Spatial reconstruction of long-term (2003-2020) sea surface pCO_2 in the South China Sea using a machine learning based regression method aided by empirical orthogonal function analysis

Zhixuan Wang[1], Guizhi Wang[1,2], Xianghui Guo[1], Yan Bai[3], Yi Xu[1] and Minhan Dai[1,*]

[1]State Key Laboratory of Marine Environmental Science and College of Ocean and Earth Sciences, Xiamen University, Xiamen, 361102, China

[2]Fujian Provincial Key Laboratory for Coastal Ecology and Environmental Studies, Xiamen University, Xiamen, 361102, China

[3]State Key Laboratory of Satellite Ocean Environment Dynamics, Second Institute of Oceanography, State Oceanic Administration, Hangzhou, 310012, China

Correspondence to: Minhan Dai (mdai@xmu.edu.cn)

Abstract. The South China Sea (SCS) is the largest marginal sea in the North Pacific Ocean, where intensive field observations including mappings of sea-surface partial pressure of CO_2 (pCO_2) have been conducted over the last two decades. It is one of the most studied marginal seas in terms of carbon cycling and could thus be a model system for marginal sea carbon research. However, the datasets of cruise-based sea surface pCO_2 are still temporally and spatially sparse. Using a machine learning-based method facilitated by empirical orthogonal function (EOF) analysis capable of constraining the spatiality, this study provides a reconstructed dataset of the monthly sea surface pCO_2 in the SCS with a reasonably high spatial resolution (0.05º×0.05º) and temporal coverage between 2003 and 2020. The input data in our reconstructed model include remote sensing-derived sea surface salinity, sea surface temperature, and chlorophyll, the spatial pattern of pCO_2 constrained by EOF, atmospheric pCO_2, and time labels (month). We validated our reconstruction with three independent testing datasets that are not involved in the model training. Among them, Test 1 includes 10% of our in situ data, Test 2 contains four independent underway datasets corresponding to four seasons, and Test 3 is an in situ monthly dataset available from 2003–2019 at the South East Asia Time-Series (SEATs) station located in the northern basin of the SCS. Our Test 1 validation demonstrated that the reconstructed pCO_2 field successfully simulated the spatial and temporal patterns of sea surface pCO_2. The root-mean-square error (RMSE) between our reconstructed data and in situ data in Test 1 averaged to ~10 μatm, which is much smaller (by ~50%) than that between the remote sensing-derived data and in situ data. Test 2 verified the accuracy of our retrieval algorithm in data months lacking observations, showing a relatively small bias (RMSE: ~8 μatm). Test 3 tested the accuracy of the reconstructed long-term trend, showing that at the SEATs Station, the difference between the reconstructed pCO_2 and in situ data ranged from -10 to 4 μatm (-2.5% to 1%). In addition to the typical machine learning performance metrics, we assessed the uncertainty resulting from the bias of the

删除[作者]: ocean
删除[作者]: of
删除[作者]: and constraints of spatiotemporal modes of
删除[作者]: The SCS
删除[作者]: and
删除[作者]: .
删除[作者]: mapping sea surface pCO_2 of this region is
删除[作者]: which
删除[作者]: The South China Sea (SCS) is the largest
删除[作者]: incomplete
删除[作者]: and selecting the remote sensing derived d
删除[作者]: O
删除[作者]: ion was initiated by using
删除[作者]: s
删除[作者]: the
删除[作者]:
删除[作者]: data (, which include
删除[作者]: and
删除[作者]: s
删除[作者]: calculated
删除[作者]:)
删除[作者]: as inputs data
删除[作者]: ,
删除[作者]: (it indicates that this part of the datawhich
删除[作者]: is completely
删除[作者]: un
删除[作者]:)
删除[作者]: where,
删除[作者]: EST
删除[作者]: .
删除[作者]: observed
删除[作者]: EST.
删除[作者]: includes
删除[作者]: EST.

reconstruction and its sensitivity to the features. These validations and uncertainty analysis strongly suggest that our reconstruction effectively captures the main spatial and temporal features of sea surface pCO_2 distributions in the SCS. Using the reconstructed dataset, we show the long-term trends of sea surface pCO_2 in 5 sub-regions of the SCS with differing physico-biogeochemical characteristics. We show that mesoscale processes such as the Pearl River plume and China Coastal Currents significantly impact sea surface pCO_2 in the SCS during different seasons. While the SCS is overall a weak source of atmospheric CO_2, the northern SCS acts as a sink, showing a trend of increasing strength over the past two decades.

Key words: Sea surface pCO_2; reconstruction; machine learning; South China Sea

1 Introduction

The ocean possesses much of the global capacity for atmospheric carbon dioxide (CO_2) sequestration and annually mitigates 22%–26% of the anthropogenic CO_2 emission associated with fossil fuel burning and land use change during the period 2012–2021 (Friedlingstein et al., 2022). Ocean margins, an essential part of the land-ocean continuum occupying only 7% of the surface area of the ocean, contributed ~ 10%-20% of the global ocean CO_2 sequestration with large uncertainties and represent a particularly challenging regime (e.g., Chen and Borges, 2009; Dai et al. 2022; Laruelle et al., 2014), often characterized by large spatial and temporal variabilities of air-sea CO_2 fluxes that lead to even larger uncertainty in their prediction than those in the open ocean (Dai et al., 2013, 2022; Cao et al., 2020; Laruelle et al., 2014; Chen and Borges, 2009 and the references therein). Limited spatiotemporal coverage of in situ observations is an important source of these uncertainties.

In recent years, many studies use numerical models or data-based approaches to improve estimates of the partial pressure of sea surface carbon dioxide (pCO_2) and the accuracy of the global carbon budget for periods and regions with poor coverage of in situ data (Rödenbeck et al., 2015; Wanninkhof et al., 2013). Numerical ocean models can successfully quantify the generally increasing trend in oceanic pCO_2 and simulate some critical processes of carbon cycling (e.g., net ecosystem production), but still suffer from regional and seasonal differences in their estimates of the ocean carbonate parameters (Luo et al., 2015; Mongwe et al., 2016; Tahata et al., 2015; Wanninkhof et al., 2013). Thus, data-based approaches have become an important complementary to numerical models (Jones et al., 2014; Lefèvre et al., 2005;Landschützer et al., 2014, 2017; Telszewski et al., 2009). The data-based approaches typically use statistical interpolations and regression methods. Statistical interpolations of data-based approaches improve the spatial coverage of in situ data, but do not work for the period without in situ data. Regression methods allow mapping of the relationship between the in situ pCO_2 data and other parameters that may drive changes in surface ocean pCO_2, and then extrapolation of this relationship to improve estimates of the spatiotemporal distribution of pCO_2. The development of machine learning methods and remote sensing-derived products (as proxy variables in regression methods) have aided the development of data-based methods (Rödenbeck et al., 2015; Bakker et al., 2016) and can improve the model results of

删除[作者]: we present a new method to assess the
删除[作者]: in both the spatial and temporal patternsour
删除[作者]: s
删除[作者]: d
删除[作者]: s
删除[作者]: 1960
删除[作者]: 19
删除[作者]: 0
删除[作者]: However, it remains largely unknown wheth
删除[作者]: with only 7% of the surface area
删除[作者]: . Thise large uncertainty is primarily attribut
删除[作者]: y
删除[作者]: occurring
删除[作者]: al data
删除[作者]: partial pressure
删除[作者]: CO_2 distribution
删除[作者]: observational
删除[作者]: of performance
删除[作者]: CO_2
删除[作者]: system
删除[作者]: popular
删除[作者]: alternative
删除[作者]: biogeochemical
删除[作者]: The former typically use statistical interpolat
删除[作者]: observational
删除[作者]: observational data
删除[作者]: observed
删除[作者]: carbon dioxide partial pressure (pCO_2)

the oceanic carbonate system by numerical assimilation methods. Consequently, machine learning has increasingly become a routine approach in reconstruction of sea surface pCO_2 in open ocean regimes (e.g., Zeng et al., 2017; Li et al., 2019). However, it remains challenging to extend this method to marginal seas featuring more dynamic changes in both time and space.

The South China Sea (SCS) is the largest marginal sea of the North Pacific Ocean with a surface area of 3.5×10^6 km^2. Although extensive field observations have been conducted of sea surface pCO_2 in the SCS in the past two decades, their spatial and temporal coverage is still limited in different physical-biogeochemical domains of the SCS and at sub-seasonal time scales (e.g., Guo et al., 2015; Li et al., 2020; Zhai et al., 2005; Zhai et al., 2013). Therefore, there is a strong need to achieve surface water pCO_2 coverage in the SCS with spatiotemporal resolution as high as possible with the aim to better estimate sea surface pCO_2 and thus to constrain air-sea CO_2 fluxes in the SCS so as to improve initial conditions of numerical models. Moreover, reasonably high spatiotemporal resolution of pCO_2 data can help identify the controlling factors of pCO_2 changes in the SCS, and reliably resolve long-term changes.

Zhu et al. (2009) presented an empirical approach to estimate sea surface pCO_2 in the northern SCS using remote sensing-derived (RS-derived) data including sea surface temperature (SST) and chlorophyll *a* (Chl *a*), and their validation results show that the reconstructed pCO_2 data were generally consistent with the in situ data. However, it should be noted that the large uncertainty of estimates from their study was caused by the limited in situ data of only two summer cruises (July 2004 data were used for algorithm tuning and those of July 2000 for validation). Jo et al. (2012) developed a neural network-based algorithm using SST and Chl *a* to estimate sea surface pCO_2 in the northern SCS. Sea surface pCO_2 data in this study were collected from May 2001 and February and July 2004. The difference between the reconstructed pCO_2 data of Jo et al. (2012) and the in situ data reflects a relatively large bias (the resultant RMSE (root-mean-square error) falls in the range 32.6 to 44.5 µatm, reported in Wang et al., 2021). Bai et al. (2015) used a 'mechanic semi-analytical algorithm (MeSAA)' to estimate satellite remote sensing-derived sea surface pCO_2 in the East China Sea during 2000–2014, and then used this algorithm to estimate sea surface pCO_2 for the whole China Seas (the South China Sea, the East China Sea, the Yellow Sea, and the Bohai Sea, 99 - 130°E & 0 - 45°N). These authors also pointed out that their MeSAA did not fully account for some local processes and therefore caused some errors (the RMSE is about 45 µatm in the SCS (Wang et al., 2021)). Yu et al. (2022) subsequently used a non-linear regression method to develop a retrieval algorithm for seawater pCO_2 in the China Seas, and the RS-derived pCO_2 data from 2003-2018 were provided by the $SatCO_2$ platform (www.SatCO2.com). In the retrieval algorithm of Yu et al. (2022), the input parameters include sea surface temperature, chlorophyll-a concentration, remote sensing reflectance of three bands (Rrs412, 443, 488 nm), the temperature anomaly in the longitude direction, and the theoretical thermodynamic background pCO_2 under corresponding SST. Although the RMSE associated with the RS-derived pCO_2 product was relatively large (21.1 µatm), it successfully showed major spatial patterns of the sea surface pCO_2 in the China Seas (Yu et al., 2022).

删除[作者]: Thus
删除[作者]: been
删除[作者]: widely used for
删除[作者]: the
删除[作者]: for the global ocean
删除[作者]: (refs?
删除[作者]: , h
删除[作者]: still
删除[作者]: Thus, machine learning has been widely usec
删除[作者]: clear
删除[作者]: a highest
删除[作者]: and
删除[作者]: help
删除[作者]: develop
删除[作者]: d
删除[作者]: a
删除[作者]: the
删除[作者]: that
删除[作者]: d
删除[作者]: in summer
删除[作者]: satellite
删除[作者]: -
删除[作者]: ,
删除[作者]: Zhu et al. (2009) presented an empirical
删除[作者]: was
删除[作者]: underway observed
删除[作者]: underway observed
删除[作者]: from
删除[作者]: 0
删除[作者]: ing
删除[作者]: by
删除[作者]: ,
删除[作者]: ,
删除[作者]: ,

To take advantage of both the high spatiotemporal resolution of the RS-derived pCO_2 data and the accuracy of the in situ data, Wang et al. (2021) reconstructed a basin–scale sea surface pCO_2 dataset in the SCS in summer using the empirical orthogonal function (EOF) based on a multi-linear regression method and demonstrated the reliability of the reconstructions. Wang et al. (2021) demonstrate that the spatial modes of RS-derived data calculated using EOF are effective in providing spatial constraints on the data reconstruction and are thus adopted in this study. However, when the spatial standard deviation of in situ data is relatively large because of the influence of outliers, the reconstruction results may be biased (Wang et al., 2021). Therefore, many studies used machine learning-based regression methods to reduce the influence of outliers in open ocean areas, with a RMSE <17 μatm in most cases (e.g., Zeng et al., 2017; Li et al., 2019).

Building upon the EOF method that significantly improved the reconstruction in terms of spatial pattern and accuracy (Wang et al., 2021), we developed a machine learning-based regression method facilitated by the EOF to fully resolve the long-term spatial distribution of sea surface pCO_2 at a resolution of 0.05º×0.05º in the SCS. And the input data in our reconstructed model include remote sensing-derived sea surface salinity, sea surface temperature, and chlorophyll, the spatial pattern of pCO_2 constrained by EOF, atmospheric pCO_2, and time labels (month). In addition to typical machine learning performance metrics, we assessed the uncertainty resulting from the bias of the reconstruction and its sensitivity to the features.

2 Study site and data sources

2.1 Study area

The SCS, located in the western Pacific, has a maximum water depth of ca. 4700 m (e.g., Gan et al., 2006, 2010). The rhombus deep-water basin with a southwest-northeast direction accounts for about half of the total area of the SCS (Figure 1). The SCS is largely modulated by the Asian monsoon and the topography, thus exhibiting seasonally varying surface circulation, river inputs, and upwelling. Forced by the northeast winds in winter, the circulation of the upper layer shows a large cyclonic circulation structure (Figure 1), while in summer it exhibits an anticyclonic circulation structure forced by southwest winds (Figure 1; Hu et al. 2010). In the northern SCS, the Pearl River discharges into the SCS with an annual freshwater input of 3.26×10^{11} m^3 (e.g., Dong et al., 2004; Dai et al., 2014). The area influenced by the Pearl River plume may extend southeastward to a few hundred kilometers from the estuary in summer because of the monsoon wind stress (Dai et al., 2014). The northern and western coastal regions of the SCS also feature summer coastal upwelling in summer, such as the Eastern Guangdong and Qiongdong upwelling systems in the northern SCS and the Vietnam upwelling systems in the western SCS (e.g., Cao et al., 2011; Chen et al., 2012; Gan et al., 2006; Gan et al., 2010; Li et al., 2020). These seasonal changes of sea surface circulation leads to a strong seasonal characteristic of sea surface pCO_2 in the SCS.

删除[作者]: s

删除[作者]: remote sensing-derived pCO_2 data (

删除[作者]:

删除[作者]:)

删除[作者]: observational data

删除[作者]: the

删除[作者]: by

删除[作者]:

删除[作者]: remote sensing

删除[作者]: observed

删除[作者]: the

删除[作者]: for

删除[作者]: of

删除[作者]: ,

删除[作者]: a

删除[作者]: the

删除[作者]: we present a novel uncertainty calculation method that incorporates the bias of both the reconstruction and the sensitivity of reconstructed models

删除[作者]: .

删除[作者]: oceanography of the

删除[作者]: red solid line in

删除[作者]: .

删除[作者]: red dashed line in

删除[作者]: .

删除[作者]: P

删除[作者]: river

删除[作者]: mainly

删除[作者]: including

删除[作者]: in SCS

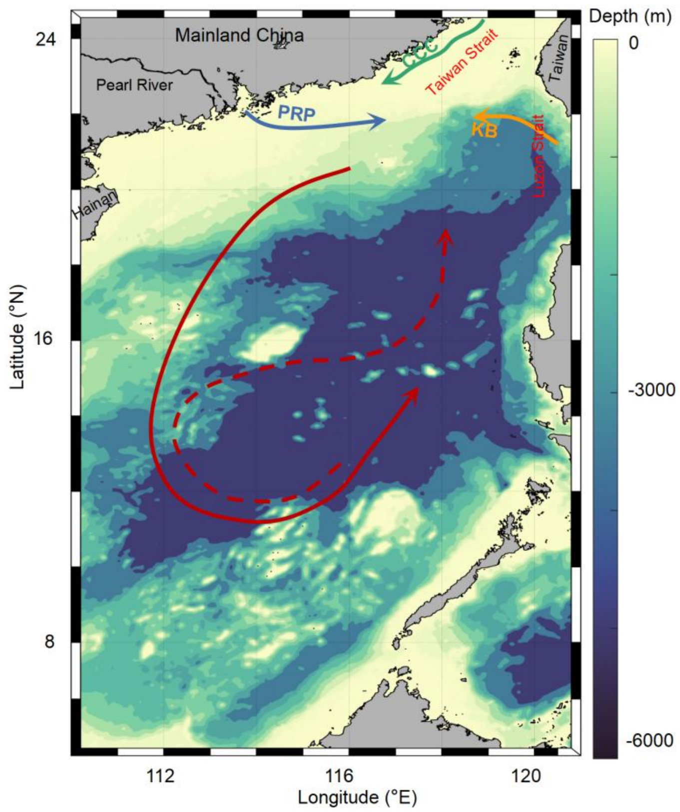


Figure 1. Topographic map of the South China Sea (SCS) showing a basin wide cyclonic circulation in winter (solid line) and an anticyclonic circulation over the southern half of the SCS in summer (dashed line). Also shown are the Kuroshio Branch (KB, orange line), the China Coastal Current (CCC, green line), and the Pearl River plume (PRP, blue line).

