedly improves this first-order estimate and, more importantly, maintains stable skill in independent years. In the training set, the reconstruction achieves an R^2 of 0.80 and an RMSE of 19.04 µatm, while the independent test set shows a similar level of performance, with an R^2 of 0.78 and an RMSE of 20.15 µatm. The test-set MAE is 12.85 µatm and the bias is reduced to −0.40 µatm, indicating that the residual model not only lowers random error but also substantially corrects systematic deviations in the prior baseline. Relative to the prior estimate, the test-set RMSE and MAE are reduced by approximately 28 % and 38 %, respectively.

The small difference between training and independent-year performance suggests that the learned residual correction is not merely fitted to the training samples but remains transferable to unseen years. These results indicate that the residual component contains predictable non-thermal information that can be constrained by environmental predictors. Therefore, the BTR framework improves reconstruction accuracy while providing a credible basis for generating a temporally extended surface-ocean pCO_2 product.

https://doi.org/10.5194/essd-2026-375
Preprint. Discussion started: 11 June 2026

Open Access Earth System Science Data Discussions

4.2 Basin-dependent predictor importance

The physical basis of residual learning can be further assessed from the predictors selected by the basin-specific models. Because the BTR framework removes the background climatological state and the direct thermal contribution before training, feature importance should be interpreted as explaining the remaining pCO_2 residual, rather than the total variability of raw pCO_2.

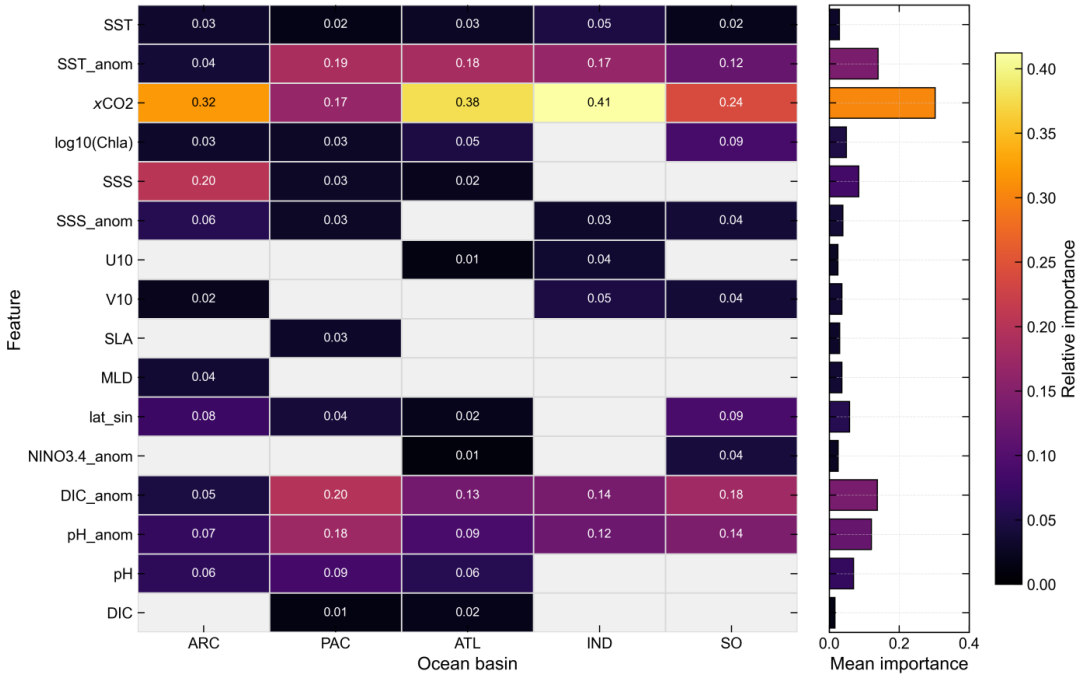

Figure 3. Relative importance of environmental predictors in basin-specific models. The bar chart on the right shows each feature's mean importance computed over the basin provinces in which that feature is available.

Figure 3 shows that the controls on residual pCO_2 variability are strongly basin dependent. Among all predictors, xCO_2 has the highest mean importance, particularly in the Indian Ocean, Atlantic Ocean, and Arctic Ocean, with values of 0.41, 0.38, and 0.32, respectively. This pattern indicates that atmospheric CO_2 growth still provides an important constraint on long-term residual variability after the background and thermal components have been removed. Carbonate-system anomalies, especially DIC_anom and pH_anom, also make substantial contributions, suggesting that the residual field retains physically meaningful signals related to internal carbon storage and acid–base perturbations. Other predictors show stronger regional specificity. SST_anom, SSS, log10(Chl-a), lat_sin, Siconc, MLD, and NINO3.4_anom contribute in specific basins where dynamical adjustment, biological activity, cryospheric processes, or remote climate forcing are more relevant.

Overall, the feature-importance analysis supports the interpretability of the BTR framework. The residual model does not simply fit unexplained noise; instead, it captures basin-dependent, physically interpretable variability that remains after removing the large-scale background state and the direct thermal effect.

https://doi.org/10.5194/essd-2026-375
Preprint. Discussion started: 11 June 2026

4.3 Boundary blending effects

Regionalized modeling is designed to improve the representation of basin-specific processes, but independently trained regional models may introduce artificial discontinuities at basin boundaries. A credible global reconstruction must therefore retain regional adaptability while maintaining spatial continuity.

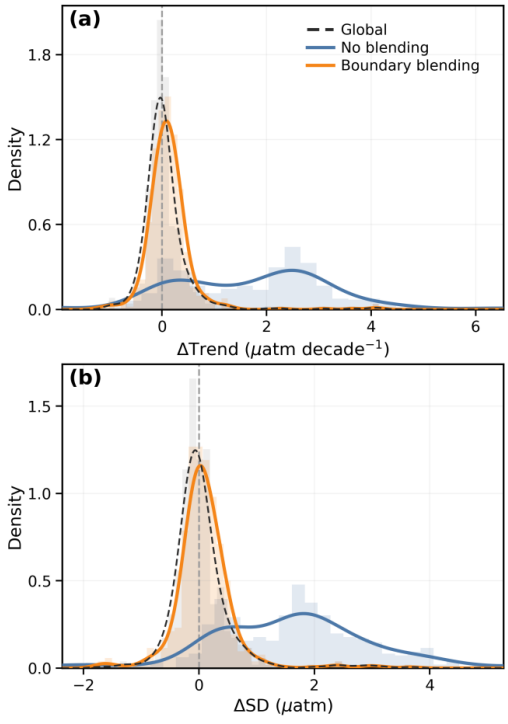

Figure 4. Effects of boundary blending on the boundary continuity of reconstructed surface-ocean pCO_2. (a) Distributions of long-term pCO_2 trend differences across regional boundaries under different modeling schemes. (b) Distributions of pCO_2 anomaly standard deviation differences across regional boundaries. The black dashed, blue, and orange curves represent the global single model, the regionalized model without boundary blending, and the regionalized model with boundary blending, respectively.

To evaluate whether boundary blending reduces discontinuities introduced by independent regional training, we compared boundary differences among three schemes: a single global model, a regionalized model without boundary blending, and a regionalized model with boundary blending (Fig. 4). Two diagnostics were calculated across regional boundaries: differences in long-term pCO_2 trends and differences in the standard deviation of pCO_2 anomalies. For each boundary, both diagnostics were estimated from the nearest grid cells on the two sides of the boundary. These metrics assess whether adjacent regions show consistent long-term change and temporal variability near model interfaces.

The single global model provides a useful reference because the entire prediction field is generated by one model. Its boundary-difference distributions are relatively narrow, but they are not exactly centered at zero, indicating that residual contrasts can also arise from real spatial gradients in the predicted field. In contrast, the unblended regionalized model shows much broader

https://doi.org/10.5194/essd-2026-375
Preprint. Discussion started: 11 June 2026

distributions and stronger positive deviations from zero for both diagnostics. This pattern indicates that directly stitching independently trained regional models introduces additional boundary contrasts in both trend and variability amplitude.

Boundary blending markedly reduces these artificial contrasts. For both diagnostics, the blended regionalized model shifts toward zero and shows much narrower distributions than the unblended regionalized model. Its distributions are closer to the single-global-model reference and, in some cases, are even more tightly concentrated near zero. This indicates that boundary blending improves the continuity of long-term trends and anomaly variability across regional interfaces while retaining the regionalized modeling structure.

These boundary differences should not be interpreted directly as model errors, because real oceanic gradients can arise from water-mass contrasts, circulation, frontal structures, and biogeochemical regime shifts. The purpose of boundary blending is therefore not to remove all spatial gradients, but to suppress non-physical discontinuities caused by independent regional training and direct model stitching. Overall, regionalized modeling preserves basin-scale heterogeneity, while boundary blending improves the spatial consistency and credibility of the global pCO_2 reconstruction.

