

Construction of a High Spatiotemporal Resolution Dataset of Satellite-Derived $p\text{CO}_2$ and Air–Sea CO_2 Flux in the South China Sea (2003–2019)

Zigeng Song¹, Shujie Yu, Yan Bai, Xianghui Guo, Xianqiang He², Weidong Zhai, and Minhan Dai³

Abstract—The South China Sea (SCS) is one of the largest marginal seas in the world. It includes a river-dominated, highly productive ocean margin on the northern shelf and an oligotrophic ocean-dominated basin along with other subregions with various features. It was a challenge to estimate the air–sea CO_2 flux in this area. We developed a retrieval algorithm for sea surface partial pressure of CO_2 ($p\text{CO}_2$) by a combination of our previously established mechanistic semianalytical method (MeSAA) and machine learning (ML) method, named MeSAA-ML-SCS, built upon a large dataset of sea surface $p\text{CO}_2$ collected from in situ measurements during 44 cruises /legs to the SCS in the last two decades. We set several semianalytical parameters, including $p\text{CO}_{2,\text{therm}}$ represented the combined effect of thermodynamics and the atmospheric CO_2 forcing on seawater $p\text{CO}_2$; upwelling index (UI_{SST}) and mixing layer depth (MLD) to characterize the mixing processes; and chlorophyll-*a* concentration (Chl-*a*) with remote sensing reflectance at 443 and 555 nm [Rrs(443) and Rrs(555)], which were proxies of biological effects and other characteristics for distinguishing shelf, basin, and subregions. We set the difference between seawater $p\text{CO}_2$ and atmospheric $p\text{CO}_2$ ($\Delta p\text{CO}_2^{\text{Sea-Air}}$) as the output, and the seawater $p\text{CO}_2$ was finally obtained by summing atmospheric $p\text{CO}_2$ and $\Delta p\text{CO}_2^{\text{Sea-Air}}$. We compared several ML models, and the XGBoost model was confirmed as the best. Independent cruise-based datasets that are not involved in the model training were used to validate the satellite products, with low root-mean-square error (RMSE = 11.69 μatm) and mean absolute percentage deviation (APD = 1.59%). The increasing trend of time-series satellite-derived $p\text{CO}_2$ ($2.44 \pm 0.24 \mu\text{atm}/\text{year}$) was validated by the in situ data at the Southeastern Asia Time-series Study (SEATS) station, showing good consistency. Results indicate that

the SCS as a whole is a source of atmospheric CO_2 , releasing an average of $12.34 \pm 3.11 \text{ Tg C}/\text{year}$ from a total area of $2.87 \times 10^6 \text{ km}^2$, while the northern shelf acts as a sink ($2.02 \pm 0.64 \text{ Tg C}/\text{year}$). With the forcing of increasing atmospheric CO_2 , the area-integrated CO_2 efflux over the entire SCS is decreasing with a rate of $0.41 \text{ Tg C}/\text{year}$ during 2003–2019. This shared long time series, high-accuracy dataset (1 km) can be helpful to further improve our understanding of the air–sea CO_2 exchange dynamics in the SCS.

Index Terms—Air–sea CO_2 flux, machine learning (ML), satellite retrieval, seawater $p\text{CO}_2$, semianalytical algorithm, South China Sea (SCS).

I. INTRODUCTION

THE South China Sea (SCS) is the largest marginal sea of the North Pacific Ocean with intense turbulent mixing, upwelling, and stratification (in the central basin) [1], [2]. It is characterized by dynamic exchanges in the upper layer with the western Pacific and deep overflow through the Kuroshio [3], [4], [3], [5]. In addition, the SCS receives abundant terrestrial inputs from the surrounding terrain. These complex dynamic processes promote diverse biogeochemical processes and carbon cycling [2], [6].

As early as 2000, measurements of the seawater partial pressure of CO_2 ($p\text{CO}_2$) were conducted in the SCS. Zhai et al. [1] summarized the seasonal distribution of the seawater $p\text{CO}_2$ and air–sea CO_2 fluxes in four subregions (off the Pearl River estuary, northern basin, the central and southern basin, and west of the Luzon Strait) of the SCS based on in situ data from 14 cruises in 2003–2008 and estimated that the annual average air–sea CO_2 flux density was $1.1 \pm 0.9 \text{ mol C}/\text{m}^2/\text{year}$ (positive value/“+” represents source/efflux and negative value/“−” represents sink/influx). Based on the in situ dataset from 47 cruises (2000–2018), Li et al. [2] demonstrated the spatiotemporal variations of seawater $p\text{CO}_2$ and air–sea CO_2 fluxes in the five subregions of the SCS (northern shelf, northern slope, SCS basin, west of the Luzon Strait, and western SCS) and updated estimations of the annual average CO_2 flux to $0.4 \pm 0.6 \text{ mol C}/\text{m}^2/\text{year}$. Li et al. [2] also preliminarily analyzed the impact of freshwater input from the Pearl River, typhoons, and El Niño–Southern Oscillation (ENSO) events on the air–sea CO_2 flux and found that the CO_2 sink in the northern SCS shelf was enhanced in the winter. Zhai et al. [1] and Li et al. [2] presented the integrated analysis of the $p\text{CO}_2$ and CO_2 flux data in the SCS over the

Manuscript received 19 April 2023; revised 1 August 2023; accepted 14 August 2023. Date of publication 18 August 2023; date of current version 31 August 2023. This work was supported by the National Natural Science Foundation of China under Grant 42141001, Grant 42141002, Grant 42176177, and Grant 42006165. (Zigeng Song and Shujie Yu contributed equally to this work.) (Corresponding author: Yan Bai.)

Zigeng Song, Yan Bai, and Xianqiang He are with the State Key Laboratory of Satellite Ocean Environment Dynamics, Second Institute of Oceanography, Ministry of Natural Resources, Hangzhou 310012, China, and also with the College of Oceanography, Hohai University, Nanjing 210098, China (e-mail: songzigeng9712@outlook.com; baiyan@sio.org.cn).

Shujie Yu is with the State Key Laboratory of Satellite Ocean Environment Dynamics, Second Institute of Oceanography, Ministry of Natural Resources, Hangzhou 310012, China, and also with Ocean College, Zhejiang University, Zhoushan 316021, China.

Xianghui Guo and Minhan Dai are with the State Key Laboratory of Marine Environmental Science, College of Ocean and Earth Sciences, Xiamen University, Xiamen 361102, China.

Weidong Zhai is with the Southern Marine Science and Engineering Guangdong Laboratory (Zhuhai), Zhuhai 519080, China.

This article has supplementary downloadable material available at <https://doi.org/10.1109/TGRS.2023.3306389>, provided by the authors.

Digital Object Identifier 10.1109/TGRS.2023.3306389

past two decades; however, the spatiotemporal coverage of the in situ data was insufficient. Thus, they were unable to cover all areas with different biogeochemical characteristics and could only present data on seasonal scales. It is necessary to obtain seawater $p\text{CO}_2$ with higher spatiotemporal resolution and broader coverage in the SCS using remote sensing to better estimate the surface $p\text{CO}_2$ and air–sea CO_2 fluxes and display the carbon source–sink pattern and variation.

Current research on SCS $p\text{CO}_2$ remote sensing retrieval or model simulation algorithms has mainly focused on the northern shelf, where abundant observations are available. Zhu et al. [7] established an empirical algorithm for $p\text{CO}_2$ based on a cruise of in situ sea surface temperature (SST) and chlorophyll-a concentration (Chl-a) in July 2004. Jo et al. [8] used in situ $p\text{CO}_2$ (May 2001, February, and July 2004) from northern SCS cruises, satellite-derived SST and Chl-a, and their longitude and latitude, to establish a seawater $p\text{CO}_2$ retrieval model based on a neural network. Lv et al. [9] used the mechanistic semianalytical method (MeSAA) of seawater $p\text{CO}_2$ retrieval proposed by Bai et al. [10] and data of six cruises during 2004–2011 to establish a $p\text{CO}_2$ algorithm in the Pearl River Estuary in summer. The MeSAA algorithm considered horizontal mixing, thermodynamics, and biological effects; however, $p\text{CO}_2$ in coastal zones was underestimated due to the lack of consideration of coastal upwelling. Zhao et al. [11] developed a 3-D physical and biogeochemical coupling model to simulate the carbonate system in the Pearl River Estuary and adjacent northern SCS under typical summer conditions and theoretically analyzed the impact of riverine input, biological processes, and air–sea exchange on seawater $p\text{CO}_2$.

Few studies have been conducted on retrieving seawater $p\text{CO}_2$ at the SCS basin scale. To obtain better spatiotemporal grid data for the SCS, Wang et al. [12] proposed a method to produce a reanalysis dataset. They used an empirical orthogonal function (EOF) method to reconstruct seawater $p\text{CO}_2$ with a grid resolution of $0.5^\circ \times 0.5^\circ$ in summer by integrating massive in situ data and satellite-derived seawater $p\text{CO}_2$ product. The satellite-derived $p\text{CO}_2$ production over the whole SCS was based on the MeSAA algorithm developed by Bai et al. [10] (SCS- $p\text{CO}_2$ V1.0), with a root-mean-square error (RMSE) of $33.8 \mu\text{atm}$ evaluated by Wang et al. [12]. With more underway cruises and updated satellite-derived $p\text{CO}_2$ products (SCS- $p\text{CO}_2$ V2.0) [13], Wang et al. [14] updated the reanalysis dataset of the SCS based on a machine learning (ML) method, which significantly improved the reconstruction performance (RMSE = $16 \mu\text{atm}$) and thus produced a monthly seawater $p\text{CO}_2$ in the SCS with a 5-km resolution from 2003 to 2020. The SCS- $p\text{CO}_2$ V2.0 was generated by a pure ML algorithm (XGBoost) with eight input parameters [SST, sea surface salinity (SSS), Chl-a, remote sensing reflectance at 443, 488, and 555 nm ($R_{rs}(443)$, $R_{rs}(488)$, and $R_{rs}(555)$, respectively), mixed layer depth (MLD), atmospheric CO_2 presented in the CO_2 mole fraction in the air ($x\text{CO}_2$), and upwelling index (UI_{SST}), and SST difference of various water masses (upwelling, alongshore currents, and so on) in the same latitude], which has an RMSE of $20.66 \mu\text{atm}$. However, in the estuarine area in summer, the RMSE reaches $29.96 \mu\text{atm}$, as evaluated by Wang et al. [12].

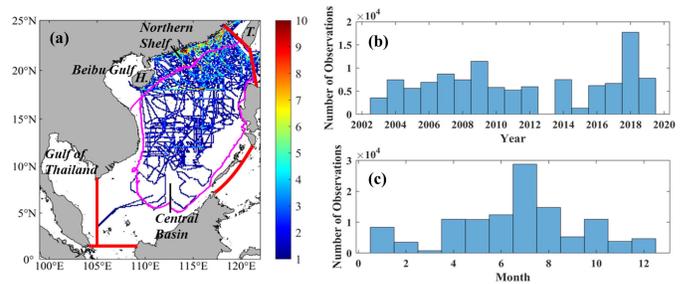


Fig. 1. (a) Number of months with underway seawater $p\text{CO}_2$ observations in the SCS in 2003–2019. The pink line indicates the continental shelf boundary and center basin, and the red lines define the entire SCS boundary. The boundary of dividing central basin and northern shelf is along the 200-m isobath. T and H mean Taiwan Island and Hainan Island, respectively. (b) and (c) In situ seawater $p\text{CO}_2$ histogram statistics observations in year and month.

Recently, most existing remote sensing and reanalysis $p\text{CO}_2$ products tend to be driven by a large volume of in situ data, and all possible influencing parameters, including SST, SSS, Chl-a, and atmospheric CO_2 concentration, are used as inputs of the ML algorithm to generate a model that fits well with the in situ seawater $p\text{CO}_2$ in a form of incomprehensible black box. The mechanistic-based algorithms can better analyze the $p\text{CO}_2$ variation and obtain stable results; however, the current capacity of analytical or semianalytical algorithms is insufficient to quantify complex spatiotemporal processes.

Thus, in this study, we proposed a new model that combined advantages of the mechanistic semianalytic and ML algorithms to retrieve $p\text{CO}_2$. First, we analyzed and parameterized the controlling mechanism of seawater $p\text{CO}_2$ in the SCS based on the MeSAA to determine input parameters more accurately and ensure its stability. Subsequently, we established a $p\text{CO}_2$ retrieval algorithm, MeSAA-ML-SCS, using ML methods based on multiyear in situ seawater $p\text{CO}_2$ data and finally produced the dataset of monthly seawater $p\text{CO}_2$ and CO_2 fluxes in the SCS with a high spatial resolution (1 km) during 2003–2019 and presented a refined estimation of the air–sea CO_2 flux in the whole SCS.