删除[作者]: **gyre**

删除[作者]: **gyre**

The SCS is a semi-enclosed sea basin with dynamic water exchanges with the East China Sea via the Taiwan Strait and Western Pacific via the Luzon Strait (Fig. 1). In winter, driven by the winter monsoon, the China Coastal Current (CCC, green line in Fig. 1; Han et al., 2013; Yang et al., 2022) flows south along the Chinese mainland through the Taiwan Strait, and occupies the northern SCS with cold, fresh, nutrient-rich waters. The strong northeast winds in winter also slow down the western boundary ocean current, forcing the intrusion of Kuroshio water, which shows high surface salinity and high total alkalinity, into the SCS via the Luzon Strait (orange line in Fig. 1; Du et al., 2013; Park, 2013; Yang et al., 2022). These water exchange processes increase the complexity of the spatial distribution of sea surface pCO_2 in the SCS. As a result, the sea surface pCO_2 in the SCS has strong seasonal characteristics and spatial variability. A high-spatial-resolution sea surface pCO_2 dataset would reveal the role of the SCS, one of the largest marginal seas, in the uptake of atmospheric CO_2.

删除[作者]: yellow

删除[作者]: Therefore, controlled by the above physical processes,

删除[作者]: pCO2

删除[作者]: pCO2

删除[作者]: helps

删除[作者]: to

删除[作者]: clarify

删除[作者]: global oceanic

删除[作者]: carbon cycle

2.2 Observational pCO_2 data

Data collected from field surveys during the study period 2003-2020 are summarized in Table 1. Most observations were made in July, and fewer observations were made in March and December of each year. The rough sea state in the SCS in winter and early spring limited the field surveys during these seasons. Data collected from July 2000 to January 2018 were originally published by Li et al. (2020). The in situ pCO_2 were collected from R/Vs *Dongfanghong-2, Tan Kah Kee (TKK)*, etc. (shown in Table 1).

During the cruises, sea surface pCO_2 was measured underway. The measurement and data processing followed the SOCAT (Surface Ocean CO2 Atlas) protocol (Li et al., 2020). More details of the data collection methods have been introduced in Li et al. (2020). The spatial coverage and frequency of the observations are shown in Figure 2 and show that there are pronounced seasonal changes and that the data cover a large spatial area. For example, the spatial coverage of the in situ data in spring and fall are relatively uniformly distributed, and the south end of the spatial coverage reaches 5 ºN in spring, whereas during other seasons the data are concentrated in the northern and central regions of the SCS. In addition, only one observation was made in the basin area in winter, while the northern coastal area was more frequently surveyed, especially in summer.

Table 1. Summary of seasonal in situ data of sea surface pCO_2 in the South China Sea for the period 2003-2020 used in this study.

Season	Spring			Summer		
	March	April	May	June	July	August
Cruise time	2004.03	2005.04 2008.04 2009.04 2012.04 2020.04*	2004.05 2011.05 2014.05 2020.05*	2006.06 2016.06 2017.06* 2019.06* 2020.06*	2004.07 2005.07 2007.07 2008.07 2009.07 2012.07 2015.07* 2019.07*	2007.08 2008.08 2019.08*
Season	**Fall**			**Winter**		
	September	October	November	December	January	February
Cruise time	2004.09 2007.09 2008.09 2020.09*	2003.10 2006.10	2006.11 2010.11	2006.12	2009.01 2010.01 2018.01	2004.02 2006.02
Data source	Li et al. (2020) *This study					

删除[作者]: continuously

删除[作者]: methods

删除[作者]: those of

删除[作者]: , http://www.socat.info/news.html

删除[作者]: protocol

删除[作者]: T

删除[作者]: used in this study

删除[作者]: observed

删除[作者]: that

删除[作者]: is

删除[作者]: **the**

删除[作者]: **observational data**

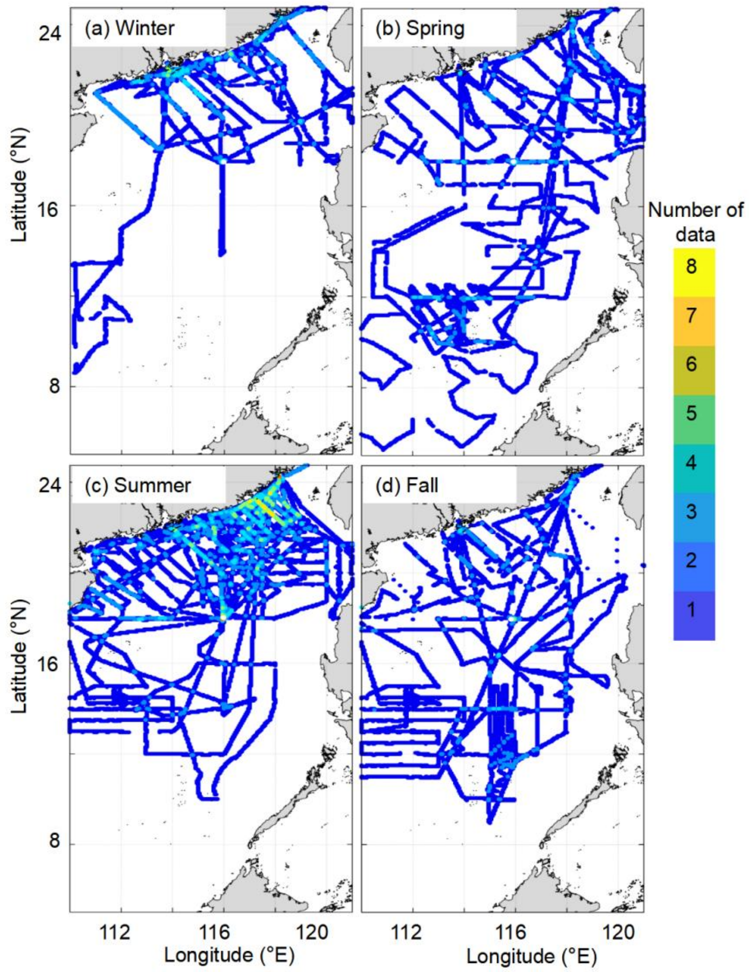
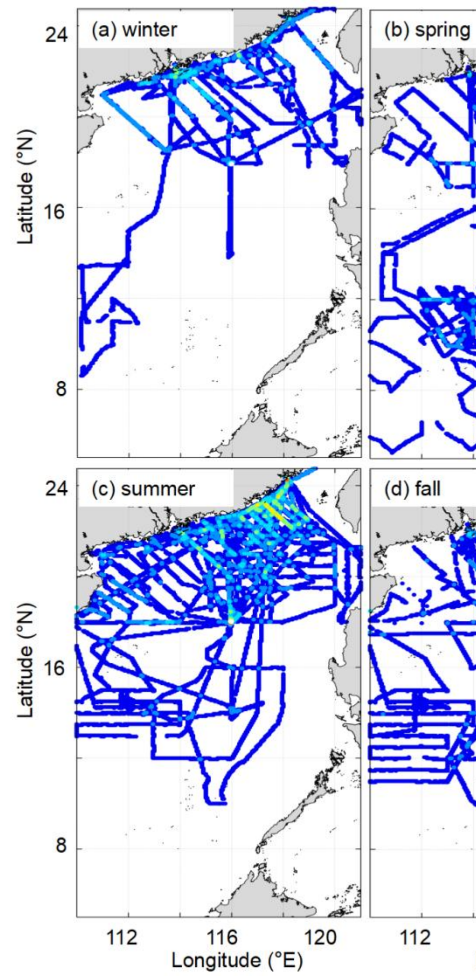


Figure 2. Cruise tracks of the observations conducted in the South China Sea in each season from 2000 to 2020: (a) Winter, (b) Spring, (c) Summer, and (d) Fall. The data collected before February 2018 are from Li et al. (2020) except those collected in July 2015 and June 2017.

Figure 3 shows the spatial and temporal distributions of in situ sea surface pCO_2. Seasonally, the lowest pCO_2 occurs in January, and the highest concentrations occur in May and June. Spatially, the pCO2 distribution in the basin is relatively homogeneous with large variability in the northern region. In the northern coastal area in summer, the in situ pCO_2 distribution is affected by the Pearl River plume (yielding low values) and coastal upwelling (yielding high values), which last into early fall. In winter and early spring, relatively low pCO_2 values (~350 μatm) were determined in the near-shore area. In addition, the high pCO_2 values recorded on the western side of the Luzon Strait in December demonstrate the influence of winter upwelling during some of the surveys.

In addition to the above in situ sea surface pCO_2 data, to verify the accuracy of our reconstruction model in extrapolation to periods lacking training datasets, we selected the in situ sea surface pCO_2 data collected in four independent surveys corresponding to four seasons, September 2018 (fall), December 2018 (winter), August 2019 (summer), and April 2020 (spring). Furthermore, we used another dataset of sea surface pCO_2 calculated from observed dissolved inorganic carbon and total

删除[作者]:



删除[作者]:

删除[作者]: w

删除[作者]: s

删除[作者]: s

删除[作者]: f

删除[作者]: were

删除[作者]: ,

删除[作者]: which are from Li et al. (2020)

删除[作者]:

删除[作者]: water

删除[作者]: of in situ measurements

删除[作者]: Spatially, the pCO_2 distribution in the basin i

删除[作者]: observed

删除[作者]: data

删除[作者]: 25

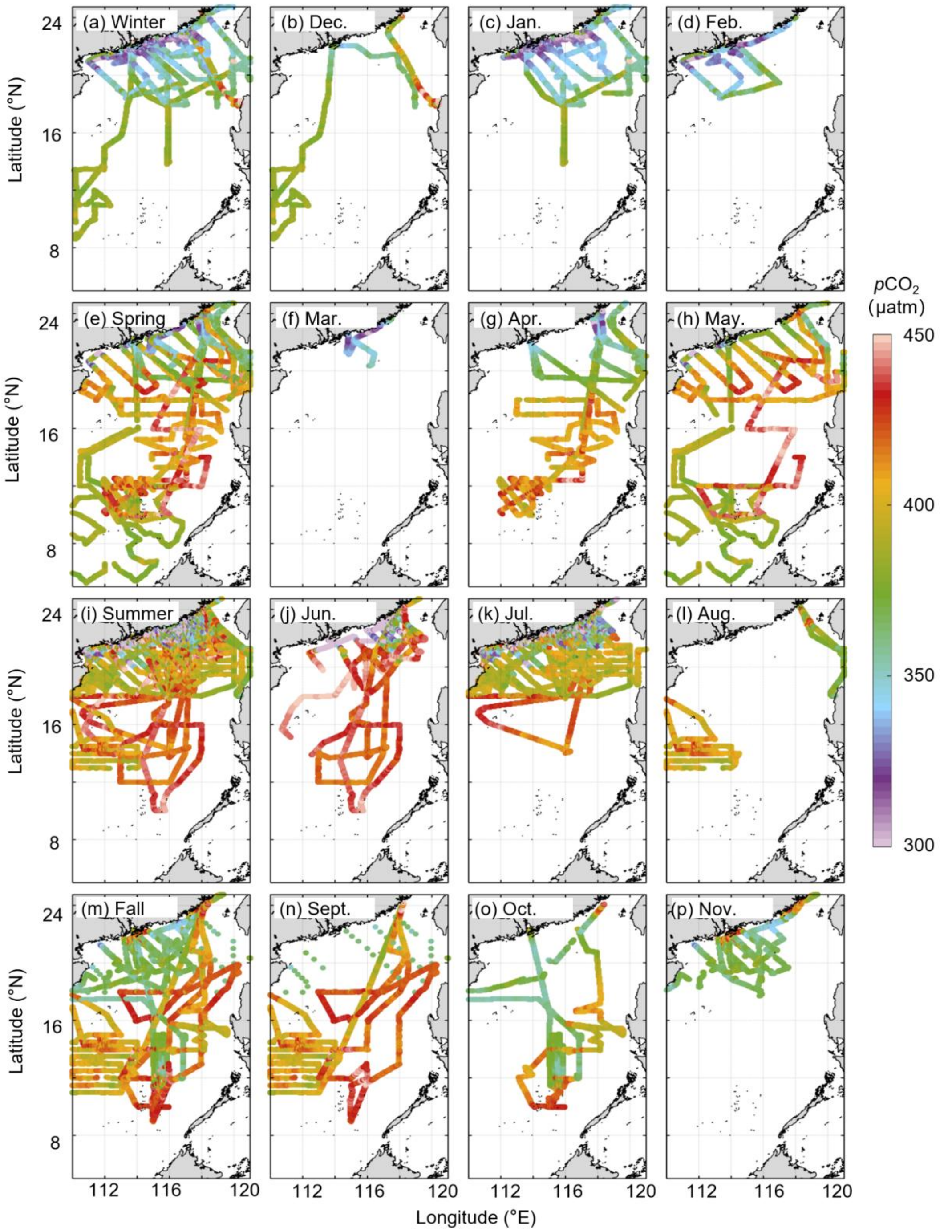
删除[作者]:

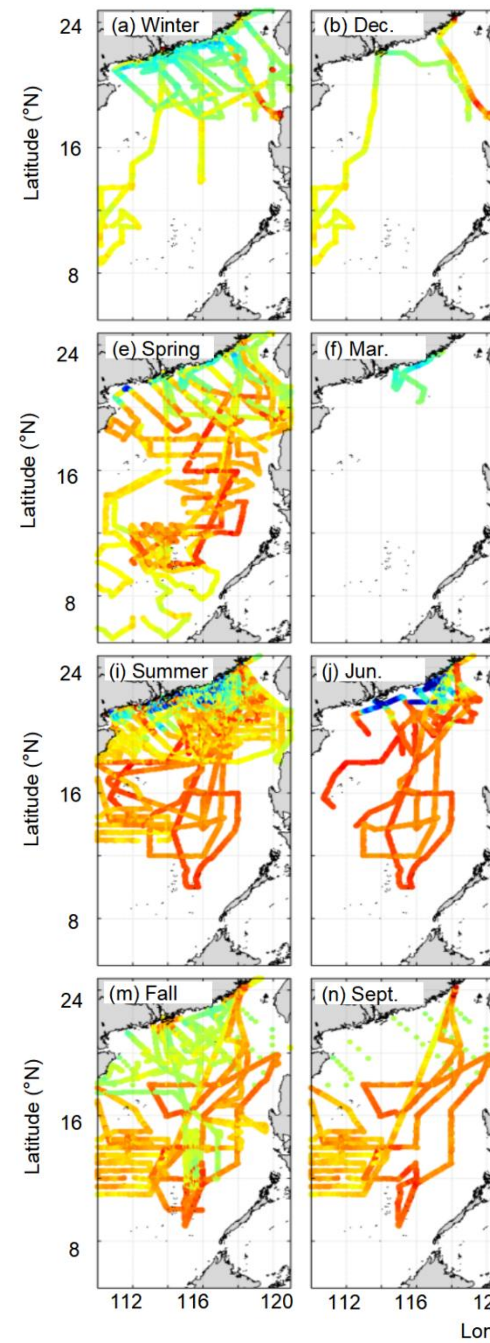
删除[作者]: coastal

删除[作者]: selected

删除[作者]: and the in situ sea surface pCO_2 data collec alkalinity during 2003–2019 at the Southeast Asia Time-Series (SEATs) station (data from Dai et al., 2022) to test the long-term consistency of the reconstruction.

删除[作者]: ,





删除[作者]:

Figure 3. Seasonal and monthly sea surface pCO$_2$ fields in the South China Sea; a. Winter; b. December; c. January; d. February; e. Spring; f. March; g. April; h. May; i. Summer; j. June; k. July; l. August; m. Fall; n. September; o. October; p. November. The data sources can be found in Table 1.

删除[作者]: .

删除[作者]: **The data sources can be found in Table 1 (**

删除[作者]: **w**

删除[作者]: **)**

2.3 Remote sensing-derived sea surface pCO$_2$ data

The gridded (0.05º×0.05º) remote sensing-derived pCO_2 data covered almost the entire SCS (5–25º N, 109–122º E), and show major variations in sea surface pCO_2 on the basin scale (Wang et al., 2021; Yu et al., 2022). Further details of the RS-derived pCO_2 data can be found on the SatCO2 platform (www.SatCO2.com).

A grid-to-grid comparison was undertaken between the RS-derived pCO_2 and the in situ pCO_2 (Table 2). This comparison shows that the difference between the RS-derived pCO_2 and the in situ pCO_2 ranges from 35 to 120 µatm in the nearshore area, and that the largest biases occur in summer. The largest RMSE is up to 29.95 µatm in summer (Table 2). Relatively large discrepancies may reflect the limitations of the current algorithm (MeSAA and non-linear regression), which considers only biological processes and the turbidity induced by the Pearl River discharge (characterized by Chl *a* and the remote sensing reflectance at 555 nm (rrs555) and does not take into account the riverine dissolved inorganic carbon and the input of other substances that may affect pCO_2 (Bai et al.,2015, Yu et al., 2022 and Wang et al.,2021).