4.4 Uncertainty diagnostics

Because observation-driven surface-ocean pCO_2 reconstructions do not have a high-resolution reference field for rigorous grid-cell-level error decomposition, this study adopts an empirical component-based uncertainty framework following Chen et al. (2025). Total uncertainty is defined as the quadratic sum of measurement, sampling, and mapping uncertainties:

$$U_{\text{total}} = \sqrt{U_{\text{meas}}^2 + U_{\text{samp}}^2 + U_{\text{map}}^2}, \tag{13}$$

where U_{meas} represents measurement and conversion error, U_{samp} represents finite-sample uncertainty within grid cells, and U_{map} represents mapping error relative to independent observations.

The measurement error U_{meas} is set to 5 μatm, following previous studies, to account for basic observational uncertainty and fCO_2-to-pCO_2 conversion uncertainty (Zhang et al., 2025). The sampling error U_{samp} is derived from the within-cell dispersion provided by the gridded observational product and converted into pCO_2 space using the same correction procedure as the training labels. Its overall scale is estimated from valid independent test samples as

$$U_{\text{samp}} = \sqrt{\frac{1}{N}\sum_{i=1}^{N} u_{\text{samp},i}^2}, \tag{14}$$

where $u_{\text{samp},i}$ is the sampling error for sample i, and $N$is the number of valid test samples.

The mapping error U_{map} quantifies the error introduced when environmental predictors are mapped to the final pCO_2 field. For each independent test sample, the residual is defined as

$$\varepsilon_i = \widehat{pCO}_{2,i} - pCO_{2,i}^{\text{obs}}, \tag{15}$$

where $\widehat{pCO}_{2,i}$ is the product value matched to the same year-month and location, and $pCO_{2,i}^{\text{obs}}$ is the observed value. Mapping

https://doi.org/10.5194/essd-2026-375
Preprint. Discussion started: 11 June 2026

Open Access Earth System Science Data Discussions

uncertainty is then calculated as the standard deviation of these residuals:

$$U_{\mathrm{map}} = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(\varepsilon_i - \bar{\varepsilon})^2}, \tag{16}$$

where $\bar{\varepsilon}$ is the mean residual. This formulation separates residual dispersion from mean bias and provides an empirical estimate of mapping skill under independent observational constraints. The resulting U_{meas}, U_{samp}, U_{map}, and U_{total} are used to analyze the magnitude and spatial structure of uncertainty. This framework provides an approximate characterization of product uncertainty and should not be interpreted as a rigorous grid-cell-level confidence interval.

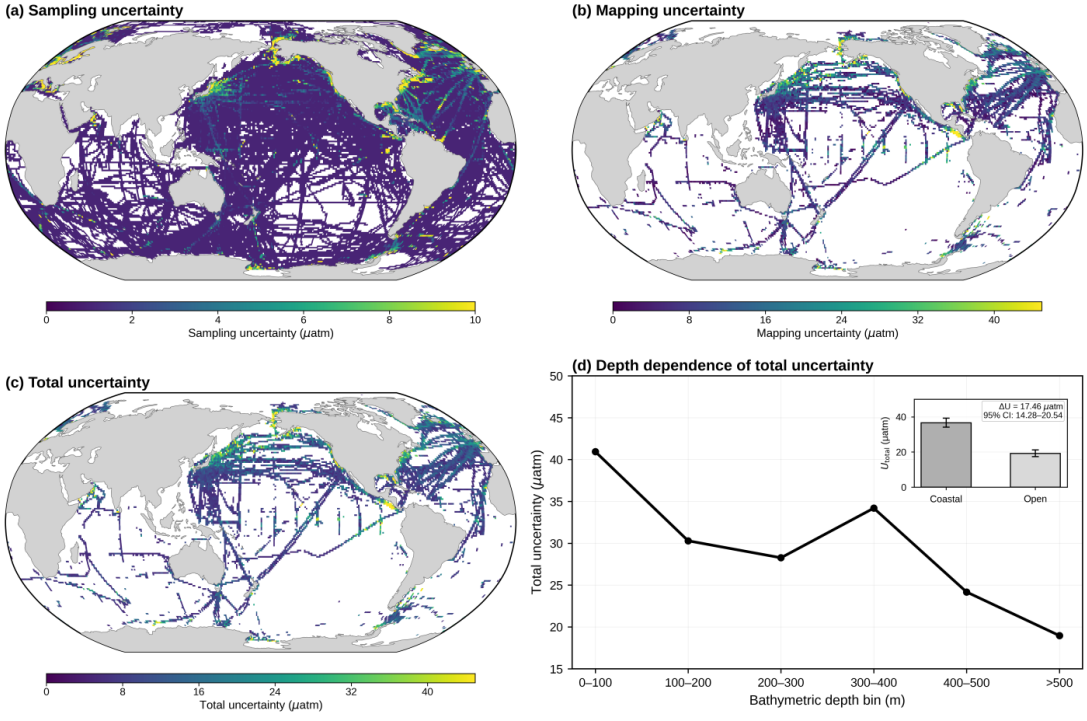

Figure 5. Spatial distribution of pCO_2 reconstruction uncertainty and water-depth-based statistics. (a) Sampling uncertainty, (b) mapping uncertainty, (c) total uncertainty, and (d) total uncertainty across water-depth classes. In (d), the main panel shows total uncertainty calculated from all valid samples within each water-depth class, while the inset compares total uncertainty between nearshore regions, defined as water depth < 200 m, and the open ocean using fixed equal-sample random sampling. Error bars indicate the 2.5 %–97.5 % percentile range.

Figure 5 indicates that mapping uncertainty dominates the total uncertainty of surface-ocean pCO_2 reconstruction. The global mean sampling uncertainty is only 1.58 μatm, suggesting that finite-sample variability within observed grid cells is not the main constraint. By contrast, mapping uncertainty reaches 10.83 μatm and largely determines the spatial structure of total uncertainty, whose global mean is 12.90 μatm. This implies that the dominant uncertainty arises from extrapolating sparse observations and environmental predictors to a continuous global pCO_2 field, rather than from instability in the observational labels themselves.

The spatial pattern of uncertainty further identifies where the reconstruction is least constrained. High uncertainty is

https://doi.org/10.5194/essd-2026-375
Preprint. Discussion started: 11 June 2026

concentrated mainly in the North Pacific, the North Atlantic, and coastal–shelf–open-ocean transition zones, where dynamical and biogeochemical heterogeneity is stronger and observational constraints are more limited. By contrast, open-ocean regions generally show lower uncertainty, consistent with broader spatial coherence and more stable relationships between pCO_2 and large-scale environmental predictors.

The depth-based comparison confirms this nearshore–open-ocean contrast. Total uncertainty is highest in the 0–100 m and 100–200 m depth classes and decreases markedly in regions deeper than 400 m. To reduce the influence of unequal sample sizes, 1000 samples are repeatedly drawn from the nearshore and open-ocean groups, and the uncertainty difference is calculated over 1000 trials. The results show that nearshore total uncertainty is approximately 17 μatm higher than open-ocean uncertainty and is nearly twice as large. This indicates that shallow and transition-zone environments remain the primary uncertainty hotspot in the current global product.

Overall, the uncertainty analysis defines the appropriate use of FDU-BTR. The product is best suited for basin-scale to global analyses of monthly to interannual pCO_2 variability, long-term trends, product intercomparison, and open-ocean air–sea CO_2 flux estimation. Caution is required in nearshore, shelf, marginal-sea, coastal-upwelling, coral-reef, and some high-latitude regions, especially in winter, where sparse observations, scale mismatch, and strong local processes increase mapping uncertainty.

5 Product validation and comparison

5.1 Independent time-series validation

The uncertainty analysis identifies where the reconstruction is expected to be less constrained. To test whether this spatial pattern is reflected in independent observations and to evaluate time-series performance, we use the NOAA NCEI surface buoy pCO_2 dataset archived by the Ocean Carbon and Acidification Data System (Sutton et al., 2018). The observations are averaged into site–year–month means and matched to the nearest 1° × 1° FDU-BTR grid cell for the corresponding month. This buoy-based validation provides an independent test of product consistency across open-ocean, coastal, and coral reef environments. The dataset, corresponding to NCEI Accession 0173932, includes observations from 40 autonomous surface buoy stations during 2004–2017. After filtering for temporal coverage, missing values, and valid grid cells, 28 stations are retained, including 17 open-ocean, 8 coastal, and 3 coral reef stations.