II. DATA AND METHODS

A. Underway $p\text{CO}_2$ Data in the SCS

We compiled in situ $p\text{CO}_2$ data from 44 cruises/legs (2003–2019), with a total 739 705 observations. The survey periods and sources are listed in Table I. A total of 43 cruises/legs were organized by Xiamen University (XMU), most of which were published by Zhai et al. [1], [15], Li et al. [2], Jo et al. [8], Lv et al. [9], and Zhai [16]. Nine unpublished cruises/legs collected by XMU (Dai et al. unpublished) and one leg by the Surface Ocean CO_2 Atlas (SOCAT, version 2021) were included. The in situ seawater $p\text{CO}_2$ was measured by R/Vs *Dongfanghong-2* and *Tan Kah Kee* (TKK). During the cruises, seawater $p\text{CO}_2$ was collected continuously. The measurement and data processing followed those of the SOCAT protocol [2]. Underway data provided by SOCAT in the fugacity form of carbon dioxide ($f\text{CO}_2$) were converted to $p\text{CO}_2$ using the corresponding in situ SST and equation reported by Takahashi et al. [17].

TABLE I

UNDERWAY SEAWATER $p\text{CO}_2$ DATA USED IN THIS STUDY. THE CRUISES WITH RED FONT COMPOSE THE INDEPENDENT VALIDATION DATASET, AND THOSE WITH BLACK FONT ARE THE TRAINING DATASET

Season	Cruise months		
	Mar.	Ari.	May.
Spring	2004.03 ^d	2004.04 ^a	2004.05 ^a
	2009.03 ^g	2005.04 ^a	2005.05 ^c
		2008.04 ^a	2011.05 ^{b,h}
		2009.04 ^g	2016.05 ^b
		2012.04 ^b	
Summer	Jun.	Jul.	Aug.
	2014.06 ^b	2004.07 ^c	2007.08 ^a
	2016.06 ^b	2007.07 ^a	2008.08 ^b
	2017.06 ^b	2008.07 ^b	2009.08 ^{b,h}
	2018.06 ^f	2009.07 ^{b,h}	2012.08 ^b
	2019.06 ^f	2012.07 ^b	2015.08 ^b
		2014.07 ^b	2018.08 ^f
		2015.07 ^b	2019.08 ^f
		2016.07 ^f	
		2017.07 ^f	
	2018.07 ^f		
	2019.07 ^f		
Fall	Sep.	Oct.	Nov.
	2006.09 ^c	2003.10 ^a	2006.11 ^a
	2007.09 ^a	2006.10 ^c	2010.11 ^b
	2008.09 ^b	2008.10 ^c	
	2018.09 ^f	2010.10 ^b	
	2017.10 ^f		
	2018.10 ^f		
Winter	Dec.	Jan.	Feb.
	2006.12 ^a	2009.01 ^b	2004.02 ^d
	2008.12 ^b	2010.01 ^b	2006.02 ^a
	2018.01 ^b	2018.02 ^b	

^a Firstly published by Zhai et al. [1]; ^b Firstly published by Li et al. [2]; ^c Firstly published by Zhai et al. [15]; ^d Firstly published by Jo et al. [8]; ^e Data released in SOCAT version 2021; ^f Dai et al. unpublished data; ^g Firstly published by Zhai [16]; ^h Partially used by Lv et al. [9].

The cruise tracks are shown in Fig. 1(a), with a concentration of data located in the northern part of the SCS, especially the northern shelf; however, the space-time coverage over the whole sea basin was still low. The in situ data histogram statistics for the year and month spans are shown in Fig. 1(b) and (c). The observations were conducted every year during 2003–2019, except for 2013; 2018 had the maximum data volume ($\sim 1.8 \times 10^4$ records). Furthermore, the in situ data were reliable seasonal representatives with at least 5000 monthly data points, except in March.

B. Satellite and Modeled Data

The ocean color data, i.e., Chl-a and Rrs(443), and remote sensing reflectance at 555 nm [Rrs(555)] employed in this study were derived from MODIS-Aqua with a spatial resolution of 4 km, associated with processing version 2018.0. The monthly SST dataset adopted was AVHRR_OI (optimal interpolation), provided by the Group for High-Resolution SST

TABLE II

SOURCE AND INFORMATION OF SATELLITE AND MODELED DATA USED IN THIS STUDY(2003–2019)

	Dataset (Source)	Version	Spatial - Resolution	Temporal-Resolution	
Rrs(443)	MODIS-Aqua (NASA)	2018.0	4km	Monthly	
Rrs(555)	MODIS-Aqua (NASA)	2018.0	4km		
Chl-a	MODIS-Aqua (NASA)	2018.0	4km		
SST	OISST (NOAA)	2.1	0.25°		
SSS	GLOBAL_REANALY SIS_PHY_001_030 (CMEMS)	-	0.083°		
MLD	GLOBAL_REANALY SIS_PHY_001_030 (CMEMS)	-	0.083°		
SLP	CarbonTracker (NOAA)	CT2019B	3° x 2°		
xCO ₂	CarbonTracker (NOAA)	CT2019B	3° x 2°		
WS	ERA5 (ECMWF)	-	0.25°		Monthly & 6-hourly

(GHRSSST), National Oceanic Atmospheric Administration (NOAA), with a 0.25° resolution, processing version 2.1. SSS and MLD obtained from the Copernicus Marine Service (CMEMS) with the product name *GLOBAL-REANALYSIS-PHY-001-030* at a spatial resolution of 0.083°. The sea-level pressure (SLP) and xCO₂ were obtained from NOAA's Carbon Tracker version CT2019B. The wind speed (WS) at 10 m above the sea surface is derived from the ERA5 dataset provided by the European Center for Medium-Range Weather Forecasts (ECMWF). The sources and information of the satellite and modeled data are listed in Table II.

C. Data Gridding and Matching-Up

To match the in situ $p\text{CO}_2$ with satellite and modeled data, all data were first gridded with a time window of the month and a spatial window of 1 km. The principle of in situ data gridding is that each sample can belong to only one grid. For the grid containing more than three samples, abnormal values were identified using the 3σ principle and eliminated. The average value of valid in situ data falling into a grid was considered as the value of the grid. The satellite and model data were also resampled to a 1-km resolution before matchup.

The matching-up processing is that, first, the nearest latitude and longitude identified the central pixel collocated with each in situ data; to avoid the impact of the noise within satellite products on the modeling, only those pixels with at least ten valid pixels in the surrounding 5×5 box satisfied the homogeneity criteria that the coefficient of variation < 0.15 is considered valid. Finally, valid central pixels and corresponding gridded in situ data are matched to create database for the modeling of seawater $p\text{CO}_2$. The data volume of the matched database was 115 273.

D. Classification of Training and Validation Datasets

The matched-up database, with a volume of 115 273 records, was further categorized into training

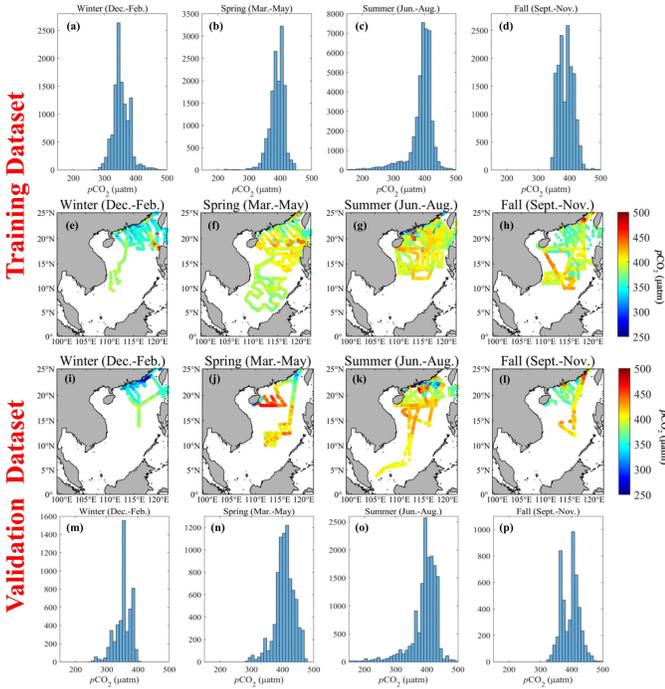


Fig. 2. Spatial distribution and corresponding histogram statistics of (a)–(h) training dataset and (i)–(p) validation dataset in four seasons. The winter months (December, January, and February); the spring months (March, April, and May); the summer months (June, July, and August); and the fall months (September, October, and November).

and validation datasets. The data were divided independently by cruise (see Table I). Moreover, we ensured that the training and validation sets had good seasonal and spatial coverage, and the validation set accounted for more than 25%. Fig. 2 shows the spatial and numerical distributions of the training and validation data for the four seasons. We obtained full coverage on the northern shelf and the central basin over four seasons on both the training and validation datasets. We only had one summer cruise in the southwest of the basin, which was included in the validation set to ensure the evaluation accuracy of the modeling predictability. The final validation set accounted for 28.64% of the total data volume, and the portion of the validation sets varied between 24% and 35% in all four seasons, although the observations were concentrated on the summer (56 049 groups of matchups) and relatively less on other seasons (approximately 20 000 groups on average).

E. Flux Calculation

After the seawater $p\text{CO}_2$ was generated, CO_2 flux between the surface water and the atmosphere (air–sea CO_2 flux, $\text{mmol C/m}^2/\text{day}$) can be calculated from

$$F\text{CO}_2 = k \times K_H^{\text{CO}_2} \times \Delta p\text{CO}_2 \quad (1)$$

where k is the gas transfer velocity (cm/h), $K_H^{\text{CO}_2}$ is the solubility of CO_2 gas in seawater (mol/kg/atm.) [18], and $\Delta p\text{CO}_2$ is the difference between atmospheric and surface seawater $p\text{CO}_2$. k was parameterized using the empirical function of Sweeney et al. [19], and nonlinear correction of gas

transfer velocity with WS was adopted following Wanninkhof et al. [20] and Jiang et al. [21]:

$$k = 0.27 \times C_2 \times U_{\text{mean}}^2 \times \left(\frac{\text{Sc}}{660} \right)^{-0.5} \quad (2)$$

$$C_2 = \frac{1}{n} \times \frac{\sum_{j=1}^n U_j^2}{U_{\text{mean}}^2} \quad (3)$$

where U_{mean} is the monthly mean wind speed at 10 m above sea level (in m/s), Sc is the Schmidt number at in situ temperatures for surface seawater [22], C_2 is the nonlinear coefficient for the quadratic term of the gas transfer relationship, U_j is the 6-hourly WS (in m/s), the subscript “mean” indicates average values, and n is the number of available WS measurements for the month.

The atmospheric $p\text{CO}_2$ ($p\text{CO}_2^{\text{Air}}$) was calculated using $x\text{CO}_2$ in the air, the SLP, and the vapor pressure of water at 100% relative humidity ($p\text{H}_2\text{O}^{\text{Air}}$), using the following formula [23]:

$$p\text{CO}_2^{\text{Air}} = x\text{CO}_2 \times (\text{SLP} - p\text{H}_2\text{O}^{\text{Air}}). \quad (4)$$

The net CO_2 flux in a certain region is estimated by multiplying the mean CO_2 flux density among the available pixels by the total area of the region. The input parameters to derive the seawater $p\text{CO}_2$ and calculate the CO_2 flux are shown in Table II.

F. Performance Evaluation

The performance evaluation was based on three statistical measures: the coefficient of determination (R^2), RMSE, and mean absolute percentage deviation (APD) as follows:

$$R^2 = \left[\frac{1}{N} \sum_{i=1}^N \left(\frac{X_i - \bar{X}}{\sigma_X} \right) \left(\frac{Y_i - \bar{Y}}{\sigma_Y} \right) \right]^2 \quad (5)$$

$$\text{RMSE} = \sqrt{\frac{\sum_{i=1}^N (X_i - Y_i)^2}{N}} \quad (6)$$

$$\text{APD} = \frac{1}{N} \sum_{i=1}^N \left| \frac{X_i - Y_i}{Y_i} \right| \times 100\% \quad (7)$$

where X_i , Y_i , and N are the algorithm retrieved values, in situ measurement values, and sample number, respectively, and σ_X and σ_Y are the standard deviations of X_i and Y_i , respectively.

III. ALGORITHM DEVELOPMENT

A. Mechanism Analysis and Input Parameter Strategy

Bai et al. [10] proposed the MeSAA algorithm for retrieving seawater $p\text{CO}_2$ using remote sensing data. In MeSAA, the variation in seawater $p\text{CO}_2$ was analytically expressed as the sum of the first-order partial differentials of individual $p\text{CO}_2$ components controlled by major factors such as the temperature-dependent thermodynamic effect, mixing effects between different water masses, biological effects, the air–sea CO_2 exchange that affects the seawater $p\text{CO}_2$ iteratively, and other individual processes that vary independently of the above-identified processes. The MeSAA algorithm has been successfully applied to the East China Sea [10]

(Bering Sea [24], Northern Shelf of the SCS [9], Gulf of Mexico [25], and Coral Sea [26]). However, the existing MeSAA algorithm still had difficulties in its analytical ability for complicated physical and biogeochemical processes.

In this study, we fully utilized the ML method advantage in solving the complex nonlinear problem and maintaining the MeSAA ability to understand the influence of the controlling factors by applying it to the input parameters selection. The parameter selection strategies are given as follows.