To remove the influence of the bias in RS-derived pCO_2 data on our reconstructed results, in this study, we used the EOF method to compute the spatial patterns of the RS-derived pCO_2 data as input data instead of directly using the RS-derived pCO_2 data. Moreover, using EOF modes of the RS-derived pCO_2 as input data in the reconstructed model can provide spatial constraints of the pCO_2 reconstruction.

Table 2. Biases between the seasonal remote sensing derived pCO_2 data and in situ pCO_2 data, and between the reconstructed and the in situ pCO_2 data. (unit: µatm; the remote sensing-derived pCO_2 data during 2003-2019 are from www.SatCO2.com and the source of in situ data can be found in Table1. The reconstructed pCO_2 data are from section 3; all data were gridded into 0.05°*0.05°; / means no data). MAE = mean absolute error; RMSE = root mean square error; R^2 = coefficient of determination; MAPE = mean absolute percentage error.

		RS-derived pCO_2 data	Training data	Testing data I	Testing data II	Testing data III
Spring	MAE	9.00	2.44	4.76	1.68	/
	RMSE	12.70	3.47	7.43	2.26	/
	R^2	/	0.98	0.92	/	/
	MAPE	/	0.01	0.01	/	/
Summer	MAE	16.75	2.48	8.46	5.73	/
	RMSE	29.95	3.54	14.69	15.18	/
	R^2	/	0.99	0.89	/	/
	MAPE	/	0.01	0.02	/	/
Fall	MAE	9.93	2.41	4.90	7.133	/
	RMSE	13.08	3.39	6.85	8.94	/

删除[作者]: the
删除[作者]: CO_2
删除[作者]: at a
删除[作者]: large
删除[作者]: Bai
删除[作者]: unpublished
删除[作者]: In the retrieval algorithm of Yu et al. (2022)
删除[作者]:
删除[作者]: remote sensing (RS) data
删除[作者]: i
删除[作者]: (Fig. 4)
删除[作者]: and the RMSE of
删除[作者]:
删除[作者]: dataRS data-derived pCO_2
删除[作者]: values were compared with
删除[作者]: observed
删除[作者]: data
删除[作者]:
删除[作者]: dataRS
删除[作者]: observed
删除[作者]: data
删除[作者]: coastal
删除[作者]: In terms of the RMSE (Table 2), t
删除[作者]: bias
删除[作者]: reaches
删除[作者]: 30.0
删除[作者]: Bai et al. (2015), Yu et al. (2022); unpublish
删除[作者]: pointed out that r
删除[作者]: pCO_2.
删除[作者]: Moreover, there are some missing values in
删除[作者]: was used
删除[作者]:
删除[作者]: of
删除[作者]:

	R^2	/	0.98	0.92	/	/
	MAPE	/	0.01	0.01	/	/
Winter	MAE	9.25	2.18	5.61	11.41	/
	RMSE	14.26	3.14	8.82	12.63	/
	R^2	/	0.98	0.89	/	/
	MAPE	/	0.01	0.01	/	/
Annual	MAE	11.95	2.41	6.30	5.27	6.19
	RMSE	20.66	3.43	10.79	11.18	8.26
	R^2	/	0.99	0.91	/	/
	MAPE	/	0.01	0.01	/	/

2.4 Other data

The RS-derived SST data produced by MODIS (https://oceancolor.gsfc.nasa.gov/) are adopted in our reconstruction. The uncertainty of this dataset in the SCS is ~0.27º (Qin et al., 2014). For sea surface salinity (SSS) data, Wang et al. (2022) found relatively large differences between different open source SSS databases (i.e., multi-satellite fusion data from https://podaac.jpl.nasa.gov/; model data from https://climatedataguide.ucar.edu/; multidimensional covariance model data from https://resources.marine.copernicus.eu/) and the in situ SSS data. Thus, Wang et al. (2022) produced a RS-derived SSS database using machine learning methods. The bias between the RS-derived SSS (Wang et al., 2022) and in situ data was near-zero (mean absolute error, MAE: ~0.25). Next, we used Chl-*a* (from https://oceancolor.gsfc.nasa.gov/) as an indicator of biological influence, which have a bias of ~0.35 on log scale and ~115% in the SCS (Zhang et al., 2006). Atmospheric pCO_2 also influences sea surface pCO_2 through air–sea CO_2 exchange. We chose the atmosphere CO_2 mole fraction (xCO_2) data from the monthly mean CO_2 concentrations measured at Mauna Loa Observatory, Hawaii (https://gml.noaa.gov/), and then calculated the atmospheric pCO_2 values from xCO_2 using the method of Li et al. (2020).

3 Methods

The pCO_2 reconstruction procedure is shown in Figure 4. It includes: (1) data processing and (2) model training and testing. For the former, we first gridded the in situ data and RS-derived pCO_2 data into 0.05°×0.05° grid boxes with monthly temporal resolution. Secondly, we filled missing pCO_2 measurements with the RS-derived pCO_2 data according to Fay et al. (2021) (see more details in Section 3.1). We then used EOF to ignore any biase in the RS-derived pCO_2 dataset itself or from the pCO_2 filling method. Thirdly, the gridded in situ pCO_2 data and their corresponding RS-derived data were divided into a training set (90%) and a testing set (10%) to calculate the pCO_2 retrieval model. To ensure that the model had sufficient training samples in the coastal

删除[作者]:
删除[作者]:
删除[作者]: here
删除[作者]: the
删除[作者]: SSS
删除[作者]: .
删除[作者]: the
删除[作者]: .Wang et al. (in preparation2022) found a
删除[作者]: in preparation
删除[作者]: reconstructedproducted
删除[作者]: remote sensing
删除[作者]:
删除[作者]: by
删除[作者]: based on based on a combination of
删除[作者]: remote sensing
删除[作者]: reconstructed
删除[作者]: observed
删除[作者]: . Chl-*a* data from MODIS
删除[作者]: water
删除[作者]: .
删除[作者]: T
删除[作者]: were calculated
删除[作者]: by
删除[作者]: 5
删除[作者]: observed
删除[作者]:
删除[作者]: RS pCO_2 data
删除[作者]: And all these data used in machine learning
删除[作者]: used the pCO_2 filling method according to F
删除[作者]: the
删除[作者]:
删除[作者]: ss
删除[作者]: Secondly, we used the pCO_2 filling method
删除[作者]: a feature engineering

area, we divided the entire SCS into two regions along the 200 m depth contour (as shown in Figure 5). The data from these two regions were divided into training and testing sets with the same ratios listed above (9:1), and then combined to obtain the final training and testing sets. Note that all these data used in machine learning have been interpolated on the same grid.

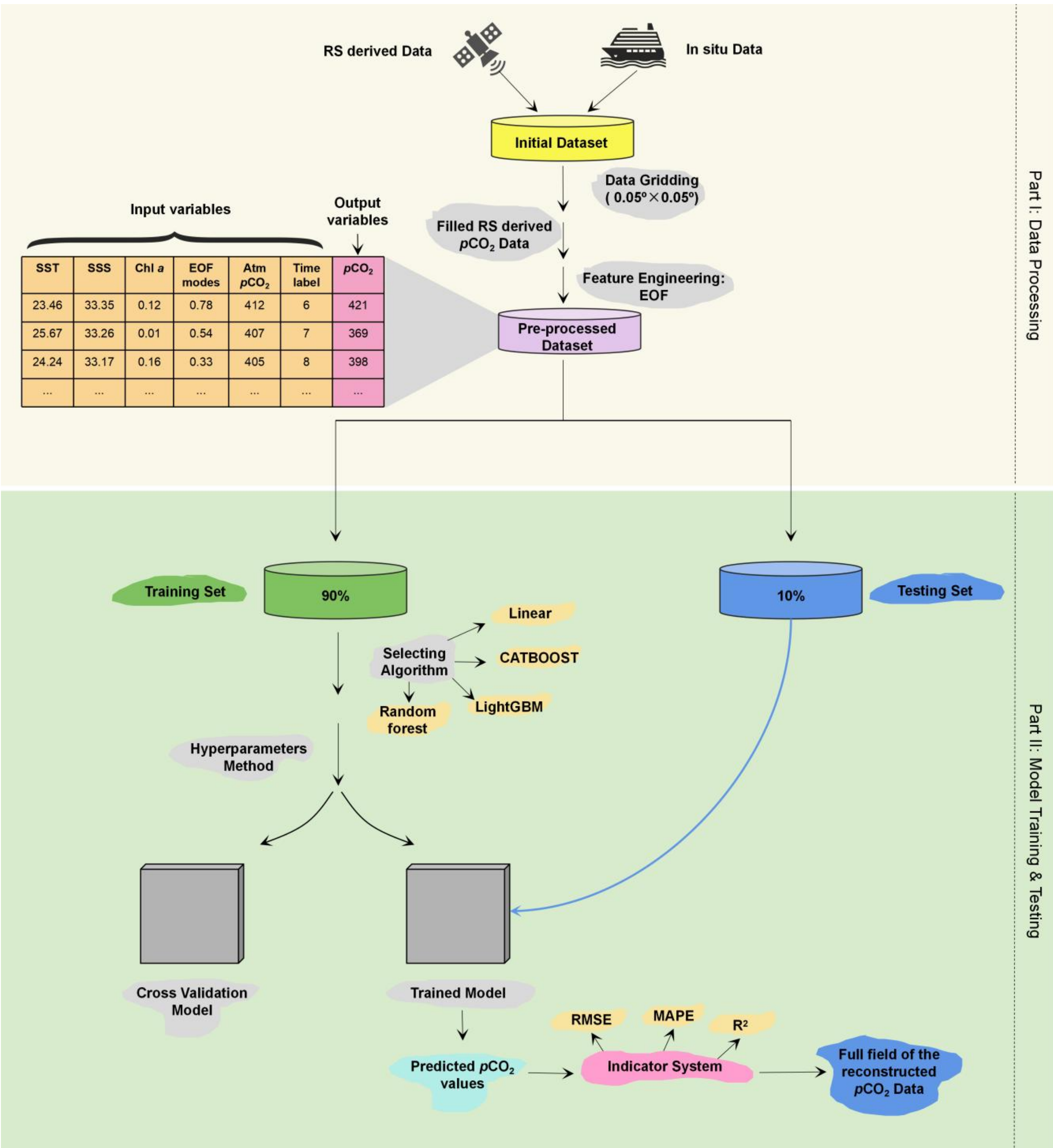
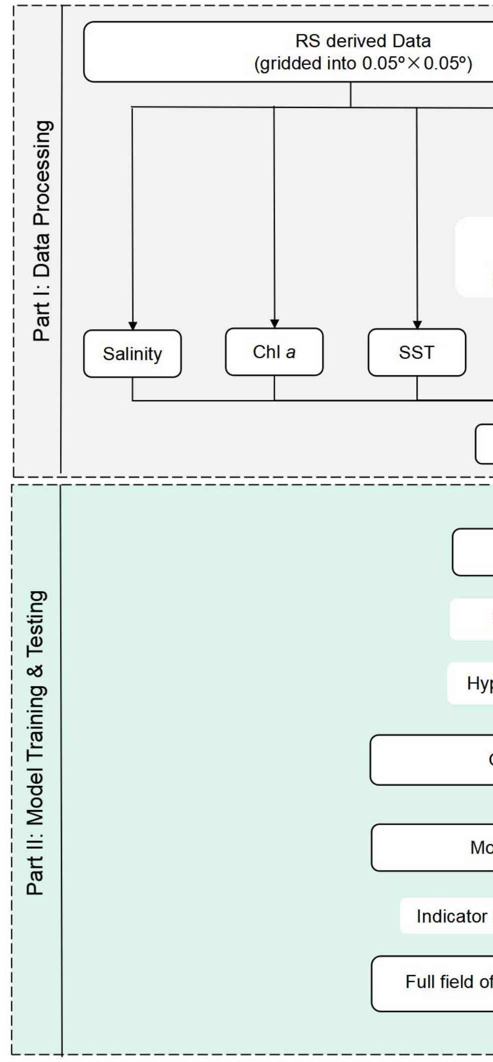


Figure 4. Procedure for the reconstruction of surface water pCO_2 using machine learning. RS derived data = remote sensing derived data, RMSE = root mean square error, MAPE= mean absolute percentage error, and R^2 = coefficient of determination, and MAE = average absolute error.

删除[作者]: as
删除[作者]: shown
删除[作者]: which were
删除[作者]: And
删除[作者]:



删除[作者]: 5
删除[作者]:
删除[作者]:

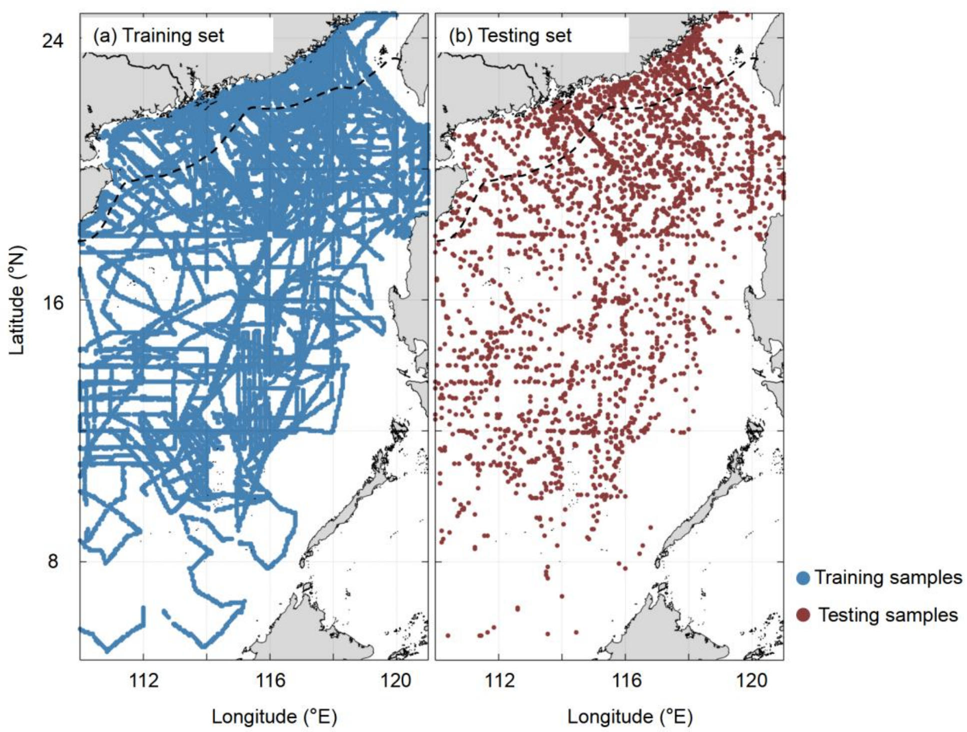


Figure 5. Spatial distributions of Training samples (a) and Testing samples (b); The black dash line stands for the 200 m depth contour.

For model training and testing, we firstly chose a relatively reliable algorithm to undertake the pCO_2 reconstruction. After that, we determined the optimal range of the parameters using hyperparameter methods (code from https://github.com/optuna/) for the training set. The final optimal parameter values were then determined using the *K*-fold and cross validation method (code from https://github.com/suryanktiwari/Linear-Regression-and-K-fold-cross-validation) for the training set. These optimal parameters were applied to the chosen algorithm. Finally, the testing set was used to verify the accuracy of the pCO_2 retrieve algorithm produced by the training set, and some indicators of the model's accuracy were calculated. More detailed methods employed in the present study are described below.

3.1 Remote sensing data filling

As mentioned in the SatCO2 platform (www.SatCO2.com), the RS-derived pCO_2 data miss some values. Thus, we used the pCO_2 filling method suggested by Fay et al. (2021) to fill in the missing portions. First, a scaling factor for a filled month was calculated according to Equation 1:

$$sf_{pCO2} = mean_{x,y}\left(\frac{pCO_2^{ens}}{pCO_2^{clim}}\right) \quad (1)$$

where sf_{pCO2} is the scaling factor, pCO_2^{ens} is the monthly RS-derived pCO_2 data, and pCO_2^{clim} is the monthly climatology RS-derived pCO_2 data; *x* and *y* indicate that we took the area-weighted average over longitude (*x*) and latitude (*y*) to produce the monthly sf_{pCO2} value. Then, the filled portion of the data can be calculated from the pCO_2^{clim} data multiplied by the sf_{pCO2} value (see Fay et al. (2021) for details of this method).

删除[作者]: **The s**

删除[作者]: **o**

删除[作者]:

删除[作者]: RS pCO_2 data

删除[作者]: are missing

删除[作者]: es

删除[作者]:

删除[作者]: RS pCO_2 datum

删除[作者]:

删除[作者]: RS pCO_2 datum

Briefly, this filling method scales the climatological monthly pCO_2 field values to fill in the missing measurements. Therefore, although specific values may be biased, the interpolated measurements still retain the main spatial distribution pattern of the filled months.