https://doi.org/10.5194/essd-2026-375
Preprint. Discussion started: 11 June 2026

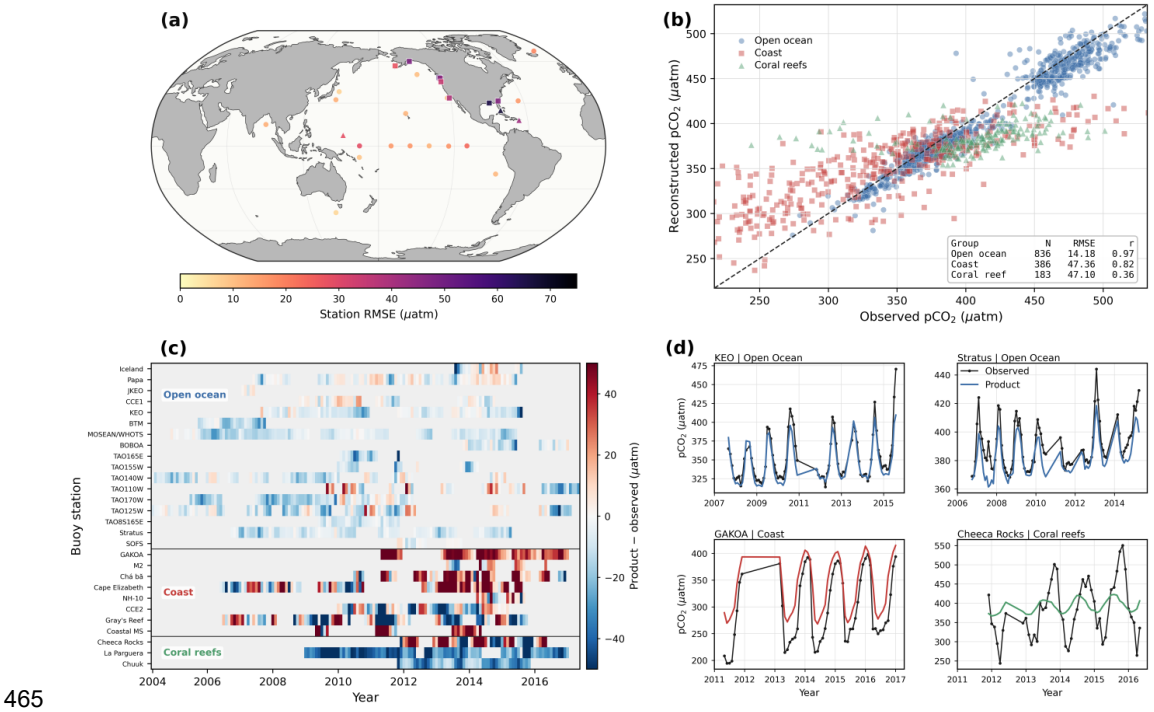

Figure 6. Independent time-series validation using NCEI surface buoy stations. (a) Global distribution of valid validation stations, with symbols denoting station type. (b) Comparison between observed and reconstructed pCO_2; the dashed line indicates the 1:1 line. (c) Monthly residual distributions at each station, where residuals are defined as product values minus observations. (d) Time-series comparisons at representative stations.

Figure 6 shows a clear environment-dependent contrast in validation skill. FDU-BTR performs best at open-ocean stations, where broad-scale variability is better resolved by a 1° monthly product. Across 836 valid monthly samples, the open-ocean stations show an RMSE of 14.18 μatm and a correlation coefficient of 0.97. Errors increase substantially in more locally heterogeneous environments. Coastal stations have an RMSE of 47.36 μatm and a correlation coefficient of 0.82 based on 386 samples, while coral reef stations show an RMSE of 47.10 μatm and a lower correlation coefficient of 0.36 based on 183 samples. This contrast indicates that FDU-BTR provides robust time-series performance in the open ocean, whereas reconstruction skill is reduced in coastal and coral reef settings.

The difference among environments mainly reflects scale representativeness. Open-ocean pCO_2 variability is strongly influenced by broad-scale temperature, circulation, mixing, and carbonate-system conditions, which are more compatible with the spatial and temporal resolution of a 1° monthly product. In contrast, coastal regions are affected by terrestrial inputs, upwelling, tidal mixing, and local biological processes, while coral reef regions are additionally influenced by reef metabolism, tidal exchange, and residence-time effects. These processes generate localized and high-frequency pCO_2 variability that cannot be fully represented by monthly grid-cell averages.

The residual distributions and representative station time series support this interpretation. Open-ocean residuals are generally

https://doi.org/10.5194/essd-2026-375
Preprint. Discussion started: 11 June 2026

small and centered near zero, and FDU-BTR captures the observed variability well at KEO and Stratus. Coastal stations show larger residual amplitudes; at GAKOA, the product captures part of the temporal evolution but underrepresents variability and extreme values. Coral reef stations show the least stable agreement. At Cheeca Rocks, observed pCO_2 varies strongly, whereas the product remains much smoother, indicating that local reef-scale processes are not resolved by the gridded reconstruction. Overall, the NCEI buoy validation confirms that FDU-BTR provides stable time-series performance in the open ocean and supports its use for large-scale carbon-cycle analyses. The larger discrepancies in coastal and coral reef environments mainly reflect the scale mismatch between point observations and 1° monthly grid values, together with strong local process variability, rather than a general failure of the product. This pattern is consistent with the uncertainty analysis, which identifies dynamically and biogeochemically complex coastal and reef environments as regions of elevated reconstruction uncertainty.

5.2 Generalization performance relative to existing products

HOT (Karl and Lukas, 1996) and BATS (Steinberg et al., 2001) are long-term open-ocean time-series stations that are widely used as benchmarks for evaluating surface-ocean pCO_2 products. Here, FDU-BTR is compared with 11 existing products listed in Table 2 after consistent temporal and spatial matching. This comparison assesses whether FDU-BTR can reproduce observed pCO_2 variability at representative subtropical open-ocean sites and how its performance compares with current community products.

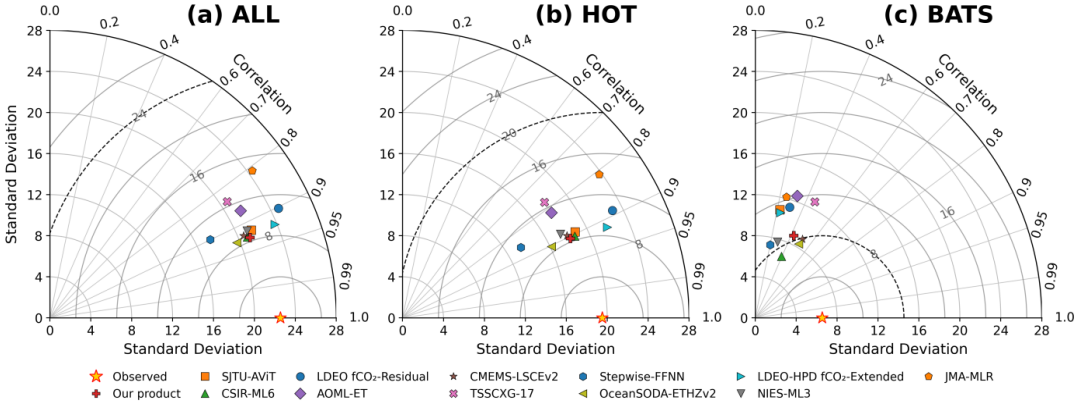

Figure 7. Taylor diagrams for the HOT/BATS independent station validation. (a) Combined HOT and BATS samples, (b) HOT station, and (c) BATS station. The yellow star denotes the observations, and the other symbols represent different sea surface pCO_2 products. Radial distance, azimuthal angle, and grey curves indicate standard deviation, correlation coefficient, and centered root-mean-square difference, respectively.

For the combined HOT and BATS samples, FDU-BTR shows strong agreement with the observations and performs comparably to leading existing products (Fig. 7). Its standard deviation is 21.07 µatm, close to the observed value of 22.54 µatm, indicating that the product reproduces the observed variability amplitude well. It also achieves a high temporal

https://doi.org/10.5194/essd-2026-375
Preprint. Discussion started: 11 June 2026

correlation of 0.93, an R^2 of 0.86, a low RMSE of 8.34 µatm, an MAE of 6.54 µatm, and a near-zero bias of −0.14 µatm. Together, these metrics indicate that FDU-BTR captures both the phase and amplitude of long-term pCO_2 variability while introducing little systematic offset.

The cross-product comparison places FDU-BTR among the better-performing products in this benchmark. In the Taylor diagram, FDU-BTR is located close to the observational reference point, with an error level comparable to those of CMEMS-LSCEv2, NIES-ML3, SJTU-AViT, and CSIR-ML6. OceanSODA-ETHZv2 also performs well, with an RMSE of 8.47 µatm and an R^2 of 0.86. In contrast, JMA-MLR, AOML-ET, TSSCXG-17, and Stepwise-FFNN show larger deviations from the observations. These differences indicate that FDU-BTR reaches a level of skill comparable to current mainstream surface-ocean pCO_2 reconstructions at widely used open-ocean validation sites.

Station-specific results further show that FDU-BTR reproduces the observed variability amplitude well at HOT. At BATS, inter-product differences are larger, partly because the smaller sample size makes the statistics more sensitive to individual months and local variability. Nevertheless, FDU-BTR maintains a reasonable error level at both stations, suggesting that its performance is not limited to a single benchmark site.

The HOT/BATS validation further demonstrates that FDU-BTR performs consistently with established pCO_2 reconstruction products. It shows high temporal correlation, realistic variability amplitude, low RMSE, and negligible systematic bias against independent long-term station records.