1) *Target Parameter: $p\text{CO}_2$ or $\Delta p\text{CO}_2^{\text{Sea-Air}}$* : Except for the northern shelf of the SCS, the vast SCS basin belongs to the oligotrophic zone, which leads to a centralized distribution of $p\text{CO}_2$ data (“narrow-type histogram”) in the sea basin. In most sea basin areas, the forcing effect of atmospheric CO_2 was evident, and the seawater and atmospheric $p\text{CO}_2$ changed at a similar rate at 2.10 ± 0.79 and $\sim 2.2 \mu\text{atm}/\text{year}$, respectively [2]. Due to the significant similarity between seawater and atmospheric $p\text{CO}_2$, if $p\text{CO}_2^{\text{Air}}$ was directly used as an input parameter to retrieve seawater $p\text{CO}_2$, the weight of $p\text{CO}_2^{\text{Air}}$ would be too large and would mask the weight of other control mechanisms, and correspondingly, the model would be unstable. We found that the generated satellite images had significant patches when $p\text{CO}_2$ was directly outputted. Therefore, we used $\Delta p\text{CO}_2^{\text{Sea-Air}}$ as target to increase the data gradient and reduce the gap between the mean and standard deviation values (mean \pm SD). For example, the mean value of seawater $p\text{CO}_2$ was $387.25 \pm 42.86 \mu\text{atm}$, but $\Delta p\text{CO}_2^{\text{Sea-Air}}$ was $-9.06 \pm 43.25 \mu\text{atm}$, which solved the model instability when the inputs and outputs were highly similar.

2) *Comparison of Various Input Parameters in Major Controlling Processes*: In previous studies, four major controlling effects on seawater $p\text{CO}_2$ variation in the SCS were identified: atmospheric CO_2 forcing, thermodynamics effect, water mixing, and biological activities [7], [8], [9], [11]. Thus, we selected the appropriate proxies to denote these three processes: SST, $x\text{CO}_2$, and $p\text{CO}_{2,\text{therm}}$ as a group to explore the atmospheric CO_2 forcing and thermodynamics effect; MLD, U_{SST} , and in situ SSS ($\text{SSS}_{\text{in situ}}$) as a group to proxy for mixing processes; and the ocean color parameters such as Chl-a, Rrs(443), and Rrs(555) for biological effects in various domains.

To determine the impact of different inputs on the model, we performed multiple comparative experiments on the selection of parameters. The above parameters were divided into eight sets (Cases 1–8) for model training (see Table III) and the relative feature importance (RFI) of each input is shown in Fig. 3. The detailed calculation of RFI was presented in Text S1. Here, we directly used the optimal XGBoost model to select the input parameters to reduce the document length (the performance of different ML methods is shown in Section III-B).

3) *Atmosphere Forcing and Thermodynamic Effect: $x\text{CO}_2$ or $p\text{CO}_{2,\text{therm}}$* : When constructing a nearly 20-year sequence of $p\text{CO}_2$ data, the atmospheric CO_2 forcing effect on seawater $p\text{CO}_2$ cannot be ignored due to the variation range of increasing atmospheric CO_2 . Similar studies have usually used $x\text{CO}_2$ directly as an input, mainly on a global scale. However, the SCS includes different subsystems with high

TABLE III
VALIDATION GROUP USING DIFFERENT PARAMETERS

Cases	Input factors	R ²	RMSE	APD
Case 1	SST, Chl-a, Rrs(443), Rrs(555), MLD	0.86	15.93	2.39%
Case 2	SST, Chl-a, Rrs(443), Rrs(555), MLD, $\text{SSS}_{\text{in situ}}$	0.91	13.02	1.64%
Case 3	SST, Chl-a, Rrs(443), Rrs(555), MLD, U_{SST}	0.90	13.76	1.92%
Case 4	SST, Chl-a, Rrs(443), Rrs(555), MLD, $x\text{CO}_2$	0.91	12.68	1.60%
Case 5	SST, Chl-a, Rrs(443), Rrs(555), MLD, $p\text{CO}_{2,\text{therm}}$	0.91	12.53	1.63%
Case 6	SST, Chl-a, Rrs(443), Rrs(555), MLD, U_{SST} , $x\text{CO}_2$	0.92	11.58	1.52%
Case 7	SST, Chl-a, Rrs(443), Rrs(555), MLD, $\text{SSS}_{\text{in situ}}$, $p\text{CO}_{2,\text{therm}}$	0.92	11.27	1.38%
Case 8	SST, Chl-a, Rrs(443), Rrs(555), MLD, U_{SST} , $p\text{CO}_{2,\text{therm}}$	0.92	11.69	1.59%

spatiotemporal variability, while the $x\text{CO}_2$ product provided by Carbon Tracker has a low spatial resolution ($3^\circ \times 2^\circ$ grid). Thus, $x\text{CO}_2$ directly used as an input will cause model instability and induce patches in the retrieved seawater $p\text{CO}_2$ images. Here, we proposed a new mechanism parameter ($p\text{CO}_{2,\text{therm}}$) to simultaneously characterize the thermodynamic effect of SST and the forcing effect of long-term atmospheric CO_2 rise on seawater $p\text{CO}_2$. $p\text{CO}_{2,\text{therm}}$ is the theoretical seawater $p\text{CO}_2$ value under the impact of only temperature-dependent thermal dynamics, assuming that seawater CO_2 can reach equilibrium with atmospheric CO_2 within one year through the air–sea exchange. It was calculated as follows [27]:

$$p\text{CO}_{2,\text{therm}} = p\text{CO}_{2(o)}^{\text{Air}} \times e^{0.0423(\text{SST} - \text{SST}_{(o)})} \quad (8)$$

where $p\text{CO}_{2(o)}^{\text{Air}}$ and $\text{SST}_{(o)}$ are the annual mean $p\text{CO}_2^{\text{Air}}$ and SST, respectively. Fig. 4(a)–(h) shows $p\text{CO}_{2,\text{therm}}$ for four typical months in 2003 and 2019; for example, features of seasonal thermodynamic effect and atmosphere forcing of increasing CO_2 are clearly depicted.

According to the RFI of Cases 1–8 (see Fig. 3), the group of SST, $p\text{CO}_{2,\text{therm}}$, and $x\text{CO}_2$ is shown in green. The results of Cases 4 and 5 in Table III show that $p\text{CO}_{2,\text{therm}}$ (RMSE = $12.53 \mu\text{atm}$ and APD = 1.63%) was better than $x\text{CO}_2$ (RMSE = $12.68 \mu\text{atm}$ and APD = 1.60%) when $p\text{CO}_{2,\text{therm}}$ and $x\text{CO}_2$ were included with Case 1. In addition, the RFIs of $p\text{CO}_{2,\text{therm}}$ ($\sim 11\%$) and $x\text{CO}_2$ ($\sim 15\%$) have similar contributions to Cases 4 and 5 (see Fig. 3), indicating that $p\text{CO}_{2,\text{therm}}$ can mechanistically replace $x\text{CO}_2$ but can avoid the image patches result from the low spatial resolution of $x\text{CO}_2$.

4) *Mixing Effects: SSS, MLD, and U_{SST}* : The hydrodynamic characteristics of the SCS northern shelf are complex and involve freshwater input from rivers, seasonal coastal currents, upwelling, and so on. Mesoscale eddies, a feature of the SCS, markedly impact ecological processes, thus changing

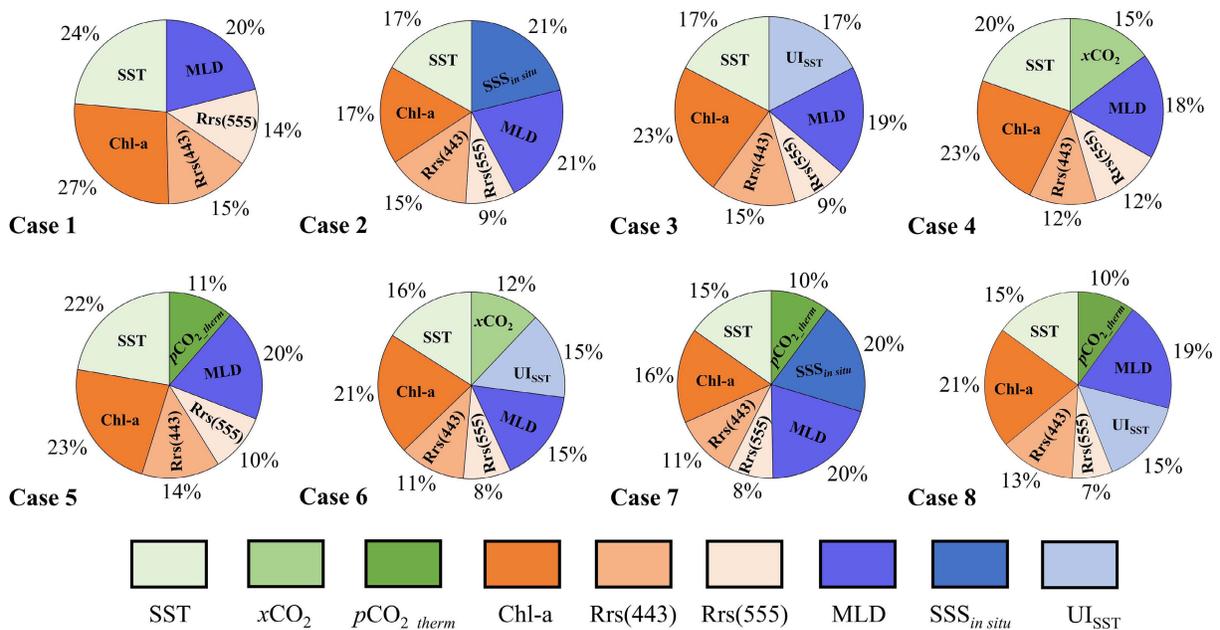


Fig. 3. Percentage pie chart for RFI of Cases 1–8 in XGBoost model.

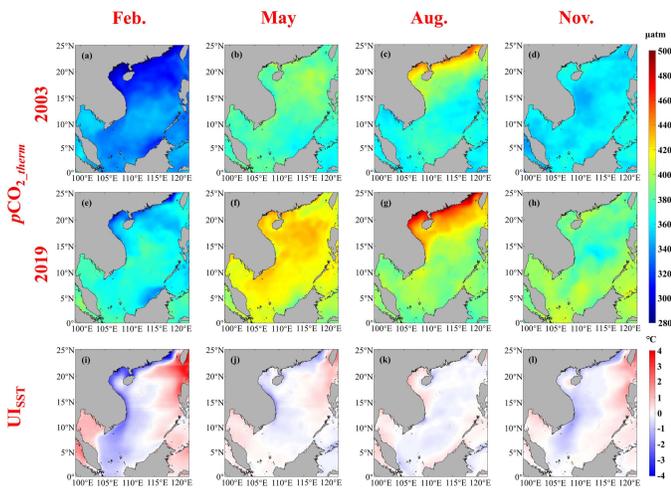


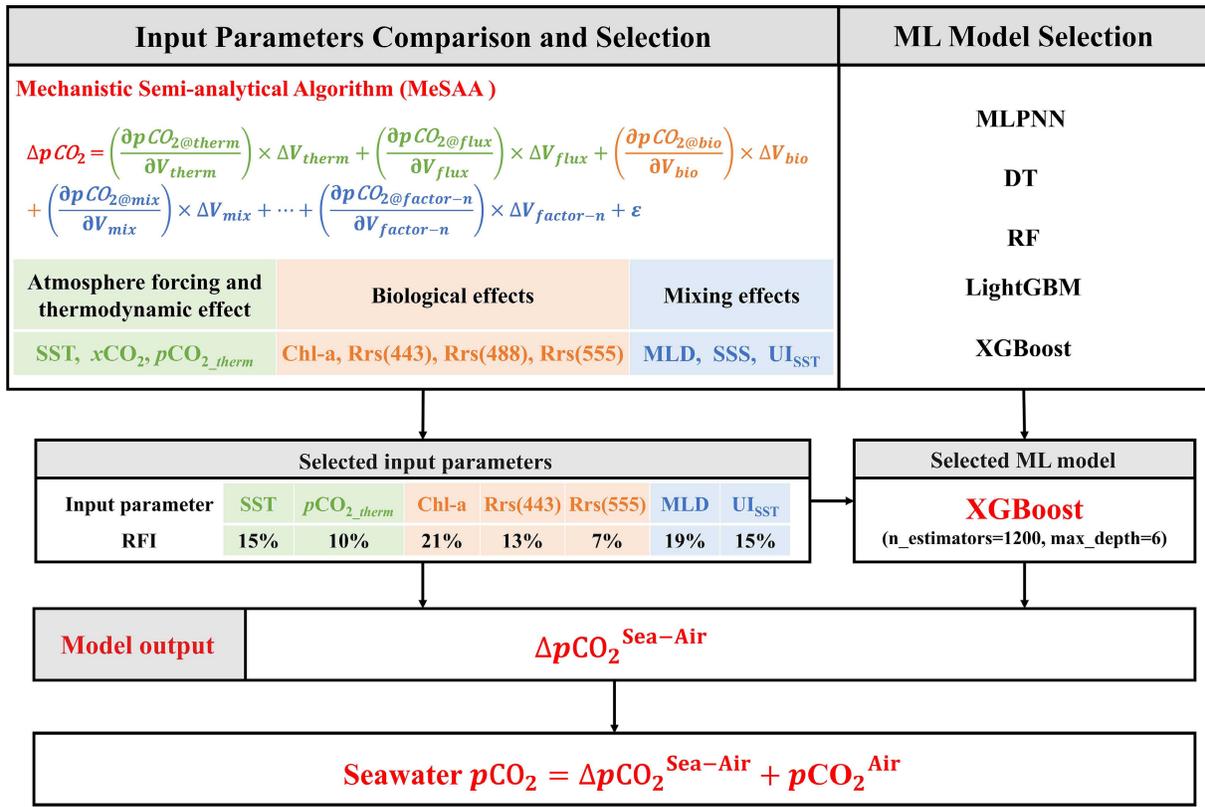
Fig. 4. Monthly $p\text{CO}_2_{\text{therm}}$ for (a) February 2003, (b) May 2003, (c) August 2003, (d) November 2003, (e) February 2019, (f) May 2019, (g) August 2019, and (h) November 2019. Multiyear (2009–2019) averaged monthly UI_{SST} for (i) February, (j) May, (k) August, and (l) November.

