Response to Referee #1

General comments:

This manuscript introduces a novel machine learning framework (SJTU-AViT) for reconstructing global sea surface pCO_2 at 1°×1° monthly resolution over the period 1982–2023. By incorporating physical–biogeochemical constraints as derived features, the approach enhances the quality of ocean carbon data reconstruction. The evaluation is comprehensive, covering mean states, seasonal cycles, and interannual variability, and shows strong skill in reproducing ENSO-related signals. This study makes a substantial contribution by providing a valuable new ocean carbon data product for the ocean carbon community and a useful machine learning framework in the field of ocean data reconstruction. The subject is highly relevant to the scope of Earth System Science Data. However, I have several general and specific comments and suggestions that should be addressed before the manuscript can be considered for publication.

We sincerely thank the reviewer for the constructive and insightful comments, which have greatly improved the quality and clarity of our manuscript. The reviewer's main concerns focused on the independence of training and validation data and its impact on the robustness of our results. To address these concerns, we supplemented analyses using independent datasets, confirming the robustness of the assessed interannual variability. We also clarified and expanded methodological details, data processing, training settings, and figures/tables to enhance transparency and reproducibility. Specifically, we have made the following revisions:

- Clarify the train/test split strategy with added spatiotemporal distribution maps, and validate model generalization against independent long-term stations (major comment #1, #5, minor comments #7).
- Address concerns on model–validation dependence by recalculating detrended/deseasoned STD using MPI-SOM-FFN trends and seasonal cycles, confirming robust spatial patterns beyond data dependence (major comment #2).
- Enhance methodological clarity by adding a process flowchart, specifying SOCAT gridded comparison and interpolation procedures, and clarifying variable standardization (major comment #3, #6, minor comments #1).
- Clarify input coverage by filling pre-1997 and 2023 gaps in Chl-a with climatology, using climatological MLD as input data with noted limitations, detailing xCO_2 (MBL) mapping (major comment #7, #8, minor comments #5).
- Implement minor edits and clarifications to enhance precision and consistency throughout the text, as detailed in the point-by-point response (major comment #4, minor comments #1-17).

Please see the detailed response below.

Major comments:

1. The Methods section (Model training & testing) should more clearly describe how the data were split into training and testing sets, along with the sample size

distribution. This information is essential for evaluating the model's generalization ability. The authors should specify whether the split was random, temporal, or spatial (e.g., by cruise lines or fixed stations). They should also report the number or proportion of samples in each subset, ideally stratified by time (e.g., decades) and/or region. Such details would improve transparency and reproducibility.

We have clarified the data partitioning procedure in the section 2.3 and have included additional analyses to support transparency and robustness of the evaluation.

Specifically, SOCAT samples were randomly split into training and test subsets, with 80% (277,528 samples) allocated for model training and 20% (69,142 samples) reserved as an independent test set. All random operations were conducted using a fixed seed (seed = 42) to ensure full reproducibility. The detailed split procedure and exact sample counts are now explicitly documented in the revised text (new Fig. 2 and lines 208-209), now as *"The SOCAT dataset was randomly divided into 80% (277,528 samples) for training and 20% (69,142 samples) for validation, using a fixed random seed (seed = 42) to ensure reproducibility."*.

The test procedure is included in the new reconstruction workflow (Fig. 2) and described in section S5 of the supplement (see explanation in our response to comment 3), and Figure S1 has also been revised to clearly illustrate the temporal and spatial distributions of the training and test sets. In addition, we additionally evaluated its performance against nine independent long-term observation stations that are not included in the SOCAT dataset. These stations provide continuous time series and serve as an independent benchmark. The results indicate that the model reliably captures both temporal variability and long-term trends, providing strong evidence of its generalization capability beyond the original training data.