5.3 Large-scale trend consistency

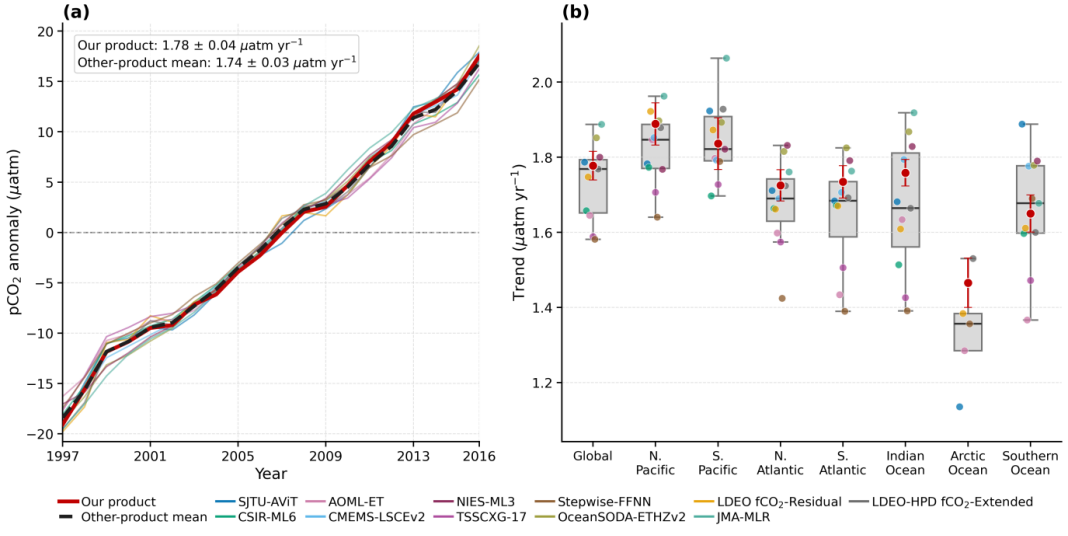

Figure 8. Regional annual deseasonalized pCO_2 anomalies and trend comparison across reconstruction products. (a) Annual deseasonalized pCO_2 anomalies for the global ocean. (b) Regional linear trends of annual deseasonalized pCO_2 anomalies. Grey boxplots show the distribution of trends from the other products, colored points show individual product trends using the same colors as in the line legend, and the red circles with error bars show the trend of our product ± standard error.

https://doi.org/10.5194/essd-2026-375
Preprint. Discussion started: 11 June 2026

We further evaluated the consistency of large-scale temporal variability between FDU-BTR and the existing surface-ocean pCO_2 reconstruction products listed in Table 2 by comparing annual deseasonalized pCO_2 anomalies across multiple products and ocean regions (Fig. 8). Before the comparison, all products were remapped to the 1° FDU-BTR grid. Products originally reported as fCO_2 were converted to pCO_2, and regional means were calculated using monthly, dynamically defined common valid grid cells to reduce differences caused by inconsistent spatial coverage. For each product and region, the monthly climatological seasonal cycle over the common comparison period was removed, and annual mean anomalies were then calculated from the monthly anomalies. Linear trends were estimated from the annual anomaly time series. For the Arctic Ocean, products with very limited or absent coverage were excluded from the regional trend calculation.

At the global scale, FDU-BTR reproduces the sustained increase in surface-ocean pCO_2 anomalies during 1997–2016. Its estimated global trend is 1.78 ± 0.04 μatm yr^{-1}, closely matching the multi-product mean of 1.74 ± 0.03 μatm yr^{-1}. This agreement indicates that FDU-BTR captures the dominant long-term increase in the global surface-ocean carbon system. Although existing products show relatively strong consistency in the global mean trend, regional differences remain substantial. Across the major ocean basins, FDU-BTR generally falls within the multi-product spread and shows no systematic high or low trend bias. At the same time, it retains regional differences in pCO_2 growth rates, indicating that the product is consistent with the existing product ensemble while preserving basin-scale trend variability.

Inter-product differences are largest in the Arctic Ocean and the Southern Ocean, where trend estimates remain more uncertain. The Arctic Ocean shows the lowest estimated trend overall and a large inter-product spread, indicating substantial disagreement among current reconstruction products. The FDU-BTR Arctic trend lies between LDEO fCO_2-Residual and LDEO-HPD fCO_2-Extended. It differs from some lower trend estimates but remains within the plausible range of existing products. The Southern Ocean also shows pronounced inter-product differences, consistent with the known difficulty of reconstructing high-latitude pCO_2 variability under sparse observations, strong seasonality, and complex physical–biogeochemical coupling.

The multi-product comparison further shows that FDU-BTR is consistent with existing reconstruction products in representing the global-scale increase in surface-ocean pCO_2, while retaining meaningful regional trend differences. Its agreement with the multi-product mean supports broader product intercomparison, whereas the regional spread highlights the Arctic and Southern oceans as key regions of persistent trend uncertainty.

6 Observational coverage constraints

6.1 Observational constraints on multi-product divergence

To diagnose whether inter-product disagreement is related to observational support, we calculated the multi-product dispersion,

https://doi.org/10.5194/essd-2026-375
Preprint. Discussion started: 11 June 2026

σ_{spread}, using six global surface-ocean pCO_2 products: NIES-ML3, LDEO fCO_2-Residual, LDEO-HPD fCO_2-Extended, CMEMS-LSCEv2, OceanSODA-ETHZv2, and SJTU-AViT. These products are described in Sect. 2.2. Let $C_i(m,g)$ denote the detrended monthly climatology of product i for calendar month $m$and grid cell g, adjusted to the reference year 2005. For the set of valid products $P(m,g)$, with $N(m,g)$ valid members, σ_{spread} is defined as

$$\sigma_{\text{spread}}(m,g) = \sqrt{\frac{1}{N(m,g)}\sum_{i\in\mathcal{P}(m,g)}\left[C_i(m,g) - \overline{C}(m,g)\right]^2}, \quad (17)$$

where

$$\overline{C}(m,g) = \frac{1}{N(m,g)}\sum_{i\in\mathcal{P}(m,g)} C_i\,(m,g). \quad (18)$$

When $N(m,g) < 2$, σ_{spread} is set to missing. This metric represents structural dispersion among multi-product monthly climatologies and should not be interpreted as direct observational error or reconstruction error.

Observational support is quantified using normalized Shannon entropy. For a discrete binning variable, the normalized entropy is defined as

$$H = -\frac{\sum_{i=1}^{B} p_i \ln p_i}{\ln B}, \quad (19)$$

where p_i is the sample proportion in bin i, and $B$is the total number of bins. We calculate temporal coverage entropy, H_{year}, and spatial coverage entropy, H_{space}, to describe the uniformity of training-sample coverage across years and 5° spatial blocks, respectively. Lower entropy indicates more uneven coverage. The effective coverage entropy is then defined as the harmonic mean of the two components:

$$H_{\text{eff}} = \frac{2H_{\text{year}}H_{\text{space}}}{H_{\text{year}}+H_{\text{space}}}, \quad (20)$$

with $H_{\text{eff}} = 0$ when $H_{\text{year}} + H_{\text{space}} = 0$. This formulation is sensitive to deficiencies in either temporal or spatial coverage and therefore provides a joint measure of observational support.

https://doi.org/10.5194/essd-2026-375
Preprint. Discussion started: 11 June 2026

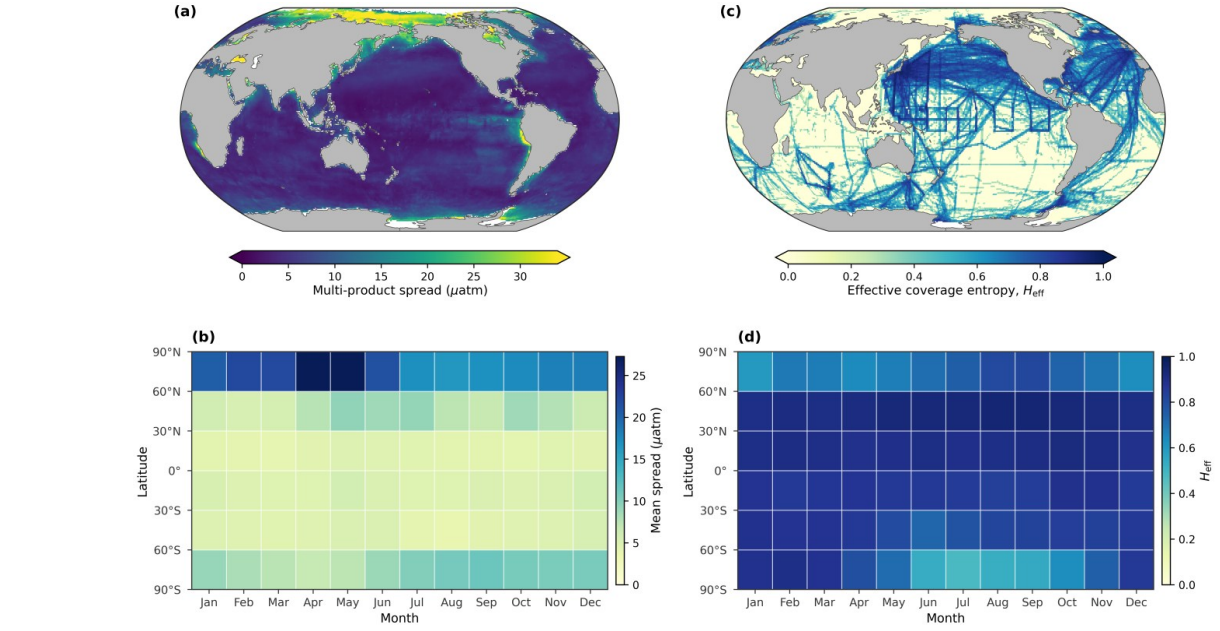

Figure 9. Spatial distribution and latitude–month variability of multi-product structural disagreement and effective coverage entropy. (a) Spatial distribution of the annual-mean structural dispersion of multi-product sea surface pCO_2 climatologies during the 1990–2020 reference period. (b) Monthly variation in multi-product spread across latitude bands. (c) Global 1° distribution of effective coverage entropy. (d) Latitude–month heatmap of effective coverage entropy.