$p\text{CO}_2$ [28]. As shown in Fig. 4(i)–(l), a UI_{SST} index, defined as the SST at each pixel minus the mean SST at the same altitude in study area [29], can describe the upwelling, alongshore currents, and Kuroshio intrusion in the SCS. The SCS is also a region dominated by oceanic processes and is affected by the Kuroshio and upwelled deep water characterized by high dissolved inorganic carbon (DIC) [4], which partly reflects the seasonal variation of the MLD. Meanwhile, as the considerable difference in MLD between the shelf and basin, MLD can also be used to distinguish these two systems.

SSS denotes the mixing between different water masses. As satellite-derived SSS products had lower accuracy and spatial resolution (0.25°), introducing significant error propagation, we used $\text{SSS}_{\text{in situ}}$ as one of the inputs to test the performance

of other mixing proxies. In Fig. 3, the blue color represents the mixing characteristics of MLD, UI_{SST} , and SSS. Comparing $\text{SSS}_{\text{in situ}}$ in Cases 2 and 7 and UI_{SST} in Cases 3 and 8, UI_{SST} achieves a similar result as $\text{SSS}_{\text{in situ}}$ although with slightly lower statistical values in model performance. In Cases 2 and 3 (see Fig. 3), the RFIs of UI_{SST} ($\sim 17\%$) and $\text{SSS}_{\text{in situ}}$ ($\sim 21\%$) are also similar, indicating that UI_{SST} can approximately replace $\text{SSS}_{\text{in situ}}$. Finally, MLD and UI_{SST} together contributed 34% and were selected as proxy of the mixing effects on $p\text{CO}_2$ in the selected Case 8.

5) *Biological Effects: Chl-a and Rrs*: Compared to reanalysis data, the high spatial resolution (4 km) and observation frequency (daily) of ocean color data reduce the gap between underway measurements and satellite data. Therefore, we chose the most common parameter, Chl-a, to characterize the biological effect. Although $p\text{CO}_2$ has a highly complex response to biological processes, such as productivity and respiration, Chl-a is still the most stable and practical parameter [10]. We added Rrs(443) and Rrs(555) to characterize additional water ecological information. Rrs(443) mainly includes the effects of chromophoric dissolved organic matter and Chl-a. Rrs(555) contains information about particulate matter (phytoplankton, nonalgal, and terrestrial particles) and turbidity. Although Chl-a was calculated using the blue-to-green band ratio algorithms (OCx algorithm) and color index (CI) algorithm [30], [31], the spectral information of each band and Chl-a as input parameters can characterize biological characteristics in different ecological areas (e.g., offshores, basins, and mesoscale processes) and algal blooms under the same Chl-a concentration. In Fig. 3, the biological effects are represented by Chl-a, Rrs(443), and Rrs(555) with orange color, whose RFIs are very large, indicating that the biological effects have irreplaceable effects on $\Delta p\text{CO}_2^{\text{Sea-Air}}$ in the SCS. We also tested the addition of more Rrs to other bands. Although addition of more inputs to the SCS slightly increases

Fig. 5. Flowchart of retrieving the seawater $p\text{CO}_2$.

the statistical value of the model, model instability will be propagated to the final satellite image with artificial patches; therefore, we used two bands of Rrs(443) and Rrs(555) as inputs.

Through the above mechanism analysis, we determined that $p\text{CO}_2_{\text{therm}}$, SST, MLD, U_{SST} , Chl-a, Rrs(443), and Rrs(555) were used as input parameters to obtain the output of $\Delta p\text{CO}_2^{\text{Sea-Air}}$ in the ML model.

B. ML Model Selection

In the published literature, many ML technologies [i.e., multilayer perceptron neural network (MLPNN), random forest (RF), and light gradient boosting machine (LightGBM)] have been successfully applied to develop satellite-based surface $p\text{CO}_2$ models [14], [32], [33], [34], [35], [36]. The MLPNN is a feedforward neural network model based on the Levenberg–Marquardt backpropagation algorithm [37], [38], while RF, LightGBM, and XGBoost are based on the decision tree (DT) algorithm. RF adopts the “bagging” method to evenly extract training samples, and XGBoost adopts the “boosting” method and samples according to the error rate [39]. LightGBM is a fast, distributed, and high-performance ML framework, but it is sensitive to the noise of datasets due to its bias-based algorithm, which might lead to worse results [40].

Recently, XGBoost has proven to be more accurate than RF and MLPNN in various ML algorithms [41], [42], [43], [44], [45]. XGBoost aggregates a weak ML model to form a more robust estimator in an iterative process [46]. XGBoost can find data separation points more precisely and reduce the

TABLE IV
MODEL COMPARISON OF DIFFERENT ML APPROACHES

Machine learning model	R^2	RMSE	APD
MLPNN	0.58	29.59	5.04%
DT	0.65	26.57	4.09%
RF	0.78	20.09	2.98%
LightGBM	0.90	13.14	1.76%
XGBoost	0.92	11.69	1.59%

influence of the extreme values on the model stability by using a “presorted feature” algorithm [47]. The publicly available Scikit-Learn and XGBoost packages were used to complete the $p\text{CO}_2$ retrieval model. The scheme of parameter settings in XGBoost was given as follows: the number of regression trees was 1200, the maximum depth of trees was 6, the learning rate was 0.1, and the weight of the L_2 regularization term was 0.01 [48]. The MLPNN comprised one input layer, three hidden layers, and one output layer. The number of neurons was 30, 20, and 10 for the three hidden layers in the MLPNN, with a tan-sigmoid function as the transfer function.

To verify the effectiveness of the different ML models, we compared MLPNN, DT, RF, LightGBM, and XGBoost based on the same dataset and inputs. The common configuration parameters (i.e., the number of regression trees is 1200 and the maximum depth of each tree is 6) of DT, RF,

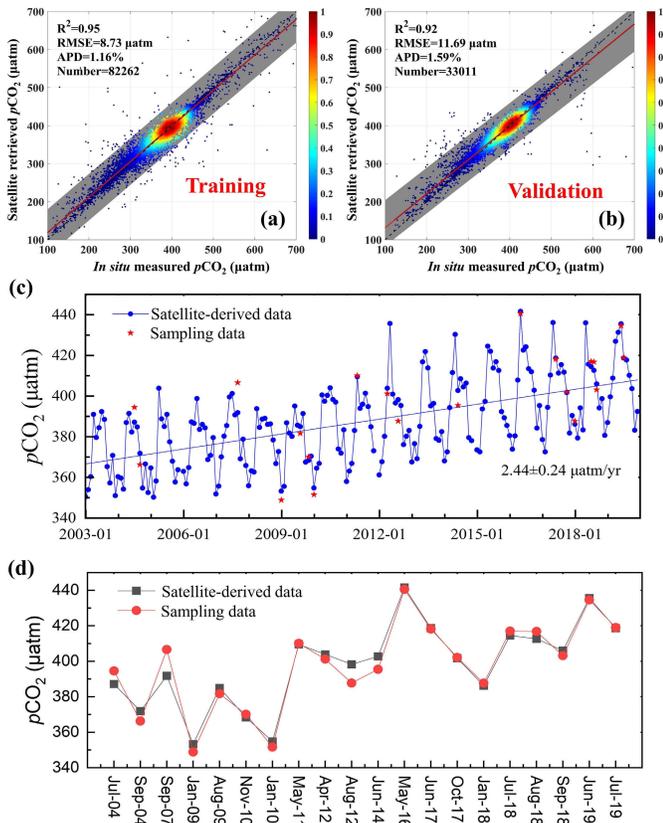


Fig. 6. Scatter density plot comparison between the in situ and satellite-derived $p\text{CO}_2$ for (a) training and (b) independent validation datasets. Gray shading represents the area within three standard deviations of the average values. (c) Monthly satellite-derived and corresponding sampling $p\text{CO}_2$ at SEATS station in 2003–2019. (d) Matching-up between satellite-derived $p\text{CO}_2$ and the sampling $p\text{CO}_2$ calculated from in situ DIC and TA at the SEATS station.

and LightGBM were all consistent with those of XGBoost in this experiment. Table IV presents the results of each ML method. XGBoost showed the best performance, with $\text{RMSE} = 11.69 \mu\text{atm}$, $R^2 = 0.92$, and $\text{APD} = 1.59\%$ in the model validation, followed by LightGBM, with $\text{RMSE} = 13.14 \mu\text{atm}$, $R^2 = 0.90$, and $\text{APD} = 1.76\%$. The performances of the MLPNN and DT were worse than those of the other three ML approaches. Therefore, XGBoost was selected.

C. Summary of the MeSAA-ML-SCS Algorithm

In summary, the flowchart of MeSAA-ML-SCS is shown in Fig. 5, which includes three parts: input parameter selection, ML model selection, and $p\text{CO}_2$ retrieval. First, we determined input parameters and ML technique based on $p\text{CO}_2$ mechanism analysis and multiple experiments. Then, $\Delta p\text{CO}_2^{\text{sea-air}}$ was estimated by the XGBoost model with input parameters [$p\text{CO}_{2,\text{therm}}$, SST, MLD, UI_{SST} , Chl-a, $\text{Rrs}(443)$, and $\text{Rrs}(555)$]. Finally, seawater $p\text{CO}_2$ was obtained by summing $p\text{CO}_2^{\text{air}}$ and $\Delta p\text{CO}_2^{\text{sea-air}}$. Furthermore, we conducted validation with independent in situ $p\text{CO}_2$ data to ensure the performance of the algorithm and satellite products in Section III-D.

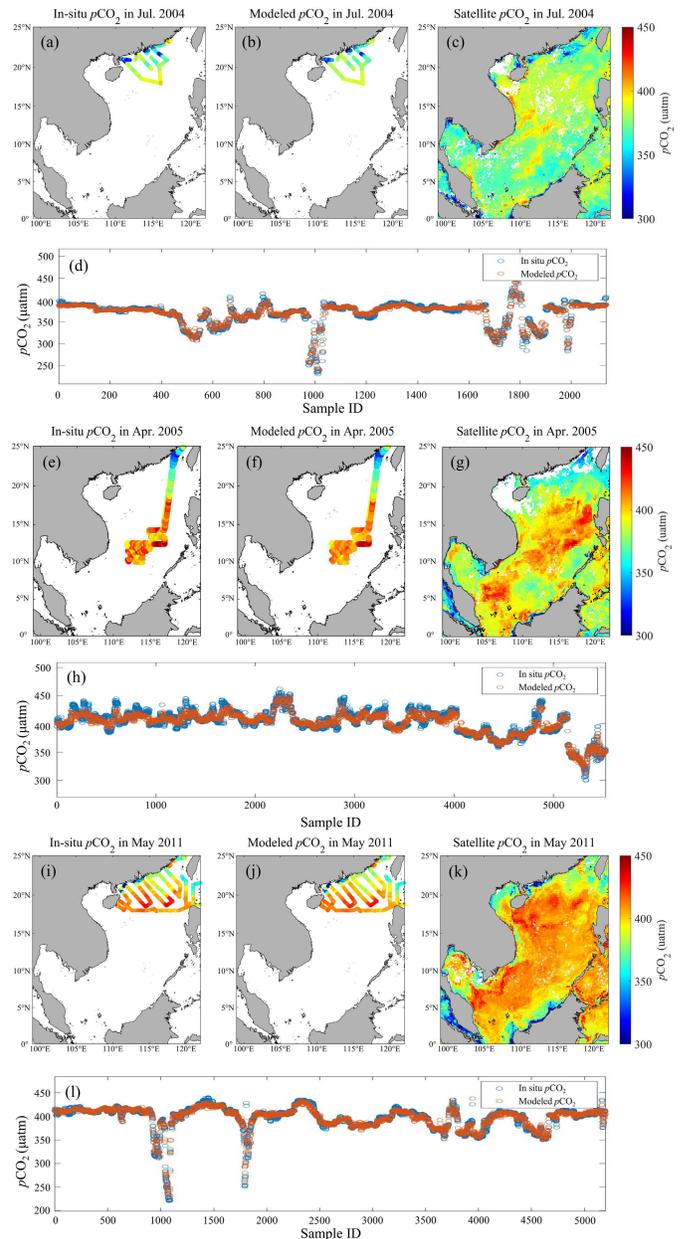


Fig. 7. Comparison between the retrieved $p\text{CO}_2$ and in situ $p\text{CO}_2$ for three cruises in (a)–(d) July 2004, (e)–(h) April 2005, and (i)–(l) May 2011.