3.2 Feature engineering and selection

As mentioned above, the pCO_2 filling method may bias some of the actual values. To avoid the influence of such biases on the reconstructed results, instead of directly using the RS-derived pCO_2 data as features in our reconstructed model, we used the EOF method to obtain the main spatiotemporal distribution patterns of the RS-derived pCO_2 data as features in our reconstructed model. The EOF reflects the spatial commonality of variables shown in the time series, thus it is widely used to calculate spatial patterns of climate variability (e.g. Levitus et al., 2005; Dye et al., 2020; McMonigal and Larson, 2022). Typically, the spatial commonality of variables, also named EOF modes, is found by computing the eigenvalues and eigenvectors of a spatially weighted anomaly covariance matrix of a field. Each EOF modes' corresponding variance represents its degree of interpretation of the spatial pattern of a variable. For each 12 months, the cumulative variance contribution of the first eight EOF values was consistently > 90%, indicating it that it could explain the main pCO_2 spatial characteristics during each month, and we therefore selected them as features.

The feature selection in our reconstructed model can be divided into two main categories. The first one is related to the underlying physicochemical mechanism controlling the pCO_2 distribution, and the other one can provide spatiotemporal information for pCO_2 reconstruction. For example, the SST dominating the seasonal variation in surface water pCO_2 in the northern SCS (Zhai et al., 2005; Chen et al., 2007; Li et al., 2020). Previous researches (Landschützer et al., 2014; Laruelle et al., 2017; Denvil et al., 2019) show that Chl-*a* plays a critical role in fitting the influence of biological activity to pCO_2, especially in the northern SCS (Landschützer et al., 2014; Laruelle et al., 2017; Denvil et al., 2019). Sutton et al. (2017) suggest that the increase in atmospheric pCO_2 controls the increase in seawater pCO_2. For the features that provide spatiotemporal information for pCO_2 reconstruction, whereas in the present study we selected the first eight EOF values of pCO_2 as the main spatial distribution feature and monthly information of the in situ datasets as the temporal feature.

3.3 Algorithm selection

Ensemble learning provides one of the most powerful machine learning techniques (e.g., Zhan et al., 2022; Chen et al., 2020). It is the process of training multiple machine learning models and combining their output to improve the reliability and accuracy of predictions (e.g., Zhan et al., 2022; Chen et al., 2020). Different models are used as the basis to develop an optimal predictive model. There are two main ways to employ ensemble learning: bagging (to decrease the model's variance) or boosting (to decrease the model's bias). The random forest algorithm (code from https://scikit-learn.org/stable/) is an extension of the bagging method as it utilizes both bagging and feature randomness to create an uncorrelated forest of decision trees. The Light Gradient Boosting Machine (LightGBM; code from https://github.com/microsoft/LightGBM/) is a gradient boosting framework that uses

删除[作者]:

删除[作者]: RS pCO_2 data

删除[作者]: feengineered featured data (via the

删除[作者]:)

删除[作者]:

删除[作者]: RS pCO_2 data

删除[作者]: are

删除[作者]: The EOF reflects the spatial commonality of variables shown in the time series, thus it is widely used to calculate spatial patterns of climate variability (refs). Typically, the spatial commonality of variables, also named EOF modes, are found by computing the eigenvalues and eigenvectors of a spatially weighted anomaly covariance matrix of a field. Each EOF modes' corresponding variance represents its degree of interpretation of spatial pattern of variable.The EOF method can reflect the spatial commonality of variables shown in the time series.

删除[作者]: s

删除[作者]: observed

tree-based learning algorithms. LightGBM can be used for regression, classification, and other machine learning tasks; it exhibits rapid and high-performance as a machine learning algorithm. CATBOOST (code from https://github.com/catboost/) is a gradient boosting algorithm, which improves prediction accuracy by adjusting weights according to the data distribution and by incorporating prior knowledge about the dataset. This can help to reduce overfitting and improve generalization performance.

From the above options, we chose three ensemble learning algorithms as the machine learning-based regression portion, and multi-linear regression methods (Wang et al., 2021) as the linear regression portion. We then used the K-fold and cross validation methods to verify the applicability of different regression algorithms in the pCO_2 reconstruction for seasonal training data. The results show that in summer the CATBOOST algorithm yields the best degree of accuracy with an RMSE of 16 µatm (Table R1). In contrast, the RMSE of LightGBM was 27 µatm, and that of Random Forest was 26 µatm. The RMSE was nearly 20 µatm using the linear regression algorithm employed by Wang et al. (2021). Thus, CATBOOST appears to provide a reliable algorithm for reconstructing pCO_2. In other three seasons, however, different algorithms resulted in minor differences (~2 µatm in RMSE).

Table 3. RMSEs associated with different algorithms in different seasons.

Season	Random Forest	LightGBM	CATBOOST	Multi-linear regression (Wang et al., 2021)
Spring	10.65 µatm	9.52 µatm	8.17 µatm	NaN*
Summer	26.53 µatm	27.83 µatm	16.15 µatm	20.13 µatm
Fall	10.34 µatm	11.56 µatm	10.35 µatm	NaN
Winter	12.48 µatm	12.75 µatm	11.52 µatm	NaN

***NaN stands for the missing value**

3.4 Evaluation metrics

It is necessary to evaluate the accuracy of any model based on certain error metrics before applying it to specific scenarios. Common model evaluation metrics include RMSE, MAPE, R^2 (coefficient of determination), and MAE.

The mean squared error (MSE) stands for the standard deviation of the residuals (prediction error), where the residuals represent the distance between the fitted line and the data point, i.e., stands for the degree of concentration of the reconstructed data around the regression line. In regression analysis, RMSE is commonly used to verify experimental results. To assess bias, the RMSE needs to combine the magnitude of the model data and is calculated as:

$$\mathrm{RMSE} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - y_{ri})^2} \quad . \qquad (2)$$

where y stands for the in situ data, y_r represents the reconstructed data, and n is the number of data.

The mean absolute percentage error (MAPE) is a statistical measure used to define the accuracy of a machine learning algorithm on a particular dataset. It is commonly used because, compared to other metrics, it uses a percentage to measure the magnitude of the bias and is easy to understand and interpret; the lower the value of MAPE, the better a model is at forecasting. MAPE is calculated as follows:

删除[作者]: the

删除[作者]: We

删除[作者]: ,

删除[作者]: ,

删除[作者]: For comparison

删除[作者]:

删除[作者]: Note that

删除[作者]: for other three seasons only

删除[作者]: From the above options, we chose three ensemble learning algorithms as the machine learning-based regression portion, and multi-linear regression methods (Wang et al., 2021) as the linear regression portion, and we then used the K-fold and cross validation methods to verify the applicability of the different regression algorithms in the pCO_2 reconstruction of summerfour seasoanl training datasets in the SCS, since the greatest temporal sampling coverage occurs in summer (Table 1; Fig. 2). Results show that in the summer, the CATBOOST algorithm yields the best degree of accuracy, with an RMSE of 16 µatm; for comparison, the RMSE of LightGBM was 27 µatm, that of Random Forest was 26 µatm, and nearly 20 µatm was found for the linear regression algorithm employed by Wang et al. (2021). In the other three seasons, the RMSE between the pCO2 producted by different algorithm and in situ pCO2 data is similar. Thus, CATBOOST appears to provide a relatively reliable algorithm for pCO_2 reconstruction.

删除[作者]: **between**

删除[作者]: **The RMSE between the CO2 different algorithm and in situ data of different seasonal (NaN stand for the missing value**

删除[作者]:

删除[作者]: .

删除[作者]: it

删除[作者]: observational data

$$\mathrm{MAPE} = \frac{1}{n}\sum_{i=1}^{n}\frac{|y_i - y_{ri}|}{|y_i|} \quad (3)$$

The regression error metric, the coefficient of determination R^2, can describe the performance of a model by evaluating the accuracy and efficiency of modeled results, i.e., it indicates the magnitude of the dependent variable calculated by the regression model that can be explained by the independent variable, and is calculated as:

$$R^2 = 1 - \frac{\sum_{i=1}^{n}(y_i - \overline{y_i})^2}{\sum_{i=1}^{n}(y_i - y_{ri})^2} \quad (4)$$

MAE is the average absolute difference between the in situ data (true values) and model output (predicted values). The sign of these differences is ignored so that cancellations between positive and negative values do not occur. It is calculated as:

$$\mathrm{MAE} = \frac{1}{n}\sum_{i}^{n} | y_i - y_{ri} | \quad (5)$$

3.5 Uncertainty

In previous studies, RMSE and MAE were mostly used to represent the uncertainties in reconstructed data. However, this expression of uncertainty ignores the sensitivity of the reconstructed model to the features; i.e., the biases that the features themselves pass to the reconstructed model are ignored. Moreover, it is clearly unreasonable to use a single RMSE or MAE value to represent the entire region because the spatial bias pattern in the coastal region clearly differs from that in the basin.

Thus, we here present a novel method of uncertainty calculation as shown below:

$$Uncertainty = MAX\left(\left[\frac{\sum_{i=1,j=1,k=1}^{n}\frac{|OR\_Monthly\_Data(i,j,k) - Obs\_Monthly\_Data(i,j,k)|}{Obs\_Monthly\_Data(i,j,k)}}{num(i)+num(j)}, \ldots, \frac{\sum_{i=1,j=1,k=n}^{n}\frac{|OR\_Monthly\_Data(i,j,k) - Obs\_Monthly\_Data(i,j,k)|}{Obs\_Monthly\_Data(i,j,k)}}{num(i)+num(j)}\right)\right) * 100\% * pCO2\_recon + \left(\frac{\partial pCO2}{\partial Feature}\right) dFeature \quad (6)$$

Equation (6) includes two terms: the first term is the conservative bias between the reconstructed pCO_2 fields and the in situ data (the first term), and the second is the sum over sensitivity of the reconstructed model to the features (the second term). For the first term in Equation 6, k stands for kth month, $OR\_Monthly\_Data(i,j,k)$ stands for the kth monthly reconstructed data at longitude(i) and latitude(j), and $Obs\_Monthly\_Data(i,j,k)$ stands for the kth monthly in situ data at longitude (i) and latitude (j). Therefore, MAX in the first term stands for the maximum of the k monthly bias ratios. And 'pCO_2_recon' stands for the reconstructed pCO_2 data. In the second term, where $dFeature$ stands for the bias of the features. We conducted a sensitivity analysis using a chain rule to evaluate the influence of these biases in the features on pCO_2. Then we estimated pCO_2 changes due to these features' variability by constraining these features based on our model, and computed $\frac{\partial pCO2}{\partial Feature}$. For example, for $\frac{\partial pCO2}{\partial SST}$, we only changed the value of SST and kept the values of the other features constant to calculate the effect of each additional unit of SST on the simulated pCO_2.

4 Results and discussion

删除[作者]: field observations
删除[作者]:
删除[作者]: the
删除[作者]: s
删除[作者]: between
删除[作者]: coastal
删除[作者]: and basin areas
删除[作者]: <math>+<math> (part 1)
删除[作者]: parts
删除[作者]: ;
删除[作者]: part 1
删除[作者]: the
删除[作者]: (part 2)
删除[作者]: F
删除[作者]: R1
删除[作者]: the
删除[作者]: that
删除[作者]: value between
删除[作者]: or part 1, <math> stands for the monthly
删除[作者]:
删除[作者]: For part
删除[作者]: 2
删除[作者]: ,
删除[作者]: w
删除[作者]: And t
删除[作者]: ing
删除[作者]: the
删除[作者]: part
删除[作者]: ,
删除[作者]: ,
删除[作者]: results
删除[作者]: of the
删除[作者]: simulation

4.1 Results

The reconstructed pCO_2 fields show relatively low values in the northern coastal study region but generally shows high values in the mid and southern basins (Fig. 6). The continuous changes of the spatiotemporal distribution can be found in the reconstruction results (Fig. 6). The reconstructed pCO_2 fields show a trend of slow but sustained increase from 2003 to 2020. Spatial patterns of pCO_2 change between 2003 and 2020, such that the coastal portion of the northern SCS shows relatively complex variability because of multiple controlling factors, such as coastal upwelling, river plumes, biological activity, etc. However, pCO_2 values in the mid and southern basin are relatively homogeneous, because they are mainly controlled by atmospheric pCO_2 forcing and SST. Temporal changes in pCO_2 between 2003 and 2020, are relatively large (~44 μatm) in summer and relatively small (~33 μatm) in winter.

删除[作者]: continuity

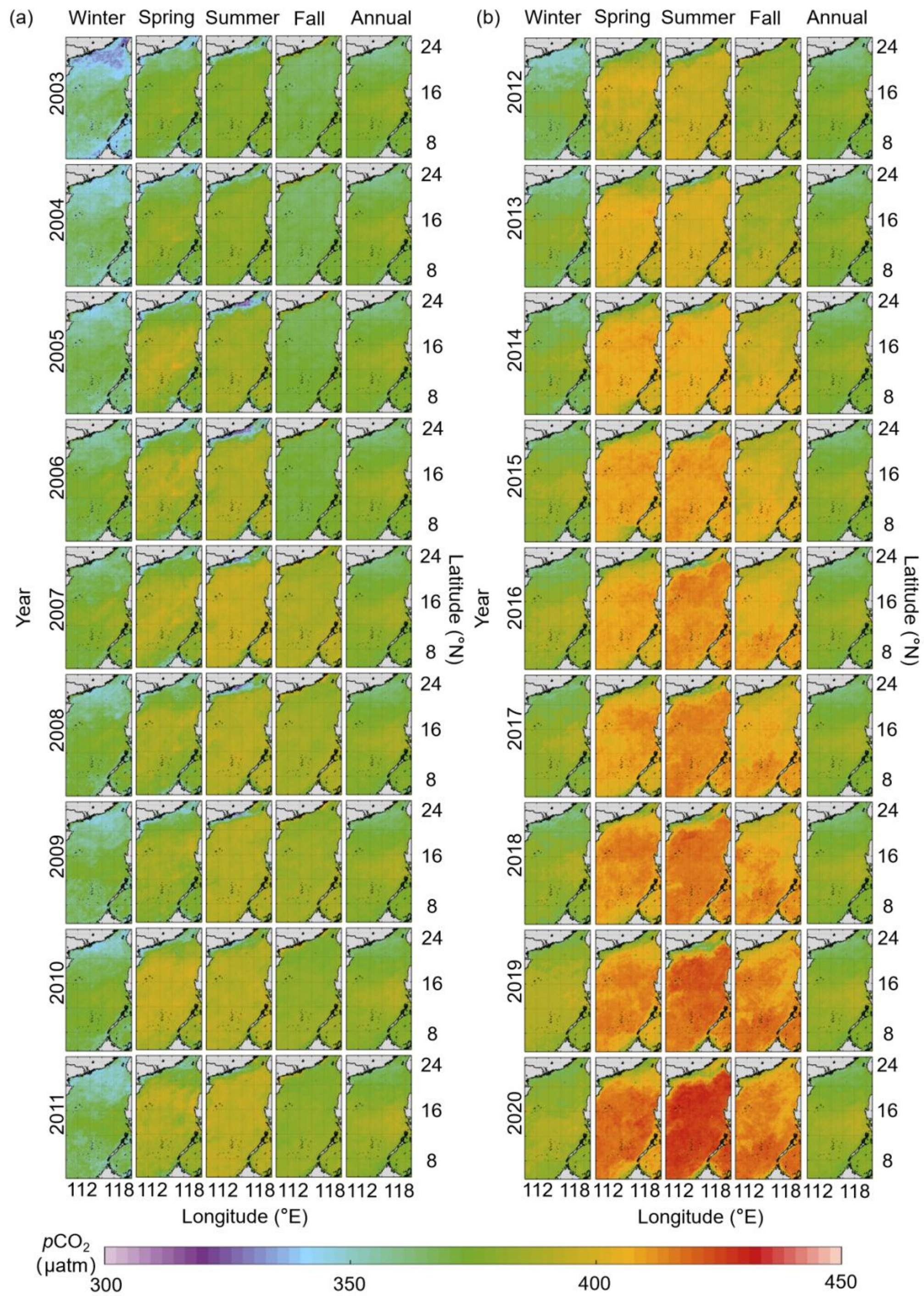
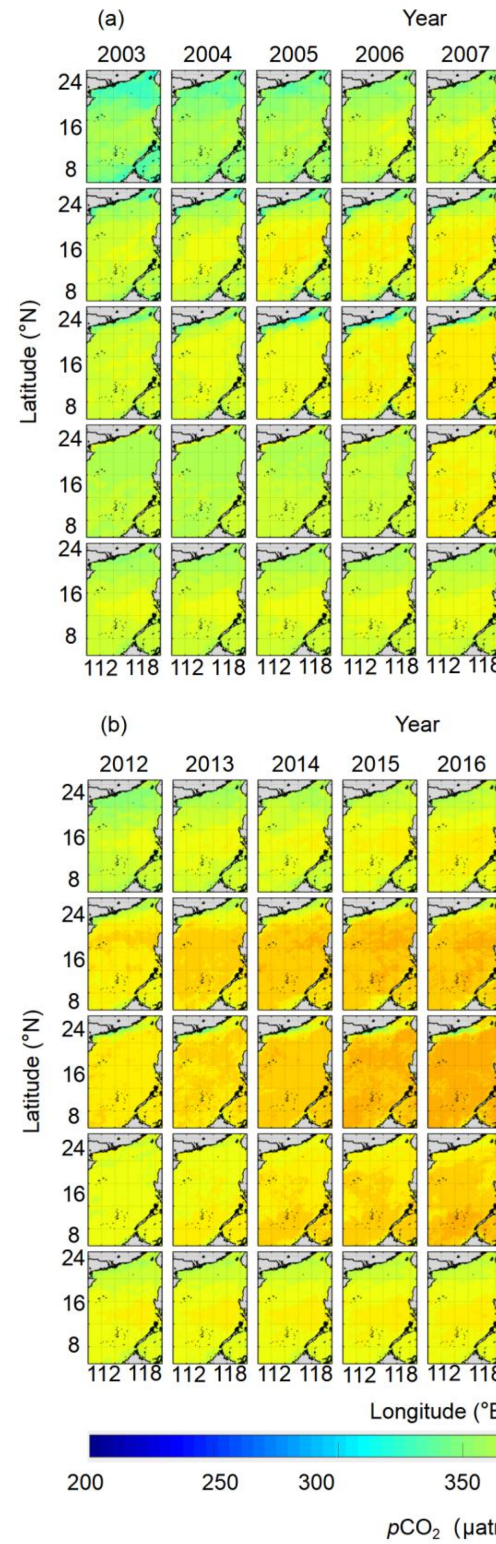


Figure 6. Reconstructed seasonal and annual pCO_2 fields in the South China Sea during the period 2003 to 2020 (a, 2003-2011; b, 2012-2020).