Figure 9(a–b) shows that multi-product dispersion is strongly heterogeneous in both space and season. High σ_{spread} occurs mainly in high-latitude oceans, marginal seas, and some coastal upwelling regions, whereas tropical and subtropical open-ocean regions generally show lower dispersion. The latitude–month statistics further reveal a pronounced high-latitude dependence. Dispersion is largest in the Northern Hemisphere high latitudes, especially around spring, and is also elevated in the Southern Hemisphere high latitudes, although with weaker seasonality. In contrast, low- and mid-latitude regions show smaller and more stable inter-product differences.

The effective coverage entropy shows an almost opposite spatial structure (Fig. 9c–d). High H_{eff} values occur mainly along long-term repeat cruise tracks and in densely sampled regions, whereas low values dominate the interior South Pacific, South Atlantic, South Indian Ocean, and much of the Southern Ocean. The latitude–month pattern shows persistently stronger observational support in the 30°N–60°N band and relatively stable coverage in tropical and subtropical regions. By contrast, the 60°S–90°S band has the lowest H_{eff}, especially during Southern Hemisphere winter, indicating the combined effects of spatial undersampling and seasonal coverage deficiency.

Together, the spatial patterns of σ_{spread} and H_{eff} indicate a broad correspondence between high inter-product disagreement and weak observational support. This relationship is particularly evident in high-latitude regions, the Southern Ocean, and marginal seas, where low joint temporal–spatial coverage coincides with large multi-product dispersion. These results suggest that reduced agreement among current pCO_2 reconstructions is closely linked to insufficient observational support.

https://doi.org/10.5194/essd-2026-375
Preprint. Discussion started: 11 June 2026

However, the relationship between observational coverage and product disagreement is not spatially uniform. The Indian Ocean has relatively low H_{eff} but only modest product dispersion, whereas the tropical southeastern Pacific also has weak observational support but much higher dispersion. This contrast indicates that observational coverage is a key constraint, but not the only factor controlling inter-product differences. In process-complex regions, high background pCO_2, upwelling, and strong biogeochemical variability can further amplify differences among reconstruction products.

Overall, current surface-ocean pCO_2 products show clearer convergence in well-observed regions, while major disagreements remain concentrated where observations are sparse and processes are complex. This indicates that the reconstruction bottleneck is jointly controlled by observational coverage structure and regional process complexity, rather than by methodological differences alone.

6.2 Response of model accuracy to temporal and spatial coverage

The spatial co-occurrence of large inter-product spread and weak observational support suggests that coverage structure is closely linked to reconstruction uncertainty. However, this spatial diagnosis does not separate the effects of temporal and spatial coverage on reconstruction error. To quantify their respective impacts, we designed two controlled thinning experiments in which temporal and spatial coverage were reduced separately while the model framework was kept unchanged. The response of independent-year pCO_2 RMSE to temporal coverage entropy, H_{year}, and spatial coverage entropy, H_{space}, was then evaluated.

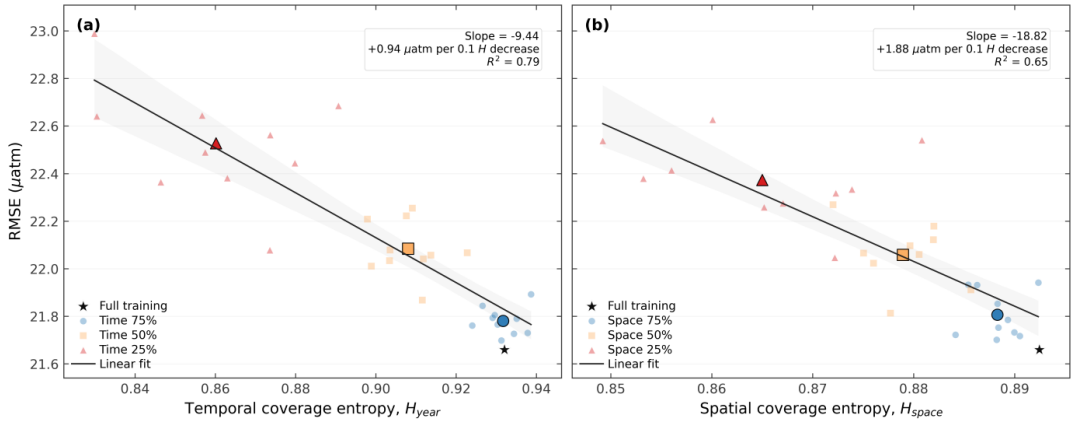

Figure 10. Coverage entropy versus independent-year reconstruction error. (a) RMSE as a function of normalized temporal coverage entropy H_{year} in the temporal thinning experiment. (b) RMSE as a function of normalized spatial coverage entropy H_{space} in the spatial thinning experiment. Light symbols denote random replicates, large outlined symbols denote thinning-level means, and black stars denote the full-training case. Black lines show linear fits, with grey shading indicating 95 % stratified-bootstrap confidence bands.

In both experiments, the BTR target, feature set, sample weighting, reconstruction procedure, and CatBoost ensemble configuration were fixed. Only the structure of the training-sample coverage was changed. To avoid interference from regional sample-size differences, a single global model was used. Model performance was evaluated using RMSE in the independent

https://doi.org/10.5194/essd-2026-375
Preprint. Discussion started: 11 June 2026

test years 1984, 1988, 1995, 2004, 2013, and 2021, which were excluded from training, feature selection, and hyperparameter optimization in all scenarios.

The temporal thinning experiment removed complete training years to simulate reduced observational continuity. Year selection was stratified by decade, and three retained-year proportions were tested: approximately 75 %, 50 %, and 25 %. Each thinning level was repeated 10 times with random sampling, together with the full-training baseline. The spatial thinning experiment removed 5° × 5° spatial blocks. Block selection was stratified by geographic region and 30° latitude band to avoid excessive spatial clustering of removed samples. Because spatial thinning also changes the number of retained samples, a row-matched design was used: for each temporal thinning case, a spatial thinning case with a comparable number of retained rows was constructed. The matching tolerance was 2 % of the target sample size, with a minimum tolerance of 50 samples. This design allows the effects of temporal and spatial coverage degradation to be compared under similar sample-size conditions.

Figure 10 shows that independent-year RMSE increases as either temporal or spatial coverage becomes less uniform. For temporal thinning, RMSE decreases with increasing H_{year}, with a fitted slope of −9.44 and an R^2 of 0.79 (Fig. 10a). This means that a 0.1 decrease in H_{year} corresponds to an average RMSE increase of about 0.94 µatm. The consistent trend across thinning levels indicates that reduced temporal continuity weakens cross-year generalization.

The error response to spatial thinning is stronger. RMSE also decreases with increasing H_{space}, but the fitted slope is −18.82, with an R^2 of 0.65 (Fig. 10b). A 0.1 decrease in H_{space} therefore corresponds to an average RMSE increase of about 1.88 µatm, approximately twice the response associated with temporal coverage loss. This stronger sensitivity indicates that independent-year reconstruction skill is more strongly constrained by the spatial representativeness of the training data than by temporal continuity alone.

In both experiments, the full-training case lies at the high-entropy and low-error end of the relationship, supporting the overall coverage–skill link. The scatter also becomes broader under low-coverage conditions, indicating that insufficient observational support not only increases mean error but also increases the variability of model outcomes.

Overall, the thinning experiments reveal an asymmetric effect of observational coverage on out-of-sample pCO_2 reconstruction skill. Both temporal and spatial coverage losses degrade independent-year performance, but spatial coverage loss imposes a substantially larger error penalty under comparable sample-size conditions. This indicates that out-of-sample pCO_2 reconstruction is limited not only by the number of observations, but also, and more critically, by their spatial representativeness.

7 Discussion

7.1 Performance gains from the BTR framework

The results demonstrate that the BTR framework systematically improves sea surface pCO_2 reconstruction relative to the prior

https://doi.org/10.5194/essd-2026-375
Preprint. Discussion started: 11 June 2026

baseline defined by the background and thermal terms. It substantially reduces RMSE, MAE, and systematic bias in both the training and independent test sets. Its performance at the HOT and BATS stations is also comparable to that of current leading products, indicating that the improvement is not limited to training samples but extends to independent years and external time-series observations.