D. Model Validation

1) Overall Performance Statistics of the Model: As we used matchup satellite data and grid data as model inputs instead of in situ data to develop the algorithm, validation of the model was equivalent to that of the final product. Fig. 6(a) and (b) compares the in situ and retrieved $p\text{CO}_2$ values for the training and independent validation datasets. In the training dataset, the retrieved $p\text{CO}_2$ values were consistent with the in situ values, with $R^2 = 0.95$, $\text{RMSE} = 8.73 \mu\text{atm}$, and $\text{APD} = 1.16\%$. Fig. 6(b) shows that the R^2 , RMSE , and APD of the independent validation dataset were 0.92, 11.69 μatm , and 1.59%, respectively, slightly lower than those of the training dataset but still within a reasonable error range. Moreover, the scatters in Fig. 6(a) and (b) outside the gray shaded area might be due to the impact of some complex

physical or biological processes (e.g., mesoscale eddies or algal blooms). Overall, the differences in the statistical measures between the training and independent validation datasets were relatively small, indicating that the established model can learn from the training dataset accurately and has strong stability and robustness.

2) *Long-Term Trend Accuracy of Satellite Products*: As the timespan (2003–2019) was long, the accuracy of the seawater $p\text{CO}_2$ trend became crucial to evaluate the product quality; therefore, we adopted observed data from long-term series stations to verify the $p\text{CO}_2$ products. Dai et al. [49] produced a long time series (2003–2019) of $p\text{CO}_2$ data at the Southeastern Asia Time-series Study (SEATS) station (116°E, 18°N), which were calculated from measured DIC and total alkalinity (TA) and were not involved to train the $p\text{CO}_2$ retrieval model. Therefore, we used these time-series data to test the accuracy of the long-term trends of our retrieved $p\text{CO}_2$ with the results. Fig. 6(c) reveals that the long-term changes in retrieved seawater $p\text{CO}_2$ at the SEATS station were consistent with observed values, and the retrieved $p\text{CO}_2$ increased at a rate of $2.44 \pm 0.24 \mu\text{atm}/\text{year}$. Fig. 6(d) further describes the slight differences in the matched-up $p\text{CO}_2$ dataset between the retrieval values and observed data during 2003–2019. Overall, the long-term trends of our $p\text{CO}_2$ retrieval model showed high reliability.

3) *Cruise-by-Cruise Comparisons*: We also validated our satellite products in detail for each cruise. Naturally, the results of the comparison along the cruise were favorable (see Fig. 7) in cases where scatter plots compared well [see Fig. 6(a) and (b)]; however, the cruise-by-cruise comparison provides more details on the satellite products reliability. In the main text, we only present the comparison of three cruises with longer routes, focusing on the cruise across the whole basin area and the complex north shelf. The $p\text{CO}_2$ comparisons for all cruises are shown in Figs. S1–S54.

Fig. 7 compares the retrieved and in situ $p\text{CO}_2$ in July 2004, April 2005, and May 2011. Among them, the in situ data in July 2004 [see Fig. 7(a)–(d)] and May 2011 [see Fig. 7(i)–(l)] were from the training datasets, and those in April 2005 [see Fig. 7(e)–(h)] were from the independent validation datasets. The consistency and continuity in Fig. 7(c), (g), and (k) indicate that the retrieval model has high applicability in the SCS basin. Comparing Fig. 7(g) with (e), the spatial variations of the retrieved $p\text{CO}_2$ were consistent with those of in situ $p\text{CO}_2$ in the southern Taiwan Strait. Furthermore, the regions with low $p\text{CO}_2$ values in Fig. 7(a)–(c), (e)–(g), and (i)–(k) indicate the potential influence of the river plume in summer and temperature effect in April on $p\text{CO}_2$ and can also be reflected in the retrieval model. Specifically, Fig. 7(g) presents an evident gradient change (from ~ 370 to $\sim 430 \mu\text{atm}$) in seawater $p\text{CO}_2$ from the waters of northern Luzon to the waters of western Luzon, which is seen in Fig. 7(e). Similarly, multiple significant gradient changes (from ~ 300 to $\sim 380 \mu\text{atm}$) also exist on the northern shelf of the SCS, shown in Fig. 7(c), which is consistent with the results in Fig. 7(a). These results indicate that $p\text{CO}_2$ products can capture high-precision spatiotemporal variation information and mesoscale and microscale changes.

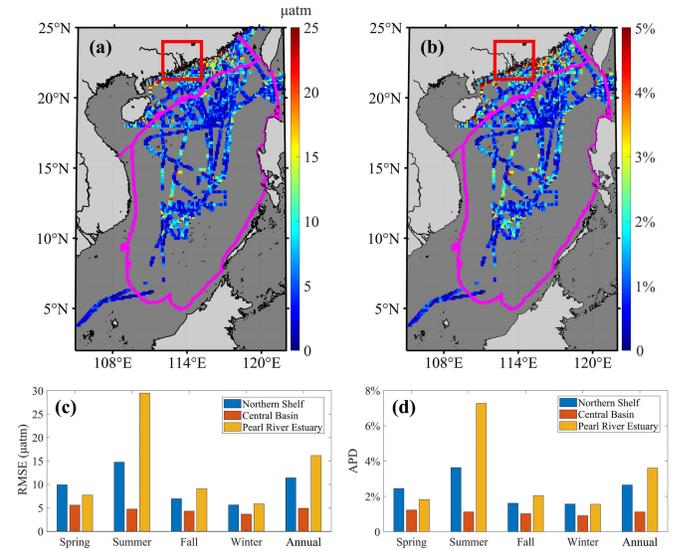


Fig. 8. Spatial distributions of (a) RMSE and (b) APD of the retrieval and in situ $p\text{CO}_2$ data for the validation dataset. Seasonal and annual (c) RMSE and (d) APD in northern shelf, central basin [pink line areas, same as that in Fig. 1(a)], and Pearl River Estuary (red box).

4) *Model Uncertainty*: We evaluated the uncertainty of the model by calculating the RMSE and APD of the retrieval and in situ $p\text{CO}_2$ data at $0.1^\circ \times 0.1^\circ$ grid for the validation dataset and the spatial distribution shown in Fig. 8(a) and (b). Then, the gridded RMSE and APD were averaged seasonally and annually over the northern shelf, central basin, and Pearl River Estuary to analyze the model uncertainty in these regions [see Fig. 8(c) and (d)]. Relatively large uncertainty (RMSE $> 20 \mu\text{atm}$ and APD $> 4\%$) might be affected by the river plumes, mesoscale eddies, short-term heavy precipitation, submarine groundwater discharge, and other events that have short timescale and small areas. The annual RMSEs are ~ 4.9 , ~ 11.4 , and $\sim 16.1 \mu\text{atm}$ in northern shelf, central basin, and Pearl River Estuary, respectively, and the annual APDs are $\sim 1.1\%$, $\sim 2.6\%$, and $\sim 3.6\%$, respectively. The model uncertainty in the central basin is relatively small (RMSE $< \sim 5.5 \mu\text{atm}$ and APD $< \sim 1.2\%$) in four seasons. In Pearl River Estuary, relatively large RMSE ($\sim 29.4 \mu\text{atm}$) and APD ($\sim 7.3\%$) mainly concentrated in summer might be due to the short-term heavy precipitation and Pearl River plume [1], [12], [14]. However, annual RMSE ($\sim 16.1 \mu\text{atm}$) and APD ($\sim 3.6\%$) indicate that the MeSAA-ML-SCS can reconstruct well the nonlinear relationship between the ecological environment characteristics and seawater $p\text{CO}_2$ in the estuary area in similar situations. Overall, our algorithm has the ability to retrieve the seawater $p\text{CO}_2$ with relatively low uncertainty in most areas of SCS.

IV. RESULTS AND DISCUSSION

A. Seasonal Variations in Seawater $p\text{CO}_2$

The monthly climatological seawater $p\text{CO}_2$ values for the entire SCS are shown in Fig. 9. The seasonal variation of seawater $p\text{CO}_2$ on the northern shelf was over $70 \mu\text{atm}$, which is much higher than that in the central basin ($\sim 30 \mu\text{atm}$) (see Fig. 9). Generally, it follows a pattern of high levels in summer

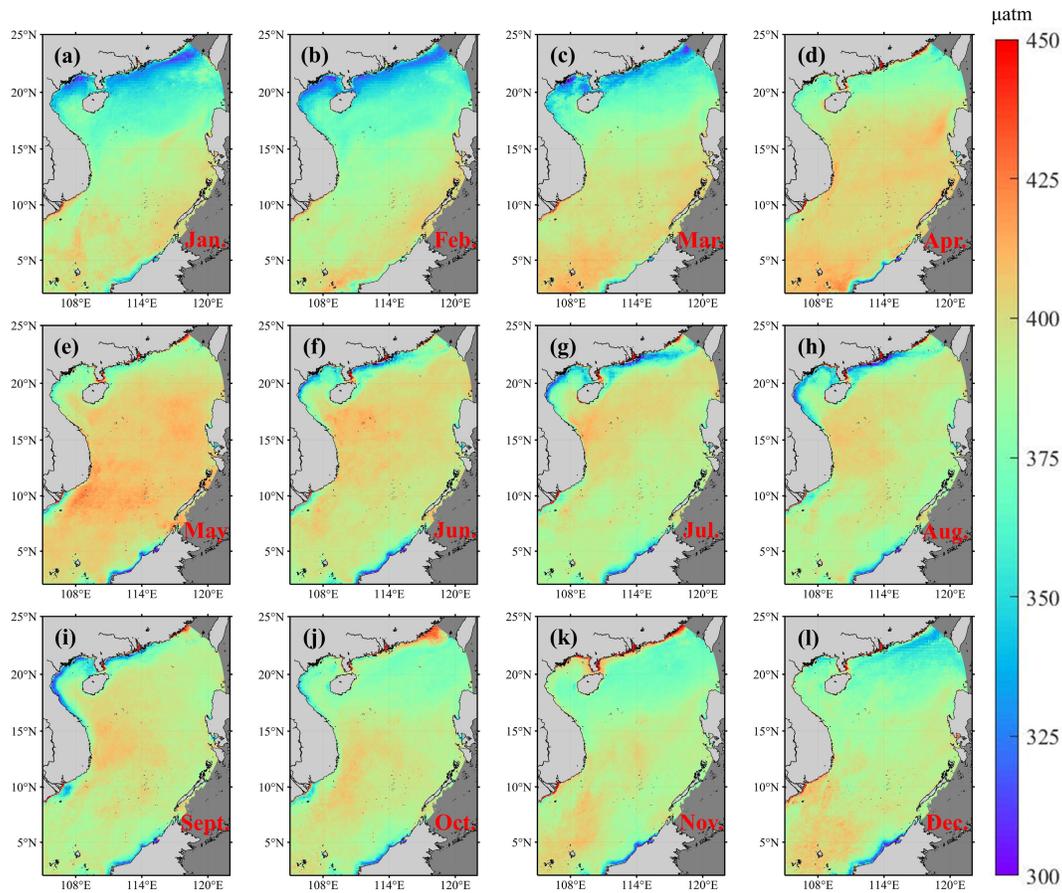


Fig. 9. Distribution of monthly average of seawater $p\text{CO}_2$ from 2003 to 2019. (a)–(l) From January to December.

and fall and low levels in winter, which aligns with Li et al. [2]. From December to March, the northern shelf is undersaturated for atmospheric CO_2 ($p\text{CO}_2$ values are less than $350 \mu\text{atm}$) and associated with low SST. The closer to the coast, the lower the $p\text{CO}_2$ value. Post $p\text{CO}_2$ reaching a peak value of $400 \mu\text{atm}$ in May, it remains oversaturated until October. Of note, data with 1-km resolution show the $p\text{CO}_2$ distribution along the coast and estuary. Far-oversaturated CO_2 was captured along the Guangdong Coast and in the Pearl River Estuary from April to December, which may be influenced by high terrestrial organic matter input and coastal upwelling, as previously observed [6], [50]. A low $p\text{CO}_2$ ($\sim 330 \mu\text{atm}$) plume emerges from the Pearl River and extends eastward from summer to mid-fall, which may be related to high biological productivity and low turbidity (compared to the estuary) [1]. Li et al. [2] also reported an extremely high $p\text{CO}_2$ (up to $650 \mu\text{atm}$) in the near-shore area and a low $p\text{CO}_2$ value (as low as $150 \mu\text{atm}$) in the area influenced by the Pearl River plume in summer, based on in situ $p\text{CO}_2$.

The $p\text{CO}_2$ in the central basin of the SCS is relatively homogeneous and primarily controlled by SST, which relatively has high $p\text{CO}_2$ in the summer (June, July, and August) and low $p\text{CO}_2$ during the winter (December, January, and February). The regional-averaged $p\text{CO}_2$ in the SCS basin is oversaturated throughout the year except in winter, which agrees with the observations of Zhai et al. [1] and Li et al. [2] based on in situ data. In the winter months, $p\text{CO}_2$ in the

southern part of the central basin ($\sim 410 \mu\text{atm}$) was higher than that in the north ($\sim 385 \mu\text{atm}$), while the pattern reversed in the summer and early fall months. Noticeably, in the spring months, especially May, the central basin witnessed the highest and most homogeneous $p\text{CO}_2$ value ($>410 \mu\text{atm}$) over the year, due to the high SST (recorded from satellite data).