删除[作者]:

4.2 Model validation

Figure 7 compares the monthly reconstructed and in situ data. For the training dataset, the reconstructed pCO_2 fields of the four seasons fit the in situ data well (Fig. 7), with an average RMSE of 3.43 μatm and an average MAE of 2.14 μatm (Table 2). For the testing sets, although there are some outliers, most of the reconstructed pCO_2 data are consistent with in situ data, with RMSE averaging 10.79 μatm and MAE averaging 6.30 μatm. The R^2 of the testing set is ca. 0.91. In terms of MAPE, the accuracies of the four seasonal models are all around 99% (Table 2), with the highest value for spring data and the lowest value for summer data.

删除[作者]: field-observed

删除[作者]: field-observed

删除[作者]: ~

删除[作者]: field-observed

The relative large bias (14.67 μatm) in the summer may be the influence of relatively complex regional processes, such as river plumes and upwelling. The four evaluation metrics indicate that our reconstructed pCO_2 field is highly accurate in simulating both the training and testing sets.

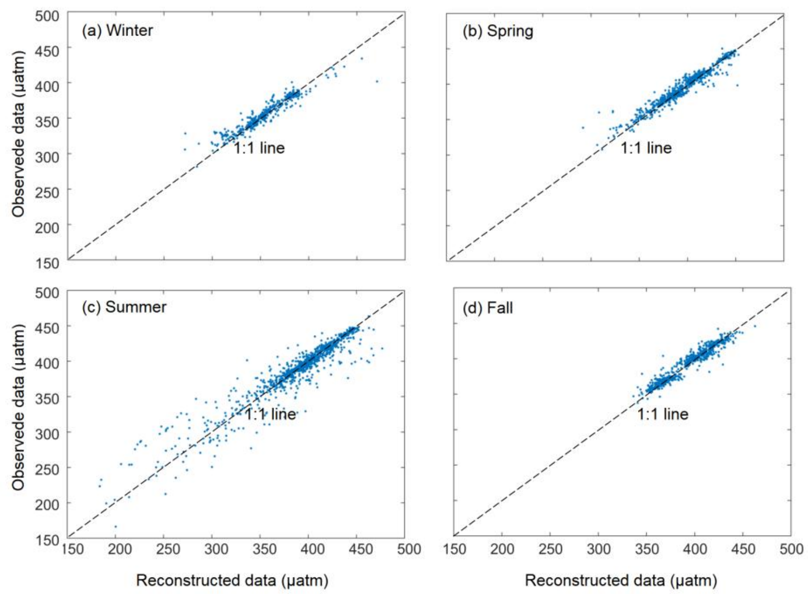
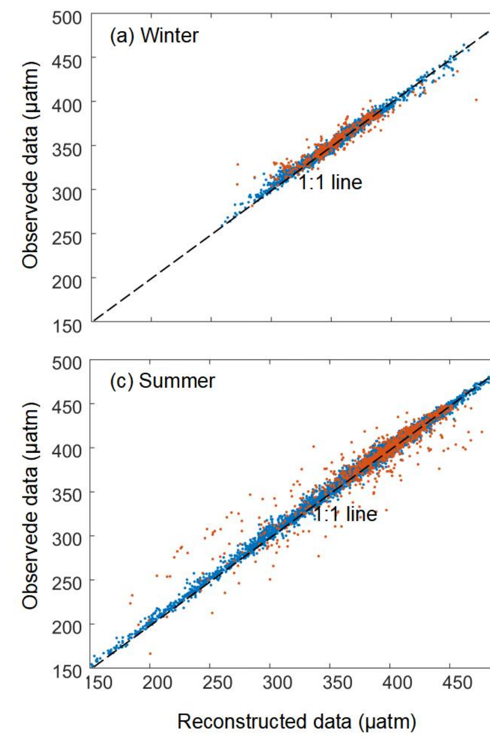


Figure 7. Comparisons between the monthly reconstructed and the in situ pCO_2 values for the testing set (monthly results were overlaid to the four seasons: (a) Winter: Dec., Jan., Feb.; (b) Spring: Mar., Apr., May; (c) Summer: Jun., Jul., Aug.; (d) Fall: Sept., Oct., Nov.).

The distribution pattern of the biases between the reconstructed fields and the in situ data in both training and testing datasets can be found in Figure 8. In terms of the temporal distribution pattern, the biases are concentrated mainly in summer. For the spatial distribution pattern, the biases in the northern coastal area are much greater than those in the basin. However, 95% of the biases are < ±10 μatm. Therefore, our reconstruction data exhibit relatively high accuracy.

删除[作者]: largestgreatest

删除[作者]:



删除[作者]: **observed**

删除[作者]: **Ttesing**

删除[作者]: field observations

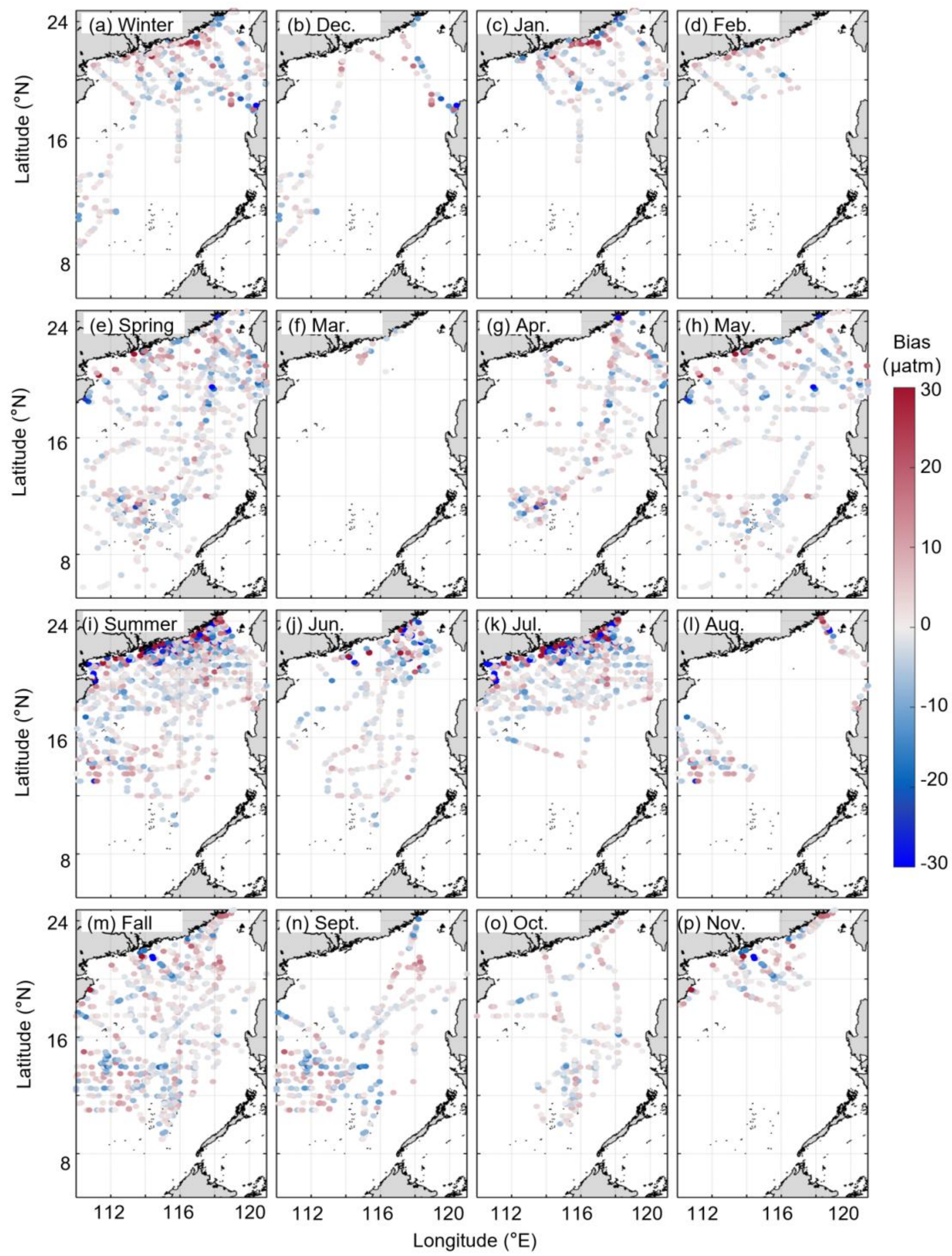
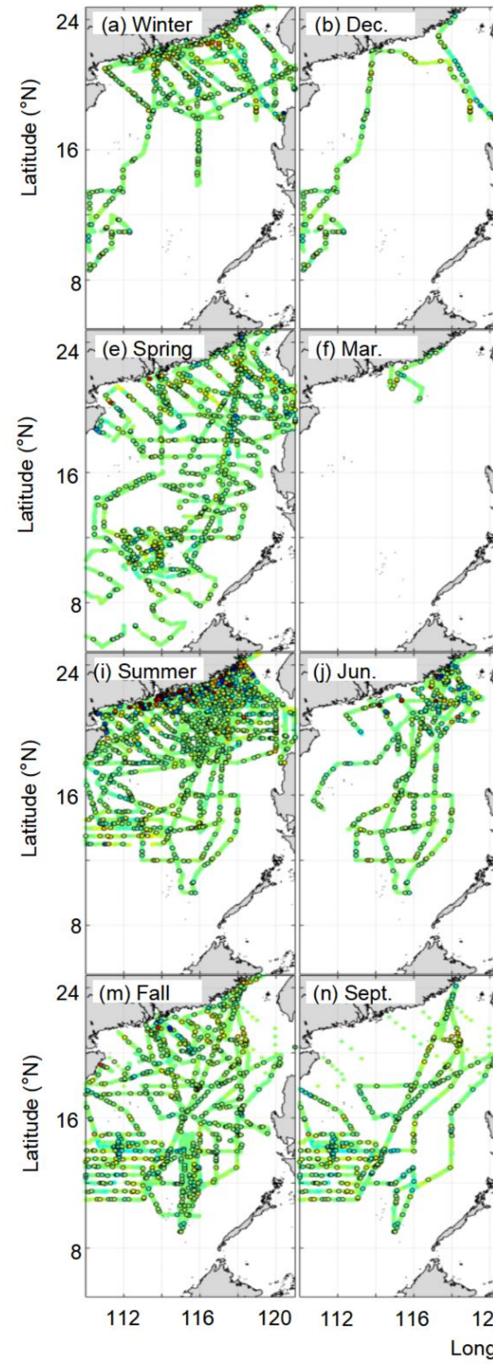


Figure 8. Differences between the seasonal and monthly reconstructed *p*CO_2 and the in situ *p*CO_2 data for the testing set (a. Winter; b. December; c. January; d. February; e. Spring; f. March; g. April; h. May; i. Summer; j. June; k. July; l. August; m. Fall; n. September; o. October; p. November).

Figure 9 shows the bias between our reconstructed fields and the four independent in situ datasets corresponding to the four seasons. This validation can verify the accuracy of the retrieval algorithm in data months without observations, namely the applicability of the retrieval algorithm extrapolation. This comparison shows that the retrieval algorithm is relatively accurate in the basin, with a near-zero bias (MAE: ~8 μatm, Fig. 9 a). The largest bias occurs in the Pearl River plume area in summer (~35 μatm). The retrieval algorithm also has high accuracy in the *p*CO_2 spatial variation trends, except in the Pearl River plume area in summer (22–20 °N), as shown in Fig. 9 b–e). The effect of the Pearl River plume on the *p*CO_2 spatial distribution in our retrieval algorithm is smaller than that shown by the in situ data. This is because at around the survey time (August 24–28, 2019), a large



删除[作者]:
删除[作者]: **observed**
删除[作者]: **w**
删除[作者]: **The open circles represent the difference**
删除[作者]: field observation
删除[作者]: e
删除[作者]: reconstruction model
删除[作者]: no
删除[作者]: e
删除[作者]: reconstruction model
删除[作者]: e
删除[作者]: reconstruction model
删除[作者]:

amount of precipitation (~30mm/day; https://psl.noaa.gov/data/gridded/data.ncep.reanalysis2.surface.html) occurred around the Pearl River estuary region (24–20 °N), which led to intensification of the Pearl River plume, such that the plume with relative low *p*CO_2 values eventually decreased the observed values. However, the monthly average runoff of the Pearl River during that month (August, 2019; http://www.pearlwater.gov.cn/; Pearl River Plume Index in Wang et al., 2022) is low, indicating that our retrieval algorithm is still highly reliable from the monthly average perspective. Thus, the inconsistency between the reconstructed (monthly average) and the in situ datasets is mainly due to the differences in the time scales of the remote sensing and the in situ data. The reconstructed data in this study were determined on a monthly scale, while the temporal resolution of the in situ data was on the order of hours. It is clear that relatively pronounced short-term changes in *p*CO_2, such as the diurnal variation caused by short-term heavy precipitation, cannot be reflected in the reconstructed data.

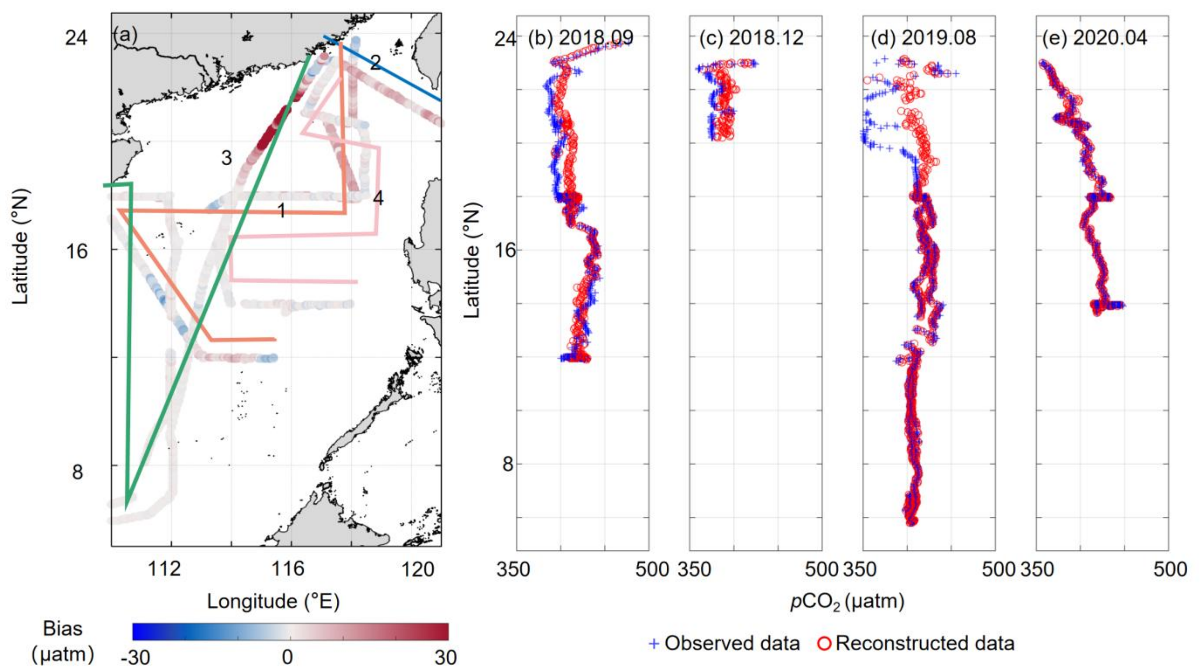


Figure 9. Difference between the reconstructed *p*CO_2 data and four independently testing in situ datasets during the four seasons. In (a), the numbers 1–4 represent September (2018.9, b), December 2018 (2018.12, c), August 2019 (2019.8, d), and April 2020 (2020.4, e), respectively.