This improvement stems primarily from the redefinition of the learning target. Conventional approaches generally regress raw sea surface pCO_2 directly and therefore require the model to simultaneously capture background gradients, seasonal variability, and local anomalies (Fay et al., 2024). In the BTR framework, by contrast, observed pCO_2 is first decomposed into a background term, a thermal term, and a residual term, and the model is trained only on the residual component. By removing first-order physically interpretable variability before model training, the framework reduces target complexity and allows the model to focus on non-thermal variability that is not captured by the prior estimate.

Although the explicit thermal term and the multi-product background climatology together account for roughly 70–80% of the total variance in surface-ocean pCO_2, the residual term, despite its smaller variance share, dominates the inter-product disagreement and controls the reproduction of interannual variability, climate-mode responses, and the long-term anthropogenic trend. By separating these background and thermal components before model training, BTR reformulates the reconstruction problem from directly fitting the full pCO_2 field to explicitly learning the residual, process-sensitive component that is otherwise difficult to isolate and constrain from sparse observations.

The non-decomposed CatBoost reconstruction, in which pCO_2 is predicted directly under the same modeling configuration, provides a strong direct-regression baseline in the independent-year test, with an RMSE of 21.62 µatm, an MAE of 13.78 µatm, and an R^2 of 0.74. Under the same configuration, BTR further lowers the independent-year RMSE by about 1.46 µatm overall, including reductions of 2.05 µatm in nearshore regions and 1.43 µatm in the open ocean, while also reducing MAE and increasing R^2. This moderate but systematic improvement is consistent with prior physics-guided ML studies and supports the interpretation that the principal value of the BTR framework is not in capturing the dominant variance itself, but in resolving the structurally most uncertain and scientifically consequential residual component of the surface-ocean pCO_2 field (Chau et al., 2024; Bennington et al., 2022a).

7.2 Limitations and implications for future product development

The BTR framework, regionalized modeling, and boundary blending improve surface-ocean pCO_2 reconstruction, but they do not remove the fundamental sources of product uncertainty. These uncertainties arise from two related limitations. First, substantial errors persist in high-latitude regions, marginal seas, coastal upwelling systems, and the Southern Ocean, where observations remain sparse, seasonal coverage is incomplete, and biogeochemical and physical processes are highly heterogeneous. In these settings, the model is required to extrapolate beyond well-sampled conditions, which weakens the stability of predictor–pCO_2 relationships and increases mapping uncertainty. Second, the effectiveness of the BTR framework

https://doi.org/10.5194/essd-2026-375
Preprint. Discussion started: 11 June 2026

depends on the quality of the prior field. Because the background climatology and thermal correction jointly define the residual target, biases in either component can propagate into the learned residual. The residual may therefore contain not only genuine non-thermal variability but also compensatory corrections for prior-field errors. Thus, BTR reduces the complexity of the learning task, but it cannot fully eliminate uncertainty inherited from sparse observations or imperfect background constraints.

This interpretation is consistent with previous findings. Observation-based products often show high mutual correlations and comparable RMSE values at subtropical stations such as HOT and BATS, where observational support is strong and seasonal variability is well defined (Gregor et al., 2019; Bennington et al., 2022a; Gloege et al., 2022). Under such conditions, different gap-filling methods tend to converge because the training data provide sufficient constraints on the dominant variability. By contrast, in poorly observed or process-complex regions, reconstruction skill becomes more strongly controlled by the spatiotemporal structure of the available observations and by the accuracy of the prior information (Heimdal et al., 2024). Methodological differences are therefore more easily amplified in these regions. This suggests that future progress should not rely solely on algorithmic refinement, but should combine improved observing-system design, more robust prior-field construction, and stronger representation of regional processes.

A key practical implication is that targeted observations in data-sparse regions may be more valuable than simply increasing the total number of samples. Even a limited number of additional shipboard, moored, or autonomous observations in poorly constrained regions could strengthen regional training constraints and reduce extrapolation uncertainty. For future SOCAT sampling strategies, this suggests that priority should be given to expanding measurements in regions with sparse coverage and large inter-product disagreement, rather than further intensifying repeated sampling along already well-observed routes.

This finding also has implications for future float-based observing systems, including potential extensions of Argo-type carbon observations. Expanded float deployments could be particularly valuable in high-latitude oceans and other process-complex regions, where broader spatial coverage is needed and ship-based sampling is difficult to sustain year-round, especially during winter. For coastal zones, shelves, and upwelling systems, additional process-relevant predictors are also needed, including river inputs, tidal mixing, shelf topography, coastal circulation, nutrients, and high-resolution ocean color. For high-latitude regions and the Southern Ocean, reducing reconstruction errors will require both broader spatial coverage and improved wintertime observations, because seasonal and spatial extrapolation uncertainties act together in these regions.

Methodologically, next-generation sea surface pCO_2 products should be developed through the joint optimization of observational structure and model structure. Unified machine-learning frameworks can already produce relatively stable reconstructions in well-observed regions. In poorly observed regions or regions with complex processes, more adaptive strategies are needed, including hierarchical modeling, regional transfer learning, the enforcement of physical constraints, and explicit uncertainty propagation. Overall, further progress depends not simply on the application of more complex algorithms, but on the deliberate integration of targeted observations, process understanding, and data-driven reconstruction.

https://doi.org/10.5194/essd-2026-375
Preprint. Discussion started: 11 June 2026

8 Data availability

The FDU-BTR dataset is publicly available from Zenodo at https://doi.org/10.5281/zenodo.20152530 (Wang and Fu, 2026). The archive contains the global monthly gridded surface-ocean pCO_2 product used in this study, provided as the NetCDF file. The file follows the CF-1.8 and ACDD-1.3 metadata conventions and is distributed under the Creative Commons Attribution 4.0 International license.

The dataset contains 516 monthly fields from January 1982 to December 2024 on a regular 1° × 1° latitude–longitude grid. The primary variable is pCO2 with dimensions (time, latitude, longitude); missing or invalid grid cells are stored as NaN.

FDU-BTR is recommended for basin-scale to global analyses of monthly to interannual pCO_2 variability, long-term trends, product intercomparison, and open-ocean air–sea CO_2 flux estimation. Users should exercise caution when interpreting values in nearshore, shelf, marginal-sea, coastal-upwelling, coral-reef, and some high-latitude regions, where mapping uncertainty is higher. The product is not intended to resolve reef-scale, estuarine, event-scale, or submonthly pCO_2 variability without additional local observational constraints.

9 Code availability

The source code supporting the generation of FDU-BTR and the main analyses in this study is available at https://github.com/layne1202/FDU-BTR_Code. The repository provides the core scripts for construction of the multi-product background climatology, BTR target decomposition, basin-specific residual model training, boundary blending, and monthly NetCDF generation.

10 Conclusions

We present FDU-BTR, a new global monthly surface-ocean pCO_2 reconstruction product for 1982–2024 at 1° × 1° resolution, developed using a Background–Thermal Residual (BTR) ensemble learning framework (Wang and Fu, 2026, https://doi.org/10.5281/zenodo.20152530). By decomposing observed pCO_2 into a multi-product background climatology, an explicit thermal term, and a residual component, the framework reduces the complexity of direct pCO_2 learning while retaining first-order physical constraints. The main conclusions are as follows.

- First, the BTR framework provides a physically interpretable basis for surface-ocean pCO_2 reconstruction. The background and thermal terms constrain large-scale pCO_2 variability, while the residual model captures predictable non-background and non-thermal signals associated with regional biogeochemical and physical processes. Relative to the prior baseline, residual learning substantially improves independent-year reconstruction skill, with an R^2 of 0.78, an RMSE of 20.15 μatm, an MAE of 12.85 μatm, and a near-zero bias of −0.40 μatm.

https://doi.org/10.5194/essd-2026-375
Preprint. Discussion started: 11 June 2026

- Second, regionalized modeling and boundary blending allow FDU-BTR to represent basin-scale differences while maintaining spatial continuity. The regional models reflect distinct controls on residual pCO_2 variability across ocean basins, and boundary blending reduces artificial contrasts in long-term trends and anomaly variability across regional interfaces.
- Third, the uncertainty analysis shows that mapping uncertainty is the dominant contributor to total uncertainty. Uncertainty is highest in nearshore regions, shelf–open-ocean transition zones, marginal seas, coastal upwelling systems, and some high-latitude oceans, with nearshore total uncertainty nearly twice that of the open ocean. This contrast highlights the reduced reconstruction stability in regions where sparse observations, scale mismatch, and complex local processes jointly affect pCO_2 variability.
- Fourth, product comparisons show that FDU-BTR is consistent with current community pCO_2 reconstructions. At the HOT and BATS long-term stations, FDU-BTR shows balanced performance across correlation, variability amplitude, RMSE, and bias, with skill comparable to leading existing products. Multi-product comparisons also show that FDU-BTR reproduces the sustained global increase in surface-ocean pCO_2, while regional disagreements remain largest in high-latitude and poorly observed regions.
- Fifth, observational coverage remains a major constraint on global pCO_2 reconstruction. Multi-product disagreement is largest where observational support is weak and processes are complex. Controlled thinning experiments show that both temporal and spatial coverage losses degrade independent-year skill, but spatial coverage loss imposes a stronger error penalty, with an RMSE response approximately twice that of temporal coverage loss. This indicates that future improvements should prioritize spatially representative observations, especially in nearshore regions, marginal seas, the Southern Ocean, high-latitude winter conditions, and other process-complex regions.