B. Seasonal Variations in Air–Sea CO_2 Fluxes

This study reports the most comprehensive dataset of CO_2 fluxes based on satellite data with complete coverage of the SCS at a monthly temporal resolution and a spatial resolution of 1 km over 17 years. We employed the same calculation approach of air–sea CO_2 flux adopted by Li et al. [2]. The climatology of the monthly mean air–sea CO_2 flux for the entire study area is shown in Fig. 10. Most parts of the SCS were weak-to-moderate sources of atmospheric CO_2 . The northern shelf and central basin showed different seasonal patterns of air–sea CO_2 fluxes. In addition, our results of the seasonal seawater $p\text{CO}_2$ and air–sea CO_2 flux are basically consistent with statistics on five subregions defined in [2] (see Tables S1 and S2 and Text S2 for details). There are certain differences in some seasons and regions, where there are very few in situ measurements covered by cruise. Also, the satellite-derived results have better spatiotemporal statistical properties than those from limited in situ observations.

The northern shelf serves as an intense CO_2 sink in typical winter and early spring months (with maximum CO_2 influx

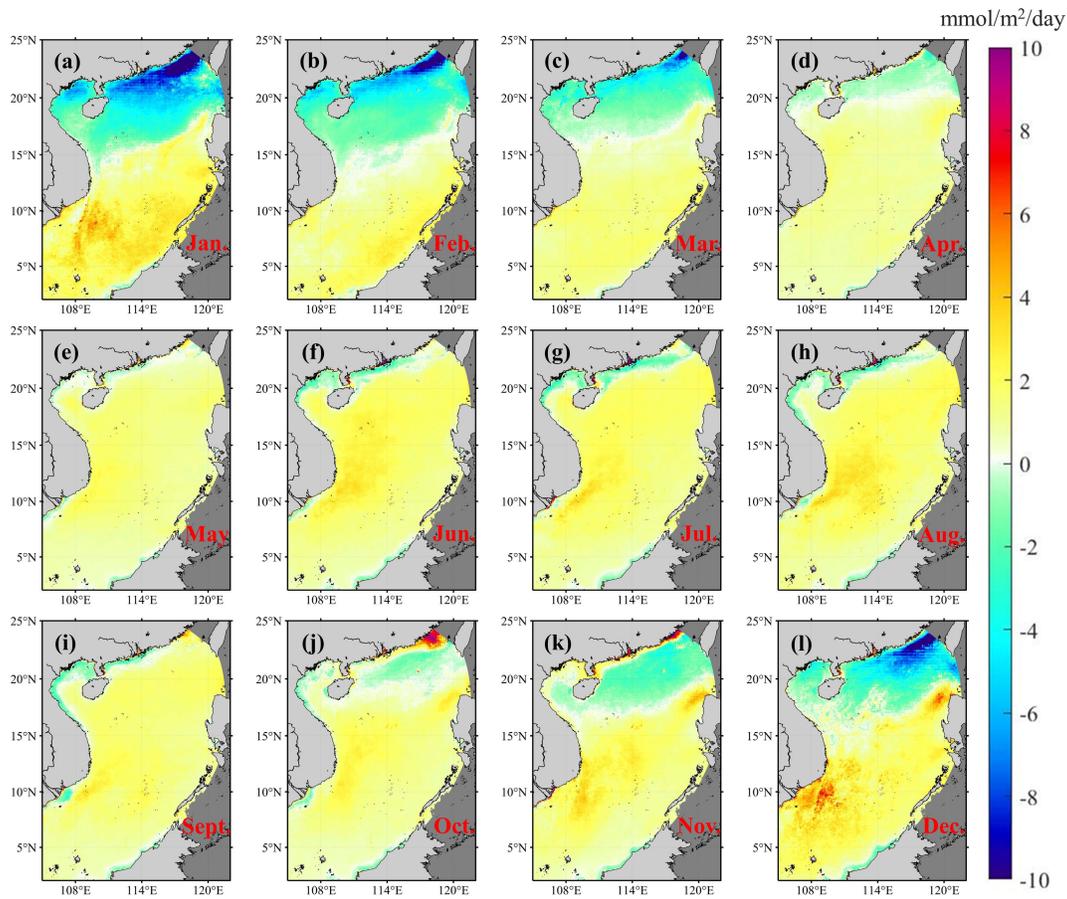


Fig. 10. Distribution of monthly averaged air-sea CO_2 flux from 2003 to 2019. (a)–(l) From January to December.

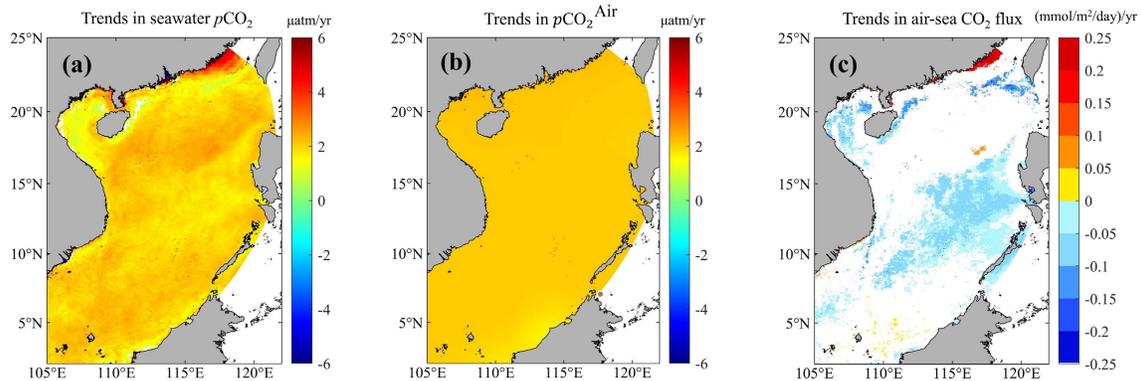


Fig. 11. Trends in (a) seawater $p\text{CO}_2$, (b) $p\text{CO}_2^{\text{Air}}$, and (c) air-sea CO_2 flux. Only pixels with significant trends ($p < 0.05$) are colored.

rates of 8–10 $\text{mmol/m}^2/\text{day}$ in January) and as a weak CO_2 source from May to September (with CO_2 efflux rates of 1–2 $\text{mmol/m}^2/\text{day}$). This pattern is consistent with that reported by Li et al. [2] based on in situ data. However, high-resolution satellite estimations showed the Pearl River plume with a CO_2 influx of 1–2 $\text{mmol/m}^2/\text{day}$ in summer. In April and October, the northern shelf showed transitional features between the typical CO_2 influx and CO_2 efflux. The very near-shore area along the Guangdong and Fujian Provinces released large amounts of CO_2 (over 8 $\text{mmol/m}^2/\text{day}$) into the atmosphere, corresponding to the far-oversaturated CO_2 from April to December, and the CO_2 efflux zone extended in October and

November. The strong carbon source here needs to be further confirmed and verified by measured data.

Overall, the central basin serves as a weak source throughout the year, with a mean efflux rate of 1 $\text{mmol/m}^2/\text{day}$. The maximum efflux occurred in the southern part of the basin in December and January, reaching 4–6 $\text{mmol/m}^2/\text{day}$. However, the northern edge of the central basin absorbs CO_2 in winter and releases CO_2 , similar to the other parts of basin, in the rest of the seasons. Noticeably, our CO_2 flux field characterizes a relatively strong CO_2 source located northwest of the Luzon Island in December and January, observed previously from in situ data [1], [2].

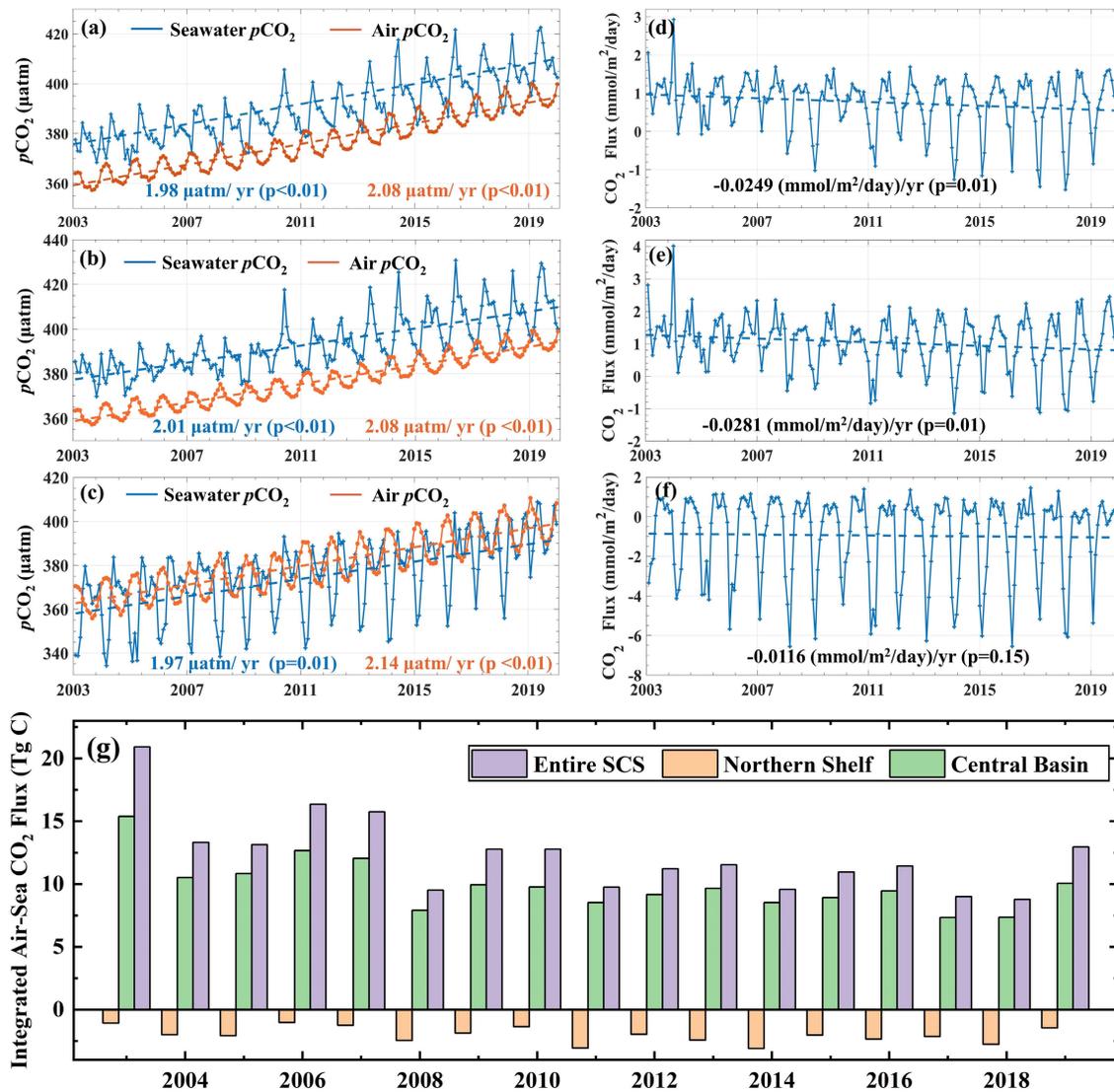


Fig. 12. (a)–(c) Long-term series of seawater and atmospheric $p\text{CO}_2$ in the entire SCS, central basin, and northern shelf. (d)–(f) Long-term series of air–sea CO_2 flux density in the entire SCS, central basin, and northern shelf. Dashed lines are linear fitting curves. (g) Annual integrated air–sea CO_2 flux in 2003–2019. The entire SCS is defined by the red lines of Fig. 1, and the northern shelf and center basin are bounded by the pink line.

Annually, the regional-averaged air–sea CO_2 flux was summarized at $-1.12 \pm 0.30 \text{ mmol/m}^2/\text{day}$ in the northern shelf and $1.34 \pm 0.23 \text{ mmol/m}^2/\text{day}$ in the central basin. The average of air–sea CO_2 flux in the entire SCS was estimated to be $0.99 \pm 0.22 \text{ mmol/m}^2/\text{day}$.

For the area-integrated CO_2 flux, the annual mean CO_2 sink of the northern shelf was $2.02 \pm 0.64 \text{ Tg C}$ (with an area of $4.15 \times 10^5 \text{ km}^2$), accumulating at 34.27 Tg C during 2003–2019. The annual mean CO_2 source from the central basin was $9.89 \pm 2.03 \text{ Tg C}$ (from an area of $1.71 \times 10^6 \text{ km}^2$) and has accumulated to 168.07 Tg C over 17 years. The entire SCS released $12.34 \pm 3.11 \text{ Tg C}$ (from an area of $2.87 \times 10^6 \text{ km}^2$) annually and accumulated 209.79 Tg C during 2003–2019 to the atmosphere.

C. Long-Term Trends in Seawater $p\text{CO}_2$ and Air–Sea CO_2 Fluxes

Figs. 11 and 12 show the interannual variation in seawater and atmospheric $p\text{CO}_2$ and air–sea CO_2 fluxes in the SCS.

The Pearl River Estuary and pixels close to the shore were not involved in the trend analysis, because of poor data coverage (data for 204 months from 2003 to 2019 were generated in this study, and the available data for these regions were less than 102 months).

The trends in seawater $p\text{CO}_2$ in the SCS during 2003–2019 were generally significant [see Figs. 11(a) and 12(a)], with a rate of $1.98 \mu\text{atm/year}$ ($p < 0.01$), which is slightly slower than that of $p\text{CO}_2^{\text{air}}$ [$2.08 \mu\text{atm/year}$, $p < 0.01$, Fig. 12(a)]. Hence, the trend in air–sea CO_2 flux density, $-0.0249 \text{ (mmol/m}^2/\text{day)/yr}$, is significant [$p = 0.01$, Fig. 12(d)]. However, the annual carbon source of the whole SCS decreased by $\sim 43\%$ from 2003 to 2019 ($p < 0.01$), and the total area-integrated carbon release fell from averaged 15.90 Tg C in 2003–2008 to 10.63 Tg C in 2015–2019 [see Fig. 12(g)].

On the northern shelf, trends in seawater and atmospheric $p\text{CO}_2$ [1.97 and $2.14 \mu\text{atm/year}$, respectively, see Fig. 12(c)] are similar; thus, the air–sea CO_2 flux did not show a significant trend. However, rapid $p\text{CO}_2$ and air–sea

CO_2 flux increased at a rate of $\sim 4 \mu\text{atm}/\text{year}$ and over $0.25 \text{ (mmol/m}^2/\text{day)}/\text{year}$ were observed along the coast of the Fujian Province. To investigate the high increase rate, we further made a support check by comparing the retrieved and in situ $p\text{CO}_2$ at four sites with multiple observations. The results indicate that this rapid increase could be reliable as the retrieved $p\text{CO}_2$ values exhibited good accuracy (see Fig. S55 for details). Significant air–sea CO_2 flux decreases were captured in sea areas to the southwest of Taiwan Island and northeast of Hainan Island, and the Beibu Gulf with the rate of $0.04\text{--}0.06 \text{ (mmol/m}^2/\text{day)}/\text{year}$ due to the relatively slow increase in seawater $p\text{CO}_2$ ($\sim 1.0\text{--}1.5 \mu\text{atm}/\text{year}$) compared to the atmospheric $p\text{CO}_2$. The annual area-integrated carbon sink fluctuated between 1.0 and 3.0 Tg C without any significant trend ($p = 0.12$) in the northern shelf.

In the central basin, seawater $p\text{CO}_2$ generally increased following $p\text{CO}_2^{\text{air}}$ (at a rate of 2.01 and $2.08 \mu\text{atm}/\text{year}$). The increase of $p\text{CO}_2$ in the north of the basin is slightly faster than that in the south, even reaching $3 \mu\text{atm}/\text{year}$ in some areas, which exceeds the rise of the atmospheric $p\text{CO}_2$, leading to a patch (16°N , 116°E) showing enhanced CO_2 source in the flux trend map [see Fig. 11(c)]. The air–sea CO_2 flux serves as a source weakened by the annual rate of $0.02\text{--}0.05 \text{ mmol/m}^2/\text{day}$ in the mid and east parts of the central basin [see Fig. 11(c)] resulting from the slightly slow increase in seawater $p\text{CO}_2$ [$\sim 1.75 \mu\text{atm}/\text{year}$, Fig. 11(a)]. Thus, the annual area-integrated CO_2 efflux in the central basin decreased from averaged $12.28 \text{ Tg C}/\text{year}$ in 2003–2007 to $8.63 \text{ Tg C}/\text{year}$ in 2015–2019 [at the rate of $0.28 \text{ Tg C}/\text{year}$, $p = 0.002$, Fig. 12(g)].

It is reported that the uptake of CO_2 in global shelves had enhanced under continually increasing atmospheric $p\text{CO}_2$ and may have switched from a source to a sink of CO_2 [51]. In the estuaries, such as the Mississippi Estuary, CO_2 uptakes increased rapidly since seawater $p\text{CO}_2$ did not show a significant positive trend ($-0.11 \mu\text{atm}/\text{year}$ in the Mississippi Estuary) [32]. In the northern shelf of the SCS [see Fig. 12(c)], the increasing rate of $p\text{CO}_2$ in seawater ($1.97 \mu\text{atm}/\text{year}$) was significant, but smaller than that of atmospheric $p\text{CO}_2$ ($2.14 \mu\text{atm}/\text{year}$) can also lead to an increase in CO_2 uptakes. For the entire SCS, long-term trends in seawater $p\text{CO}_2$ are more similar to that in the open ocean. Take the Hawaii Ocean Time-series (HOT) and Bermuda Atlantic Time-series Study (BATS) for example, the $p\text{CO}_2$ increasing trends of which were 1.88 ± 0.16 (1988–2007) [52] and 1.78 ± 0.28 (1970–2011) [53] $\mu\text{atm}/\text{year}$, respectively, close to (still slower than) the atmospheric $p\text{CO}_2$ increases. As a CO_2 source in the entire SCS, the increasing seawater $p\text{CO}_2$ ($1.86 \mu\text{atm}/\text{year}$) was slower than that of the atmospheric $p\text{CO}_2$ ($2.08 \mu\text{atm}/\text{year}$), resulting in gently reducing the CO_2 release of the SCS. If this trend continues, the SCS will likely reach equilibrium in the future.

V. CONCLUSION

In this study, the MeSAA-ML-SCS model was developed to retrieve seawater $p\text{CO}_2$ based on controlling mechanism analysis and ML method. Then, we produced high spatial resolution (1 km, monthly) seawater $p\text{CO}_2$ and air–sea CO_2

flux maps of the SCS during 2003–2019, which were validated with independent in situ data, including the SEATS time-series site data. The $p\text{CO}_2$ product showed high accuracy with an $11.69\text{-}\mu\text{atm}$ RMSE and 1.59% APD.

Sea surface $p\text{CO}_2$ and air–sea CO_2 fluxes in the SCS presented substantial spatiotemporal variations, with significant long-term trends during 2003–2019. Our results suggested the SCS as a whole to be a source of atmospheric CO_2 , releasing an average of $12.34 \pm 3.11 \text{ Tg C}/\text{year}$, and the northern shelf to be an atmospheric CO_2 sink, absorbing an average of $2.02 \pm 0.64 \text{ Tg C}/\text{year}$. Seawater $p\text{CO}_2$ in the SCS increased following the trend of $p\text{CO}_2^{\text{air}}$ but was slightly slower; thus, the air–sea CO_2 flux decreased significantly ($p = 0.01$). However, the area-integrated CO_2 efflux in the entire SCS and central basin decreased significantly ($p < 0.01$) in 2003–2019, with rates of 0.41 and $0.28 \text{ Tg C}/\text{year}$, respectively. In addition to the seawater $p\text{CO}_2$ and air–sea CO_2 fluxes, the high spatial resolution and long temporal span dataset helps to reveal the controlling factors of the air–sea CO_2 fluxes in the SCS and predict the CO_2 source and sink patterns.

Our satellite-derived seawater $p\text{CO}_2$ and air–sea CO_2 flux data are shared at Zenodo (<https://doi.org/10.5281/zenodo.7743187>). Although only preliminary analysis was conducted in this study, our dataset has the potential to capture variations in $p\text{CO}_2$ induced by mesoscale processes, such as typhoons and eddies, due to its high resolution. We hope that our work will inspire and encourage the community and scholars to use our dataset for further studies. As the MeSAA-ML-SCS algorithm may have some uncertainty in the area without in situ data, we will collect more field measured data to validate the satellite products and improve the algorithm by introducing more mechanism analysis and parameterizing controlling factors to the ML model, and reveal more detail features in the carbonate system and their mechanisms.

ACKNOWLEDGMENT

The authors thank all the members of the satellite ground station, satellite data processing, and sharing center, and marine satellite data online analysis platform (SatCO2) of the State Key Laboratory of Satellite Ocean Environment Dynamics (SOED)/Second Institute of Oceanography (SIO)/Ministry of Natural Resources (MNR) for assistance with data collection and processing.

REFERENCES

- [1] W.-D. Zhai et al., “Seasonal variations of sea-air CO_2 fluxes in the largest tropical marginal sea (South China Sea) based on multiple-year underway measurements,” *Biogeosciences*, vol. 10, no. 11, pp. 7775–7791, 2013.
- [2] Q. Li, X. Guo, W. Zhai, Y. Xu, and M. Dai, “Partial pressure of CO_2 and air–sea CO_2 fluxes in the South China Sea: Synthesis of an 18-year dataset,” *Prog. Oceanogr.*, vol. 182, Mar. 2020, Art. no. 102272.
- [3] C.-T. A. Chen, S.-L. Wang, B.-J. Wang, and S.-C. Pai, “Nutrient budgets for the South China Sea basin,” *Mar. Chem.*, vol. 75, no. 4, pp. 281–300, 2001.
- [4] M. Dai et al., “Why are some marginal seas sources of atmospheric CO_2 ?” *Geophys. Res. Lett.*, vol. 40, no. 10, pp. 2154–2158, 2013.
- [5] C. Du et al., “Impact of the Kuroshio intrusion on the nutrient inventory in the upper northern South China Sea: Insights from an isopycnal mixing model,” *Biogeosciences*, vol. 10, no. 10, pp. 6419–6432, 2013.

- [6] H. Cao, Y. Hong, M. Li, and J.-D. Gu, "Diversity and abundance of ammonia-oxidizing prokaryotes in sediments from the coastal Pearl River estuary to the South China Sea," *Antonie Van Leeuwenhoek*, vol. 100, no. 4, pp. 545–556, 2011.
- [7] Y. Zhu, S. Shang, W. Zhai, and M. Dai, "Satellite-derived surface water $p\text{CO}_2$ and air-sea CO_2 fluxes in the northern South China Sea in summer," *Prog. Natural Sci.*, vol. 19, no. 6, pp. 775–779, 2009.
- [8] Y.-H. Jo, M. Dai, W. Zhai, X.-H. Yan, and S. Shang, "On the variations of sea surface $p\text{CO}_2$ in the northern South China Sea: A remote sensing based neural network approach," *J. Geophys. Res. Oceans*, vol. 117, no. C8, pp. 1–13, 2012.
- [9] H. Lv, Y. Bai, Q. Li, and H. Jiang, "Satellite remote sensing retrieval of aquatic $p\text{CO}_2$ in summer in the Pearl River Estuary (in Chinese)," *J. Mar. Sci.*, vol. 36, no. 2, pp. 1–11, 2018.
- [10] Y. Bai et al., "A mechanistic semi-analytical method for remotely sensing sea surface $p\text{CO}_2$ in river-dominated coastal oceans: A case study from the East China Sea," *J. Geophys. Res. Oceans*, vol. 120, no. 3, pp. 2331–2349, 2015.
- [11] H. Zhao et al., "River-dominated $p\text{CO}_2$ dynamics in the northern South China Sea during summer: A modeling study," *Prog. Oceanogr.*, vol. 190, Jan. 2021, Art. no. 102457.
- [12] G. Wang et al., "Feasibility of reconstructing the summer basin-scale sea surface partial pressure of carbon dioxide from sparse in situ observations over the South China Sea," *Earth Syst. Sci. Data*, vol. 13, no. 3, pp. 1403–1417, 2021.
- [13] Y. Bai, S. Yu, Z. Song, and X. He. (2021). *Remote Sensing Based Sea Surface Partial Pressure of CO_2 ($p\text{CO}_2$) in China Seas (2003–2019) version 2.0*. [Online]. Available: <https://doi.org/10.5281/zenodo.7372479>
- [14] Z. Wang, G. Wang, X. Guo, Y. Bai, Y. Xu, and M. Dai, "Spatial reconstruction of long-term (2003–2020) sea surface $p\text{CO}_2$ in the South China Sea using a machine-learning-based regression method aided by empirical orthogonal function analysis," *Earth Syst. Sci. Data*, vol. 15, no. 4, pp. 1711–1731, 2023.
- [15] W. D. Zhai, M. Dai, and W.-J. Cai, "Coupling of surface $p\text{CO}_2$ and dissolved oxygen in the northern South China Sea: Impacts of contrasting coastal processes," *Biogeosciences*, vol. 6, no. 11, pp. 2589–2598, 2009.
- [16] W. Zhai, "Sea surface partial pressure of CO_2 and its controls in the northern South China Sea in the non-bloom period in spring," *Haiyang Xuebao*, vol. 37, no. 6, pp. 31–37, 2015, doi: [10.3969/j.issn.0253-4193.2015.06.004](https://doi.org/10.3969/j.issn.0253-4193.2015.06.004).
- [17] T. Takahashi, S. C. Sutherland, and A. Kozyr, "Global ocean surface water partial pressure of CO_2 database: Measurements performed during 1957–2016 (version 2016)," Carbon Dioxide Information Analysis Center, Oak Ridge National Lab., U.S. Dept. Energy, Oak Ridge, TN, USA, Tech. Rep. NDP-088, 2015. [Online]. Available: https://www.ncei.noaa.gov/access/ocean-carbon-acidification-data-system/oceans/LDEO_Underway_Database/NDP-088_V2016.pdf
- [18] R. F. Weiss, "Carbon dioxide in water and seawater: The solubility of a non-ideal gas," *Mar. Chem.*, vol. 2, no. 3, pp. 203–215, 1974.
- [19] C. Sweeney et al., "Constraining global air-sea gas exchange for CO_2 with recent bomb ^{14}C measurements," *Global Biogeochem. Cycles*, vol. 21, no. 2, pp. 1–10, 2007.
- [20] R. Wanninkhof, S. C. Doney, T. Takahashi, and W. R. Mcgillis, "The effect of using time-averaged winds on regional air-sea CO_2 fluxes," *Geophys. Monograph-Amer. Geophys. Union*, vol. 127, pp. 351–356, Jan. 2002.
- [21] L.-Q. Jiang, W.-J. Cai, R. Wanninkhof, Y. Wang, and H. Lüger, "Air-sea CO_2 fluxes on the U.S. South Atlantic Bight: Spatial and seasonal variability," *J. Geophys. Res. Oceans*, vol. 113, no. C7, pp. 1–17, Jul. 2008.
- [22] R. Wanninkhof, "Relationship between wind speed and gas exchange over the ocean," *J. Geophys. Res. Oceans*, vol. 97, no. 5, pp. 7373–7382, 1992.
- [23] R. F. Weiss and B. A. Price, "Nitrous oxide solubility in water and seawater," *Mar. Chem.*, vol. 8, no. 4, pp. 347–359, 1980.
- [24] X. Song et al., "Remote sensing of sea surface $p\text{CO}_2$ in the Bering Sea in summer based on a mechanistic semi-analytical algorithm (MeSAA)," *Remote Sens.*, vol. 8, no. 7, p. 558, 2016.
- [25] C. Le, Y. Gao, W.-J. Cai, J. C. Lehrter, Y. Bai, and Z.-P. Jiang, "Estimating summer sea surface $p\text{CO}_2$ on a river-dominated continental shelf using a satellite-based semi-mechanistic model," *Remote Sens. Environ.*, vol. 225, pp. 115–126, May 2019.
- [26] S. Zhang et al., "The carbon sink of the Coral Sea, the world's second largest marginal sea, weakened during 2006–2018," *Sci. Total Environ.*, vol. 872, May 2023, Art. no. 162219.
- [27] T. Takahashi, J. Olafsson, J. G. Goddard, D. W. Chipman, and S. C. Sutherland, "Seasonal variation of CO_2 and nutrients in the high-latitude surface oceans: A comparative study," *Global Biogeochem. Cycles*, vol. 7, no. 4, pp. 843–878, 1993.
- [28] N. Jiao et al., "Revisiting the CO_2 'source' problem upwelling areas—A comparative study eddy upwellings South China Sea," *Biogeosciences*, vol. 11, no. 9, pp. 2465–2475, 2014.
- [29] I. Alvarez, M. Gomez-Gesteira, M. DeCastro, M. N. Lorenzo, A. J. C. Crespo, and J. M. Dias, "Comparative analysis of upwelling influence between the western and northern coast of the Iberian Peninsula," *Continental Shelf Res.*, vol. 31, no. 5, pp. 388–399, 2011.
- [30] C. Hu et al., "Improving satellite global chlorophyll a data products through algorithm refinement and data recovery," *J. Geophys. Res. Oceans*, vol. 124, no. 3, pp. 1524–1543, 2019.
- [31] J. E. O'Reilly and P. J. Werdell, "Chlorophyll algorithms for ocean color sensors—OC4, OC5 & OC6," *Remote Sens. Environ.*, vol. 229, pp. 32–47, Aug. 2019.
- [32] S. Chen et al., "A machine learning approach to estimate surface ocean $p\text{CO}_2$ from satellite measurements," *Remote Sens. Environ.*, vol. 228, pp. 203–226, Jul. 2019.
- [33] Z. Fu et al., "Estimating spatial and temporal variation in ocean surface $p\text{CO}_2$ in the Gulf of Mexico using remote sensing and machine learning techniques," *Sci. Total Environ.*, vol. 745, Nov. 2020, Art. no. 140965.
- [34] L. Gloege, M. Yan, T. Zheng, and G. A. McKinley, "Improved quantification of ocean carbon uptake by using machine learning to merge global models and $p\text{CO}_2$ data," *J. Adv. Model. Earth Syst.*, vol. 14, no. 2, 2022, Art. no. e2021MS002620.
- [35] Z. Jiang et al., "Remote sensing of global sea surface pH based on massive underway data and machine learning," *Remote Sens.*, vol. 14, no. 10, p. 2366, 2022.
- [36] A. P. Joshi, V. Kumar, and H. V. Warrior, "Modeling the sea-surface $p\text{CO}_2$ of the central Bay of Bengal region using machine learning algorithms," *Ocean Model.*, vol. 178, Oct. 2022, Art. no. 102094.
- [37] C. M. Bishop, *Neural Networks for Pattern Recognition*. Oxford, U.K.: Oxford Univ. Press, 1995.
- [38] L. Gross, S. Thiria, and R. Frouin, "Applying artificial neural network methodology to ocean color remote sensing," *Ecol. Model.*, vol. 120, nos. 2–3, pp. 237–246, 1999.
- [39] T. Chen and C. Guestrin, "Xgboost: A scalable tree boosting system," in *Proc. 22nd ACM SIGKDD Int. Conf. Knowl. Discovery Data Mining*, 2016, pp. 785–794.
- [40] L. Peng, Y. Tu, L. Huang, Y. Li, X. Fu, and X. Chen, "DAESTB: Inferring associations of small molecule-miRNA via a scalable tree boosting model based on deep autoencoder," *Briefings Bioinf.*, vol. 23, no. 6, 2022, Art. no. bbac478.
- [41] D. Carrión et al., "A 1-km hourly air-temperature model for 13 northeastern U.S. states using remotely sensed and ground-based measurements," *Environ. Res.*, vol. 200, Sep. 2021, Art. no. 111477.
- [42] Y. Qiu, J. Zhou, M. Khandelwal, H. Yang, P. Yang, and C. Li, "Performance evaluation of hybrid WOA-XGBoost, GWO-XGBoost and BO-XGBoost models to predict blast-induced ground vibration," *Eng. Comput.*, vol. 38, no. 5, pp. 1–18, 2021.
- [43] T. Tiyasha et al., "Functionalization of remote sensing and on-site data for simulating surface water dissolved oxygen: Development of hybrid tree-based artificial intelligence models," *Mar. Pollut. Bull.*, vol. 170, Sep. 2021, Art. no. 112639.
- [44] Y. Xia et al., "Rapid assessments of light-duty gasoline vehicle emissions using on-road remote sensing and machine learning," *Sci. Total Environ.*, vol. 815, Apr. 2022, Art. no. 152771.
- [45] Y. Zhang et al., "Improving remote sensing estimation of Secchi disk depth for global lakes and reservoirs using machine learning methods," *GISci. Remote Sens.*, vol. 59, no. 1, pp. 1367–1383, 2022.
- [46] D. Zhang, L. Qian, B. Mao, C. Huang, B. Huang, and Y. Si, "A data-driven design for fault detection of wind turbines using random forests and XGboost," *IEEE Access*, vol. 6, pp. 21020–21031, 2018.
- [47] A. G. Smirnov et al., "Medium energy electron flux in earth's outer radiation belt (MERLIN): A machine learning model," *Space Weather*, vol. 18, no. 11, 2020, Art. no. e2020SW002532.
- [48] S. Ramraj, N. Uzir, R. Sunil, and S. Banerjee, "Experimenting XGBoost algorithm for prediction and classification of different datasets," *Int. J. Control Theory Appl.*, vol. 9, no. 40, pp. 651–662, 2016.
- [49] M. Dai et al., "Carbon fluxes in the Coastal Ocean: Synthesis, boundary processes, and future trends," *Annu. Rev. Earth Planet. Sci.*, vol. 50, pp. 593–626, May 2022.

- [50] W. Zhai, M. Dai, W.-J. Cai, Y. Wang, and H. Hong, "The partial pressure of carbon dioxide and air-sea fluxes in the northern South China Sea in spring, summer and autumn," *Mar. Chem.*, vol. 96, nos. 1–2, pp. 87–97, 2005.
- [51] G. G. Laruelle, W.-J. Chai, X. Hu, N. Gruber, F. T. Mackenzie, and P. Regnier, "Continental shelves as a variable but increasing global sink for atmospheric carbon dioxide," *Nature Commun.*, vol. 9, no. 1, p. 454, 2018.
- [52] J. E. Dore, R. Lukas, D. W. Sadler, M. J. Church, and D. M. Karl, "Physical and biogeochemical modulation of ocean acidification in the central North Pacific," *Proc. Nat. Acad. Sci. USA*, vol. 106, no. 30, pp. 12235–12240, 2009.
- [53] J. F. Tjiputra et al., "Long-term surface $p\text{CO}_2$ trends from observations and models," *Tellus B Chem. Phys. Meteorol.*, vol. 66, no. 1, p. 23083, 2014.



Zigeng Song received the B.S. degree in marine and navigation technology from Jimei University, Xiamen, China, in 2019. He is currently pursuing the Ph.D. degree in marine science with Hohai University, Nanjing, China.

He is also with the State Key Laboratory of Satellite Ocean Environment Dynamics, Second Institute of Oceanography, Ministry of Natural Resources, Hangzhou, China. His research interests include atmospheric correction of satellite ocean color remote sensing and atmospheric aerosol remote sensing.



Shujie Yu received the B.S. degree from the Ocean University of China, Qingdao, China, in 2015, the M.S. degree from the Second Institute of Oceanography, Ministry of Natural Resources, Hangzhou, China, in 2018, and the Ph.D. degree in ocean engineering and technology from Zhejiang University, Zhoushan, China, in 2023.

She is currently with the Polar and Marine Research Institute, College of Harbor and Coastal Engineering, Jimei University, Xiamen, China. Her research interests include satellite remote sensing of marine ecology and carbon cycle.



Yan Bai received the B.S. and M.S. degrees from Sun Yat-sen University, Guangzhou, China, in 2000 and 2003, respectively, and the Ph.D. degree in electromagnetic fields and microwave techniques from the Shanghai Institute of Technical Physics, Chinese Academy of Sciences, Shanghai, China, in 2007.

She is currently with the State Key Laboratory of Satellite Ocean Environment Dynamics, Second Institute of Oceanography, Ministry of Natural Resources, Hangzhou, China, the Southern Marine Science and Engineering Guangdong Laboratory (Guangzhou), Guangzhou, and the School of Oceanography, Shanghai Jiao Tong University, Shanghai. Her research interests include satellite remote sensing of marine carbon cycle.



Xianghui Guo received the B.S. degree from Hebei University, Baoding, China, in 1998, the M.S. degree from the Huazhong University of Science and Technology, Wuhan, China, in 2001, and the Ph.D. degree in chemical oceanography from Xiamen University, Xiamen, China, in 2009.

She is currently with the State Key Laboratory of Marine Environmental Science, College of Ocean and Earth Sciences, Xiamen University. Her research interests include air-sea CO_2 fluxes and controls in the ocean and carbon and nitrogen cycles in the upper ocean.



Xianqiang He received the B.S. degree in marine and ocean engineering from the Huazhong University of Science and Technology, Wuhan, China, in 1999, the M.S. degree in physical oceanography from the Second Institute of Oceanography, Ministry of Natural Resources, Hangzhou, China, in 2002, and the Ph.D. degree in physical electronics from the Shanghai Institute of Technical Physics, Chinese Academy of Sciences, Shanghai, China, in 2007.

He is currently with the State Key Laboratory of Satellite Ocean Environment Dynamics, Second Institute of Oceanography, Ministry of Natural Resources, the Southern Marine Science and Engineering Guangdong Laboratory (Guangzhou), Guangzhou, China, and the School of Oceanography, Shanghai Jiao Tong University, Shanghai. His research interests include radiative transfer in the coupled ocean and atmosphere system, atmospheric correction of satellite ocean color remote sensing, and oceanography research using ocean color remote sensing data.



Weidong Zhai received the B.S. degree from Nankai University, Tianjin, China, in 1993, the M.S. degree from Northwest University, Xi'an, China, in 1999, and the Ph.D. degree in environmental science from Xiamen University, Xiamen, China, in 2002.

He is currently with the Southern Marine Science and Engineering Guangdong Laboratory (Zhuhai), Zhuhai, China. His research interests include marine carbon cycling, the responses of coastal ocean carbon cycle, and ecosystem to a changing terrestrial export of carbon and nutrients.



Minhan Dai received the B.S. degree from Xiamen University, Xiamen, China, in 1987, and the Ph.D. degree in Earth science from the University of Paris VI, Paris, France, in 1995.

He is currently with the State Key Laboratory of Marine Environmental Science, College of Ocean and Earth Sciences, Xiamen University. He is also an Academician with the Chinese Academy of Science, Beijing, China. His research interests include carbon and nutrient biogeochemistry in marginal sea and estuarine systems, biogeochemistry of the oligotrophic ocean, geochemistry of radioactive elements (plutonium and thorium) in surface and groundwater, geochemistry of trace metals in marine environments, and interdisciplinary research on coupled physical-biogeochemical processes.