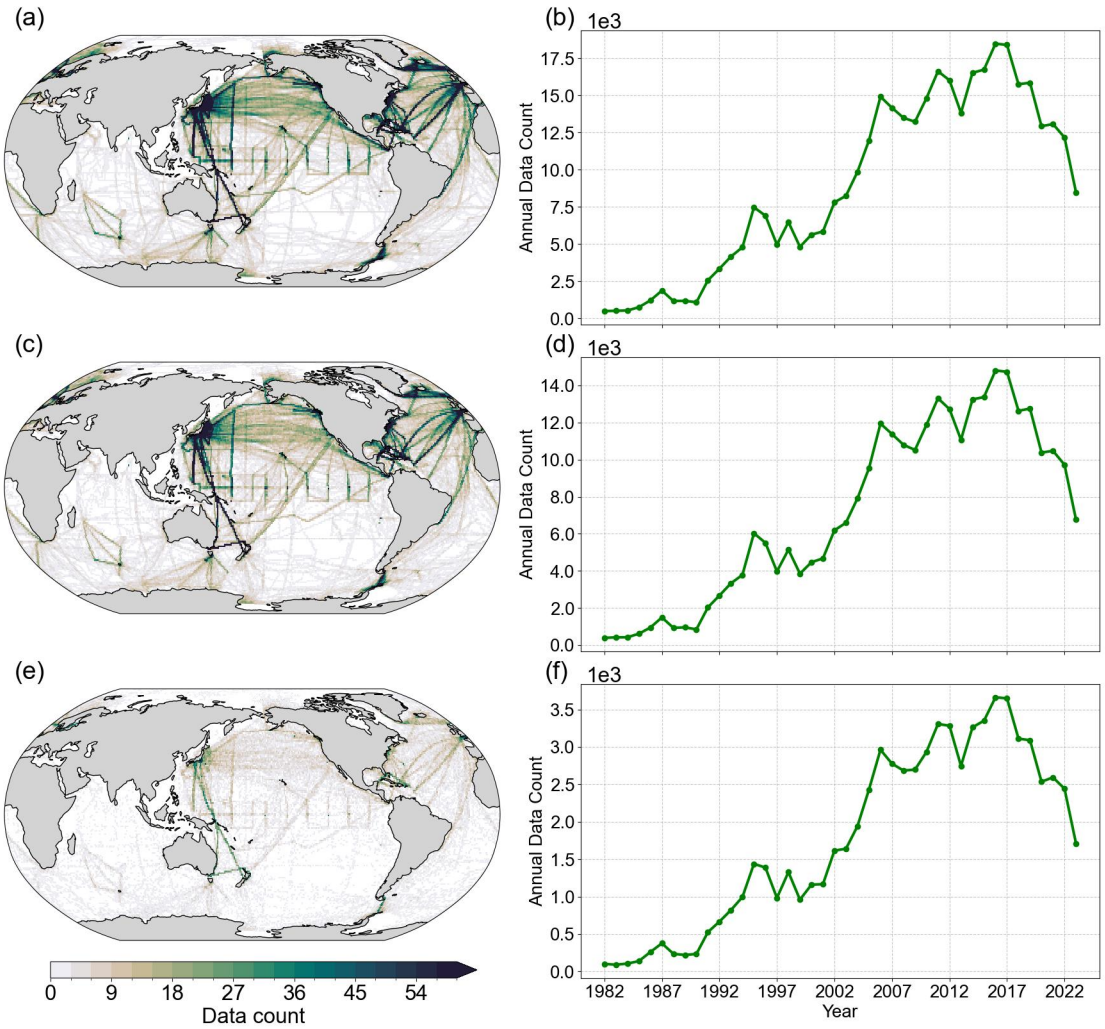

Figure R1. (Figure S1 in supplement section S4). Data availability for $spCO_2$ reconstruction. (a) Spatial distribution of the number of all $spCO_2$ data points. (b) Annual data count of all $spCO_2$ data points over the period from 1982 to 2023. (c) Spatial distribution of the number of $spCO_2$ data points used for training. (d) Annual data count used for training over the period from 1982 to 2023. (e) Spatial distribution of the number of $spCO_2$ data points used for validation. (f) Annual data count used for validation over the period from 1982 to 2023.

2. The manuscript fills gaps in SOCAT observations using long-term trends and seasonal cycles from SJTU-AViT, followed by residual analysis to assess interannual variability. This procedure raises concerns about the lack of independence between the model and the validation data, since part of the evaluation relies on model-derived estimates. The authors should clarify and quantify the impact of this approach. For instance, they could limit the analysis to grid points or stations with continuous records, or apply long-term trends and seasonal cycles from an independent product to compare robustness. Demonstrating consistent results across methods would enhance the credibility of the conclusions.

To address the concern regarding the potential lack of independence between the model and the validation data, we conducted an additional analysis using an

independent reconstructed data product from MPI-SOM-FFN (Landschützer et al. 2016). This validation results were provided in the supplement material (section S5.8). Specifically, when calculating the detrended and deseasonalized SOCAT STD, we applied the long-term trends and seasonal cycles derived from the MPI-SOM-FFN data product instead of the SJTU-AViT estimates. The results, shown in the Fig. R2, demonstrate that the overall spatial distribution of SOCAT STD remains highly consistent, with only minimal deviations (1.68 μatm). This indicates that the small deviations observed between SJTU-AViT and SOCAT are not artifacts of model-data dependence. Therefore, the analysis confirms the robustness of our methodology and supports the credibility of the interannual variability assessment.

.

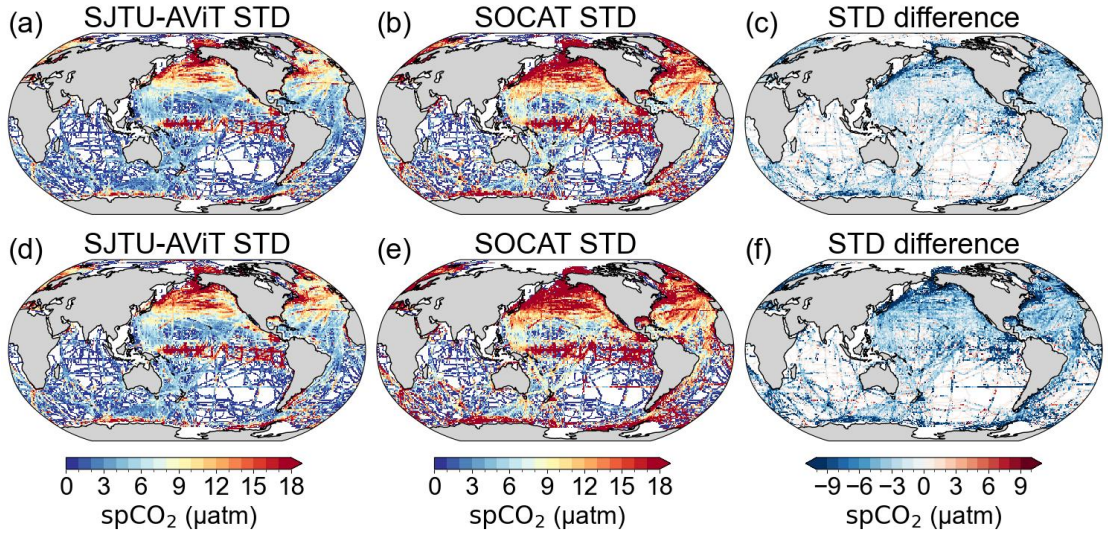

Figure R2. (Figure S15 in supplement section S5). Comparison of $spCO_2$ standard deviations on timescales longer than one year between SJTU-AViT, SOCAT, and MPI-SOM-FFN data product. (a) Standard deviation of $spCO_2$ from the SJTU-AViT at SOCAT observation grid points. (b) Standard deviation of $spCO_2$ from SOCAT data (the long-term trends and seasonal cycles derived from the SJTU-AViT). (c) Standard deviation bias between SJTU-AViT and SOCAT (panel a minus panel b). (d) Standard deviation of $spCO_2$ from the SJTU-AViT at SOCAT observation grid points. (e) Standard deviation of $spCO_2$ from SOCAT data (the long-term trends and seasonal cycles derived from the MPI-SOM-FFN). (f) Standard deviation bias between SJTU-AViT and SOCAT (panel d minus panel e).