Dai et al. (2022) produced a time-series of in situ data from 2003 to 2019 at the SEATs station, which we used here to validate the accuracy of the long-term trends of our model data (results shown in Fig. 10). The long-term trend of reconstructed *p*CO_2 data at the SEATs station are largely consistent with the in situ data, with differences mainly found before 2005. Thus, the long-term trend of our reconstructed model is also highly reliable.

删除[作者]: in preparation

删除[作者]: e

删除[作者]: reconstruction model

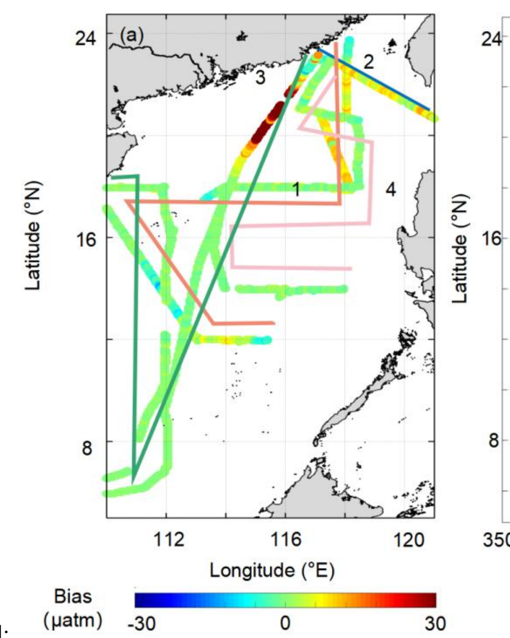
删除[作者]: a

删除[作者]: field-observed

删除[作者]: field-observed

删除[作者]: field-observed

删除[作者]:



删除[作者]: **observed**

删除[作者]:

删除[作者]: observed data

删除[作者]:

删除[作者]: observations

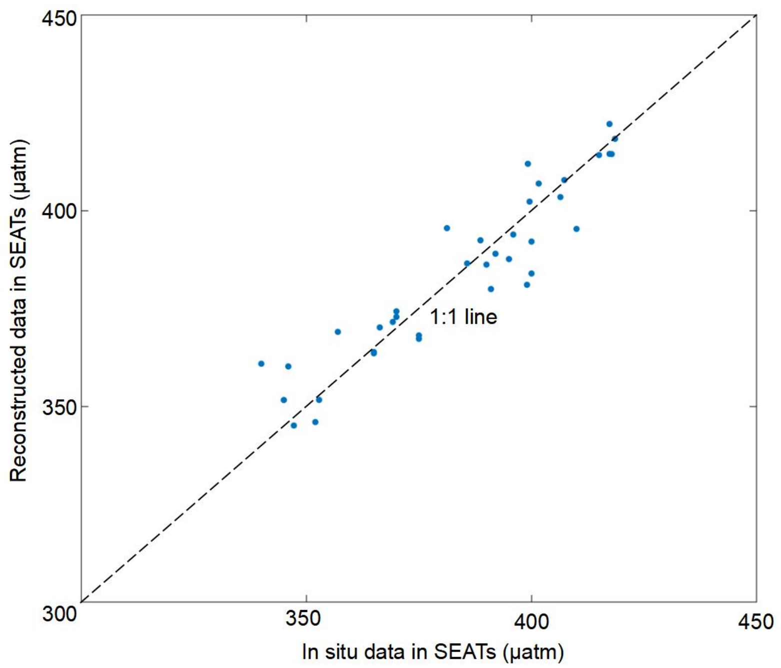


Figure 10. Comparison of the reconstructed *p*CO_2 with the in situ data at the Southeast Asia Time Series (SEATs) station (116◦ E, 18◦ N). The in situ data are from Dai et al. (2022), which were calculated from dissolved inorganic carbon and total alkalinity.

4.3 Uncertainties

As shown in Table 2, our reconstruction data have a high degree of accuracy, with an RMSE of ~10 μatm and MAE of ~6 μatm. For the uncertainty according to Equation 6, the bias of RS derived *p*CO_2 data used in the second term of Equation 6 is ~21 μatm (Table 2), the bias of SST is ~ 0.27°C (Qin et al., 2014), the bias of SSS is ~0.33 (Wang et al., 2022), and the bias of Chl-a is ~115% (Zhang et al., 2006). We then estimated *p*CO_2 changes due to these features' variability by constraining these features based on our model, and computed $\frac{\partial pCO2}{\partial Feature}$.

The overall uncertainty is greater in the coastal area (~13 μatm) than in the basin (~10 μatm) (Fig. 11 a). And this spatial pattern is mainly determined by the second term of Equation 6. The spatial distribution of the first term in Equation 6 (Fig. 11 b) calculated from a "max bias ratio" is consistent with that of *p*CO_2 (Fig. 11 b). The second term in Equation 6 (Fig. 11 c) is calculated from the propagation of bias of each variable (Fig. 11 c). The bias of Chl *a* (Fig. 11 f) shows the greatest effect on the reconstruction among these features (Fig. 11 f). Although the bias of the RS-derived *p*CO_2 data has relatively large bias, the final influence of its bias on the results from the retrieval algorithm is negligible due to the EOF method (Fig. 11 g).

删除[作者]: **observations**

删除[作者]: **observed data**

删除[作者]: In previous studies, RMSE and MAE were mostly used to represent the uncertainties in the reconstructed data.

删除[作者]: calculations,

删除[作者]: R1

删除[作者]: RS derived *p*CO_2

删除[作者]: , However, this expression of uncertainty ignores the sensitivity of the reconstructed model to the features; i.e., the bias that the features themselves pass to the reconstructed model are ignored. Moreover, it is clearly unreasonable to use a single RMSE or MAE value to represent the entire regio

删除[作者]: ing

删除[作者]: For example, for the <math> part, we only changed the value of SST, and kept the value of the other features constant, to calculate the effect of each additional unit of SST on the results of the pCO2 simulation.

删除[作者]: results of uncertainty can be found in Fig. 11. of the

删除[作者]: (Fig. 11 a)

删除[作者]: pCO2

删除[作者]: between

删除[作者]: R

删除[作者]:

删除[作者]: S derived pCO2

删除[作者]: retrieve algorithmreconstruction model

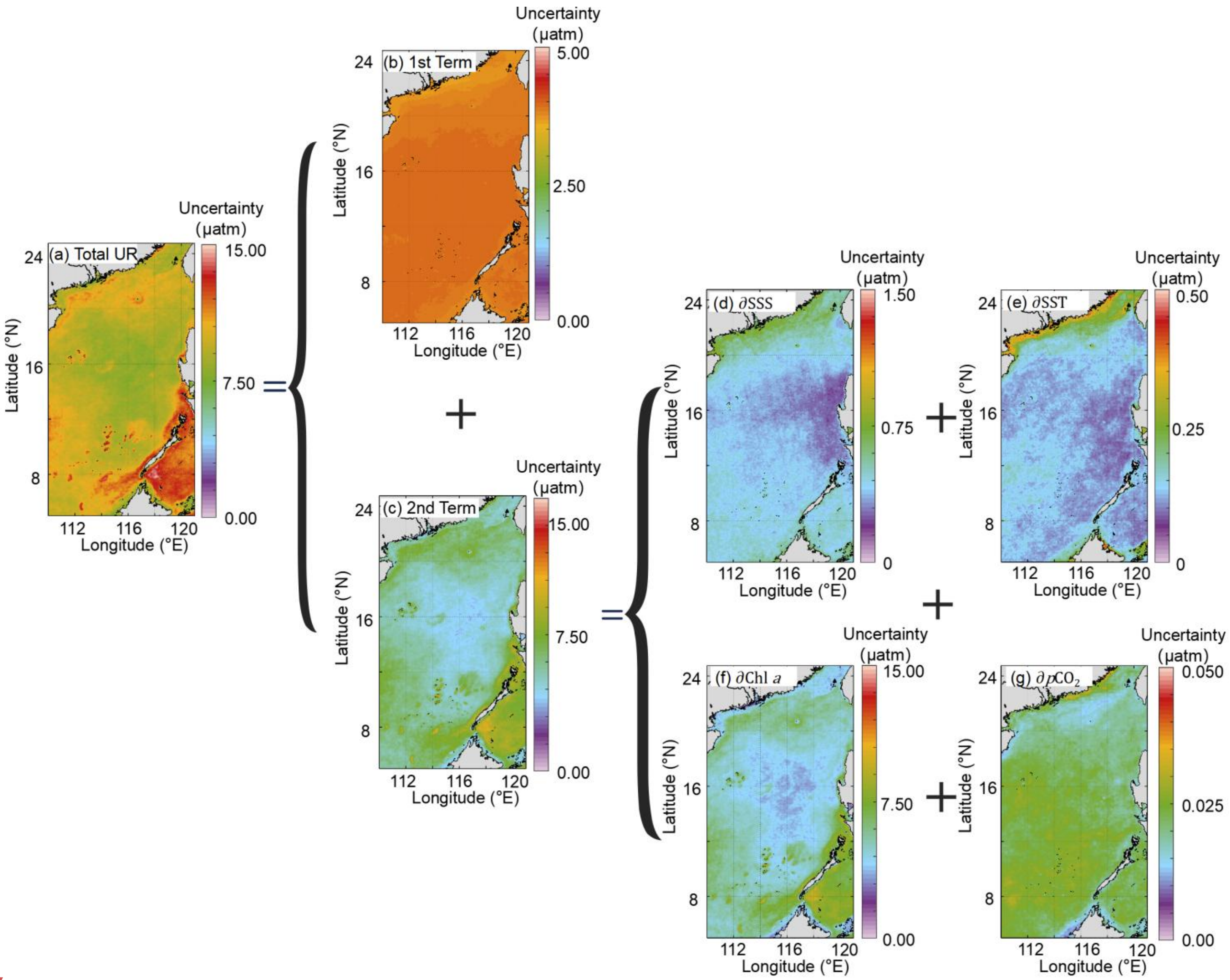
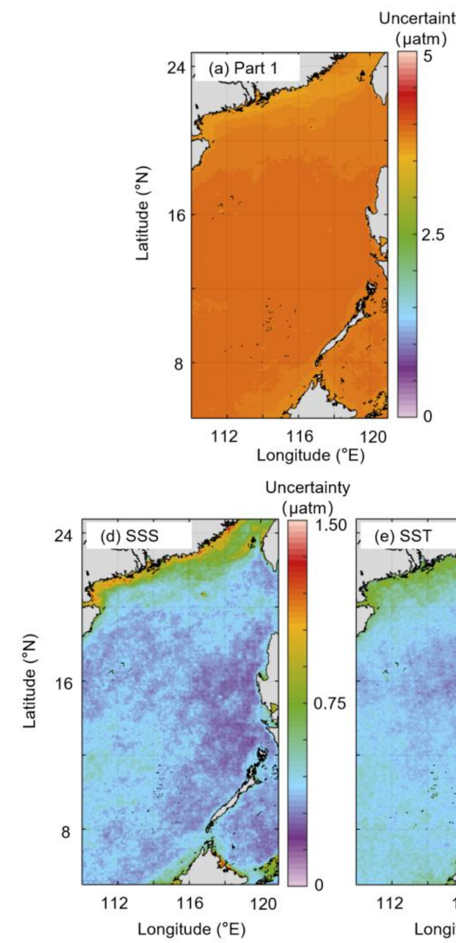


Figure 11. Uncertainties of the reconstructed pCO_2 fields (a, Total uncertainty in Equation 6; b. the first term of Equation 6; c. the second term of Equation 6; d. $(\frac{\partial pCO2}{\partial SSS})dSSS$ in the the second term of Equation 6; e. $(\frac{\partial pCO2}{\partial SST})dSST$ in the the second term of Equation 6; f. $(\frac{\partial pCO2}{\partial Chl\,a})dChl\,a$ in the the second term of Equation 6; g. $(\frac{\partial pCO2}{\partial RS\ derived\ pCO2})dRS\ derived\ pCO2$ in the the second term of Equation 6.

4.4 Spatial and temporal pCO_2 features

The climatological monthly reconstructed pCO_2 fields are shown in Figure 12. The highest values of the reconstructed pCO_2 fields occur in May and June, and the lowest value occurs in January. In winter, pCO_2 first decreases in December and then increases after January; the pCO_2 value is ca. 325 µatm in the northern coastal area, and ca. 350 µatm in the basin. In spring, pCO_2 gradually increases from the basin to the northern coastal area, and the basin high-value center gradually expands outward starting in April. In summer, pCO_2 gradually declines starting in June. In fall, pCO_2 increases from north to south, and the southern region shows consistently high values.

删除[作者]: These two parts were then added together to obtain the final uncertainty, and results are displayed in Figure 11. The uncertainties are greater in the coastal area (~13 µatm), than in the basin (~10 µatm). The spatial pattern of the uncertainty is consistent with that shown in Section 4.2.



删除[作者]:

删除[作者]: **stands for the**

删除[作者]: **stands for the**

删除[作者]: **stands for the**

删除[作者]: **stands for the**

删除[作者]: <math>

删除[作者]: **Figure 11. Uncertainties of the reconstructed pCO_2 fields.**

删除[作者]: in

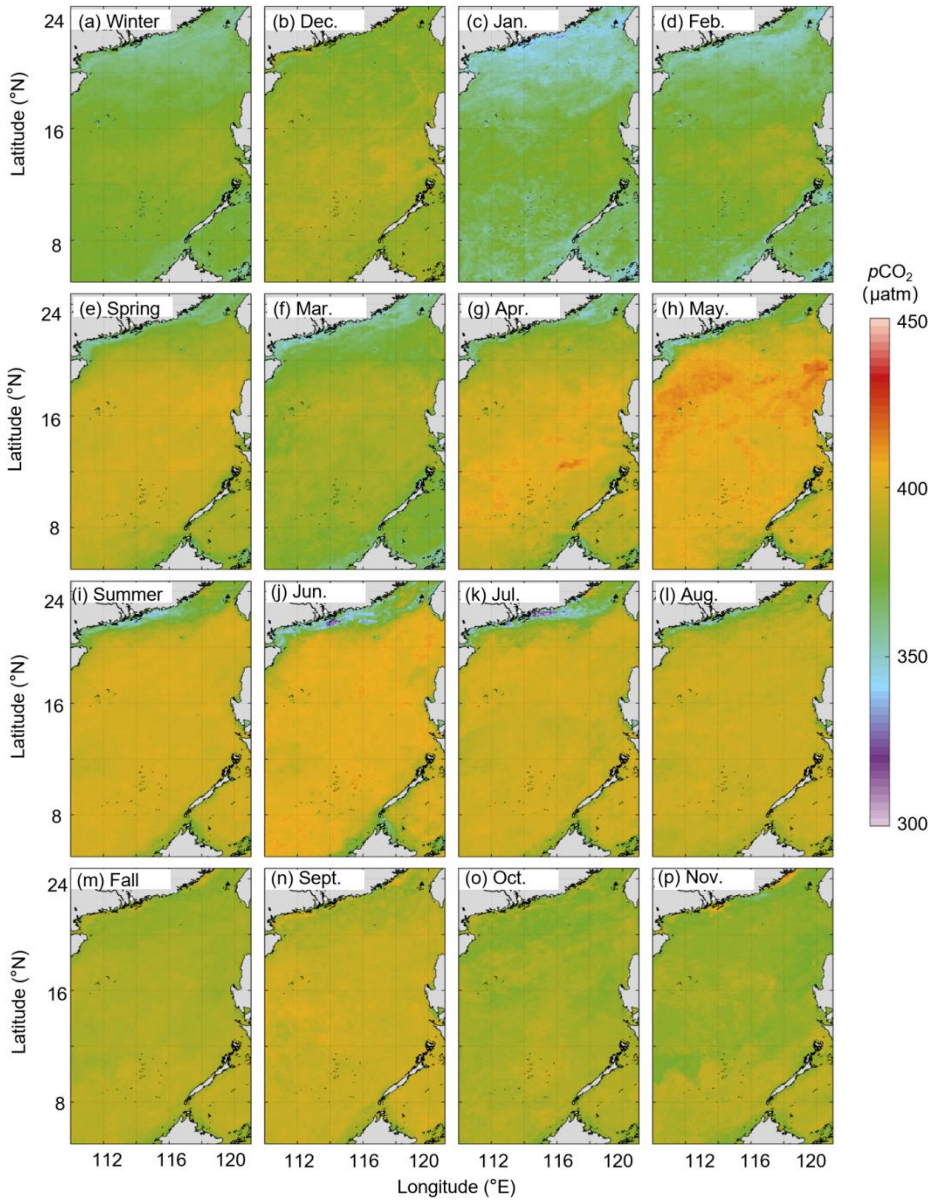
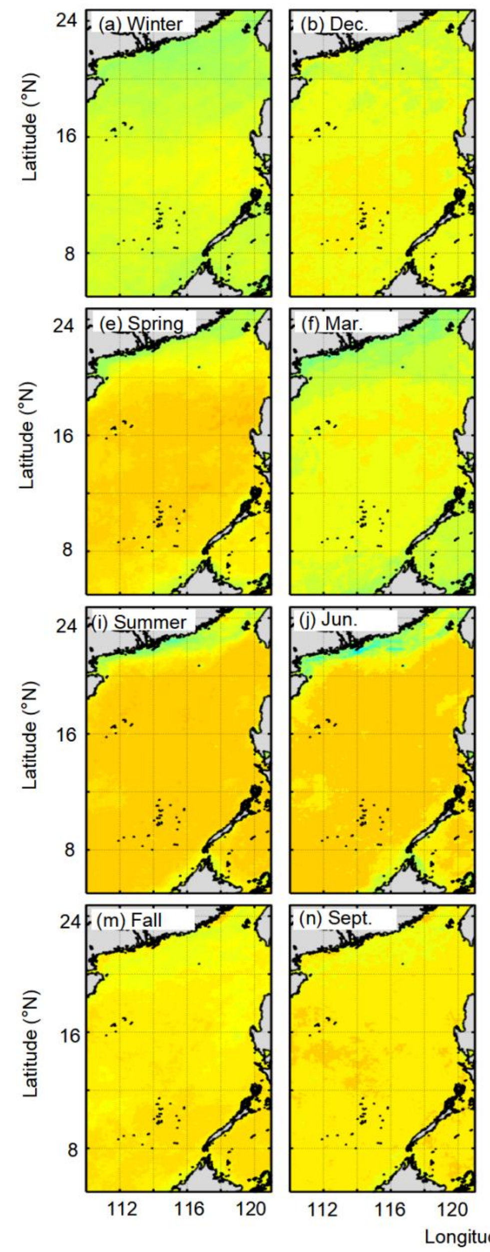


Figure 12. Long-term (2003–2020) seasonal and monthly average pCO_2 field (unit: μatm) (a. Winter; b. December; c. January; d. February; e. Spring; f. March; g. April; h. May; i. Summer; j. June; k. July; l. August; m. Fall; n. September; o. October; p. November).