Overall, FDU-BTR provides a physically constrained and uncertainty-diagnosed basis for analyzing long-term surface-ocean pCO_2 variability and estimating air–sea CO_2 fluxes. Because pCO_2 directly controls the air–sea CO_2 gradient, improved pCO_2 reconstruction is directly relevant to assessing the magnitude, variability, and regional structure of the ocean carbon sink. These results indicate that reducing carbon-sink uncertainty not only requires better reconstruction algorithms, but more importantly spatially representative observations and stronger process constraints in nearshore, marginal, high-latitude, and other process-complex regions.

Author contribution

Z. W. and W. F. conceived the project. Z. W. collected and processed the data, developed the reconstruction framework, performed the analyses, and prepared the initial manuscript. W. F. supervised the project and contributed to interpretation and manuscript revision.

https://doi.org/10.5194/essd-2026-375
Preprint. Discussion started: 11 June 2026

775 **Competing interests**

The authors declare that they have no conflict of interest.

Acknowledgements

We thank the SOCAT community for providing the core surface-ocean CO_2 observations used in this study. We also acknowledge the NOAA NCEI Ocean Carbon and Acidification Data System for the surface buoy pCO_2 records and the HOT
780 and BATS programs for the long-term station observations used in the independent evaluation. This research has been supported by the Natural Science Foundation of Shanghai Municipality (grant no. 24ZR1404500), the Science and Technology Commission of Shanghai Municipality (STCSM; grant no. 25DZ3102200), and the National Natural Science Foundation of China (NSFC; grant no. 42476011). AI-assisted language-editing tools were used to improve grammar, readability, and consistency. All scientific analyses, results, interpretations, and conclusions were produced, checked, and approved by the
785 authors.

References

Bakker, D. C., Pfeil, B., Landa, C. S., Metzl, N., O'Brien, K. M., Olsen, A., Smith, K., Cosca, C., Harasawa, S., and Jones, S. D.: A multi-decade record of high-quality fCO_2 data in version 3 of the Surface Ocean CO_2 Atlas (SOCAT), Earth Syst. Sci. Data, 8, 383-413, https://doi.org/10.5194/essd-8-383-2016, 2016.

790 Bennington, V., Galjanic, T., and McKinley, G. A.: Explicit physical knowledge in machine learning for Ocean Carbon flux reconstruction: The pCO_2-residual method, J. Adv. Model. Earth Syst., 14, e2021MS002960, https://doi.org/10.1029/2021MS002960, 2022a.

Bennington, V., Gloege, L., and McKinley, G. A.: Variability in the global ocean carbon sink from 1959 to 2020 by correcting models with observations, Geophys. Res. Lett., 49, e2022GL098632, https://doi.org/10.1029/2022GL098632, 2022b.

795 Chau, T.-T.-T., Gehlen, M., Metzl, N., and Chevallier, F.: CMEMS-LSCE: a global, 0.25°, monthly reconstruction of the surface ocean carbonate system, Earth Syst. Sci. Data, 16, 121-160, https://doi.org/10.5194/essd-16-121-2024, 2024.

Chau, T. T. T., Gehlen, M., and Chevallier, F.: 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, 1087-1109, https://doi.org/10.5194/bg-19-1087-2022, 2022.

800 Chen, Y., Wang, S., and He, W.: Global air-sea CO_2 flux inversion based on multi-source data fusion and machine learning, Chaos, Solitons & Fractals, 192, 115963, https://doi.org/10.1016/j.chaos.2024.115963, 2025.

Copernicus Climate Change Service, C. D. S.: Sea level gridded data from satellite observations for the global ocean from

https://doi.org/10.5194/essd-2026-375
Preprint. Discussion started: 11 June 2026

1993 to present, Copernicus Climate Change Service (C3S) Climate Data Store (CDS) [dataset], https://doi.org/10.24381/cds.4c328c78, 2018.

Crisp, D., Dolman, H., Tanhua, T., McKinley, G. A., Hauck, J., Bastos, A., Sitch, S., Eggleston, S., and Aich, V.: How well do we understand the land-ocean-atmosphere carbon cycle?, Reviews of Geophysics, 60, e2021RG000736, https://doi.org/10.1029/2021RG000736, 2022.

Denvil-Sommer, A., Gehlen, M., Vrac, M., and Mejia, C.: LSCE-FFNN-v1: A two-step neural network model for the reconstruction of surface ocean pCO_2 over the global ocean, Geosci. Model Dev., 12, 2091-2105, https://doi.org/10.5194/gmd-12-2091-2019, 2019.

Dickson, A. G., Sabine, C. L., and Christian, J. R.: Guide to best practices for ocean CO_2 measurements, North Pacific Marine Science Organization, http://dx.doi.org/10.25607/OBP-1342, 2007.

Fay, A. R., Munro, D. R., McKinley, G. A., Pierrot, D., Sutherland, S. C., Sweeney, C., and Wanninkhof, R.: Updated climatological mean ΔfCO_2 and net sea–air CO_2 flux over the global open ocean regions, Earth Syst. Sci. Data, 16, 2123-2139, https://doi.org/10.5194/essd-16-2123-2024, 2024.

Friedlingstein, P., O'Sullivan, M., Jones, M. W., Andrew, R. M., Bakker, D. C., Hauck, J., Landschützer, P., Le Quéré, C., Li, H., and Luijkx, I. T.: Global carbon budget 2025, Earth Syst. Sci. Data, 2025, 1-139, https://doi.org/10.5194/essd-2025-659, 2025.

Garcia, H., Wang, Z., Bouchard, C., Cross, S., Paver, C., Reagan, J., Boyer, T., Locarnini, R., Mishonov, A., and Baranova, O.: World Ocean Atlas 2023, Volume 3: Dissolved Oxygen, Apparent Oxygen Utilization, Dissolved Oxygen Saturation and 30-year Climate Normal [dataset], https://doi.org/10.25923/rb67-ns53, 2024.

Gloege, L., Yan, M., Zheng, T., and McKinley, G. A.: Improved quantification of ocean carbon uptake by using machine learning to merge global models and pCO_2 data, J. Adv. Model. Earth Syst., 14, e2021MS002620, https://doi.org/10.1029/2021MS002620, 2022.

Good, S. A., Martin, M. J., and Rayner, N. A.: EN4: Quality controlled ocean temperature and salinity profiles and monthly objective analyses with uncertainty estimates, J. Geophys. Res.-Oceans, 118, 6704-6716, https://doi.org/10.1002/2013JC009067, 2013.

Gregor, L. and Gruber, N.: 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, 777-808, https://doi.org/10.5194/essd-13-777-2021, 2021.

Gregor, L., Shutler, J., and Gruber, N.: High-resolution variability of the ocean carbon sink, Glob. Biogeochem. Cycles, 38, e2024GB008127, https://doi.org/10.1029/2024GB008127, 2024.

Gregor, L., Lebehot, A. D., Kok, S., and Scheel Monteiro, P. M.: 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, 5113-5136, https://doi.org/10.5194/gmd-12-5113-2019, 2019.

https://doi.org/10.5194/essd-2026-375
Preprint. Discussion started: 11 June 2026

Gruber, N., Clement, D., Carter, B. R., Feely, R. A., Van Heuven, S., Hoppema, M., Ishii, M., Key, R. M., Kozyr, A., and Lauvset, S. K.: The oceanic sink for anthropogenic CO_2 from 1994 to 2007, Science, 363, 1193-1199, https://doi.org/10.1126/science.aau5153, 2019.

Guinehut, S., Dhomps, A.-L., Larnicol, G., and Le Traon, P.-Y.: High resolution 3-D temperature and salinity fields derived from in situ and satellite observations, Ocean Sci., 8, 845-857, https://doi.org/10.5194/os-8-845-2012, 2012.

Heimdal, T. H., McKinley, G. A., Sutton, A. J., Fay, A. R., and Gloege, L.: Assessing improvements in global ocean pCO_2 machine learning reconstructions with Southern Ocean autonomous sampling, Biogeosciences, 21, 2159-2176, https://doi.org/10.5194/bg-21-2159-2024, 2024.

Hersbach, H.: ERA5 monthly averaged data on single levels from 1940 to present [dataset], https://doi.org/10.24381/cds.f17050d7, 2023.

Huang, B., Thorne, P. W., Banzon, V. F., Boyer, T., Chepurin, G., Lawrimore, J. H., Menne, M. J., Smith, T. M., Vose, R. S., and Zhang, H.-M.: Extended reconstructed sea surface temperature, version 5 (ERSSTv5): upgrades, validations, and intercomparisons, J. Clim., 30, 8179-8205, https://doi.org/10.1175/JCLI-D-16-0836.1, 2017.

Iida, Y., Takatani, Y., Kojima, A., and Ishii, M.: Global trends of ocean CO_2 sink and ocean acidification: an observation-based reconstruction of surface ocean inorganic carbon variables, J. Oceanogr., 77, 323-358, https://doi.org/10.1007/s10872-020-00571-5, 2021.

Jiang, L.-Q., Fay, A., Müller, J. D., Gregor, L., Roobaert, A., Keppler, L., Carroll, D., Lauvset, S. K., DeVries, T., and Hauck, J.: Synthesis of data products for ocean carbonate chemistry, Earth Syst. Sci. Data, 18, 1405-1462, https://doi.org/10.5194/essd-18-1405-2026, 2026.

Karl, D. M. and Lukas, R.: The Hawaii Ocean Time-series (HOT) program: Background, rationale and field implementation, Deep-Sea Res. II: Top. Stud. Oceanogr., 43, 129-156, https://doi.org/10.1016/0967-0645(96)00005-7, 1996.

Lan, X., Tans, P., Thoning, K., and Laboratory, N. G. M.: NOAA greenhouse gas marine boundary layer reference—CO_2 [dataset], https://doi.org/10.15138/DVNP-F961, 2024.

Landschützer, P., Gruber, N., and Bakker, D. C.: Decadal variations and trends of the global ocean carbon sink, Glob. Biogeochem. Cycles, 30, 1396-1417, https://doi.org/10.1002/2015GB005359, 2016.

Landschützer, P., Gruber, N., Bakker, D. C., and Schuster, U.: Recent variability of the global ocean carbon sink, Glob. Biogeochem. Cycles, 28, 927-949, https://doi.org/10.1002/2014GB004853, 2014.

Ocean Colour Thematic Assembly Centre (OCTAC): Global ocean colour (Copernicus-GlobColour), bio-geo-chemical, L4 (monthly and interpolated) from satellite observations (1997–ongoing), Copernicus Marine Environment Monitoring Service, https://doi.org/10.48670/moi-00281, 2023.

Peters, G. P., Le Quéré, C., Andrew, R. M., Canadell, J. G., Friedlingstein, P., Ilyina, T., Jackson, R. B., Joos, F., Korsbakken, J. I., and McKinley, G. A.: Towards real-time verification of CO_2 emissions, Nat. Clim. Change, 7, 848-850,

https://doi.org/10.5194/essd-2026-375
Preprint. Discussion started: 11 June 2026

https://doi.org/10.1038/s41558-017-0013-9, 2017.

Prokhorenkova, L., Gusev, G., Vorobev, A., Dorogush, A. V., and Gulin, A.: CatBoost: unbiased boosting with categorical features, Advances in neural information processing systems, 31, https://doi.org/10.5555/3327757.3327770, 2018.

Reynolds, R. W., Smith, T. M., Liu, C., Chelton, D. B., Casey, K. S., and Schlax, M. G.: Daily high-resolution-blended analyses for sea surface temperature, J. Clim., 20, 5473-5496, https://doi.org/10.1175/2007JCLI1824.1, 2007.

Rödenbeck, C., Bakker, D. C., Metzl, N., Olsen, A., Sabine, C., Cassar, N., Reum, F., Keeling, R. F., and Heimann, M.: Interannual sea–air CO_2 flux variability from an observation-driven ocean mixed-layer scheme, Biogeosciences, 11, 4599-4613, https://doi.org/10.5194/bg-11-4599-2014, 2014.

Rödenbeck, C., Bakker, D. C., Gruber, N., Iida, Y., Jacobson, A. R., Jones, S., Landschützer, P., Metzl, N., Nakaoka, S., and Olsen, A.: Data-based estimates of the ocean carbon sink variability–first results of the Surface Ocean p CO_2 Mapping intercomparison (SOCOM), Biogeosciences, 12, 7251-7278, https://doi.org/10.5194/bg-12-7251-2015, 2015.

Sabine, C. L., Feely, R. A., Gruber, N., Key, R. M., Lee, K., Bullister, J. L., Wanninkhof, R., Wong, C., Wallace, D. W., and Tilbrook, B.: The oceanic sink for anthropogenic CO_2, Science, 305, 367-371, https://doi.org/10.1126/science.1097403, 2004.

Steinberg, D. K., Carlson, C. A., Bates, N. R., Johnson, R. J., Michaels, A. F., and Knap, A. H.: Overview of the US JGOFS Bermuda Atlantic Time-series Study (BATS): a decade-scale look at ocean biology and biogeochemistry, Deep-Sea Res. II: Top. Stud. Oceanogr., 48, 1405-1447, https://doi.org/10.1016/S0967-0645(00)00148-X, 2001.

Sutton, A., Feely, R., Maenner Jones, S., Musielewicz, S., Osborne, J., Dietrich, C., Monacci, N., Cross, J., Bott, R., and Kozyr, A.: Autonomous seawater partial pressure of carbon dioxide (pCO_2) and pH time series from 40 surface buoys between 2004 and 2017 (NCEI Accession 0173932), Version 1.1, NOAA National Centers for Environmental Information, Dataset [dataset], https://doi.org/10.7289/v5db8043, 2018.

Takahashi, T., Olafsson, J., Goddard, J. G., Chipman, D. W., and Sutherland, S.: Seasonal variation of CO_2 and nutrients in the high-latitude surface oceans: A comparative study, Glob. Biogeochem. Cycles, 7, 843-878, https://doi.org/10.1029/93GB02263, 1993.

Takahashi, T., Sutherland, S. C., Sweeney, C., Poisson, A., Metzl, N., Tilbrook, B., Bates, N., Wanninkhof, R., Feely, R. A., and Sabine, C.: Global sea–air CO_2 flux based on climatological surface ocean pCO_2, and seasonal biological and temperature effects, Deep-Sea Res. II: Top. Stud. Oceanogr., 49, 1601-1622, https://doi.org/10.1016/S0967-0645(02)00003-6, 2002.

Wang, Z., Fu, W., Xue, C., and Wang, G.: GEOXYGEN: a global long-term dissolved oxygen dataset based on biogeochemistry-aware machine learning framework and multi-source observations, Earth Syst. Sci. Data, 18, 3125–3146, https://doi.org/10.5194/essd-18-3125-2026, 2026.

Wang, Z., Fu, W.: FDU-BTR: A global monthly gridded pCO_2 product (1982-2024), Zenodo [data set], https://doi.org/10.5281/zenodo.20152530, 2026.

Wanninkhof, R., Trinanes, J., Pierrot, D., Munro, D. R., Sweeney, C., and Fay, A. R.: Trends in sea-air CO_2 fluxes and

https://doi.org/10.5194/essd-2026-375
Preprint. Discussion started: 11 June 2026

sensitivities to atmospheric forcing using an extremely randomized trees machine learning approach, Glob. Biogeochem. Cycles, 39, e2024GB008315, https://doi.org/10.1029/2024GB008315, 2025.

Weiss, R. F.: Carbon dioxide in water and seawater: the solubility of a non-ideal gas, Mar. Chem., 2, 203-215, https://doi.org/10.1016/0304-4203(74)90015-2, 1974.

Wu, H., Ji, Y., Wang, L., Liu, X., Zhou, W., Cui, L., Wang, Y., Liu, M., and Li, Z.: Reconstruction and spatiotemporal analysis of global surface ocean pCO_2 considering sea area characteristics, Biogeosciences, 23, 967-994, https://doi.org/10.5194/bg-23-967-2026, 2026.

Zeng, J., Iida, Y., Matsunaga, T., and Shirai, T.: Surface ocean CO_2 concentration and air-sea flux estimate by machine learning with modelled variable trends, Front. Mar. Sci., 9, 989233, https://doi.org/10.3389/fmars.2022.989233, 2022.

Zeng, J., Nojiri, Y., Landschützer, P., Telszewski, M., and Nakaoka, S.-I.: A global surface ocean fCO_2 climatology based on a feed-forward neural network, Journal of Atmospheric and Oceanic Technology, 31, 1838-1849, https://doi.org/10.1175/JTECH-D-13-00137.1, 2014.

Zhang, X., Liao, E., Lu, W., Wu, Z., Wang, G., Zhu, X., and Liang, S.: A surface ocean pCO_2 product with improved representation of interannual variability using a vision transformer-based model, Earth Syst. Sci. Data, 17, 6071-6095, https://doi.org/10.5194/essd-17-6071-2025, 2025.

Zhong, G., Li, X., Song, J., Qu, B., Wang, F., Wang, Y., Zhang, B., Sun, X., Zhang, W., and Wang, Z.: Reconstruction of global surface ocean pCO_2 using region-specific predictors based on a stepwise FFNN regression algorithm, Biogeosciences, 19, 845-859, https://doi.org/10.5194/bg-19-845-2022, 2022.