3. To help readers better understand the implementation of the $spCO_2$ data reconstruction, I recommend adding a schematic figure in the main text that illustrates the reconstruction process based on the ViT model. Such a figure would improve both the readability of the manuscript and the clarity of the methodology.

We have added a schematic framework figure (Fig. 2) in the section 2.2, along with detailed description in the main text (lines 186-188) and supplement (section S5.1, lines 128-140), to illustrate the $spCO_2$ reconstruction workflow based on the ViT framework.

The summary description is presented in the main text (lines 186-188) as *"The*

overall workflow of this multi-stage training strategy is summarized in Fig. 2, which also provides a schematic overview of the $spCO_2$ reconstruction workflow based on the ViT framework. The figure clearly visualizes the main steps, from data preprocessing through model training to evaluation (see detailed description in section S5.1).".

The detailed description is presented in the supplement (section S5, lines 128-140) as *"The $spCO_2$ reconstruction workflow based on the ViT framework is organized into four main stages—Data Processing, Model Architecture, Training & Validation, and Evaluation & Analysis—as illustrated in Fig. 2 (in main text). At the top, the data processing panel shows the input sources (CMIP6, MOM6, SOCAT) and the preprocessing steps: temporal harmonization to a monthly cadence, spatial regridding to a 1°×1° grid, and feature normalization. These boxes indicate that all inputs are brought to a common spatio-temporal grid and scale before being passed to the model. The model architecture panel depicts how physical variables are converted into model inputs: variable tokenization, variable aggregation, and then fed into a Transformer backbone that learns spatial and temporal dependencies. The model output block illustrates that the network predicts monthly $spCO_2$ on the same 1° grid. The training & validation panel summarizes our multi-stage training strategy: (i) pretraining on CMIP6-derived fields, (ii) fine-tuning using MOM6 plus 80% of SOCAT, and (iii) evaluation using a withheld 20% SOCAT validation split and independent tests at long-term station sites. Finally, the evaluation & analysis panel shows the main evaluation products derived from the reconstruction: model performance metrics, climatology, seasonal cycle, interannual variability, and downstream analyses (air-sea CO_2 flux calculation and uncertainty analysis).".*

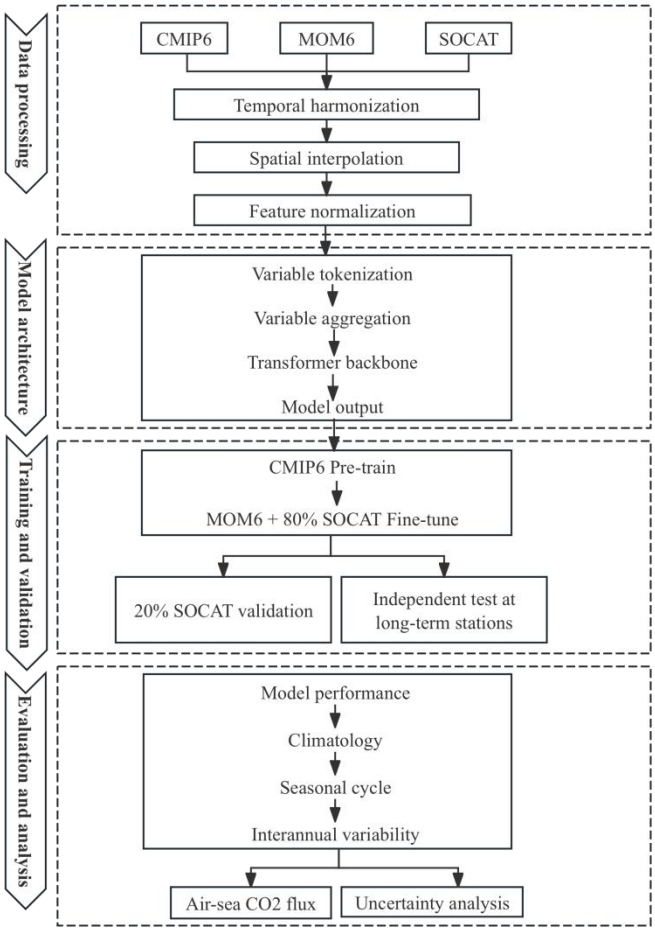

Figure R3. (Figure 2 in main text). Workflow of the $spCO_2$ reconstruction using the ViT-based framework. The workflow consists of four major stages: (a) Data processing, where CMIP6, MOM6, and SOCAT inputs are temporally harmonized, spatially interpolated, and normalized; (b) Model architecture, where variables are tokenized, aggregated into spatio-temporal embeddings, and processed by a Transformer backbone to predict monthly $spCO_2$; (c) Training and validation, involving CMIP6 pretraining, MOM6 and SOCAT fine-tuning, and evaluation against withheld SOCAT data and long-term stations; and (d) Evaluation and analysis, where model performance metrics, climatology, seasonal cycles, and interannual variability are assessed, leading to downstream analyses such as air-sea CO_2 flux estimation and uncertainty analysis (see detailed description in supplement section S5.1).

4. It is recommended that the authors include skill distribution tables in the supplement, stratified by ocean basin and latitude band. These tables should report, for each group, the sample size (N), R^2, RMSE, MAE, and MBE. Such quantitative evidence would support the statement that "biases are larger at high latitudes" and clearly demonstrate regional and latitudinal variations in model performance.

We have added skill distribution tables in the supplement (Table S4-S5), stratified by ocean basin and latitude band. For each group, we report the sample size (N), R^2, RMSE, MAE, and MBE. These tables provide quantitative evidence supporting the

statement that biases are larger at high latitudes and clearly illustrate the regional and latitudinal variations in model performance, as followed.

Based on the statistics, the skill distribution reveals distinct regional and latitudinal differences. Among ocean basins, the Pacific shows the best performance (N=159,783; R^2=0.94), while the Indian Ocean, despite its smaller sample size (N=6,354), also exhibits strong skill (R^2=0.95; RMSE=5.31). In contrast, the Atlantic performs relatively weaker with a lower correlation and a slight negative bias (N=111,326; R^2=0.81). The Arctic (N=10,316; RMSE=8.80) and Southern Ocean (N=48,636; RMSE=8.20) show notably larger errors and systematic negative biases, indicating a tendency of underestimation in polar regions. When stratified by latitude bands, errors are markedly larger at high latitudes, particularly in 60°S-90°S (N=16,602; R^2=0.86) and 60°N-90°N (N=30,802; R^2=0.92). By comparison, the tropics and subtropics exhibit smaller errors, such as 0–30°S (N=35,804; R^2=0.97). The 0-30°N band shows moderate error levels (RMSE=6.13) but a lower correlation (R^2=0.72), likely reflecting observational variance and sample characteristics. Overall, these quantitative results directly support our conclusion that biases are more pronounced at high latitudes. As discussed in the main text, this pattern can be attributed to the complexity of seasonal amplitudes and boundary processes (e.g., sea-ice cover and mixed layer variability), the limited representativeness and accuracy of forcing fields and input data in polar regions, and uneven observational coverage, all of which can amplify errors and biases.

Table R1 (Table S4 in supplement section S3). Skill metrics of the reconstructed $spCO_2$ by ocean basin.

Ocean basin	N	R^2	RMSE	MAE	MBE
Pacific ocean	159783	0.94	6.79	5.29	0.30
Atlantic ocean	111326	0.81	7.10	5.31	-0.31
Indian ocean	6354	0.95	5.31	4.75	-0.08
Arctic ocean	10316	0.90	8.80	7.58	-0.24
Southern ocean	48636	0.88	8.20	6.76	-0.55

Table R2 (Table S5 in supplement section S3). Skill metrics of the reconstructed $spCO_2$ by latitude band.

latitude band	N	R^2	RMSE	MAE	MBE
60°N-90°N	30802	0.92	9.23	7.58	-0.56
30°N-60°N	123357	0.91	9.13	6.40	0.07
0-30°N	96608	0.72	6.13	4.74	0.04
0-30°S	35804	0.97	5.70	4.96	-0.07
30°S-60°S	43497	0.90	6.13	5.29	-0.20
60°S-90°S	16602	0.86	11.80	9.29	-1.03

5. Regarding the independent test sites, it is recommended to provide a clear description in the main text along with detailed information. In the appendix, the nine observation stations used for independent testing should be listed, including their

names, geographic locations, observation periods, and the number of samples at each site. Since the BAT site does not have direct pCO_2 observations, please clarify the method used to calculate its monthly mean pCO_2 and specify the data sources for all sites.

(1) In the revised manuscript, we have added an appendix table and provided a clearer description of the independent test sites (section 2.3 (lines 215-220), and Table S3). We explicitly detail the nine long-term observation stations used for independent testing, including their names, geographic locations, observation periods, number of samples, and data sources. To facilitate visual interpretation, we have also included a map (supplement Fig. S2) showing the locations of the stations.

Lines 215-220 now reads as *"For the independent test at long-term stations, reconstructed values were extracted at the corresponding station locations using bilinear spatial interpolation, which incorporates information from surrounding grid cells to provide smoother and more representative estimates, and skill metrics were subsequently computed to evaluate model performance. Detailed information for these stations, including their names, geographic locations, observation periods, number of samples, and data sources, is provided in supplement Table S3, and their locations are shown in supplement Fig. S2 to facilitate visual interpretation."*.

(2) For the BAT station, which does not have direct pCO_2 observations, the monthly mean pCO_2 was calculated using the Python version of CO2SYS (PyCO2SYS; Humphreys et al., 2022). In the calculation, we used the carbonate dissociation constants from Waters and Millero (2013) (k1k2 = 15), the KSO_4 constant from Dickson (1990) (kso4 = 1), the HF dissociation constant from Perez and Fraga (1987) (hf = 2), and the total pH scale (pHscale = 1). pCO_2 was then estimated from measurements of dissolved inorganic carbon (DIC) and total alkalinity (ALK), together with sea surface temperature (SST), sea surface salinity (SSS), silicate, and phosphate concentrations. Monthly means were obtained by averaging all available estimates within each month. This approach ensures a consistent and physically based estimation of pCO_2 at the BAT site.

Table R3 (Table S3 in supplement section S3). List of selected independent test stations with long-term observations.

Station	Coordinates	Time range	Number of samples	URL
BAT	31.67°N, 295.83°E	10/1991-6/2022	324	https://bios.asu.edu/bats
HOT	22.75°N, 202°E	10/1988-12/2023	325	https://hahana.soest.hawaii.edu/hot/hotco2
ESTOC	29.07°N, 344.17°E	10/1995-11/2009	115	https://www.ncei.noaa.gov/access/ocean-carbon-acidification-data-system/oceans/Coastal/ESTOC.html
CCE1	33.50°N, 237.50°E	11/2008-12/2023	144	https://www.ncei.noaa.gov/access/ocean-carbon-acidification-data-system/oceans/Moorings/Pacific.html
TAO	-0.51°N, 189.98°E	2/2010-8/2016	45	https://www.ncei.noaa.gov/access/ocean-carbon-acidification-data-system/oceans/Moori

				ngs/Pacific.html
BOBOA	15°N, 90°E	11/2013-11/2018	53	https://www.ncei.noaa.gov/access/ocean-carbon-acidification-data-system/oceans/Moorings/Indian.html
Papa	50.13°N, 215.17°E	6/2007-4/2023	168	https://www.pmel.noaa.gov/co2/story/Papa
Iceland	68°N, 347.40°E	2/1985-11/2021	158	https://www.ncei.noaa.gov/access/ocean-carbon-acidification-data-system/oceans/Moorings/Atlantic.html
Irminger	64.30°N, 332°E	3/1983-11/2012	99	https://www.ncei.noaa.gov/access/ocean-carbon-acidification-data-system/oceans/Moorings/Atlantic.html

6. The manuscript states that SJTU-AViT outputs were interpolated to the spatiotemporal locations of SOCAT for comparison, but it does not specify the interpolation method used (e.g., bilinear, nearest neighbor, or other), nor whether any temporal or spatial smoothing was applied. The authors should provide these details. For example, "When comparing with SOCAT, model values were interpolated to observation locations using bilinear interpolation in space and linear interpolation in time."

The Surface Ocean CO_2 Atlas (SOCAT) provides two forms of data: synthesis files and gridded (binned) products. The gridded SOCAT product is generated by interpolating individual observations onto a regular grid with a spatial resolution of 1° × 1° and a monthly temporal resolution. Only SOCAT observations with quality control (QC) flags of A–D and WOCE flags of 2 are included in this product. The arithmetic mean is first calculated for each cruise passing through a given grid cell, and these cruise-level means are then averaged to obtain the final gridded value. The resulting product provides fields with valid values in grid cells and months where observations are available, while grid cells without observational coverage are assigned NaN values.

We used the gridded SOCAT product, which share the same longitude and latitude grid as SJTU-AViT data. Therefore, there is no need for additional spatial or temporal interpolation. To account for gaps in the SOCAT data (NaNs), we mask the corresponding reconstructed values at the same grid–time points before computing any statistics. This ensures that all comparisons are performed only where SOCAT provides valid data. For the independent test at long-term stations, we use bilinear spatial interpolation to extract the reconstructed values at the corresponding station locations. This approach allows us to account for the surrounding grid cell information rather than relying solely on the nearest neighbor, thereby providing a smoother and more representative estimate of $spCO_2$ at the station sites.

To make this clearer, we have revised the section 2.3 (lines 212-220) to explicitly state our comparison procedure. The added text reads :

"For comparison with SOCAT, we used the monthly 1° gridded SOCAT product and evaluated our SJTU-AViT reconstruction on the same grid, without applying any additional spatial interpolation. Reconstructed values were masked where SOCAT is

missing, and all skill metrics were computed only at grid-time points with valid SOCAT data. For the independent test at long-term stations, reconstructed values were extracted at the corresponding station locations using bilinear spatial interpolation, which incorporates information from surrounding grid cells to provide smoother and more representative estimates, and skill metrics were subsequently computed to evaluate model performance. Detailed information for these stations, including their names, geographic locations, observation periods, number of samples, and data sources, is provided in supplement Table S3, and their locations are shown in supplement Fig. S2 to facilitate visual interpretation." .

7. The temporal coverage of Chl-a spans 1997–2022, whereas the product extends from 1982 to 2023. It is recommended to clarify how the periods prior to 1997 and for 2023 were handled (e.g., climatology, interpolation, gap-filling, or inference from other variables) to avoid any misunderstanding that the time spans are fully consistent.

The temporal coverage of the Chl-a dataset spans 1997-2022. For periods prior to 1997 and for 2023, we applied a climatology derived from the 1997-2022 record. We recognize that the use of a climatological mean does not capture interannual variability and may introduce a slight bias into the reconstruction. However, this effect is expected to be minor, while inferring Chl-a from other variables or extrapolating beyond the observational record could introduce substantially larger uncertainties. The use of a climatology therefore represents a pragmatic balance between competing sources of uncertainty, ensuring a stable and physically reasonable baseline. A similar approach has been adopted in other surface ocean pCO_2 reconstruction efforts, including Landschützer et al. (2016) and Gregor et al., (2021).

We have clarified this methodological choice in the revised manuscript (section 2.1, lines 115-118) to avoid any misunderstanding regarding the temporal consistency of the input variables, now as *"Chl-a data were derived from the European Space Agency Climate Change Initiative (ESA CCI) Ocean Colour (version 5.0) dataset, spanning 1997 to 2022 with daily resolution and a spatial resolution of 4 km (Jackson et al., 2017). For periods prior to 1997 and for 2023, we employed a climatology computed from the 1997-2022 Chl-a record to ensure full temporal coverage."*.

8. The manuscript employs a 2° monthly climatological MLD (WOCE). It is recommended to explain why a climatological mean was used instead of incorporating interannual and monthly variability, as this choice may affect the representation of temporal dynamics.

One of the most widely used and high-quality MLD dataset is the 2° global climatological MLD product by de Boyer et al. (2004), based on global observed temperature and salinity profiles. It accurately represents the climatological mean and provides robust physical constraints for long-term, large-scale $spCO_2$ reconstruction. During 1982-2023, high-resolution, continuous interannual MLD data are lacking globally. We acknowledge that climatological MLD does not capture interannual

variability, which may slightly underestimate $spCO_2$ interannual variability. However, introducing time-varying MLD from numerical ocean models could potentially introduce additional uncertainties, potentially larger than the bias introduced by using a climatology. Consequently, the use of a climatological MLD represents a pragmatic balance between different sources of uncertainty, ensuring a stable and reliable baseline for identifying large-scale, long-term patterns. A similar choice has been adopted in other ocean pCO_2 reconstruction, including Landschützer et al., (2016) and Gregor et al., (2021).

This limitation has been explicitly stated in the revised manuscript (section 4, lines 638-642), now as *"It should be noted that the climatological MLD used in this study cannot capture interannual or monthly variability, which may slightly underestimate local or short-term impacts on $spCO_2$. Nevertheless, it provides adequate physical constraints for reconstructing long-term and large-scale spatiotemporal patterns. Future work will explore incorporating high-quality time-varying MLD data as it becomes available to improve model fidelity at regional and seasonal scales."*.

Minor comments:

1. Please clarify whether the input data were standardized during the model training process, and specify the method used (e.g., variable-wise mean–variance normalization, min–max scaling, or other approaches).

During the model training process, the input data were standardized using variable-wise mean–variance normalization. Detailed descriptions of this procedure have been added in section 2.2 (lines 153-154) of the manuscript, now as *"These input variables are standardized using variable-wise mean-variance normalization and formatted into a multi-channel input to ensure feature extraction occurs on a unified scale."*.

2. Please clearly indicate the flux sign convention in the caption of Figure 13, for example: "Negative = ocean uptake (sink), Positive = release to the atmosphere (source)." Ensure that this convention is consistent with the main text, equations, and color bar.

Done as suggested. We have updated the figure caption (now is Fig. 14) to clearly indicate the flux sign convention (lines 576-577): *"Negative = ocean uptake (sink), Positive = release to the atmosphere (source)."* In addition, we have thoroughly checked the main text, relevant equations, and color bar to ensure that this convention is applied consistently throughout the manuscript.

3. In supplement Figure S6, the legend is labeled as "spCO2" which should be "$spCO_2$". Please ensure consistency of the symbol and formatting throughout the manuscript (e.g., uniformly using the subscript form "$spCO_2$" instead of "spCO2") and apply the same convention across all figures, captions, and text. In addition, please indicate the appropriate units (e.g., μatm) where relevant to avoid confusion.

Done as suggested. We have corrected the legend in supplement Figure S6 from "spCO2" to "$spCO_2$", as followed. In addition, we have carefully reviewed the entire

manuscript to ensure that the subscript formatting for "CO_2" is used consistently across all figures, captions, and text. We have also added the appropriate units (e.g., μatm) where relevant to avoid any potential confusion.

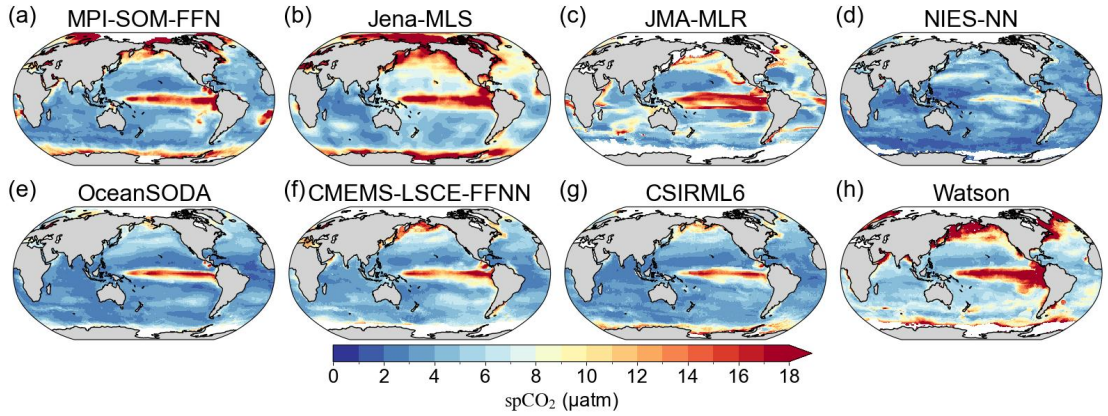

Figure R4. (Figure S6 in supplement section S4). Spatial distribution of standard deviation in interannual time scale of reconstructed $spCO_2$ at multiple data products from 1985 to 2018. All the panels show the standard deviation of residuals after removing long-term trends and seasonal cycles. The color scale represents the magnitude of variability in $spCO_2$, with higher values (red) indicating greater variability.

4. Line 113 — abbreviation usage: The term "Sea surface salinity (SSS)" repeats a definition already given earlier. Please use the abbreviation SSS here. A full-text check is recommended to correct similar inconsistencies.

Done as suggested. At the revised text (now is line 114), we have replaced "Sea surface salinity (SSS)" with the abbreviation SSS to avoid redundant definitions. In addition, we conducted a comprehensive review of similar cases throughout the manuscript to ensure that abbreviations are fully defined upon first appearance and consistently used thereafter.

5. Regarding xCO_2 (MBL), please clarify how the meridional band product was mapped onto the 1° × 1° grid (e.g., through band replication, interpolation, or another approach). Providing this detail would improve the transparency of the data processing procedure.

In our study, the meridional band xCO_2 (MBL) product, which is provided at discrete latitude bands, was mapped onto the 1° × 1° global grid using a two-step procedure:

(1) Latitudinal Interpolation: For each time step, the original xCO_2 values at discrete latitude bands were interpolated to the model's target latitudes using one-dimensional linear interpolation along the meridional direction (implemented with MATLAB's interp1 function). This ensures a smooth transition of xCO_2 values between the original latitude bands.

(2) Longitudinal Replication: Because the original xCO_2 product does not contain longitudinal variations, the interpolated latitudinal profile was replicated along all longitudes to produce a complete 2D global field at 1° × 1° resolution. This approach

preserves the meridional gradient while assuming longitudinal uniformity, consistent with the original dataset.

This mapping procedure was applied to all time steps from 1982 to 2023. These details have now been explicitly added to the section 2.1 (lines 122-124) to improve the transparency and reproducibility of the data processing, now as *"In this study, the meridional band product was mapped onto the model's 1° × 1° global grid using latitudinal interpolation and longitudinal replication, generating continuous 2D fields suitable for model simulations."*

6. In the Methods section, please specify the training setup, including the maximum number of epochs and/or the early stopping patience (e.g., "trained for up to 200 epochs with early stopping, patience = 20"), to improve the reproducibility of the approach.

In our study, the ViT-based model was trained for a maximum of 200 epochs with early stopping applied, using a patience of 10 epochs. This means that training would terminate if the validation loss did not improve for 10 consecutive epochs. Each training epoch required roughly 10 minutes. These details have now been explicitly added to section 2.2 (lines 176-178) to improve the clarity and reproducibility of our approach, now as *"The ViT-based model contains approximately 115 million parameters and was trained in parallel on eight NVIDIA RTX 4090 GPUs for up to 200 epochs with early stopping (patience = 10); each training epoch required roughly 10 minutes."*

7. It is recommended to indicate the sample size for each data point or category in Figure 3, allowing readers to more clearly understand the data coverage and the reliability of the statistics.

Following the comment, we have indicated the sample size for each data station in revised figure (now is Fig. 4). The addition is shown in the figure below.

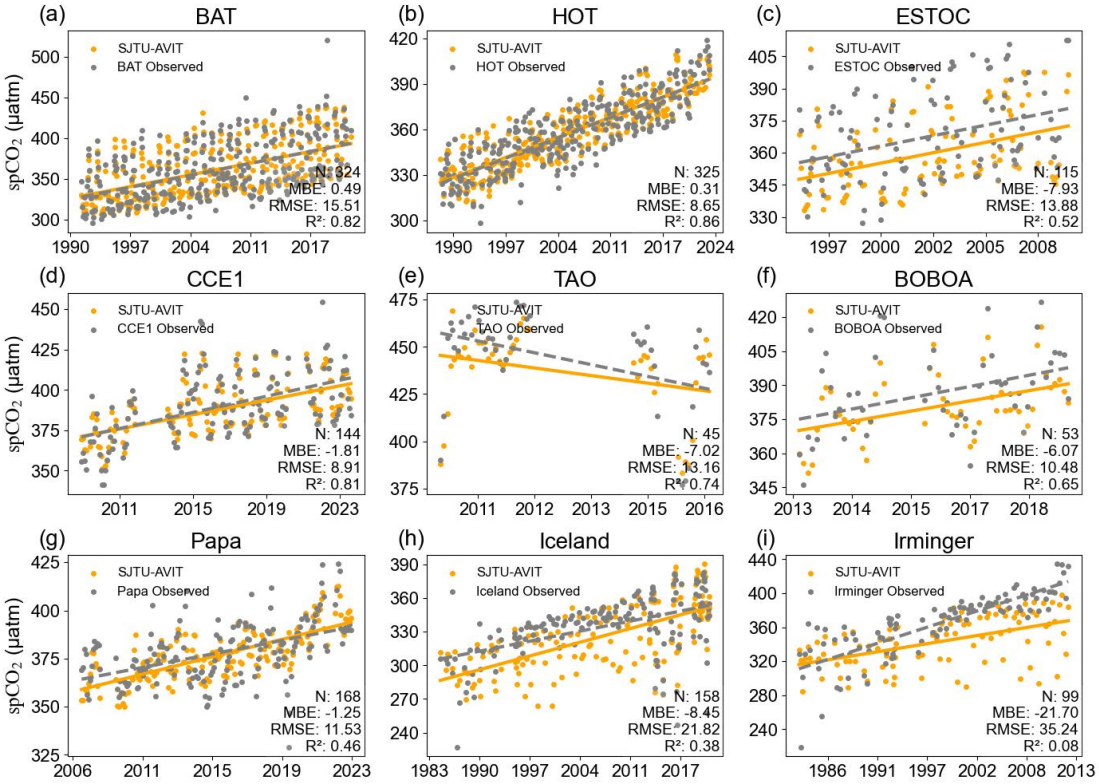

Figure R5. (Figure 4 in main text). Independent test of $spCO_2$ variability between SJTU-AViT and in situ observations at different stations. These in situ data are independent data and are not used to train the model. The station description and location refer to supplement section S2 and Fig. S2. The $spCO_2$ in SJTU-AViT is interpolated to match the station locations and times in the comparison. For each panel, the number of samples (N), the mean bias error (MBE), root mean square error (RMSE), and correlation coefficient (R^2) between the reconstructed and observed $spCO_2$ are displayed. The dashed and solid lines show the linear trend of SJTU-AViT and in situ data.

8. It is recommended to review the entire manuscript and ensure that all instances of "CO2" use a subscript for the number 2, maintaining consistency and adhering to scientific writing conventions.

Done as suggested. We have carefully reviewed the entire manuscript and ensured that all instances of "CO2" now use a subscript for the number 2, maintaining consistency throughout the text and adhering to standard scientific writing conventions.

9. Line 31: In the abstract, change "This study not only provide…" to "This study not only provides…". It is recommended to review the entire manuscript for program errors.

Done as suggested. We have corrected the sentence in the abstract (now is line 32) to *"This study not only provides…"*. In addition, we have carefully reviewed the entire manuscript to identify and correct similar grammatical errors, ensuring the accuracy

and readability of the text.

10. Line 31: In the abstract, change “Earth system” to “Earth-system” when used as a compound adjective for clarity

Done as suggested. We have revised the abstract (now is line 34) to change “Earth system” to “Earth-system” when used as a compound adjective.

11. Line 170: It is recommended to revise the sentence to: “The ViT-based model contains approximately 115 million parameters and was trained in parallel on eight NVIDIA RTX 4090 GPUs; each training epoch required roughly 10 minutes.”

Done as suggested. Incorporating the suggestion from minor comment 6, we have revised the sentence (now is lines 176-178) to: *“The ViT-based model contains approximately 115 million parameters and was trained in parallel on eight NVIDIA RTX 4090 GPUs for up to 200 epochs with early stopping (patience = 10); each training epoch required roughly 10 minutes.”*

12. Line 260: It is recommended to revise the sentence to: “Most predicted values lie close to the 1:1 line, particularly within the climatologically common $spCO_2$ range (300–420 µatm), as indicated by the high-density regions in Fig. 2.”

Done as suggested. We have revised the sentence (now is lines 287-289) to: *“Most predicted values lie close to the 1:1 line, particularly within the climatologically common $spCO_2$ range (300-420 µatm), as indicated by the high-density regions in Fig. 3.”*

13. Line 312: Ensure there is a space before “µatm,” e.g., “-12 µatm to +10 µatm.”

Done as suggested. We have corrected the formatting issue by adding a space before “µatm” (e.g., “-12 µatm to +10 µatm”) in the revised text (now is line 340). We have also carefully checked the entire manuscript to ensure consistent formatting of units throughout.

14. It is recommended to standardize the number of decimal places throughout the manuscript (e.g., consistently using two or three decimal places).

We have carefully checked all numerical values reported in the manuscript and standardized the format to retain two decimal places throughout. We note, however, that for prescribed data such as Niño 3.4 and the scaling factor (set as 0.251), no modification was made.

15. For the air–sea flux calculation, the parameterization of Wanninkhof (2014) requires the Schmidt number, wind speed source, and resolution (you used ERA5). It is recommended to specify in section 2.4 the temporal and spatial resolution of ERA5 and the formula or reference used for computing the Schmidt number.

Done as suggested. In our study, the air-sea CO_2 flux was calculated using the parameterization of Wanninkhof (2014). The wind speed data were sourced from the ERA5 reanalysis, with a 6-hourly temporal resolution covering 1982-2023 and a

horizontal spatial resolution of 1°.

The gas transfer velocity of CO_2 (k_w) is computed using a quadratic dependence on wind speed:

$$k_w = 0.251 \cdot (Sc/660)^{-\frac{1}{2}} \cdot u^2$$

where u is the wind speed at 10 m above the sea surface, and Sc is the Schmidt number of CO_2 in seawater. The Schmidt number is calculated from the temperature-dependent empirical formula:

$$Sc = 2116.8 - 136.25 \cdot T + 4.7353 \cdot T^2 - 0.092307 \cdot T^3 + 0.0007555 \cdot T^4$$

where T is the sea surface temperature in °C. This formulation accounts for the effect of sea surface temperature on CO_2 diffusivity in seawater.

These details, including the information on ERA5 data and the reference used for computing the Schmidt number, have been added to section 2.4 (lines 257-259) of the manuscript. They are now described as: *"The Schmidt number (Sc) required in this formulation is calculated following the temperature-dependent empirical formula provided by Wanninkhof (2014). The wind speed data is sourced from ERA5, with a 6-hourly temporal resolution spanning 1982-2023 and a 1° spatial resolution."*

16. Table 1: It is recommended to change "12 month" to "12-month."

Done as suggested. This correction has been applied in Table 1 and carefully reviewed throughout the manuscript to ensure consistency.

17. In the Abstract, it is stated that the model shows a correlation of 0.81 with the Niño 3.4 index, whereas section 3.4 reports a correlation of –0.81. This inconsistency in the sign of the correlation may confuse readers. Please verify the original calculation and ensure that the values and their signs are reported consistently throughout the manuscript.

Done as suggested. The value reported in the Abstract (0.81) was a typo and the correct correlation coefficient is -0.81. This has been corrected in the revised manuscript (Abstract, line 29).

References not in manuscript:

Dickson, A. G.: Standard potential of the reaction: , and the standard acidity constant of the ion HSO in synthetic sea water from 273.15 to 318.15 K, J. Chem. Thermodyn., 22, 113-127, https://doi.org/10.1016/0021-9614(90)90074-Z, 1990.

Humphreys, M. P., Lewis, E. R., Sharp, J. D., and Pierrot, D.: PyCO2SYS v1.8: marine carbonate system calculations in Python, Geosci. Model Dev., 15, 15-43, https://doi.org/10.5194/gmd-15-15-2022, 2022.

Waters, J. F. and Millero, F. J.: The free proton concentration scale for seawater pH, Mar. Chem., 149, 8-22, https://doi.org/10.1016/j.marchem.2012.11.003, 2013.

Perez, F. F. and Fraga, F.: Association constant of fluoride and hydrogen ions in seawater, Mar. Chem., 21, 161-168,

https://doi.org/10.1016/0304-4203(87)90036-3, 1987.