删除[作者]:

删除[作者]: w

删除[作者]: .

To better show specific regions in the northern coastal area, we zoomed in on the reconstructed pCO_2 fields at locations north of 18°N (Fig. 13). The reconstructed pCO_2 fields successfully reflect the influence of the meso-small scale processes on pCO_2 in this northern coastal area of the SCS. For example, in winter, the relatively low pCO_2 values, which last into early spring, are mainly controlled by the low SST, and the high pCO_2 around Luzon Strait affected by winter upwelling. In summer, the reconstructed pCO_2 field shows that the influence of the Pearl River plume on pCO_2 is the strongest in July and lasts until September; it also effectively shows the influence of coastal upwelling in the northeastern shelf (~23°N, 117°E). Thus, our reconstructed pCO_2 fields clearly reflect the spatial pattern of the in situ pCO_2 (Fig. 3), which are generally consistent with previously reported patterns (Li et al., 2020; Zhai et al., 2013; Gan et al., 2010).

删除[作者]: field observed

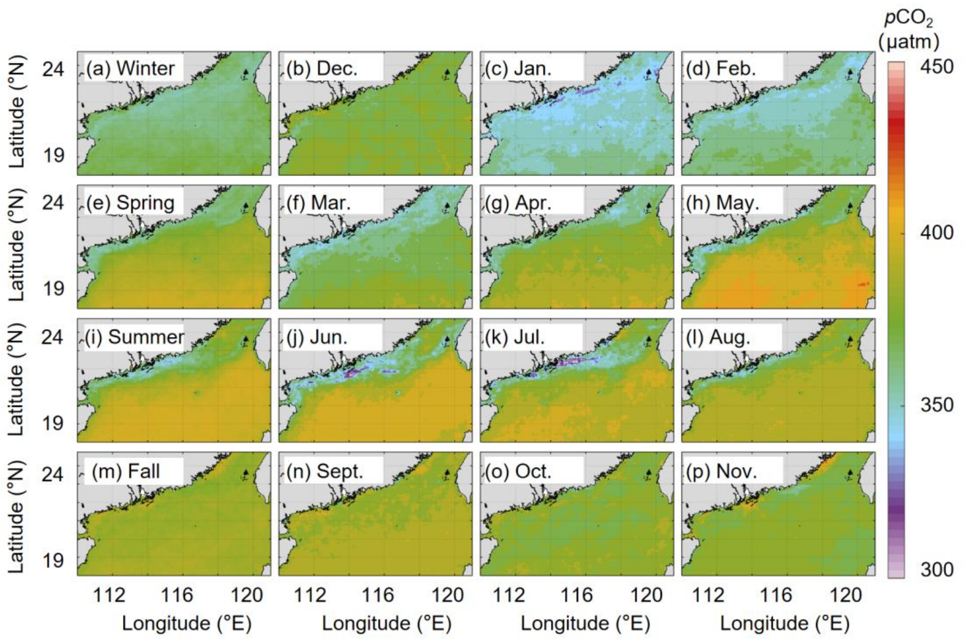


Figure 13. Long-term (2003–2020) seasonal and monthly averaged pCO$_2$ field in the region north of 18°N (unit: µatm) (a. Winter; b. December; c. January; d. February; e. Spring; f. March; g. April; h. May; i. Summer; j. June; k. July; l. August; m. Fall; n. September; o. October; p. November).

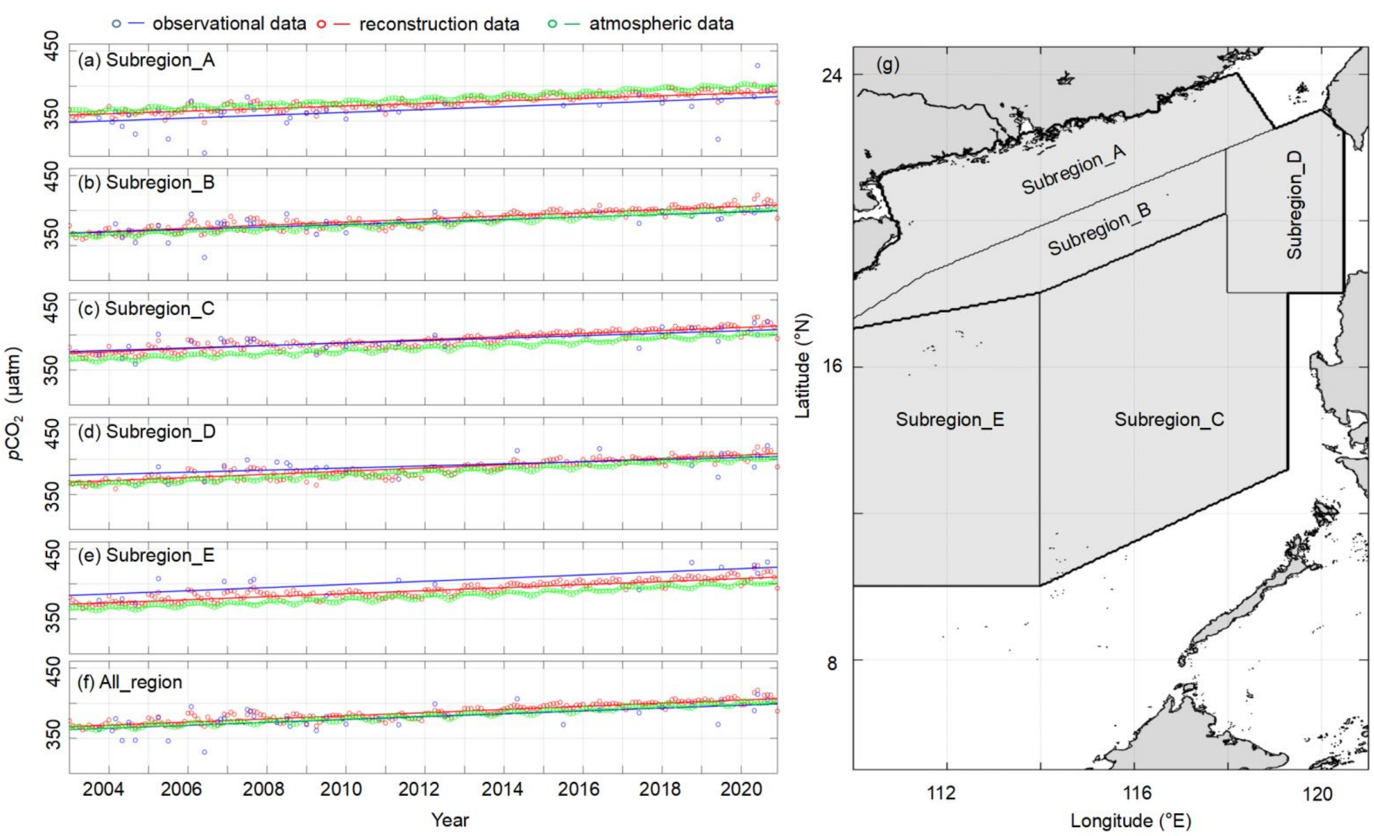
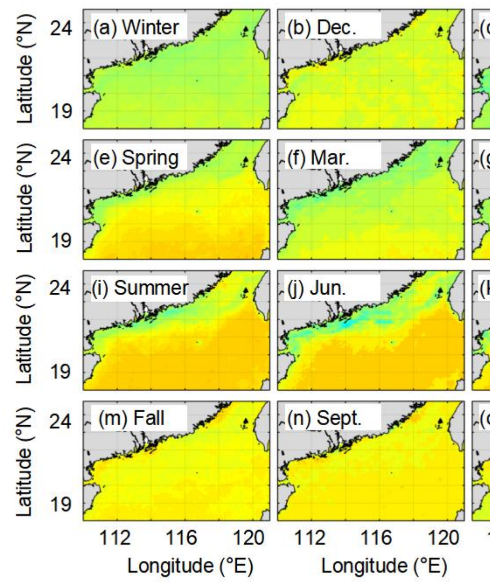


Figure 14. Time series of spatially averaged monthly pCO$_2$ data in five subregions (a-e) and the entire South China Sea (f)



删除[作者]:

删除[作者]: w

删除[作者]: S

under study. The sub-regions are shown in (g). The lines indicate the deseasonalized long-term trend of the spatially averaged monthly pCO_2 data for each sub-region with the slopes shown in Table 3. The deseasonalized method can be found in Landschützer et al. (2016).

Table 4. Deseasonalized long-term trend of the spatially averaged monthly pCO_2 data for each sub-region of the South China Sea. (unit: µatm yr^{-1}).

	All_region	Subregion_A	Subregion_B	Subregion_C	Subregion_D	Subregion_D
Reconstructed Data	2.12±0.17	1.82±0.14	2.23±0.12	2.17±0.12	2.20±0.13	2.16±0.13
In situ Data	2.10±0.79	1.80±0.86	1.73±0.84	1.81±0.85	1.41±1.16	2.13±1.10

We divided SCS into five sub-regions according to Li et al. (2020). In Fig.14, Subregion_A stands for the northern coastal area of the SCS, Subregion_B stands for the slope area of the northern SCS, Subregion_C stands for the SCS basin, Subregion_D stands for the region west of the Luzon Strait, and Subregion_E stands for the slope and basin area of the western SCS. "All_region" indicates the whole region containing the five sub-regions described above. We then calculated the deseasonalized long-term trend of spatially averaged monthly data for each sub-region, and the results are shown in Figure 14 and Table.3. This deseasonalized trend is consistent with that of in situ data, and its uncertainty is on the 95% confidence interval much lower than that shown by the in situ data. We can thus also infer that the long-term trend of our reconstructed data shows high reliability in all sub-regions, and that our data can serve as an important basis for predicting future changes of pCO_2 in the SCS.

In Fig.14 a-e, we found that the sea surface pCO_2 of the entire SCS is slightly higher than the atmospheric pCO_2, indicating that the SCS is a weak source of atmospheric CO_2. This conclusion is consistent with previous studies (e.g., Li et al., 2020). Moreover, compared to the rate of atmospheric CO_2 increase (~2.2 µatm yr-1), for Subregion_A, the pCO_2 trend is much slower than that of atmospheric pCO2, and the spatially averaged monthly mean pCO_2 is lower than the atmospheric pCO2. Thus, carbon accumulation in this region is expected to increase in the future. For rSubregion_C and Subregion_E, the spatially averaged monthly mean pCO_2 is higher than the atmospheric pCO2; thus, these two regions will still provide a weak source of atmospheric CO_2 in the future. Finally, whether Subregion_B and Subregion_D act as a source or sink of the atmospheric CO_2 is influenced by seasonal changes and physical processes. Subregion_B can be a zone of significant sink of atmospheric CO_2 as demonstrated by its low sea surface pCO_2 when the Pearl River plume spreads more widely in summer. In contrast, in winter when the Kuroshio intrusion is strong, both Subregions B and D have high sea surface pCO_2, indicating both subregions are sources of atmospheric CO_2.

5 Data availability

删除[作者]: ,

删除[作者]: .

删除[作者]: 3

删除[作者]: Observation

删除[作者]: region A

删除[作者]: region

删除[作者]: region

删除[作者]: region

删除[作者]: W

删除[作者]: region

删除[作者]: observational data

删除[作者]: observational data

删除[作者]: region

删除[作者]: egions

删除[作者]: regions

删除[作者]: zone

删除[作者]: is spreading

删除[作者]: into a wider spatial coverage

删除[作者]: s

删除[作者]: When the Pearl River plume is relatively strong in summer, resulting in relatively low pCO_2 in Sub_region B, this sub_region turns into a sink of atmospheric CO2. When the Kuroshio invasion or water mixing is strong in winter, resulting in relatively high pCO_2 in Sub_region B and Sub_region D, both two sub_regional turn into a source of atmospheric CO2.

The data (the reconstructed CO_2 data, the in situ CO_2 data before 2018 (0.5°*0.5°), and the remote sensing derived CO_2 data) for this paper are available under the link https://doi.org/10.57760/sciencedb.02050 (Wang & Dai, 2022).

删除[作者]: Observational

删除[作者]: https://github.com/Elricriven/co2data

删除[作者]: et al.,

6 Conclusions

Based on the machine learning method, we reconstructed the sea surface pCO_2 fields in the SCS with high spatial resolution (0.05*0.05°) over the last two decades (2003-2020) by calculating the statistical relationship between the in situ pCO_2 data and remote sensing-derived data. The input data we used in machine learning include remote sensing derived data (sea surface salinity, sea surface temperature, chlorophyll), the spatial patterns of pCO_2 calculated by EOF, atmospheric CO_2, and time labels (month). The machine learning method (CATBOOST) used in this study was facilitated by the EOF method, because the latter can provide spatial constraints for the data reconstruction. In addition to the typical machine learning performance metrics, we present a novel uncertainty calculation method that incorporates the bias of both the reconstruction and the sensitivity of reconstructed models to its features. This method effectively shows the spatiotemporal patterns of bias, and makes up for the spatial representation of the typical performance metrics.

删除[作者]:

删除[作者]: underway observational

删除[作者]:

删除[作者]: s

We validate our reconstruction with three independent testing datasets, and the results show that the bias between our reconstruction and in situ pCO_2 data in the SCS is relatively small (about 10 μatm). Our reconstruction successfully shows the main features of the spatial and temporal patterns of pCO_2 in the SCS, indicating that we can use these reconstructed data to further analyze the effect of meso-microscale processes (e.g., the Pearl River plume, and CCC) on sea surface pCO_2 in the SCS.

删除[作者]: observational

删除[作者]: s

We divided the SCS into five sub-regions and separately calculated the deseasonalized long term trend of pCO_2 in each subregion, and compared them with the long-term trend of atmospheric pCO_2. Our results show that the reconstructed data are consistent with those of in situ data. Moreover, the strength of the CO_2 sink in the northern SCS shows an increasing trend, whereas pCO_2 trends in other subregions are essentially the same as that of atmospheric pCO_2.

删除[作者]: observational data

This high spatiotemporal resolution of sea surface pCO_2 data is helpful to clarify the controlling factors of pCO_2 change in the SCS and may be useful to predict changes of CO_2 source or sink patterns in this system.

删除[作者]: pCO_2

Author contribution

Minhan Dai conceptualized and directed the field program of in situ observations. Xianghui Guo and Yi Xu participated in the in situ data collection. Yan Bai provided the remote sensing-derived pCO_2 data. Minhan Dai, Guizhi Wang and Zhixuan Wang developed the reconstruction method, wrote the codes, analyzed the data, and plotted the figures. Zhixuan Wang wrote the manuscript. Minhan Dai, Xianghui Guo and Guizhi Wang contributed to the writing, editing and revision of the original manuscript.

Competing interests

The authors declare that they have no conflict of interest.

Acknowledgements

We thank the support of the National Natural Science Foundation of China (grant No. 42188102, 42141001, and 41890800), and the National Basic Research Program of China (973 Program, grant No. 2015CB954000).

References

Bai, Y., Cai, W., He, X., Zhai, W., Pan, D., Dai, M., and Yu, P.: A mechanistic semi-analytical method for remotely sensing sea surface pCO_2 in river-dominated coastal oceans: A case study from the East China Sea, J. Geophys. Res.: Oceans, 120, 2331-2349, 2015.

Bakker, D., Pfeil, B., Landa, C., Metzl, N., and Xu, S.: A multi-decade record of high-quality fCO_2 data in version 3 of the Surface Ocean CO2 Atlas (SOCAT), Earth Syst. Sci. Data, 8, 383-413, 2016.

Borges, A. V., Delille, B., and Frankignoulle, M.: Budgeting sinks and sources of CO_2 in the coastal ocean: Diversity of ecosystems counts, Geophys. Res. Lett., 32, L14601, 2005.

Cao, Z. and Dai, M.: Shallow-depth $CaCO_3$ dissolution: Evidence from excess calcium in the South China Sea and its export to the Pacific Ocean, Global Biogeochem. Cy., 25, GB2019, 2011.

Cao, Z., Dai, M., Zheng, N., Wang, D., Li, Q., Zhai, W., Meng, F., and Gan, J.: Dynamics of the carbonate system in a large continental shelf system under the influence of both a river plume and coastal upwelling, J. Geophys. Res.: Biogeo., 116, G02010, 2011.

Cao, Z., Yang, W., Zhao, Y., Guo, X., Yin, Z., Du, C., Zhao, H., and Dai, M.: Diagnosis of CO_2 dynamics and fluxes in global coastal oceans, Natl. Sci. Rev., 7, 786-797, 2020.

Chen, C. and Borges, A. V.: Reconciling opposing views on carbon cycling in the coastal ocean: Continental shelves as sinks and near-shore ecosystems as sources of atmospheric CO_2, Deep-Sea Res. I, 56, 578-590, 2009.

Chen, C., Lai, Z., Beardsley, R. C., Xu, Q., Lin, H., and Viet, N. T.: Current separation and upwelling over the southeast shelf of Vietnam in the South China Sea, J. Geophys. Res.: Oceans, 117, C03033, 2012.

Chen, F., Cai, W. J., Benitez-Nelson, C., and Wang, Y.: Sea surface pCO_2？-SST relationships across a cold-core cyclonic eddy: Implications for understanding regional variability and air-sea gas exchange, Geophys. Res. Lett., 341, 265-278, 2007.

Cheng, C., Xu, P. F., Cheng, H., Ding, Y., Zheng, J., Ge, T., and Xu, J.: Ensemble learning approach based on stacking for unmanned surface vehicle's dynamics. Ocean Eng., 207, 107388, 2020.

Dai, M. H., Cao, Z., Guo, X., Zhai, W., Liu, Z., Yin, Q., Xu, Y., Gan, J., Hu, J., and Du, C.: Why are some marginal seas sources of atmospheric CO_2?, Geophys. Res. Lett., 40, 2154-2158, 2013.

Dai, M., Gan, J., Han, A., Kung, H., and Yin, Z.: "Physical Dynamics and Biogeochemistry of the Pearl River Plume"" in Biogeochemical Dynamics at Large River-Coastal Interfaces. Eds. T. Bianchi, M. Allison and W. J. Cai (Cambridge University Press, Cambridge), 321-352, 2014.

Dai, M., J. Su, Zhao,Y., Hofmann, E. E., Cao, Z., Cai, W., Gan, J., Lacroix, F., Laruelle, G., Meng, F., Müller, J., Regnier, P., Wang, G., and Wang, Z.: Carbon fluxes in the coastal ocean: Synthesis, boundary processes and future trends, Annu. Rev. Earth Pl. Sc., 50, 593-626, 2022.

Du, C., Liu, Z., Dai, M., Kao, S. J., and Li, Y.: Impact of the Kuroshio intrusion on the nutrient inventory in the upper northern South China Sea: insights from an isopycnal mixing model, Biogeosciences, 10, 6419-6432, 2013.

Dong, L., Su, J. Wong, L. Cao, Z. and Chen, J.: Seasonal variation and dynamics of the Pearl River plume, Cont. Shelf Res., 24, 1761-1777, 2004.

Dye, A. W., Rastogi, B., Clemesha, R. E. S., Kim, J. B., Samelson, R. M., Still, C. J., & Williams, A. P.: Spatial patterns and trends of summertime low cloudiness for the Pacific Northwest, 1996–2017. Geophysical Research Letters, 47, e2020GL088121, 2020.

Fassbender, A. J., Rodgers, K. B., Palevsky, H. I., and Sabine, C. L.: Seasonal Asymmetry in the Evolution of Surface Ocean *p*CO_2 and pH Thermodynamic Drivers and the Influence on Sea - Air CO_2 Flux, Global Biogeochem. Cy., 32, 1476-1497, 2018.

Fay, A., Gregor, L., Landschützer, P., McKinley, G., Gruber, N., Gehlen, M., Iida, Y., Laruelle, G., Rödenbeck, C., Roobaert, A., and Zeng, J.: SeaFlux: harmonization of air-sea CO2 fluxes from surface *p*CO_2 data products using a standardized approach, Earth Syst. Sci. Data, 13, 4693-4710, 2021

Friedlingstein, P., Jones, M. W., O'Sullivan, M., Andrew, R. M., Hauck, J., Peters, G. P., Peters, W., Pongratz, J., Sitch, S., Le Quéré, C., Bakker, D. C. E., Canadell, J. G., Ciais, P., Jackson, R. B., Anthoni, P., Barbero, L., Bastos, A., Bastrikov, V., Becker, M., Bopp, L., Buitenhuis, E., Chandra, N., Chevallier, F., Chini, L. P., Currie, K. I., Feely, R. A., Gehlen, M., Gilfillan, D., Gkritzalis, T., Goll, D. S., Gruber, N., Gutekunst, S., Harris, I., Haverd, V., Houghton, R. A., Hurtt, G., Ilyina, T., Jain, A. K., Joetzjer, E., Kaplan, J. O., Kato, E., Klein Goldewijk, K., Korsbakken, J. I., Landschützer, P., Lauvset, S. K., Lefèvre, N., Lenton, A., Lienert, S., Lombardozzi, D., Marland, G., McGuire, P. C., Melton, J. R., Metzl, N., Munro, D. R., Nabel, J. E. M. S., Nakaoka, S.-I., Neill, C., Omar, A. M., Ono, T., Peregon, A., Pierrot, D., Poulter, B., Rehder, G., Resplandy, L., Robertson, E., Rödenbeck, C., Séférian, R., Schwinger, J., Smith, N., Tans, P. P., Tian, H., Tilbrook, B., Tubiello, F. N., van der Werf, G. R., Wiltshire, A. J., and Zaehle, S.: Global Carbon Budget 2019, Earth Syst. Sci. Data, 11, 1783-1838, 2019.

Gan, J., Li, H., Curchitser, E. N., and Haidvogel, D. B.: Modeling South China sea circulation: Response to seasonal forcing

删除[作者]:

删除[作者]: pCO_2

regimes, J. Geophys. Res.: Oceans, 111, C06034, 2006.

Gan, J., Li, L., Wang, D., and Guo, X.: Interaction of a river plume with coastal upwelling in the northeastern South China Sea, Cont. Shelf Res., 29, 728-740, 2009.

Gan, J., Lu, Z., Dai, M., Cheung, A. Y. Y., Liu, H., and Harrison, P.: Biological response to intensified upwelling and to a river plume in the northeastern South China Sea: A modeling study, J. Geophys. Res.: Oceans, 115, C09001, 2010.

Guo, X. and Wong, G.: Carbonate chemistry in the Northern South China Sea shelf-sea in June 2010, Deep Sea Res. II, 117, 119-130, 2015.

Han, A. Q., Dai, M. H., Gan, J. P., Kao, S. J., Zhao, X. Z., Jan, S., Li, Q., Lin, H., Chen, C. T. A., and Wang, L.: Inter-shelf nutrient transport from the East China Sea as a major nutrient source supporting winter primary production on the northeast South China Sea shelf, Biogeosciences, 10, 8159-8170, 2013.

Hu, J., Kawamura, H., Li, C., Hong, H., and Jiang, Y.: Review on current and seawater volume transport through the Taiwan Strait, J. Oceanogr., 66, 591-610, 2010.

Jones, S. D., Quéré, C., and Rödenbeck, C.: Spatial decorrelation lengths of surface ocean fCO_2 results in NetCDF format, Global Biogeochem. Cy., 26, GB2042, 2014.

Jo, Y., Dai, M., Zhai, W., Yan, X., and Shang, S.: On the Variations of Sea Surface pCO_2 in the Northern South China Sea - A Remote Sensing Based Neural Network Approach, J. Geophys. Res.: Oceans, 117, C08022, 2012.

删除[作者]: pCO2

Landschützer, P., Gruber, N., and Bakker, D.: Decadal variations and trends of the global ocean carbon sink, Global Biogeochem. Cy., 30, 1396-1417, 2016.

Landschützer, P., Gruber, N., and Bakker, D. C. E.: An updated observation-based global monthly gridded sea surface pCO_2 and air-sea CO_2 flux product from 1982 through 2015 and its monthly climatology, Dataset, 2017.

Laruelle, G., Lauerwald, R., Pfeil, B., and Regnier, P.: Regionalized global budget of the CO_2 exchange at the air-water interface in continental shelf seas, Global Biogeochem. Cy., 28, 1199-1214, 2015.

Landschützer, P., Bakker, D. C. E., Gruber, N., and Schuster, U.: Recent variability of the global ocean carbon sink, Global Biogeochem. Cy., 28, 927-949, 2014.

Lefèvre, N., Watson, A., and Waston, A.,: A comparison of multiple regression and neural network techniques for mapping in situ pCO_2 data, Tellus B, 57, 375-384, 2005.

Lefévre, N., Watson, A. J., and Watson, A. R.: A comparison of multiple regression and neural network techniques for mapping in situ pCO_2 data, Tellus B: Chemical and Physical Meteorology, Dataset, 2017.

Levitus, S., Antonov, J. I., Boyer, T. P., Garcia, H. E., and Locarnini, R. A.: EOF analysis of upper ocean heat content, 1956–2003, Geophys. Res. Lett., 32, L18607, 2005.

Li, Y., Xie, P., Tang, Z., Jiang, T., and Qi, P.: SVM-Based Sea-Surface Small Target Detection: A False-Alarm-Rate-Controllable

Approach, IEEE Geosc.i Remote S., 16, 1225-1229, 2019.

Li, H., Wiesner, M. G., Chen, J., Lin, Z., Zhang, J., and Ran, L.: Long-term variation of mesopelagic biogenic flux in the central South China Sea: Impact of monsoonal seasonality and mesoscale eddy, Deep Sea Res. I, 126, 62-72, 2017.

Li, Q., Guo, X., Zhai, W., Xu, Y., Dai, M.: Partial pressure of CO_2 and air-sea CO_2 fluxes in the South China Sea: Synthesis of an 18-year dataset, Prog. Oceanogr., 182, 102272, 2020.

Luo, X., Hao, W., Zhe, L., and Liang, Z.: Seasonal variability of air-sea CO_2 fluxes in the Yellow and East China Seas: A case study of continental shelf sea carbon cycle model, Cont. Shelf Res., 107, 69-78, 2015.

McMonigal, K., & Larson, S. M.: ENSO explains the link between Indian Ocean dipole and Meridional Ocean heat transport. Geophysical Research Letters, 49, e2021GL095796, 2022.

Mongwe, N. P., Chang, N., and Monteiro, P.: The seasonal cycle as a mode to diagnose biases in modelled CO_2 fluxes in the Southern Ocean, Ocean Model., 106, 90-103, 2016.

Park, J. H.: Effects of Kuroshio intrusions on nonlinear internal waves in the South China Sea during winter, J. Geophys. Res.: Oceans, 118, 7081-7094, 2013.

Qin, H., Chen, G., Wang, W., Wang, D., and Zeng, L.: Validation and application of MODIS-derived SST in the South China Sea, Int. J. Remote Sens., 35, 4315-4328, 2014.

Rödenbeck, C., Bakker, D. C. E., 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, 14-49, 2015.

Sheu, D. D., Chou, W. C., Wei, C. L., Hou, W. P., Wong, G., and Hsu, C. W.: Influence of El Niño the sea-to-air CO_2 flux at the SEATs time-series site, northern South China Sea, J. Geophys. Res.: Oceans, 115, C10021, 2010.

Tahata, M., Sawaki, Y., Ueno, Y., Nishizawa, M., Yoshida, N., Ebisuzaki, T., Komiya, T., and Maruyama, S.: Three-step modernization of the ocean: Modeling of carbon cycles and the revolution of ecological systems in the Ediacaran/Cambrian periods, Geosci. Front., 6, 121-136, 2015.

Telszewski, M., Chazottes, A., Schuster, U., Watson, A. J., and Wanninkhof, R.: Estimating the monthly $p$$CO_2$ distribution in the North Atlantic using a self-organizing neural network, Biogeosciences, 6, 1405-1421, 2009.

Wang, G., Shen, S. S. P., Chen, Y., Bai, Y., Qin, H., Wang, Z., Chen, B., Guo, X., and Dai, M.: Feasibility of reconstructing the basin-scale sea surface partial pressure of carbon dioxide from sparse in situ observations over the South China Sea, Earth Syst. Sci. Data, 13, 1403-1417, 2021.

Wanninkhof, R., Park, G. H., Takahashi, T., Sweeney, C., Feely, R., Nojiri, Y., Gruber, N., Doney, S. C., Mckinley, G. A., and Lenton, A.: Global ocean carbon uptake: magnitude, variability and trend, Biogeosciences, 10, 1983-2000, 2013

Wang, Z. and Dai, M.: Datasets of reconstructed sea surface pCO2 in the South China Sea, Science Data Bank,

删除[作者]: , Wang, G., Guo, X., Bai, Y., Xu, Y.,

删除[作者]: Spatial reconstruction of long-term (2003-2020) sea surface pCO2 in the South China Sea using a machine learning based regression method aided by empirical orthogonal function analysis

删除[作者]: Github

https://doi.org/10.57760/sciencedb.02050

Wang, Z., Wang, G., Guo, X., Hu, J., and Dai, M. Reconstruction of High-Resolution Sea Surface Salinity over 2003–2020 in the South China Sea Using the Machine Learning Algorithm LightGBM Model. Remote. Sens., 14, 6147, 2022. https://doi.org/10.3390/rs14236147.

Xu, X., Zang, K., Zhao, H., Zheng, N., Huo, C., and Wang, J.: Monthly CO_2 at A4HDYD station in a productive shallow marginal sea (Yellow Sea) with a seasonal thermocline: Controlling processes, J. Marine Syst., 159, 89-99, 2016.

Jo, Y., Dai, M., Zhai, W., Yan, X., and Shang, S.: On the variations of sea surface pCO_2 in the northern South China Sea: A remote sensing based neural network approach, J. Geophys. Res.: Oceans, 117, C08022, 2012.

Yang, W., Guo, X., Cao, Z., Wang, L., Guo, L., Huang, T., Li, Y., Xu, Y., Gan, J., and Dai, M.: Seasonal dynamics of the carbonate system under complex circulation schemes on a large continental shelf: The northern South China Sea, Prog Oceanogr., 197, 1026-1045, 2021.

Yu, S., Song, Z., Bai, Y., and He, X.: Remote Sensing based Sea Surface partial pressure of CO2 (pCO_2) in China Seas (2003-2019) (2.0). Zenodo, 2022. https://doi.org/10.5281/zenodo.7372479.

Yu, Z., Shang, S., Zhai, W., and Dai, M.: Satellite-derived surface water pCO_2 and air-sea CO_2 fluxes in the northern South China Sea in summer, Prog. Nat. Sci., 19, 775-779, 2009.

Zeng, J., Matsunaga, T., Saigusa, N., Shirai, T., Nakaoka, S. I., and Tan, Z. H.: Technical note: Evaluation of three machine learning models for surface ocean CO_2 mapping, Ocean Sci., 13, 303-313, 2017.

Zhai, W., Dai, M., Cai, W. J., Wang, Y., and Hong, H.: The partial pressure of carbon dioxide and air-sea fluxes in the northern South China Sea in spring, summer and fall, Mar. Chem., 96, 87-97, 2005.

Zhai, W. D., Dai, M. H., Chen, B. S., Guo, X. H., Li, Q., Shang, S. L., Zhang, C. Y., Cai, W. J., and Wang, D. X.: Seasonal variations of sea-air CO_2 fluxes in the largest tropical marginal sea (South China Sea) based on multiple-year underway measurements, Biogeosciences, 10, 7775-7791, 2013.

Zhang, C., Hu, C., Shang, S., Müller-Karger, F., Yan, L., Dai, M., Huang, B., Ning, X., and Hong, H.: Bridging between SeaWiFS and MODIS for continuity of chlorophyll-a concentration assessments off Southeastern China, Remote Sens. Environ., 102, 250-263, 2006.

Zhan, Y., Zhang, H., Li, J., and Li, G.: Prediction Method for Ocean Wave Height Based on Stacking Ensemble Learning Model. J. Mar. Sci. Eng., 10, 1150, 2022.

Zhu, Y., Shang, S., Zhai, W., and Dai, M.: Satellite-derived surface water pCO_2 and air-sea CO_2 fluxes in the northern South China Sea in summer, Prog. Nat. Sci., 19, 775-779, 2009.

删除[作者]: https://github.com/Elricriven/co2data

删除[作者]: