Science of the Total Environment 904 (2023) 166804

Contents lists available at ScienceDirect

Science of the Total Environment

journal homepage: www.elsevier.com/locate/scitotenv

Satellite-estimated air-sea CO_2 fluxes in the Bohai Sea, Yellow Sea, and East China Sea: Patterns and variations during 2003–2019

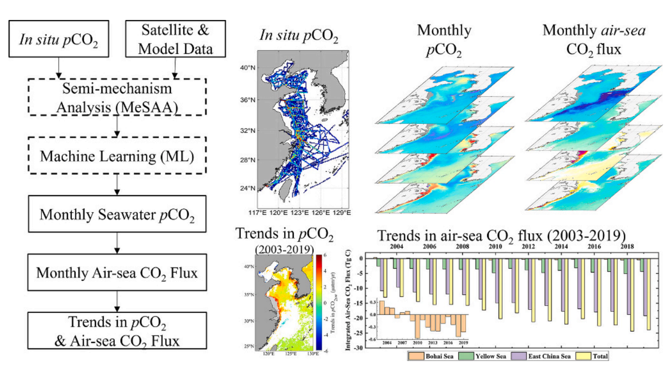
Shujie Yu [a,b,1], Zigeng Song [a,1], Yan Bai [a,b,*], Xianghui Guo [c], Xianqiang He [a,b], Weidong Zhai [d], Huade Zhao [e], Minhan Dai [c]

[a] State Key Laboratory of Satellite Ocean Environment Dynamics, Second Institute of Oceanography, Ministry of Natural Resources, Hangzhou 310012, PR China
[b] Ocean College, Zhejiang University, Zhoushan 316021, PR China
[c] State Key Laboratory of Marine Environmental Science and College of Ocean and Earth Sciences, Xiamen University, Xiamen 361102, PR China
[d] Southern Marine Science and Engineering Guangdong Laboratory (Zhuhai), Zhuhai 519080, PR China
[e] National Marine Environmental Monitoring Center, Dalian 116023, PR China

HIGHLIGHTS

- A pCO_2 algorithm combining semi-mechanistic and machine learning methods.
- The East China Sea was a CO_2 sink that annually absorbed 14.80 Tg C.
- The Yellow Sea was a weak CO_2 sink on an annual basis but a source in summer.
- The Bohai Sea is nearly neutral in annual scale with strong seasonal variations.

GRAPHICAL ABSTRACT



ARTICLE INFO

Editor: Martin Drews

Keywords:
Seawater pCO_2
Air-sea CO_2 flux
Semi-analytical algorithm (MeSAA)
Machine learning
Bohai Sea-Yellow Sea-East China Sea

ABSTRACT

The Bohai Sea (BS), Yellow Sea (YS), and East China Sea (ECS) together form one of the largest marginal sea systems in the world, including enclosed and semi-enclosed ocean margins and a wide continental shelf influenced by the Changjiang River and the strong western boundary current (Kuroshio). Based on *in situ* seawater pCO_2 data collected on 51 cruises/legs over the past two decades, a satellite retrieval algorithm for seawater pCO_2 was developed by combining the semi-mechanistic algorithm and machine learning method (MeSAA-ML-ECS). MeSAA-ML-ECS introduced semi-analytical parameters, including the temperature-dependent seawater pCO_2 ($pCO_{2,therm}$) and upwelling index (UI_{SST}), to characterise the combined effect of atmospheric CO_2 forcing, thermodynamic effects, and multiple mixing processes on seawater pCO_2. The best-selected machine learning algorithm is XGBoost. The satellite-derived pCO_2 achieved excellent performance in this complicated marginal sea, with low root mean square error (RMSE = 20 μatm) and mean absolute percentage deviation (APD = 4.12 %) for independent *in situ* validation dataset. During 2003–2019, the annual average CO_2 sinks in the BS, YS,

* Corresponding author at: State Key Laboratory of Satellite Ocean Environment Dynamics, Second Institute of Oceanography, Ministry of Natural Resources, Hangzhou 310012, PR China.
E-mail address: baiyan@sio.org.cn (Y. Bai).
[1] These authors contributed equally to this work.

https://doi.org/10.1016/j.scitotenv.2023.166804
Received 16 March 2023; Received in revised form 1 August 2023; Accepted 2 September 2023
Available online 7 September 2023

ECS, and entire study area were 0.16 ± 0.26, 3.85 ± 0.68, 14.80 ± 3.09, and 18.81 ± 3.81 Tg C/yr, respectively. Under continuously increasing atmospheric CO_2 concentration, the BS changed from a weak source to a weak sink, the YS experienced interannual fluctuations but did not show significant trend, while the ECS acted as a strong sink with CO_2 absorption increased from ~10 Tg C in 2003 to ~19 Tg C in 2019. In total, CO_2 uptake in the entire study area increased by 85 % in 17 years. For the first time, we present the most refined variation in the satellite-derived pCO_2 and air-sea CO_2 flux dataset. These complete ocean carbon sink statistics and new insights will benefit further research on carbon fixation and its potential capacity.

1. Introduction

The importance of marginal seas in the global carbon cycle has been widely recognised (Gruber, 2015; Liu et al., 2018; Dai et al., 2022) despite their small total area, ~ 8 % of the global ocean surface area. With the rapid growth of *in situ* ocean carbon observations (especially the underway sea surface partial pressure of CO_2 (pCO_2) measurements) in the past decades, the estimates of the air-sea CO_2 absorption in the global marginal seas have converged to 0.2 to 0.5 PgC/yr (Borges et al., 2005; Cai et al., 2006; Chen et al., 2013; Chen and Borges, 2009; Dai et al., 2013; Dai et al., 2022; Gruber, 2015; Laruelle et al., 2010; Laruelle et al., 2014; Liu et al., 2018; Roobaert et al., 2019), which accounts for 10 % to 20 % of the annual CO_2 absorption in the global ocean (Cai et al., 2006; Chen and Borges, 2009; Laruelle et al., 2010; Roobaert et al., 2019). However, this estimated value still has significant uncertainty. Due to the complexity of the marginal seas, determining the refined temporal and spatial variations in seawater pCO_2 and air-sea CO_2 flux is still a considerable challenge (Dai et al., 2022; Gruber, 2015). These marine systems receive a large amount of carbon and nutrients from river runoff, upwelling, and the open ocean, which promote high biological production and respiration, and the interactions of physical and biogeochemical processes significantly affect the carbon cycle processes in the marginal seas (Chen and Borges, 2009; Dai et al., 2022; Hales et al., 2005; Huang et al., 2015; Roobaert et al., 2019; Wang et al., 2013).

The Bohai Sea (BS), Yellow Sea (YS), and East China Sea (ECS) system, adjacent to the Chinese mainland and the Korean Peninsula, account for approximately 3 % (1.27×10^6 km^2) of the global marginal sea area, of which 70 % (0.9×10^6 km^2) is located on the continental shelf with a water depth of <200 m. The carbon cycle here is quite challenging to decipher due to the complicated hydrodynamic and biogeochemical processes (Bauer et al., 2013; Dai et al., 2004; Shim et al., 2007; Zhai et al., 2014). The average water depth of the BS is only 18 m, and it is surrounded by land on three sides and many rivers along the coast (including the Yellow River, which has a high sediment concentration), so it is significantly affected by terrestrial inputs and human activities. The YS, with an average water depth of 44 m, is a semi-enclosed marginal sea that is strongly influenced by anthropogenic disturbance, terrestrial inputs, and the East Asian monsoon, thus forming a unique internal circulation current system with a long residence time of the water mass (Su, 1998). The ECS is featured by a large amount of material transport from the Changjiang River (the fourth largest river in the world in terms of runoff) on the west side, and the east side is subject to dynamic exchanges with the oligotrophic Kuroshio (western boundary current) (Chen and Wang, 1999).

Many studies have been conducted to determine whether the complex marginal seas like BS, YS, and ECS are sources or sinks of atmospheric CO_2 (Choi et al., 2019; Guo et al., 2015; Liu et al., 2018; Shim et al., 2007; Tseng et al., 2014; Tseng et al., 2011; Tsunogai et al., 1999; Wang et al., 2000; Wang and Zhai, 2021; Xue et al., 2012; Xue et al., 2011; Zhai and Dai, 2009). Due to the high heterogeneity of the coastal and continental shelf regions, the biogeochemical process and the air-sea CO_2 exchange flux have significant spatial and temporal variability (Deng et al., 2021; Takahashi et al., 2009). Marine CO_2 dynamics in the BS have not been well documented owing to a paucity of observations. Based on four surveys conducted by the State Oceanic Administration of China (SOA) during 2011–2012, the *Bulletin of Marine Environmental Status of China in 2012* (SOA, 2013) reported that the BS is a net source of atmospheric CO_2 (0.55 mmol/m^2/d) on an annual scale, with a strong carbon efflux (3.9 mmol/m^2/d) in autumn, but a weak sink in spring and winter (−0.2 and −1.9 mmol/m^2/d). Yin et al. (2012) collected *in situ* pCO_2 from 30 stations around the BS and reported that it acted as a weak source (averaged 5.03 mmol/m^2/d) in September 2009. There are more studies in the YS than in the BS, but considerable uncertainties remain in estimating the net CO_2 flux. Xue et al. (2011 & 2012), based on data collected in the South and North YS (SYS and NYS) during nine cruises covering all seasons in 2001 and 2005–2007, concluded that the YS acted as a net CO_2 source (3.16 mmol/m^2/d). However, recent reports have reconsidered the YS as a sink of CO_2 with a yearly mean influx rate of 2.39–2.79 mmol/m^2/d (Choi et al., 2019; Wang and Zhai, 2021; Xu et al., 2016). Even on a seasonal scale, a variety of source–sink patterns still exist, as Qu et al. (2014, 2015 & 2017) noted that the SYS was a source of atmospheric CO_2 in the summer of 2011 (2.8 mmol/m^2/d) and a CO_2 sink in the summers of 2012 and 2013 (−2.63 mmol/m^2/d). The ECS has the most abundant observations among the three seas. Previous studies have shown that the ECS is generally a net sink of atmospheric CO_2 on an annual scale and is one of the strongest carbon sinks in the world, with significant seasonal variations (Chou et al., 2011; Chou et al., 2009; Deng et al., 2021; Guo et al., 2015; Kim et al., 2013; Liu et al., 2022; PENG et al., 1999; Qu et al., 2013; Qu et al., 2017; Qu et al., 2015; Shim et al., 2007; Tseng et al., 2014; Tseng et al., 2011; Tsunogai et al., 1999; Wang et al., 2000; Zhai and Dai, 2009). Guo et al. (2015) collected 24 sampling surveys in the ECS from 2006 to 2011 and reported that the current air-sea CO_2 flux estimates range from −3.3 to −6.5 mmol/m^2/d in spring, −2.4 to −4.8 mmol/m^2/d in summer, 0.4 to 2.9 mmol/m^2/d in autumn and −13.7 to −10.4 mmol/m^2/d in winter. However, it has also been reported that near-shore areas outside the Changjiang River estuary acted as a source of atmospheric CO_2 in the summers of 2003 and 2018, with observed air-sea CO_2 fluxes of 3.29 and 1.75 mmol/m^2/d, respectively (Liu et al., 2022; Zhai and Dai, 2009).

Generally, previous estimates of the air-sea CO_2 flux in the BS, YS, and ECS were based mainly on limited surveys with spatial and observation frequency restrictions; thus, they are unable to reveal long-term changes in the carbon cycle process. To obtain observations with higher frequency and broader coverage, remote sensing data have become an essential tool for estimating seawater pCO_2 (Chen et al., 2011; Chen et al., 2019; Olsen et al., 2004; Sarma, 2003; Sarma et al., 2006; Stephens et al., 1995). Few remote sensing algorithms for seawater pCO2 in the Changjiang River-ECS system have been reported thus far, owing to the high spatial heterogeneity and complex biogeochemistry characteristics of various water masses. Tseng et al. (2011 & 2014) proposed an empirical algorithm using the Changjiang discharge and sea surface temperature (SST) to calculate the monthly average seawater pCO_2 from 1998 to 2011. Their results showed that the entire ECS continental shelf is a significant CO_2 sink, and the annual CO_2 flux is -4.9 ± 1.4 mmol/m^2/d. To overcome the poor applicability of a single empirical method to multiple regions with different biogeochemical features, Bai et al. (2015) proposed a mechanistic semi-analytical method (MeSAA), summarised the seawater pCO_2 variations resulting from a series of control processes, including thermodynamics, water mass mixing, and biological effects; and carried out an item by item parameterisation, thus retrieving summer seawater pCO_2 values on the ECS shelf. Since MeSAA is more effective and mechanistic than the empirical regression method, it has been applied successfully to the Bering Sea (Song et al., 2016b), the Northern Shelf off the Pearl River estuary in the South China Sea (Lv et al., 2018), the Gulf of Mexico (Le et al., 2019), and the Coral Sea (Zhang et al., 2023).

Although the semi-mechanical model has interpretational advantages, the existing MeSAA algorithm still has considerable uncertainty in its ability to analyse complicated biogeochemical processes owing to its insufficient parameterisation capability. Recently, machine learning algorithms such as neural networks and random forests have been applied to retrieve seawater pCO_2 values (Chen et al., 2019; Joshi et al., 2022; Landschützer et al., 2014; Laruelle et al., 2017; Roobaert et al., 2019). The continuous accumulation of underway pCO_2 measurements also promotes the use of machine learning methods. Although big data-driven methods have statistical advantages, we still need to fully use the mechanisms we understand and add more mechanistic or semi-mechanistic algorithms into the machine learning network, which is also a development trend in the new generation of artificial intelligence.

Therefore, this study proposes a model combining mechanical analytic and machine learning methods to estimate the seawater pCO_2 in complicated marginal seas. We collected *in situ* data on 51 cruises/legs in the BS, YS, and ECS, the most abundant data thus far, to support the development of machine learning algorithms. We also updated the MeSAA algorithm by considering the impact of thermodynamics, water mass mixing, biological effects, and atmospheric CO_2 forcing on seawater pCO_2, analysed the parameterisation and input combination strategies of these mechanisms, and compared the performance of various machine learning methods, thus establishing the pCO_2 retrieval algorithm, MeSAA-ML-ECS (MeSAA-Machine Learning). We produced monthly average pCO_2 and CO_2 fluxes with a high spatial resolution (1 km) for 17 years from 2003 to 2019 and showed the high-precision spatial and seasonal distributions, as well as long-term trends, of pCO_2 and carbon source–sink patterns in the BS, YS, and ECS for the first time.

This paper is arranged as follows. In Section 2, we introduced the sources and processing methods of the *in situ*, satellite, and model data we used, as well as the performance evaluation indicators and air-sea CO_2 flux calculation methods. In Section 3, based on the MeSAA algorithm, we described the parameterisation of the main control factors of seawater pCO_2, as well as the input strategy and machine learning model selection. Then we verified the final MeSAA-ML-ECS model with *in situ* pCO_2 data from independent cruise data. In Section 4, we displayed satellite-based pCO_2 and air-sea CO_2 flux maps in the BS, YS, and ECS, compared our results with previous studies to illustrate the advantages of our products in mechanism understanding and statistics, and finally showed the most refined long-term trends in pCO_2 and CO_2 fluxes.

2. Materials and methods

2.1. In situ data

A total of 1,048,166 underway pCO_2 records, collected on 51 cruises/legs from 2003 to 2019, are compiled in this study and shown in Table S1. Data from 35 cruises/legs were collected from Guo et al. (2015); Guo et al. (2021); Wang et al. (2014); Wang and Zhai (2021); Zhai et al. (2014); Zhai and Dai (2009). Underway data from ten cruises/legs were provided by the Surface Ocean CO₂ Atlas (SOCAT, version 2021) in the form of fCO_2 (fugacity of carbon dioxide), which was converted to pCO_2 using the corresponding *in situ* sea surface temperature (SST) and equation reported in Takahashi et al. (2017):

$$pCO_2 = fCO_2 \times \left(1.00436 - 4.66910^{-5}\text{SST}\right) \tag{1}$$

Seven unpublished surveying cruises were included, one of which was conducted in the eastern YS in June 2009 by the State Key Laboratory of Marine Environmental Science (MEL), Xiamen University, and the others were conducted in the BS and provided by Dr. Huade Zhao. The pCO_2 measurement of these unpublished surveys followed the procedure described by Guo et al. (2015). pCO_2/fCO_2 in all cruises were continuously measured every 80 s (Guo et al., 2015).

Fig. 1(a) shows the tracks of all the cruises/legs, whose data covered most of the BS, western YS, and ECS, with fewer data collected on the east side (deep ECS). The annual and seasonal spans of the *in situ* data are shown in Fig. 1(b) and (c). The observations were conducted every year from 2004 to 2019; 2007 had the maximum data volume (~1.9×10^4 records), and few data were collected in 2013, 2015, and 2019. These *in situ* data were soundly representative of seasonal patterns, with at least 1000 records collected each month.

2.2. Satellite and modeled data

The ocean colour data, *i.e.*, chlorophyll concentration (CHL) and remote sensing reflectance at 443 nm (Rrs(443)), 488 nm (Rrs(488)), and 555 nm (Rrs(555)), employed in this study were derived from MODIS/Aqua with a spatial resolution of 4 km and associated with processing version 2018.0. The adapted monthly SST dataset was the AVHRR_OI (optimal interpolation) dataset provided by the Group for High Resolution Sea Surface Temperature (GHRSST), NOAA, with a 0.25-degree resolution and processing version 2.1. The sea surface salinity (SSS) was obtained from the *GLOBAL-REANALYSIS-PHY-001-030* product of the Copernicus Marine Service (CMEMS) with a spatial resolution of 0.083°. The monthly sea level pressure (SLP) and atmospheric CO_2 presented in the mole fraction of CO_2 in the dry air (xCO_2) were obtained from NOAA's CarbonTracker, version CT2019B (Jacobson et al., 2020), with a spatial grid of 3° × 2°. The wind speed at 10 m (WS) above the sea surface was derived from the ERA5 dataset provided by the European Centre for Medium-Range Weather Forecasts (ECMWF) with a special resolution of 25 km and a temporal resolution of a month and 6 h.

2.3. Data gridding, match-up, and sub-dataset classification

To match the underway pCO_2 measurements with satellite and model data and to ensure a sufficient number of match-ups to support the machine learning algorithm, all the data were first gridded to a time window of one month and a spatial window of 1 km. The principle of *in situ* data gridding is that each *in situ* sample can only belong to one grid. For a grid containing more than three samples, abnormal values were identified by the 3σ principle and eliminated. The averaged value of valid sample data in each grid was considered the value of the grid. The satellite and model data were also resampled to 1 km resolution before the match-up.

The match-up processing between the gridded *in situ* data and satellite data involved the following process: First, the nearest latitude and longitude identified the central pixel co-located with each *in situ* data; to avoid the impact of the noise in the satellite products on the subsequent algorithm development, only those pixels with at least ten valid pixels in the surrounding 5 pixels×5 pixels box and that satisfied the homogeneity criteria of a coefficient of variation <0.15 were considered valid; Finally, valid central pixels and corresponding gridded *in situ* data were matched up to compile the database for retrieving seawater pCO_2.

In total, we obtained a match-up database with a volume of 63,883 groups, which was further divided into training and validation datasets. In contrast to traditional random data classification, we divided the data by cruises to ensure their independence of temporal/spatial representation. To ensure that both two datasets (especially the validation dataset) have good spatiotemporal representativeness, at least one cruise for each region is included for each season. Fig. 1(d) and (e) show the spatial and histogram statistics of the training and validation datasets in the four seasons. We obtained full coverage in the ECS and west YS over the four seasons in both the training and validation datasets. The BS cruises were mainly conducted in summer and fall, without winter

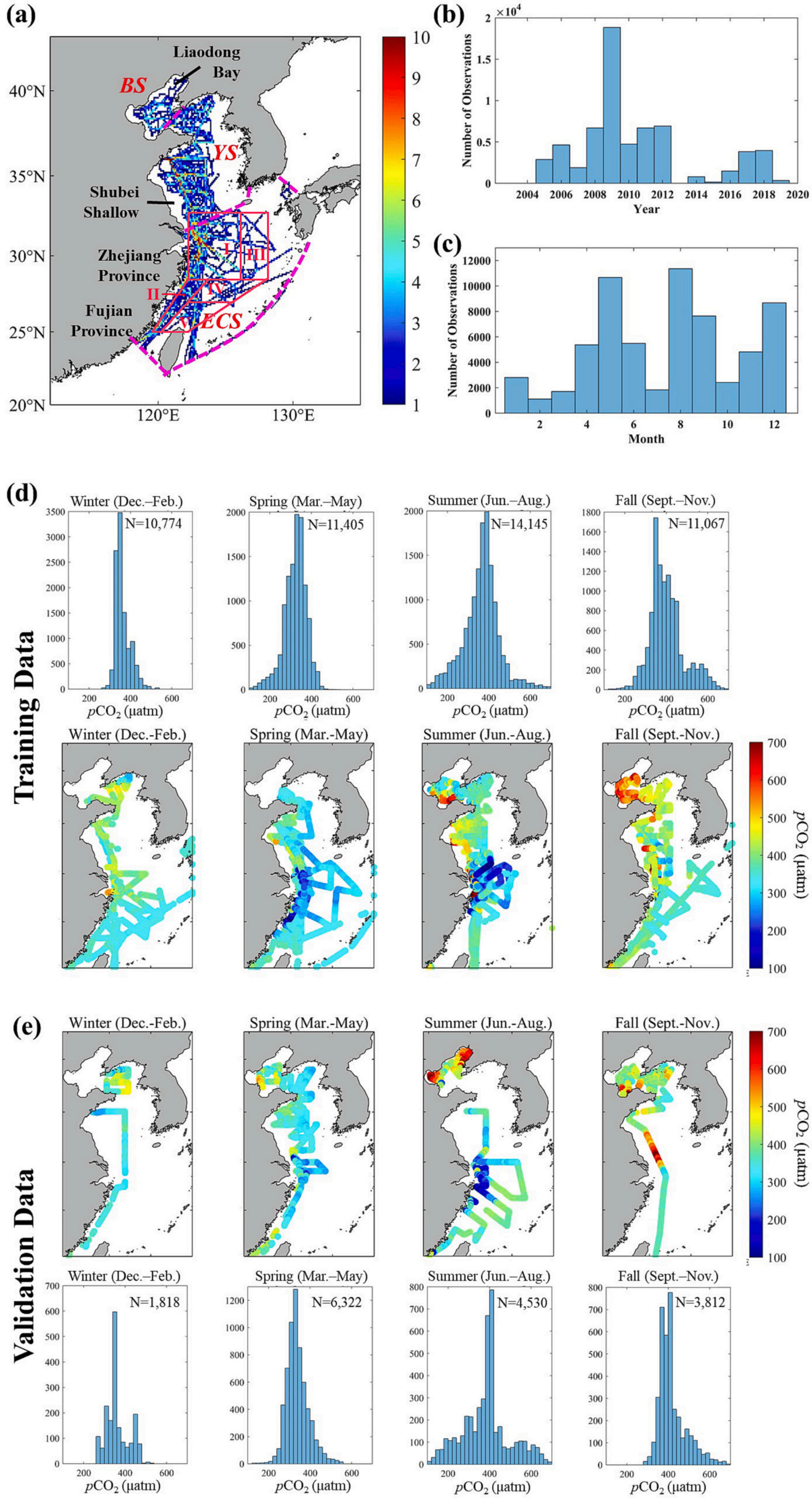


Fig. 1. (a) The number of months with underway seawater pCO_2 observations in the Bohai, Yellow, and East China Seas in 2003–2019. (b) and (c) Histogram statistics of underway seawater pCO_2 observations in years and months. (d) and (e) Spatial and histogram statistics of the training and validation datasets in the four seasons. The boundaries of the Bohai Sea (BS), Yellow Sea (YS), and East China Sea (ECS) are indicated by pink dashed lines and the enclosed coastal lines in panel a. Areas framed with pink solid lines in panel (a) indicate the five domains (Table 1) categorized in Guo et al. (2015).

cruises. Only one spring cruise was conducted and included in the validation dataset to ensure the retrieving predictability. Then, the validation dataset should have ~25 %–30 % records of total data volume to provide enough statistics for model performance. The final validation set accounted for 25.82 % of the total data volume, and a proportion of 15–35 % was also maintained in all four seasons.

2.4. Performance evaluation

The performance evaluation of the satellite-derived pCO$_2$ algorithm and satellite products was based on three statistical measures, including the coefficient of determination (R^2), root mean square error (RMSE), and mean absolute percentage deviation (APD), as follows:

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

$$RMSE = \sqrt{\frac{\sum_{i=1}^{N}(X_i - Y_i)^2}{N}} \tag{3}$$

$$APD = \frac{1}{N}\sum_{i=1}^{N}\left|\frac{X_i - Y_i}{Y_i}\right| \times 100\% \tag{4}$$

where X_i, Y_i, and N are the algorithm-retrieved values, *in situ* measurements, and sample number, respectively. $\overline{X}$ and $\overline{Y}$ are the average of all algorithm-retrieved values and *in situ* measurements, respectively. σ_X and σ_Y are the standard deviations of X_i and Y_i, respectively.

2.5. Air-sea CO_2 flux calculation

After the seawater pCO$_2$ field was generated, the air-sea CO_2 flux between the surface water and the atmosphere can be calculated from:

$$FCO_2 = k \times K_H^{CO2} \times \Delta pCO_2 \tag{5}$$

where k is the gas transfer velocity, K_H^{CO2} is the solubility of CO_2 gas in seawater (Weiss, 1974), and ΔpCO$_2$ is the difference between atmospheric and surface seawater pCO$_2$. k was parameterised using the empirical function of Sweeney et al. (2007), and nonlinear correction of gas transfer velocity with wind speed was adopted following Wanninkhof et al. (2002) and Jiang et al. (2008):

$$k = 0.27 \times C_2 \times U_{mean}^2 \times \left(\frac{Sc}{660}\right)^{-0.5} \tag{6}$$

$$C_2 = \frac{1}{n} \times \frac{\sum_{j=1}^{n} U_j^2}{U_{mean}^2} \tag{7}$$

where U_{mean} is the monthly mean wind speed at 10 m above sea level, and Sc is the Schmidt number at *in situ* temperatures for surface seawater (Wanninkhof, 1992). C_2 is the nonlinear coefficient for the quadratic term of the gas transfer relationship, U_j is the 6-hourly wind speed, the "mean" subscript indicates the average values, and n is the number of available wind speed measurements for the month.

The atmospheric pCO$_2$ ($p{CO_2}^{Air}$) was calculated using xCO$_2$ in the air, the barometric pressure (SLP), and the vapour pressure of water at 100 % relative humidity (pH_2O^{Air}) using the following formula (Weiss and Price, 1980):

$$p{CO_2}^{Air} = xCO_2 \times \left(SLP - pH_2O^{Air}\right) \tag{8}$$

To quantify how much carbon was taken up/released in a particular region, we calculated the integrated CO_2 flux (in Tg C/yr) by multiplying the mean CO_2 flux density among the available pixels by the total area of the region. For the air-sea CO_2 fluxes, positive values mean CO_2 release to the atmosphere, whereas negative values mean CO_2 uptake by the sea.

3. pCO$_2$ algorithm development and validation

3.1. Parameterisation based on mechanism analysis

We collected and collated massive amounts of matched *in situ* pCO$_2$ and remote sensing data. Considering the ocean characteristics of the study area and the complexity of pCO$_2$ changes, we built a pCO$_2$ retrieval algorithm using a data-driven machine learning (ML) method and considered more mechanistic variation in pCO$_2$ in the selection of input parameters. As described by Bai et al. (2015), seawater pCO$_2$ variation (ΔpCO$_2$) can be expressed as follows:

$$\Delta pCO_2 = \left(\frac{\partial pCO_{2@flux}}{\partial V_{flux}}\right) \times \Delta V_{flux} + \left(\frac{\partial pCO_{2@therm}}{\partial V_{therm}}\right) \times \Delta V_{therm} + \left(\frac{\partial pCO_{2@mix}}{\partial V_{mix}}\right) \times \Delta V_{mix} + \left(\frac{\partial pCO_{2@bio}}{\partial V_{bio}}\right) \times \Delta V_{bio} + \ldots + \left(\frac{\partial pCO_{2@factor-n}}{\partial V_{factor-n}}\right) \times \Delta V_{factor-n} + \varepsilon \tag{9}$$

where ΔpCO$_2$ is analytically expressed as the sum of individual pCO$_2$ components associated with each process or controlling factor $\left(\frac{\partial pCO_{2@factor-n}}{\partial V_{factor-n}}\right) \times \Delta V_{factor-n}$, in which ($V_{factor-n}$) is the independent variable characterising that process. ε is the residual error of the total ΔpCO$_2$. Based on first principles and our knowledge of ocean pCO$_2$ research, in our study area, we analysed the factors controlling pCO$_2$ variations and identified the input parameters as $\left(\frac{\partial pCO_{2@flux}}{\partial V_{flux}}\right) \times \Delta V_{flux}$ —air-sea CO_2 exchange/atmospheric pCO$_2$ forcing, $\left(\frac{\partial pCO_{2.therm}}{\partial V_{therm}}\right) \times \Delta V_{therm}$—temperature-dependent thermodynamic term, $\left(\frac{\partial pCO_{2@mix}}{\partial V_{mix}}\right) \times \Delta V_{mix}$—mixing between various water masses with different carbonate components, and $\left(\frac{\partial pCO_{2@bio}}{\partial V_{bio}}\right) \times \Delta V_{bio}$—biological effect.

1) the air-sea CO_2 exchange/atmospheric pCO$_2$ forcing $\left(\frac{\partial pCO_{2@flux}}{\partial V_{flux}}\right) \times \Delta V_{flux}$

Many early regional studies were limited by the number of cruises and their brief time spans and did not consider the impact of air-sea CO_2 exchange on seawater pCO$_2$. However, since this study aimed to build a decadal-scale (17 years) seawater pCO$_2$ product, during which atmospheric CO_2 has increased significantly (~34 μatm), the forcing effect of this rapid CO_2 increase on seawater pCO$_2$ was non-negligible. Previous studies have usually used the atmospheric pCO$_2$ measured at nearby monitoring stations or xCO$_2$ products as the inputs of their pCO$_2$ retrieval models (Gloege et al., 2022; Landschützer et al., 2014; Roobaert et al., 2019; Wang et al., 2022). However, as the BS, the YS, and the ECS are closed/semi-closed marginal seas that are greatly affected by terrestrial effects (*i.e.*, the atmospheric pCO$_2$ on the land side (east and north) is significantly higher than that on the side close to the Pacific Ocean), the nearby monitoring stations alone cannot reflect the spatial variability of atmospheric pCO$_2$. In addition, the BS, the YS, and the ECS have various complex subsystems, and the seawater pCO$_2$ has high temporal and spatial variability. However, the current prevailing global grid xCO$_2$ product provided by NOAA Carbontracker has a spatial resolution of 3° × 2°, which is insufficient compared to the spatial scale of the study area. If we directly use this xCO$_2$ product as an input, model instability will occur, resulting in artificial discontinuities in the spatial distribution of the seawater pCO$_2$ product.

2) Temperature-dependent thermodynamic term $\left(\frac{\partial pCO_{2.therm}}{\partial V_{therm}}\right) \times \Delta V_{therm}$

Bai et al. (2015) proposed a temperature-dependent thermal dynamic pCO_2 and used it in the satellite retrieval algorithm by combining the SST-dependent thermal dynamic pCO_2 with mixing effect. In this study, we define a new thermodynamic term, $pCO_{2,therm}$, simultaneously characterises the thermodynamic effect of temperature and the forcing effect of the long-term atmospheric CO_2 increase on seawater. The $pCO_{2,therm}$ term is the theoretical pCO_2 value controlled only by the temperature-dependent thermodynamic effect, assuming that seawater CO_2 can reach equilibrium with atmospheric CO_2 within one year through the air-sea exchange. This term is calculated as follows (Takahashi et al., 1993):

$$\mathrm{pCO}_{2,\mathrm{therm}} = \mathrm{pCO}_{2(\mathrm{o})} \times \mathrm{e}^{0.0423\left(\mathrm{SST}-\mathrm{SST}_{(\mathrm{o})}\right)} \tag{10}$$

where $pCO_{2(o)}$ and $SST_{(o)}$ are the yearly mean atmospheric pCO_2 (pCO_{2air}) and SST in each pixel, respectively; note that $pCO_{2,air}$ has been gridded to 1 km in its calculation from xCO2 (see Section 2.3). Fig. 2 (a~h) show the seasonal distribution of the thermal dynamic seawater pCO_2 in 2003 and 2019. Under the effect of SST, $pCO_{2,therm}$ has an evident seasonal variation pattern; that is, the $pCO_{2,therm}$ value in summer is much higher than that in winter. For the interannual variation, the $pCO_{2,therm}$ in 2019 is generally higher than that in 2003, indicating that $pCO_{2,therm}$ can reflect the forcing of the increasing atmospheric CO_2 concentration on the seawater pCO_2. Moreover, $pCO_{2,therm}$ also illustrates the spatial gradient among the BS, the YS, the ECS shelf, and the outer sea of the ECS, showing its spatial resolution advantage compared to $x CO_2$ data.

3) Mixing between various water masses with different carbonate components $\left(\frac{\partial pCO_{2@mix}}{\partial V_{mix}}\right) \times \Delta V_{mix}$

The dynamic processes in this study area are complicated, including the mixing of riverine runoff with seawater, Kuroshio intrusion, seasonal coastal currents, and coastal upwelling. Bai et al. (2015) quantified the impact of mixing processes on seawater pCO_2 by ocean colour-retrieved sea surface salinity (SSS) in summer. However, the existing microwave satellite SSS data (available from 2010) do not have a sufficient time span, and the accuracy of model-derived SSS data in the BS, YS, and ECS is strongly affected by the highly turbid coastal water. Therefore, we employed an upwelling index (UI_{SST}) to approximately reflect the impact of most mixing processes on seawater pCO_2. Alvarez et al. (2011)

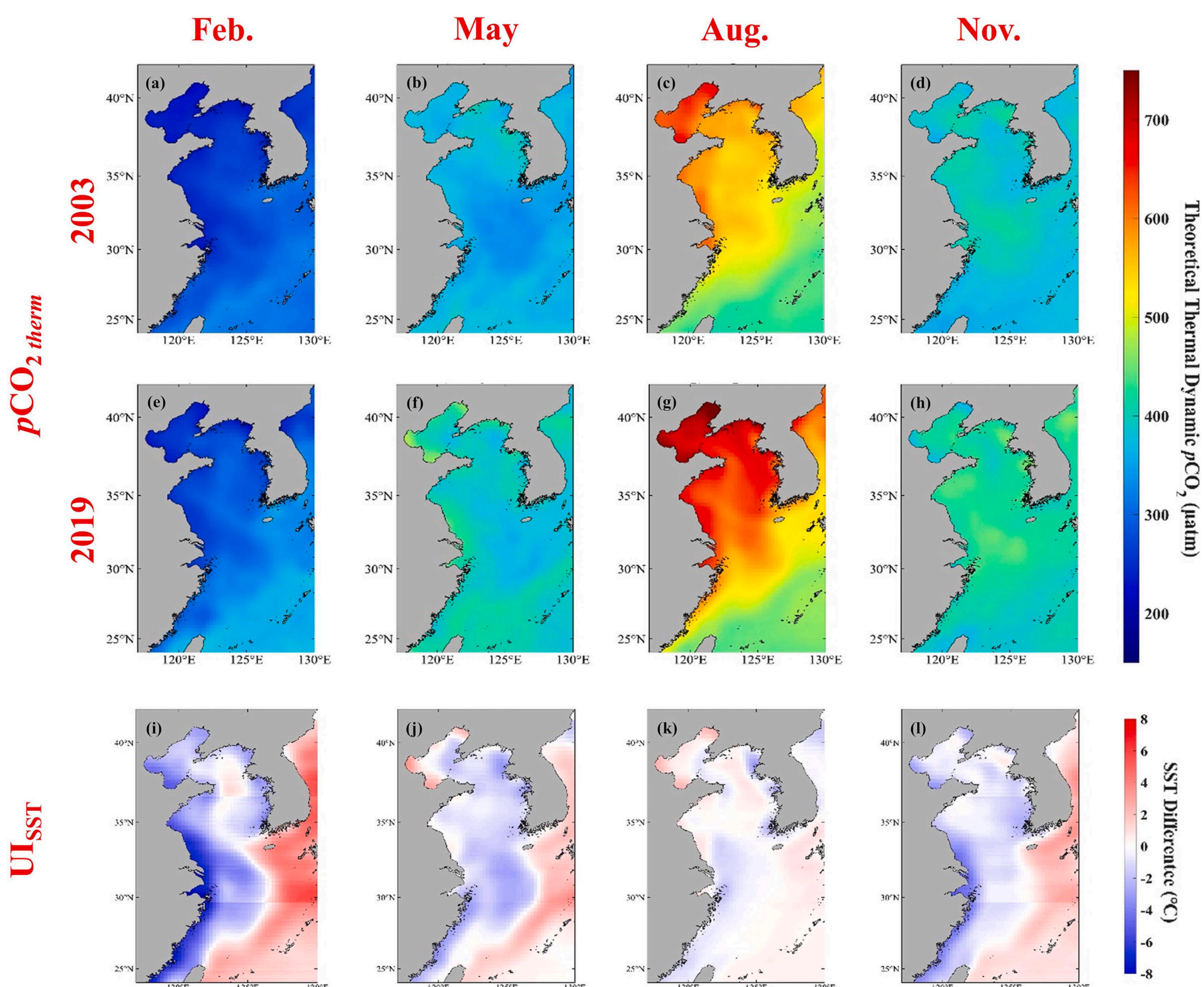
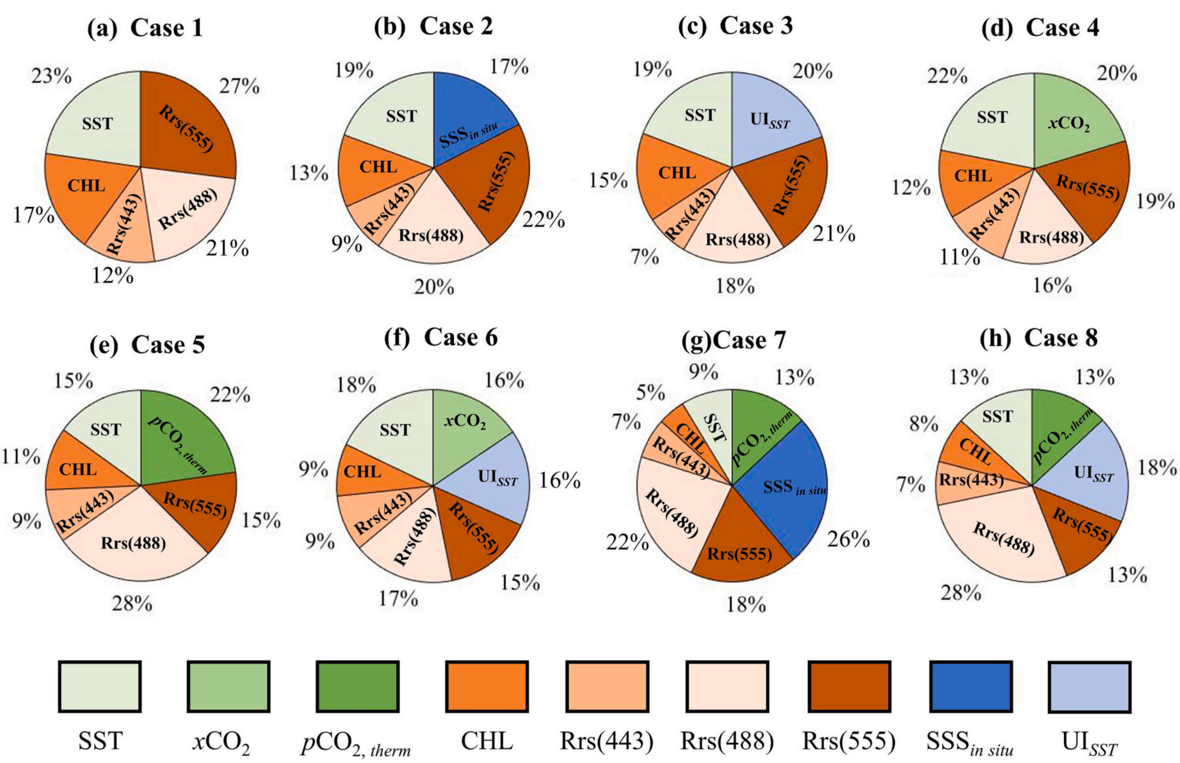


Fig. 2. Seasonal variation in mechanical parameters, with monthly thermal dynamic seawater pCO_2 for (a) February 2003, (b) May 2003, (c) August 2003, (d) November 2003, (e) February 2019, (f) May 2019, (g) August 2019, and (h) November 2019. (i-l) Monthly mean UI_{SST} over multiple years (2003–2019) for (i) February, (j) May, (k) August, and (l) November.

used UI_{SST} to describe the strength of coastal upwelling, which was calculated as the SST difference between coastal and oceanic points at the same latitude. Here, we calculated the UI_{SST} of each pixel as the SST at that point minus the mean SST at the same latitude in area 117–135°E.

Fig. 2 (i~l) show the seasonal distribution of UI_{SST}. In the winter, represented by February, due to the low temperature on the continents, the SST in the western part of the study area (BS, YS, and ECS shelf) is far lower than that on the eastern side (close to the Pacific Ocean), and the SST near-shore is lower than that offshore; thus, the coastal feature can be easily distinguished by a negative UI_{SST} value. Additionally, UI_{SST} can describe the shape of the Yellow Sea Warm Current (YSWC). In spring (May) and summer (August), the SST difference between the continental shelf and the ocean continues to decrease, and a positive UI_{SST} value appears in the semi-enclosed bays of the BS under the terrestrial influence. In the fall (November), the UI_{SST} distribution shows that



Cases	Input factors	R^2	RMSE	APD
(a) Case 1	SST, CHL, Rrs(443), Rrs(488), Rrs(555)	0.86	27	6.34%
(b) Case 2	SST, CHL, Rrs(443), Rrs(488), Rrs(555), $SSS_{in\ situ}$	0.91	23	5.02%
(c) Case 3	SST, CHL, Rrs(443), Rrs(488), Rrs(555), UI_{SST}	0.90	24	5.26%
(d) Case 4	SST, xCO_2, CHL, Rrs(443), Rrs(488), Rrs(555)	0.93	22	4.53%
(e) Case 5	SST, $pCO_{2,\ therm}$, CHL, Rrs(443), Rrs(488), Rrs(555)	0.93	22	4.45%
(f) Case 6	SST, xCO_2, CHL, Rrs(443), Rrs(488), Rrs(555), UI_{SST}	0.94	20	4.29%
(g) Case 7	SST, $pCO_{2,\ therm}$, CHL, Rrs(443), Rrs(488), Rrs(555), $SSS_{in\ situ}$	0.94	19	4.08%
(h) Case 8	SST, $pCO_{2,\ therm}$, CHL, Rrs(443), Rrs(488), Rrs(555), UI_{SST}	0.94	19	4.12%

Fig. 3. Percentage pie chart for 8 groups of inputs in cases 1 to 8 (a~h) in the XGBoost model simulations. The pie charts show 8 cases of input combination and their relative feature importance (RFI). In the table, the numbers show the model performance (derived from validation dataset) in each case. Colours in the pie chart and table indicate parameters that denote different controlling effects, with green presenting thermodynamic and atmospheric forcing, orange as the biological effect, and blue as the mixing effect.

transitional features from summer to winter are similar to those in spring, but the UI_{SST} on the BS coast switches to a negative value. Spatially, the UI_{SST} pattern is identical to that of salinity, both of which show lower values on the shelf than in the deep ECS and open ocean. UI_{SST} can also reflect the spatial distribution of fresh water in the BS in spring and summer, as well as the characteristics of the coastal current, coastal upwelling, and some water masses in winter. Therefore, UI_{SST} is considered to be a good input parameter to characterise the moderating effect of the water mass.

4) Biological effect $\left(\frac{\partial pCO_{2@bio}}{\partial V_{bio}}\right) \times \Delta V_{bio}$

Chlorophyll-a concentration (CHL), a primary ocean satellite product, is widely used in seawater $p\mathrm{CO}_2$ remote sensing retrieval to represent the biological effect. In addition to CHL, we used remote sensing reflections at three bands, 443 nm, 488 nm, and 555 nm (Rrs(443), Rrs (488), and Rrs(555), respectively) in our seawater $p\mathrm{CO}_2$ model to describe more ecological characteristics of the seawater. Specifically, Rrs(443) and Rrs(488) reflect the chlorophyll-a and chromophoric dissolved organic matter (CDOM) signals in the seawater, and Rrs(555) reflects the presence of suspended matter (from both phytoplankton and terrain). Although the CHL we adapted is calculated from remote sensing reflection (Rrs) at these bands by the standard OC3M band ratio algorithm merged with the colour index (CI) of Hu et al. (2012), the input of single Rrs can distinguish different biological domains and processes with the same chlorophyll concentration, such as seasonal high productivities in the eutrophic coastal zones and algal blooms caused by mesoscale phenomena in the sea basins. In addition to biological effects, bands in ocean colour satellite images can express the influence of offshore transport and water masses with different biochemical characteristics. There are other Rrs bands, and we believe that the more bands of input parameters there are, the better the effect will be; however, considering that the characteristics and data quality of the different bands may introduce propagation errors and uncertainties (Hu et al., 2013), we only input the Rrs at three bands (443 nm, 488 nm, and 555 nm).

3.2. Input parameter selection

The above input parameters were preliminarily selected based on the mechanism of $p\mathrm{CO}_2$ variation in our study area, namely, SST, CHL, Rrs (443), Rrs(488), Rrs(555), UI_{SST}, and $pCO_{2,therm}$, and the output was $p\mathrm{CO}_2$. As discussed in the section above, the satellite-derived SSS and modeled $x\mathrm{CO}_2$ are unsuitable inputs. Here, we still used them as inputs to test our selection, as well as the underway SSS data ($\mathrm{SSS}_{in\ situ}$) that accompanied the underway $p\mathrm{CO}_2$ data, which had the best accuracy. The above nine parameters were divided into eight combinations as eight sets of inputs and tested by the output performance of the machine learning model (Fig. 3). The statistic values about the model performance are derived from validation dataset in each case. We also tested several machine learning models (see Section 3.4). For the sake of brevity, we only present the results of the best model, the XGBoost model.

We can compare and analyse the pie chart of the relative feature importance (RFI) of cases 1 to 8 in Fig. 3. Colours were used to denote the group of parameters of mechanistic features. SST, $x\mathrm{CO}_2$, and $pCO_{2,therm}$ are shown in green in the pie chart and reflect the thermodynamic characteristics and atmospheric $p\mathrm{CO}_2$ forcing in MeSAA-ML-ECS. The biological effects of seawater are represented by CHL, Rrs(443), Rrs(488), and Rrs(555) in orange. The blue colours represent the mixing effects between different water masses with different carbonate components from UI_{SST} and $\mathrm{SSS}_{in\ situ}$.

Fig. 3(a) shows the basic parameters as inputs, such as, SST, CHL, and Rrs (Case 1). Comparing Case 2–8 and Case 1, inducing more control mechanism inputs can improve the model performance. In the aspect of mixing effects (Case 2 and 3), the RFIs of UI_{SST} (light blue, ~20 %, Case 2) and $\mathrm{SSS}_{in\ situ}$ (dark blue, ~17 %, Case 3) are almost equal, indicating that UI_{SST} can replace $\mathrm{SSS}_{in\ situ}$ in the XGBoost model, as shown in Fig. 3 (b) and (c). In the control mechanism of the thermodynamic characteristics and atmospheric $p\mathrm{CO}_2$ forcing (Case 4 and 5), the RFI of $pCO_{2,therm}$ (dark green, ~22 %, Case 5) is larger than that of $x\mathrm{CO}_2$ (green, ~20 %, Case 4), as compared in Fig. 3(d) and (e), indicating that $pCO_{2,therm}$ can replace $x\mathrm{CO}_2$. When $x\mathrm{CO}_2$ and $pCO_{2,therm}$ were added to the SST, CHL, Rrs(443), Rrs(488), and Rrs(555), the model performance with $pCO_{2,therm}$ (RMSE = 22 μatm and APD = 4.45 % for the validation dataset) as an input was better than that with $x\mathrm{CO}_2$ (RMSE = 22 μatm and APD = 4.53 % for the validation dataset) as an input. The RFIs of Cases 6 to 8 in Fig. 3(f~h) further demonstrate the rationality of substituting the input parameter $pCO_{2,therm}$ in this part.

The results of Fig. 3 show that the parameters of UI_{SST} and $pCO_{2,therm}$ have good performance in our study area, and thus, we selected SST, CHL, Rrs(443), Rrs(488), Rrs(555), UI_{SST}, and $pCO_{2,therm}$ as the input parameters (Case 8).

Regarding the RFI of Cases 6–8, we can understand that in our $p\mathrm{CO}_2$ model, the mixing effect contributes approximately 16–28 %, the temperature-dependent thermodynamic effect contributes approximately 21–34 %, and the biological effect combined with the other ocean colour effect contributes approximately 50–56 %, which is rational in such high productivity marginal seas.

3.3. Machine learning model selection

Recently, various machine learning models have been widely applied in geological research mainly due to their superior ability to capture the characteristics of geological variables and solve complicated non-linear problems among them (Carrión et al., 2021; Tiyasha et al., 2021; Xia et al., 2022; Zhang et al., 2022). In the published literature, many machine learning technologies have been successfully applied to develop satellite-based surface $p\mathrm{CO}_2$ models, such as the multilayer perceptron neural network (MLPNN), random forest (RF), and light gradient boosting machine (LightGBM) models (Chen et al., 2019; Fu et al., 2020; Gloege et al., 2022; Joshi et al., 2022; Wang et al., 2022). In our study, we will test the mainstream ML models.

Chen and Guestrin (2016) proposed an optimised distributed gradient-boosting machine learning model, XGBoost, based on a gradient-boosting decision tree (GBDT). XGBoost is a gradient-boosting algorithm considering more gradient-boosting information of training datasets with high efficiency and flexibility (Carrión et al., 2021; Qiu et al., 2021; Xia et al., 2022; Zhang et al., 2022) and adds a regularisation term to the loss function to prevent overfitting. A second-order Taylor expansion of the loss function is conducted to approximate this objective function in the XGBoost model. Detailed optimisation processes of the objective function were described in Chen and Guestrin (2016). This study used the publicly available Scikit-Learn and XGBoost packages to complete the $p\mathrm{CO}_2$ retrieval model development process. Referring to previous similar studies and our multiple tests, the parameter setting scheme in this XGBoost model is as follows: the number of regression trees is 1200, the maximum depth of trees is 6, the learning rate is 0.05, and the weight of the L2 regularisation term is 0.01 (Ramraj et al., 2016).

RF is also a commonly used ML method. Both binary decision trees in RF and XGBoost models, including classification and regression trees, are usually selected as basic learning models. RF adopts the “bagging” method to evenly extract training samples, while XGBoost adopts the “boosting” method and samples according to the error rate (Chen and Guestrin, 2016). LightGBM is also a fast, well-distributed, and high-performance machine learning framework, but unlike XGBoost, LightGBM employs a “histogram” algorithm to speed up the training process by segmenting continuous feature values into discrete bins (Dorogush et al., 2018). Previous studies have shown that LightGBM is sensitive to dataset noise due to its bias-based algorithm, which might lead to bad training results (Peng et al., 2022). XGBoost can more precisely find data separation points and reduce the influence of extreme values on model stability by its "pre-sorted feature" algorithm (Smirnov et al., 2020). XGBoost has been proven to be more accurate than RF and MLPNN in recent algorithm competitions (Carrión et al., 2021; Qiu et al., 2021; Tiyasha et al., 2021; Xia et al., 2022; Zhang et al., 2022).

We test the MLPNN, DT (decision tree), RF, LightGBM, and XGBoost with the same model inputs of SST, CHL, Rrs(443), Rrs(488), Rrs(555), UI_{SST}, and $pCO_{2,therm}$ based on the same training and validating datasets. In particular, the typical configuration parameters (*i.e.*, the number of regression trees and a maximum depth of each tree) of the DT, RF, LightGBM, and XGBoost are all basically consistent in this test. The MLPNN model is a feed-forward neural network based on the Levenberg–Marquardt back-propagation algorithm (Bishop, 1995; Gross et al., 1999). The MLPNN comprises 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, respectively, with a tan-sigmoid function as the transfer function. Table S3 shows the results of each approach. The XGBoost model showed the best performance for the validation dataset, with RMSE = 20 μatm, R^2 = 0.94, and APD = 4.12 %. The LightGBM model also showed relatively good (but slightly worse than XGBoost) performance (RMSD = 21 μatm, R^2 = 0.93, and APD = 4.33 %), followed by RF (RMSD = 31 μatm, R^2 = 0.82, and APD = 7.37 %). The performances of MLPNN and DT were worse than that of the other three machine learning approaches.

3.4. Independent validation of satellite products

Rather than using *in situ* data to develop the algorithm, we use match-up satellite data and grid data as model inputs, which are also used for satellite product generation; thus, the validation of the model is equivalent to the final satellite product validation. As shown in Fig. 1(d) and (e), the training and independent validation datasets have full spatial coverage in the ECS and west YS over the four seasons, and fewer data are available in the BS; the *in situ* dataset is classified cruise by cruise to ensure independency. Fig. 4(a) and (b) show the scatter density plot comparison between the *in situ* pCO_2 values and the pCO_2 values retrieved by the XGBoost model for the training and independent validation datasets. In the training dataset, the retrieved pCO_2 values were consistent with the *in situ* values, with an R^2 of up to 0.96, a RMSE of 14 μatm, and an APD of <3.11 %. Fig. 4(b) shows that the R^2, RMSE, and APD of the independent validation dataset were 0.94, 20 μatm, and 4.12 %, respectively, which are slightly lower than those of the training dataset but still within a reasonable error range. In addition, the differences in statistical results (R^2, RMSE, and APD) between the training and independent validation datasets are relatively small, indicating that the established XGBoost model can learn the training dataset accurately and has strong stability and robustness.

Fig. 4(c~j) display the *in situ* and retrieved pCO_2 in August 2009 and May 2012. The *in situ* pCO_2 of these two cruises are from the independent validation dataset. Comparisons for the rest of the cruises/legs can be found in Figs. S1~S41 in the Supporting information. The high consistency of pCO_2 values indicates that our pCO_2 retrieval model has good accuracy, high applicability, and stability in the whole region. Moreover, MeSAA-ML-ECS can retrieve pCO_2 in high-heterogeneity water bodies. The spatial variations in the satellite pCO_2 match perfectly with those of the *in situ* pCO_2 (Fig. 4c), featuring evident ascending gradients (from ~200 to ~300 μatm) from the Changjiang estuary to the ECS shelf (Fig. 4e). The gradient change (from ~450 to ~350 μatm) in the seawater pCO_2 from the Yellow River estuary to the BS (Fig. 4e) is also well reconstructed by the satellite data (Fig. 4f and g). Compared with the MeSAA algorithm (Bai et al., 2015), the current MeSAA-ML-ECS can very well describe the gradient changes in the seawater pCO_2 in the Changjiang estuary-ECS system. The current MeSAA-ML-ECS algorithm can better characterise more seawater pCO_2 details and produces good results in some regions where other algorithms cannot characterise the whole BS, YS, and ECS well.

4. Results and discussion

4.1. Seasonal variation in seawater pCO_2

The monthly mean seawater pCO_2 during 2003–2019 in the entire study area is shown in Fig. 5. Seawater pCO_2 in the study area ranges from 270 to over 600 μatm and has a pattern of high values in summer and fall and low values in winter in most of the area. In terms of the climatological mean pCO_2 value, the BS has the highest value (398 μatm), followed by the YS (364 μatm), and the ECS has the lowest value (347 μatm).

The BS, an enclosed marginal sea with shallow bathymetry, has a large seasonal pCO_2 variation with a range of ~250 μatm. In the winter and early spring (March and April), CO_2 in almost the entire area of the BS is undersaturated with pCO_2 values of <350 μatm associated with low SST. After that, the pCO_2 begins to rise in the bays and alongshore until September, accompanied by the increase of SST, reaching a value of 500–700 μatm resulting from terrestrial runoff with high pCO_2 as well as net community respiration induced by frequent algal blooms in spring and summer (Song et al., 2016a; Zhai et al., 2019; Zheng et al., 2021).

The YS is characterized by spatial heterogeneity with complicated seasonal variations. In the summer and early fall months, seawater CO_2 in the central YS is slightly oversaturated with a pCO_2 value of 400–450 μatm. In comparison, highly oversaturated CO_2 with pCO_2 values that reach over 500 μatm are found at the Subei Shallow. In October, with the decreased SST and possible autumn bloom, the pCO_2 values in the central area become undersaturated, while those in the coastal zones completely switch in December, with pCO_2 values continuing to decrease until they become lower than those in the central area. Notably, in the Yellow Sea Warm Current-influenced area in the north YS (Xue et al., 2012), a high pCO_2 (~450 μatm) water mass is observed in January, associated with relatively high SST values of ~7 °C. The extreme turbidity in the Subei Shallow may lead to the low quality of ocean colour data, resulting in uncertainty of the seawater pCO_2 retrieval, and meanwhile, the *in situ* cruises rarely cover this area. Thus, the extremely high pCO_2 in the Subei Shallow during summertime needs to be further verified by more observations.

The ECS is a large-river-dominated marginal sea in which the seasonal variation in pCO_2 is relatively small, within 150 μatm, except for coasts along Zhejiang and Fujian Provinces. The seasonal patterns of pCO_2 are significantly different on the outside and inside of the continental shelf. The pCO_2 outside the continental shelf is relatively homogeneous and basically controlled by SST, presenting a seasonal pattern of high values in summer and fall and low values in winter and spring. In comparison, the pCO_2 on the ECS shelf exhibits substantial spatial variations (Fig. 5). In winter and early spring, the pCO_2 on the ECS shelf is generally undersaturated, with pCO_2 of ~300 μatm on the outer shelf and ~380 μatm in the coastal zones and outer Changjiang estuary where coastal currents may carry terrestrial waters rich in CO_2 southwards. In late spring and summer, pCO_2 in the coastal zones and Changjiang estuary increased rapidly, reaching over 700 μatm in August. In the meantime, the pCO_2 on the outer shelf also increased to 400 μatm in August. In comparison, that in the outer Changjiang estuary, where dominated by the river plume, significantly decreased to a minimum value of under 270 μatm in June and August. In the fall, water with higher pCO_2 levels spreads offshore, and then the pCO_2 on the whole ECS shelf becomes homogeneous in November and December, with a value of ~375 μatm.

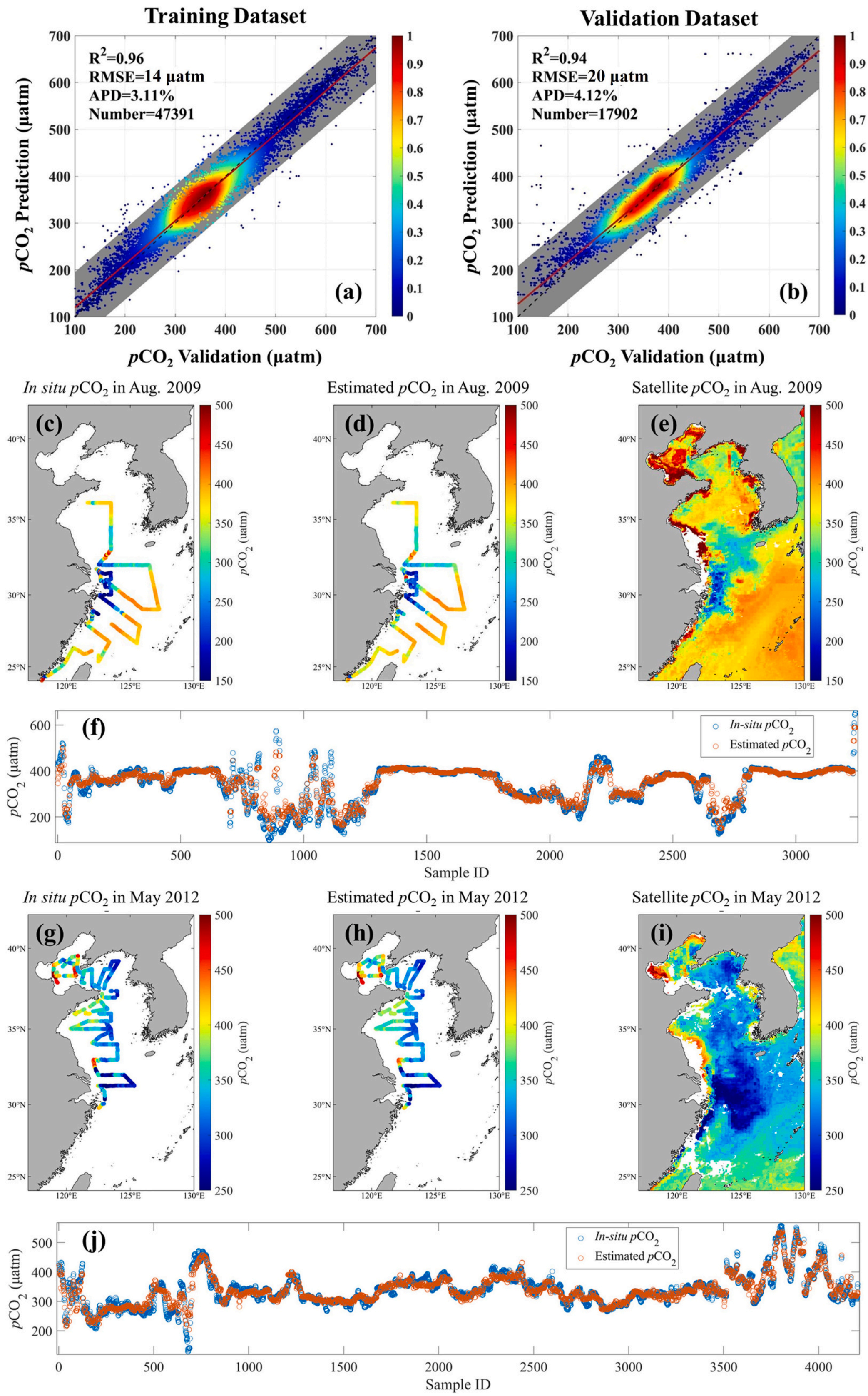


Fig. 4. Satellite results and validation. Scatter density plot comparison between the *in situ* pCO_2 values and the pCO_2 values retrieved by the XGBoost model for the (a) training and (b) validation datasets. (c) and (g) Distributions of underway *in situ* pCO_2 in August 2009 and May 2012, respectively, with the corresponding satellite-derived pCO_2 along cruise in (d) and (h). (e) and (i) Satellite-derived pCO_2 in August 2009 and May 2012, respectively. (f) and (j) Comparisons between the satellite results and the *in situ* underway measurements along the two cruise track. These results are from two samples, and the comparisons for the rest of the cruises/legs can be found in Figs. S1~S41 in the Supporting information.

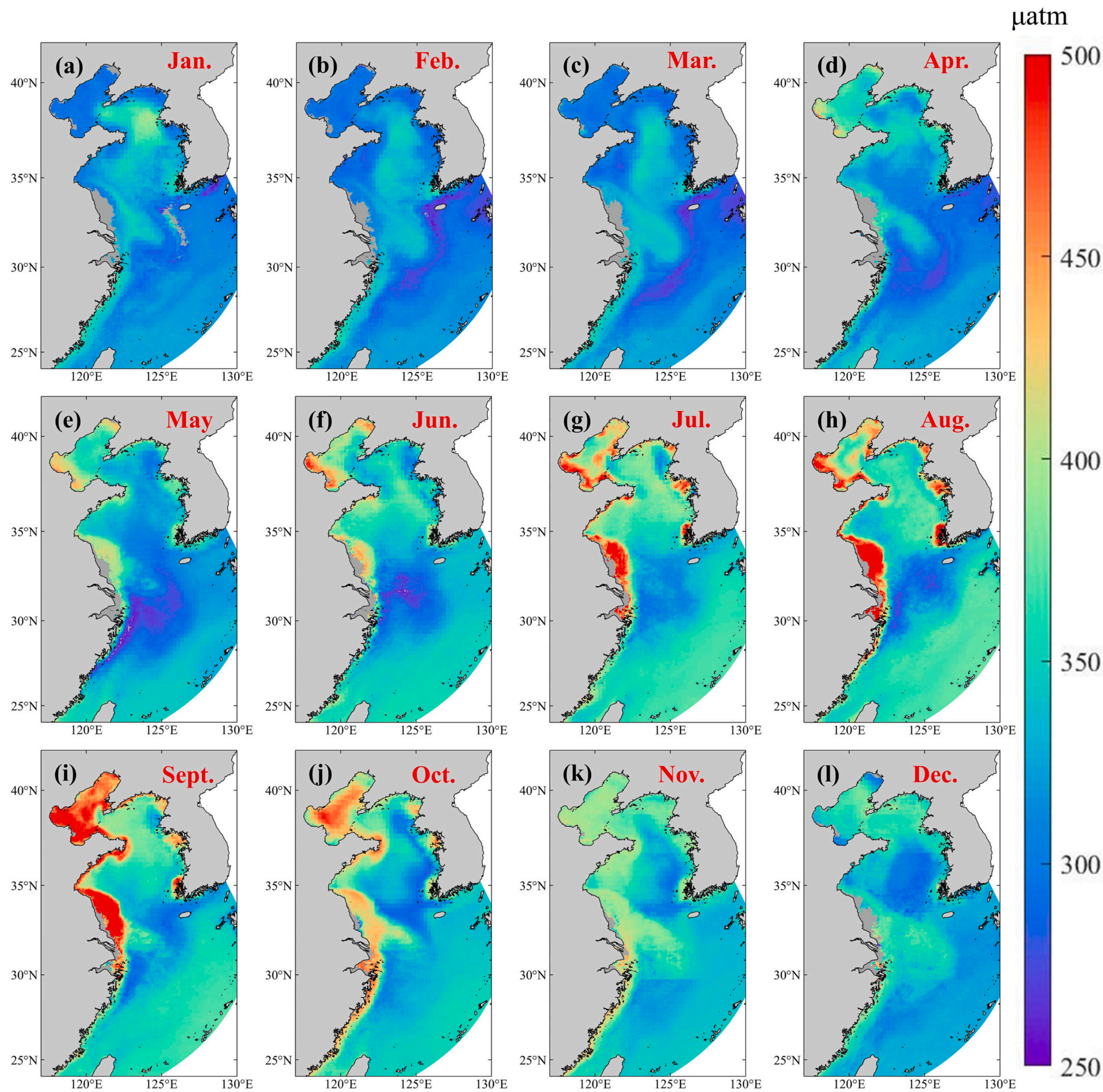


Fig. 5. Distribution of monthly average seawater pCO_2 from 2003 to 2019.

4.2. Seasonal variation in air-sea CO_2 flux

This study reports what we believe to be the most comprehensive dataset of the air-sea CO_2 fluxes, based on satellite data with complete coverage of the BS, YS, and ECS, at a monthly temporal resolution and a 1 km spatial resolution over 17 years. For comparison with previous studies in the air-sea CO_2 flux estimation, we employ the gas transfer velocity algorithms from Sweeney et al. (2007) and the C_2 coefficient for the gas transfer relationship adopted by Guo et al. (2015). The multiyear monthly mean air-sea CO_2 fluxes in the entire study area are shown in Fig. 6.

The BS is near neutralized with a mean air-sea CO_2 flux density of -0.48 ± 5.20 mmol/m^2/d. The seasonal variation range of air-sea CO_2 flux density reaches 30 mmol/m^2/d due to large pCO_2 fluctuations. The BS acts as an evident CO_2 sink in typical winter and early spring months (with maximum CO_2 influx rates of 12 mmol/m^2/d in January) but as an intense CO_2 source in summer to fall, especially in September and October (with CO_2 efflux rates reaching 18 mmol/m^2/d). In late spring (May) and early winter (December), the BS shows transitional features between typical CO_2 influx and CO_2 efflux.

The YS serves as a CO_2 sink at the annual scale, with an air-sea CO_2 flux density of -2.47 ± 2.74 mmol/m^2/d. However, as a whole, the YS tends to release CO_2 into the atmosphere from July to September and absorb CO_2 in the remaining months of the year. Remarkably, the Subei Shallow acts as a very intense CO_2 source in August and September, with an efflux reaching 20 mmol/m^2/d. The other notable phenomenon is that the centre of the south YS absorbs CO_2 at an immense flux of over 15 mmol/m^2/d, resulting from the co-effect of the low pCO_2 and strong

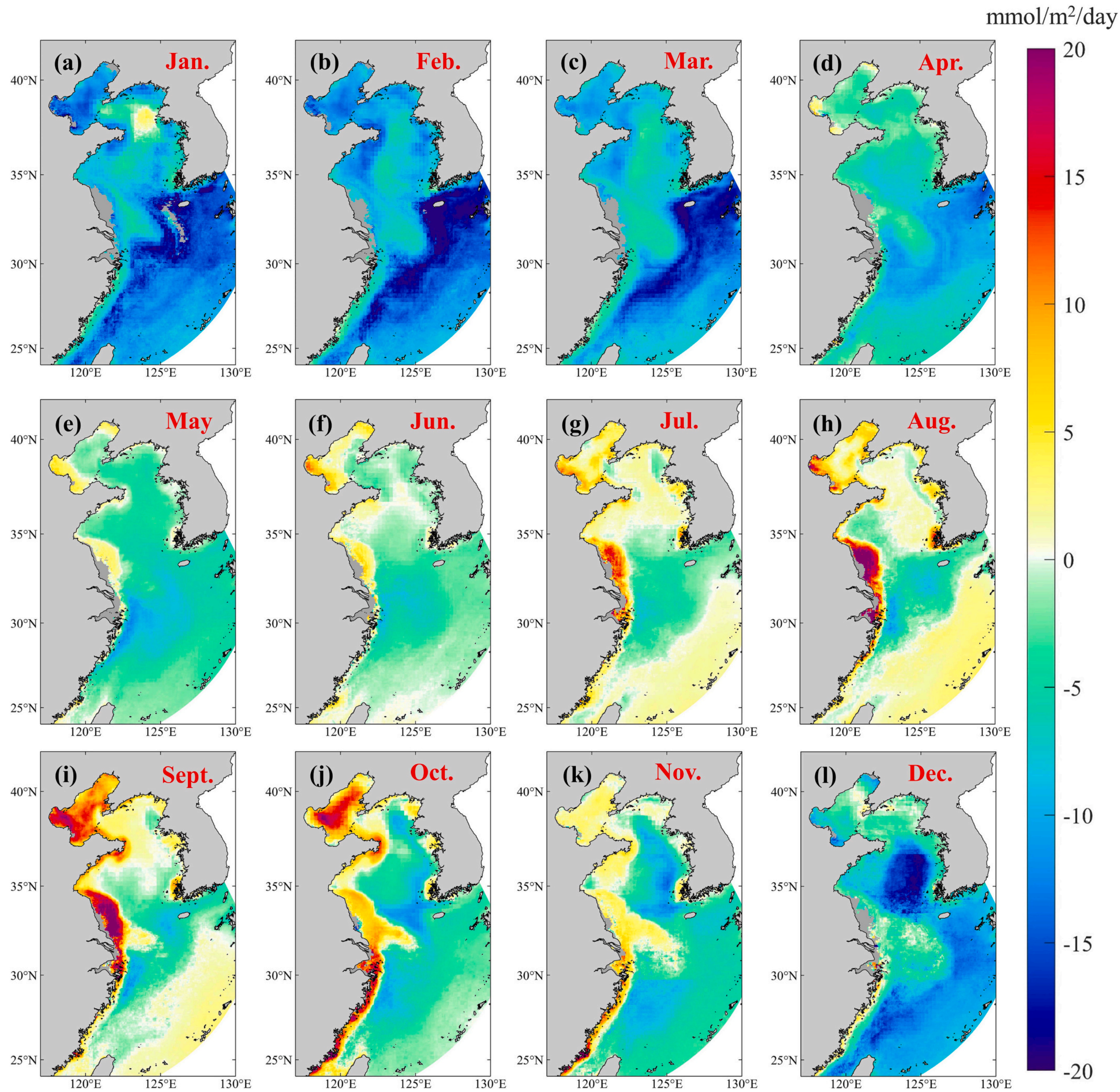


Fig. 6. Distribution of monthly average air-sea CO_2 flux from 2003 to 2019.

wind in December.

The CO_2 influx rate in the ECS is 4.10 ± 3.61 mmol/m^2/d. Seasonal patterns of air-sea CO_2 flux in different areas in the ECS vary. The area outside of the ECS shelf absorbs CO_2 in winter and early spring (March), with an influx rate ranging from 8 to 13 mmol/m^2/d, releases CO_2 (reaching 5 mmol/m^2/d) in summer and shows transitional features between typical CO_2 influx and CO_2 efflux in the remaining months. The outer ECS shelf acts as a very intense CO_2 sink in winter and early spring, with a maximum influx rate of over 20 mmol/m^2/d (with the peak occurring in February), and it remains a weak CO_2 sink in the other seasons. The Changjiang plume area acts as a weak sink with an influx of ~5 mmol/m^2/d throughout the year except in the fall months, resulting from the strong vertical mixing of the subsurface CO_2-enriched water and spread of oversaturated water from the Subei Shallow. The coastal zone and Changjiang estuary are intense sources of atmospheric CO_2 in summer and fall and weak sinks in winter and spring. Peak efflux rates reaching over 20 mmol/m^2/d in the coastal zone and Changjiang estuary are observed in August and October, respectively.

Annually, regional average of these CO_2 influx rates is estimated at -3.42 ± 3.36 mmol/m^2/d. Although the three seas display different seasonal variations in air–sea CO_2 fluxes, they all serve as sinks of atmospheric CO_2 annually. The annual CO_2 sinks of the ECS (with an area of 8.4×10^5 km^2), YS (with an area of 3.61×10^5 km^2), and BS (with an area of 7.65×10^4 km^2) are 14.80 ± 3.09, 3.85 ± 0.68, and 0.16 ± 0.26 Tg C, respectively. Combining these three seas together (accounting for an area of 1.27×10^6 km^2), the entire domain absorbs an annual amount of 18.81 ± 3.81 Tg C and a cumulative amount of 319.78 Tg C during 2003–2019 from the atmosphere. Liu et al. (2018) reported that the area-integrated CO_2 flux in three seas are -23.3 ± 13.5, -1.0 ± 0.3, and 0.2 ± 0.1 Tg C/yr, respectively, based on the synthesised estimate from *in situ* data. The difference in values may be caused by the difference in the spatial and temporal span of the statistics.

Generally, the whole study area serves as a sink of atmospheric CO_2, and CO_2 influx in the ECS is much more intense than the YS and the BS. Owing to high productivity and strong marine dynamic processes, the ECS surface has considerable CO_2 uptake capacity and can move them from the shallow coastal sea to an adjacent deep ocean *via* a continental shelf pump (Tsunogai et al., 1999; Zhai and Dai, 2009; Guo et al., 2015). The phytoplankton production in the BS and YS is also at a high level, and the SST is lower than that in the ECS; however, the limited water exchange in the BS and YS makes the accumulated carbon difficult to be removed *via* transportation and other mechanisms (Xue et al., 2012; Wang and Zhai, 2021). Besides, the shallow bathymetry and vertical mixing of water columns enable the accumulated carbon in the subsurface layer to be released back into the atmosphere (Wang and Zhai, 2021). Hence the CO_2 absorption in the BS and YS was limited. There are basic understandings of the carbon cycle process in the ECS and YS (Guo et al., 2015; Zhai and Dai, 2009; Tseng et al., 2011; Wang and Zhai, 2021), but still room to achieve a clear understanding of the deep mechanism. The comprehensive coverage and high-resolution data produced in this study can provide a new perspective for in-depth mechanism research.

4.3. Comparison with previous air-sea CO_2 flux estimates

As remote sensing data has better spatial resolution and coverage, as well as temporal continuity, it can provide more information on spatial and temporal variability. In this section, we compared our estimation of air-sea CO_2 flux with typical observations in the same time period and region to further understand the impact of temporal and spatial coverage on the CO_2 flux estimation. Table 1 shows the comparisons of the CO_2 fluxes estimated in this study with previous reports in the BS,

Table 1
Comparison of air–sea CO_2 fluxes on the Bohai Sea, Yellow Sea, and East China Sea shelf.

References	Study seasons	Sampling times	Flux (mmol/m^2/d)		Domain
			Reference (number of cruises)	This study	
SOA (2013)	Spring	2011–2012	−0.2	−2.37	Bohai Sea
	Summer		0.08	0.87	
	Fall		3.9	4.23	
	Winter		−1.9	−5.33	
Xu et al. (2016)[a]	Spring	Mar 2011 & 2013	−14.2 ± 5.0	−13.93	North Yellow Sea (38°40′N, 122°10′E)
		Apr 2011 & 2012	−7.7 ± 0.6	−6.04	
		May 2011 & 2013	−2.6 ± 1.5	−4.92	
	Summer	Jun 2011 & 2012	−1.2 ± 1.7	−3.31	
		Jul 2011 & 2013	0.0 ± 0.9	1.32	
		Aug 2011 & 2012	−0.5 ± 0.3	0.34	
	Fall	Sept 2011 & 2013	−3.6 ± 1.2	1.46	
		Oct 2011 & 2012	−1.2 ± 0.3	−1.51	
		Nov 2011 & 2013	−5.0 ± 4.3	−0.2	
	Winter	Dec 2011 & 2012	8.4 ± 0.4	7.83	
		Feb-12	−3.1 ± 0.8	−10.65	
Choi et al. (2019)[c]	Spring	May 2017	−7.5	−2	Southeast Yellow Sea (34–37°N, 124–126°E)
	Summer	Jul 2014	−1.9	−2.53	
	Fall	Nov 2015	−2.3	−4.25	
	Winter	Feb 2015	0.6	−5.67	
Guo et al. (2015)[b]	Spring	2006–2011 Average	−10.7 ± 8.2 (6)	−7.84 ± 1.46	ECS Domain I
	Summer		−6.5 ± 10.7 (5)	−3.98 ± 1.94	(28.5–33°N, 122–126°E)
	Fall		2.2 ± 6.8 (4)	−1.94 ± 1.82	
	Winter		−9.8 ± 4.7 (6)	−8.43 ± 3.31	
	Annual		−6.2 ± 9.1 (21)	−5.55 ± 3.49	
	Spring	2006–2011 Average	−10.7 ± 3.5 (5)	−6.93 ± 2.13	ECS Domain II
	Summer		−2.4 ± 3.3 (5)	0.76 ± 1.81	(25–28.5°N, 119.33–123.5°E)
	Fall		0.7 ± 4.1 (4)	−0.69 ± 2.24	
	Winter		−8.9 ± 1.4 (5)	−8.90 ± 3.78	
	Annual		−5.3 ± 3.7 (19)	−3.94 ± 4.82	
	Spring	2006–2011 Average	−17.8 ± 3.1 (1)	−10.87 ± 4.57	ECS Domain III
	Summer		−4.6 ± 4.0 (2)	−1.50 ± 1.77	(28.5–33°N, 126–128°E)
	Fall		−3.7 ± 5.1 (1)	−2.41 ± 2.45	
	Winter		−10.8 ± 1.4 (2)	−13.51 ± 4.24	
	Annual		−9.2 ± 4.2 (6)	−7.16 ± 6.34	
	Spring	2006–2011 Average	−11.2 ± 2.2 (5)	−9.20 ± 4.37	ECS Domain IV
	Summer		1.0 ± 1.5 (5)	0.98 ± 2.30	(27–28.5°N, 126–128°E)
	Fall		−9.3 ± 0.5 (2)	−3.69 ± 4.29	
	Winter		−10.6 ± 1.3 (4)	−11.79 ± 3.41	
	Annual		−7.5 ± 1.7 (16)	−6.05 ± 6.22	
	Spring	2006–2011 Average	−6.8 ± 4.3 (5)	−6.34 ± 3.50	ECS Domain V
	Summer		1.8 ± 2.8 (4)	0.77 ± 1.02	(25–27°N, 120–125.42°E)
	Fall		−8.4 ± 2.0 (1)	−2.29 ± 3.12	
	Winter		−10.0 ± 2.5 (3)	−10.96 ± 3.80	
	Annual		−5.9 ± 3.4 (13)	−4.86 ± 5.52	
	Spring	2006–2011 Average	−11.7 ± 2.5 (7)	−8.44 ± 2.65	ECS Shelf
	Summer		−3.5 ± 4.6 (5)	−1.44 ± 1.39	
	Fall		−2.3 ± 3.1 (5)	−2.28 ± 1.72	
	Winter		−10.0 ± 2.0 (7)	−10.71 ± 2.51	
	Annual		−6.9 ± 4.0 (24)	−5.72 ± 4.49	

[a] Use the equation of Wanninkhof et al. (2009) to obtain the gas transfer velocity from wind speed.
[b] Use the equation of Sweeney et al. (2007) to obtain the gas transfer velocity from wind speed.
[c] Use the equation of Wanninkhof (2014) to obtain the gas transfer velocity from wind speed.

YS, and on the ECS shelf.

In the BS, the study of air-sea CO_2 fluxes has been very limited. Four surveys conducted by the State Oceanic Administration of China (SOA, 2013) in the BS during 2011–2012 displayed the gridded difference between atmospheric and seawater pCO_2 and the regional integrated air-sea CO_2 flux and showed that the BS is a source of atmospheric CO_2 in fall, a sink in winter and spring, and at near-equilibrium with the atmosphere in summer. Our results in 2011–2012 highly agree with that report regarding the source and sink diagnoses and air-sea pCO_2 differences but disagree regarding the value of the CO_2 flux density (Table 1).

Xu et al. (2016) reported the complete monthly variation in air-sea CO_2 flux at the A4HDYD station (38°40′N, 122°10′E) located in the North YS based on 21 field surveys conducted from March 2011 to November 2013, providing a valuable baseline for understanding the temporal variability in air-sea CO_2 fluxes. The air-sea CO_2 fluxes we estimated were highly consistent with Xu et al. (2016), especially in spring (APD of <17.5 %). Our results suggested that the area around the A4HDYD station was a weak source, with a mean efflux rate of 1.04 mmol/m^2/d from July to September, which differed from the weak sink reported by Xu et al. (2016). However, the CO_2 flux density observations based on surveys in July 2006 (Xue et al., 2012), July 2016, and September 2017 (Wang and Zhai, 2021) in the North YS (37–39°N, 121–124°E) were 3.4, 7.2, and 3.7 mmol/m^2/d, respectively, illustrating that the North YS was very likely to be a source of atmospheric CO_2, which supports our estimation. Choi et al. (2019) reported the seasonal air-sea CO_2 fluxes in the south-eastern YS drawing from four surveys and identified this region as a sink of atmospheric CO_2 (−2.8 mmol C/m^2/d). Our result indicates that the corresponding area absorbed atmospheric CO_2 at a rate of 3.61 mmol C/m^2/d but with some differences from Choi et al. (2019) during winter and spring. Although there is a lack of *in situ* data on the eastern YS near the Korean Peninsula, and the uncertainty of retrieved pCO_2 and CO_2 flux cannot be quantified, we believe that remote sensing data has reliability here due to the similarity in the water mass characteristics (Kim et al., 1991) and control mechanism of seawater pCO_2 (Choi et al., 2019) between the eastern YS and the rest of the study area which was well captured by MeSAA-ML.

In the ECS, Guo et al. (2015) reported a comprehensive dataset of surface seawater pCO_2 and the associated air-sea CO_2 fluxes on the continental shelf based on 24 surveys from 2006 to 2011. Our results and the air-sea CO_2 fluxes observed by Guo et al. (2015) based on the multiple cruises highly agreed at the scale of the entire ECS shelf (Table 1). Regarding the comparison associated with the sub-domains and seasons, the agreement could be summarised as directly proportional to the number of observations/cruises. In sea areas with a high observation coverage and frequency, *e.g.*, Domains I (lower estuary and inner shelf influenced by river plume) and II (inner shelf dominated by turbid coastal waters) in Table 1, remote sensing data have been proven to estimate the air-sea CO_2 flux accurately. On the other hand, in domains with limited *in situ* observations, such as the outer shelf of the ECS, significant differences exist between retrieved and observed data.

Multiple factors may contribute to the differences between retrieved and observation-based air-sea CO_2 flux estimation. Firstly, there is a difference in spatiotemporal resolution and representativeness. The SOA cruises conducted during winter, spring, and fall did not cover the Liaodong Bay, which has a considerable area located in the northeast BS. Xu et al. (2016) conducted fixed-point measurements at a specific location with 25-hour observations taken during each survey to represent data for that month. Guo et al. (2015) took limited cruises to calculate averaged air-sea CO_2 fluxes during 2006–2011 in Mid- and outer shelf as Domain III, IV, and V (with Domain III characterized by visible river plume signals in flood seasons and Domain V characterized by upwelling northern Taiwan) of the ECS for some seasons, which may somewhat lack representativeness. However, retrieved data can offer more representative and statistically reliable estimates by providing higher observation frequency (monthly) and full spatial coverage. Moreover, the difference may have been contributed by the different gas-transfer velocities with wind speed algorithms adopted in the air-sea CO_2 flux calculation. Xu et al. (2016) used the equation of Wanninkhof et al. (2009), which may induce 11 % lower than the Sweeney et al. (2007) that we adopted in gas-transfer velocity (Tseng et al., 2014), and Choi et al. (2019) employed the equation of Wanninkhof (2014). Additionally, we used C_2 (∼ 1.2) correction in the air-sea CO_2 flux calculation, which may induce ∼13 % higher fluxes.

4.4. Long-term trends in seawater pCO_2 and air-sea CO_2 fluxes

The inter-annual trends (2003–2019) in seawater and atmospheric pCO_2, as well as air-sea CO_2 fluxes, are shown in Fig. 7. The carbon sink of the whole study area increased by 85 % from 2003 to 2019, and the total carbon absorption rose from 13 to 24 Tg C per year (Fig. 7f). Laruelle et al. (2018) also reported that the uptake of atmospheric CO_2 in global shelves had enhanced as pCO_2 in shelf waters lagged the rise in atmospheric CO_2, and some might have switched from a source to a sink during the last century, which employed the SOCAT database spanning over 35 years. The Subei Shallow, Changjiang estuary, and areas very close to the coastal line were not involved in the trend analysis because of the poor data coverage (there were 204 months of data from 2003 to 2019 generated in this study, but there were fewer than 102 months of available data for these regions).

The trends in seawater pCO_2 in the BS during 2003–2019 were rarely significant (Fig. 7a), while the atmospheric pCO_2 increased continuously at a mean rate of 2.17 μatm/yr in the whole domain (Fig. 7b and d). Although facing the continuous increase in atmospheric CO_2 and SST (Figs. 7b and S42b), the lack of substantial change in pCO_2 in the BS may be attributed to increasing biomass and bloom intensity that reported by He et al. (2013). The air-sea CO_2 flux in the BS did not present a statistically significant trend as a whole owing to large seasonal variation (ranging from −20 to 20 mmol/m^2/d). However, the annually-integrated CO_2 flux in the BS switched from being a CO_2 source of 0.2 Tg C/yr before 2006 to a sink of −0.3 Tg C/yr after 2008 (Fig. 7j).

For the whole YS, the increasing rate of seawater pCO_2 (1.74 μatm/yr) is close to that of atmospheric pCO_2 (2.18 μatm/yr, Fig. 7e). Seawater pCO_2 in the near-shore area of the eastern YS and the edge of the Subei Shallow increased at an annual rate of 3–4 μatm/yr, and the rising rate of pCO_2 in most of the central YS area was approximately 1.5–2 μatm/yr (Fig. 7a). Accordingly, the air-sea CO_2 flux density increased slightly (did not pass the significance test) during the study period (Fig. 7h), but the annually-integrated CO_2 influx escalated from 2.5 Tg C in 2003 to 4.5 Tg C in 2019 (at a rate of 0.094 Tg C per year, $p < 0.002$, Fig. 7j).

In the ECS, the overall rising rate of seawater pCO_2 was 0.78 μatm/yr, and that of atmospheric pCO_2 was 2.22 μatm/yr (Fig. 7f). The seawater pCO_2 did not show a significant upwards trend on the ECS continental shelf, which accounts for most of the area, while that in the Changjiang plume area and off the shelf increased at 3–4 and ∼1 μatm/yr, respectively (Fig. 7a). Significant CHL increases (Fig. S43a) maybe the key factor that prevented pCO_2 rises on the ECS continental shelf. Tsao et al. (2023) investigated nine years of time-series data sampled at the stations on the "PN-line" (126°E, 29°N-128.3°E, 27.5°N) within the mainstream Kuroshio Current during 2010–2018 and found trends of surface seawater pCO_2 was 3.70 ± 0.57 μatm/yr. On a two-decade scale, pCO_2 may not achieve such a rapid increase. Laruelle et al. (2018) reported that seawater pCO_2 at the region to the northeast of the "PN-line" (130°E, 30°N) is increasing but at a rate that is moderately slower than that of atmospheric pCO_2 (∼1–1.5 μatm/yr) during 1980–2015. Since the seawater pCO_2 increase was much lower than the atmospheric pCO_2, the CO_2 flux density showed significant decreasing trends (more negative, stronger sink) in the ECS except for the Changjiang plumes and the north area beyond the continental shelf (Fig. 7c). The mean air-sea CO_2 flux density of the whole ECS changed gradually from −5 mmol/m^2/d in 2003 to −8.5 mmol/m^2/d in 2019 (Fig. 7i), with an average annual trend of −0.2 mmol/m2/d ($p < 0.01$). The sink enhancing trend on the

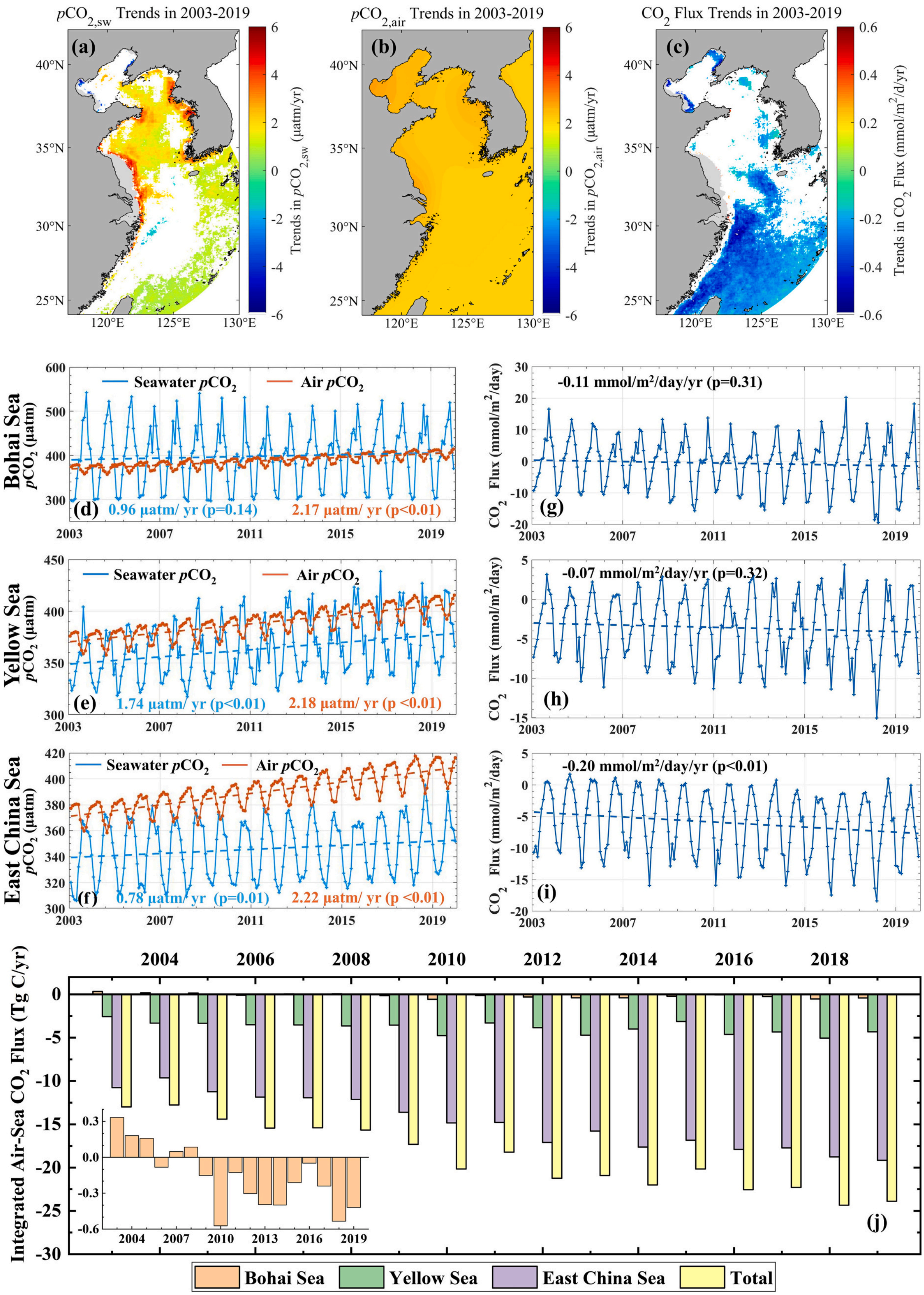


Fig. 7. Spatial distribution of trends during 2013–2019 in a) seawater pCO$_2$, b) atmospheric pCO$_2$, and c) air-sea CO_2 flux; only pixels with significant trends are coloured on the map, and the bright grey colour represents insufficient data for estimating a trend (*i.e.*, the number of valid data collected in the statistical period (204 months) does not reach 50 %). d)-f) Long-term series of seawater and atmospheric pCO$_2$ in the Bohai Sea, Yellow Sea, and East China Sea. g)-i) Long-term series of air-sea CO_2 flux density in the Bohai Sea, Yellow Sea, and East China Sea. Dashed lines are linear fitting curves. j) Annual area-integrated air-sea CO_2 flux in 2003–2019. The inset plot shows the Bohai Sea with an enlarged coordinate axis.

continental shelf south of 30°N could reach 0.3–0.4 mmol/m^2/d per year (Fig. 7c). The annual integrated CO_2 absorption of the ECS increased from 10 Tg C in 2003 to 19 Tg C in 2019 (Fig. 7j), with an average yearly trend of 0.73 Tg C ($p < 0.01$).

5. Conclusion and implications

This study established a MeSAA-Machine Learning (MeSAA-ML-ECS) algorithm innovatively combining the parameterized mechanism and machine learning model, XGBoost, based on a large amount of *in situ* pCO_2 measurements and remote sensing parameters in the BS, YS, and ECS. Satellite-derived monthly seawater pCO_2 and air-sea CO_2 flux datasets over 17 years (2003–2009) were retrieved for the first time. The results showed that the pCO_2 product had high accuracy, with an RMSE of 19.6 μatm and an APD of 4.12 % for the independent validation set. Based on this dataset, we obtained the spatiotemporal distributions and long-term changes of surface water pCO_2 and air-sea CO_2 fluxes in the BS, YS, and ECS. In particular, it makes up for the relative lack of research and observations in the BS, the eastern YS, and the outer shelf of the ECS.

Our pCO_2 and air-sea CO_2 flux data have been shared on Zenodo (doi: https://doi.org/10.5281/zenodo.7701112). Although only initial analyses were conducted in this study, our dataset has the potential to capture variations in the pCO_2 induced by mesoscale processes such as typhoons and eddies owing to its high resolution. Hopefully, our work can inspire and encourage the community and researchers to use this dataset for further studies.

The MeSAA-ML-ECS model represents a development trend to retrieve biogeochemical parameters in complicated marginal seas. For now, we introduced the semi-analytic parameter $pCO_{2,therm}$ and UI_{SST} to reflect the thermodynamics and atmospheric forcing in the air-sea equilibrium state and complicated water mass mixing effects, respectively. We will continue to develop new versions of these pCO_2 and air-sea CO_2 flux datasets, introducing more analytical or semi-analytic parameters based on the MeSAA algorithm and handing over parameterisations to machine learning methods to make the most of the potential and the advantages of machine learning, thus upgrading simple data-driven machine learning algorithms to a more stable and interpretational AI version 2.0.

CRediT authorship contribution statement

Shujie Yu: Writing – original draft, Writing – review & editing, Data Processing, Formal analysis;

Zigeng Song: Writing – review & editing, Methodology, Data Processing, Validation;

Yan Bai: Conceptualisation, Formal analysis, Writing – review & editing, Supervision, Funding acquisition;

Xianghui Guo: Formal analysis, Writing – review, Resources, Data curation;

Xianqiang He: Formal analysis, Writing – review & editing, Supervision;

Weidong Zhai: Writing – review & editing, Resources, Data curation;

Huade Zhao: Resources, Data curation;

Minhan Dai: Writing – review & editing, Supervision.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

The retrieved sea surface pCO2 and air-sea CO2 flux data are openly and freely available at Zenodo under the link https://doi.org/10.5281/zenodo.7701112.

Acknowledgments

This study was supported by the National Natural Science Foundation of China (Grants# 42141001, #42176177, #41825014) and the "Pioneer" R&D Program of Zhejiang (#2023C03013).

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.scitotenv.2023.166804.

References

Alvarez, I., Gomez-Gesteira, M., DeCastro, M., Lorenzo, M., Crespo, A., Dias, J., 2011. Comparative analysis of upwelling influence between the western and northern coast of the Iberian Peninsula. Cont. Shelf Res. 31 (5), 388–399.

Bai, Y., Cai, W.J., He, X., Zhai, W., Pan, D., Dai, M., Yu, P., 2015. 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 (3), 2331–2349.

Bauer, J.E., Cai, W.-J., Raymond, P.A., Bianchi, T.S., Hopkinson, C.S., Regnier, P.A., 2013. The changing carbon cycle of the coastal ocean. Nature 504 (7478), 61–70.

Bishop, C.M., 1995. Neural Networks for Pattern Recognition. Oxford university press.

Borges, A.V., Delille, B., Frankignoulle, M., 2005. Budgeting sinks and sources of CO_2 in the coastal ocean: diversity of ecosystems counts. Geophys. Res. Lett. 32 (14).

Cai, W.J., Dai, M., Wang, Y., 2006. Air-sea exchange of carbon dioxide in ocean margins: a province-based synthesis. Geophys. Res. Lett. 33 (12).

Carrión, D., Arfer, K.B., Rush, J., Dorman, M., Rowland, S.T., Kioumourtzoglou, M.-A., Kloog, I., Just, A.C., 2021. A 1-km hourly air-temperature model for 13 northeastern US states using remotely sensed and ground-based measurements. Environ. Res. 200, 111477.

Chen, C.-T.A., Borges, A.V., 2009. 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. II Top. Stud. Oceanogr. 56 (8–10), 578–590.

Chen, T., Guestrin, C., 2016. Xgboost: a scalable tree boosting system. In: Proceedings of the 22nd acm sigkdd International Conference on Knowledge Discovery and Data Mining.

Chen, C.T.A., Wang, S.L., 1999. Carbon, alkalinity and nutrient budgets on the East China Sea continental shelf. J. Geophys. Res. Oceans 104 (C9), 20675–20686.

Chen, L., Xu, S., Gao, Z., Chen, H., Zhang, Y., Zhan, J., Li, W., 2011. Estimation of monthly air-sea CO_2 flux in the southern Atlantic and Indian Ocean using in-situ and remotely sensed data. Remote Sens. Environ. 115 (8), 1935–1941.

Chen, C.-T., Huang, T.-H., Chen, Y.-C., Bai, Y., He, X., Kang, Y., 2013. Air–sea exchanges of CO 2 in the world's coastal seas. Biogeosciences 10 (10), 6509–6544.

Chen, S., Hu, C., Barnes, B.B., Wanninkhof, R., Cai, W.-J., Barbero, L., Pierrot, D., 2019. A machine learning approach to estimate surface ocean pCO_2 from satellite measurements. Remote Sens. Environ. 228, 203–226.

Choi, Y., Kim, D., Cho, S., Kim, T.-W., 2019. Southeastern Yellow Sea as a sink for atmospheric carbon dioxide. Mar. Pollut. Bull. 149, 110550.

Chou, W.C., Gong, G.C., Sheu, D.D., Jan, S., Hung, C.C., Chen, C.C., 2009. Reconciling the paradox that the heterotrophic waters of the East China Sea shelf act as a significant CO_2 sink during the summertime: evidence and implications. Geophys. Res. Lett. 36 (15).

Chou, W.-C., Gong, G.-C., Tseng, C.-M., Sheu, D.D., Hung, C.-C., Chang, L.-P., Wang, L.-W., 2011. The carbonate system in the East China Sea in winter. Mar. Chem. 123 (1–4), 44–55.

Dai, M., Zhai, W.-D., Lu, Z., Cai, P., Cai, W.J., Hong, H., 2004. Regional studies of carbon cycles in China: progress and perspectives (in Chinese). Adv. Earth Sci. 19 (1), 120–130.

Dai, M., Cao, Z., Guo, X., Zhai, W., Liu, Z., Yin, Z., Xu, Y., Gan, J., Hu, J., Du, C., 2013. Why are some marginal seas sources of atmospheric CO_2? Geophys. Res. Lett. 40 (10), 2154–2158.

Dai, M., Su, J., Zhao, Y., Hofmann, E.E., Cao, Z., Cai, W.-J., Gan, J., Lacroix, F., Laruelle, G.G., Meng, F., 2022. Carbon fluxes in the Coastal Ocean: synthesis, boundary processes, and future trends. Annu. Rev. Earth Planet. Sci. 50, 593–626.

Deng, X., Zhang, G.-L., Xin, M., Liu, C.-Y., Cai, W.-J., 2021. Carbonate chemistry variability in the southern Yellow Sea and East China Sea during spring of 2017 and summer of 2018. Sci. Total Environ. 779, 146376.

Dorogush, A.V., Ershov, V., Gulin, A., 2018. CatBoost: Gradient Boosting With Categorical Features Support. *arXiv preprint arXiv:1810.11363*.

Fu, Z., Hu, L., Chen, Z., Zhang, F., Shi, Z., Hu, B., Du, Z., Liu, R., 2020. Estimating spatial and temporal variation in ocean surface pCO_2 in the Gulf of Mexico using remote sensing and machine learning techniques. Sci. Total Environ. 745, 140965.

Gloege, L., Yan, M., Zheng, T., McKinley, G.A., 2022. Improved quantification of ocean carbon uptake by using machine learning to merge global models and pCO_2 data. J. Adv. Model. Earth Syst. 14 (2), e2021MS002620.

Gross, L., Thiria, S., Frouin, R., 1999. Applying artificial neural network methodology to ocean color remote sensing. Ecol. Model. 120 (2–3), 237–246.

Gruber, N., 2015. Carbon at the coastal interface. Nature 517 (7533), 148–149.

Guo, X.-H., Zhai, W.-D., Dai, M.-H., Zhang, C., Bai, Y., Xu, Y., Li, Q., Wang, G.-Z., 2015. Air–sea CO_2 fluxes in the East China Sea based on multiple-year underway observations. Biogeosciences 12 (18), 5495–5514.
Guo, X., Yao, Z., Gao, Y., Luo, Y., Xu, Y., Zhai, W., 2021. Seasonal variability and future projection of ocean acidification on the East China Sea shelf off the Changjiang Estuary. Front. Mar. Sci. 8.
Hales, B., Takahashi, T., Bandstra, L., 2005. Atmospheric CO_2 uptake by a coastal upwelling system. Global Biogeochem. Cycles 19 (1).
He, X., Bai, Y., Pan, D., Chen, C.-T., Cheng, Q., Wang, D., Gong, F., 2013. Satellite views of the seasonal and interannual variability of phytoplankton blooms in the eastern China seas over the past 14 yr (1998–2011). Biogeosciences 10 (7), 4721–4739.
Hu, C., Lee, Z., Franz, B., 2012. Chlorophyll aalgorithms for oligotrophic oceans: a novel approach based on three-band reflectance difference. J. Geophys. Res. Oceans 117 (C1).
Hu, C., Feng, L., Lee, Z., 2013. Uncertainties of SeaWiFS and MODIS remote sensing reflectance: implications from clear water measurements. Remote Sens. Environ. 133, 168–182.
Huang, W.J., Cai, W.J., Wang, Y., Lohrenz, S.E., Murrell, M.C., 2015. The carbon dioxide system on the M ississippi R iver-dominated continental shelf in the northern G ulf of M exico: 1. Distribution and air-sea CO_2 flux. J. Geophys. Res. Oceans 120 (3), 1429–1445.
Jacobson, A.R., Schuldt, K.N., Miller, J.B., Oda, T., Tans, P., Andrews, A., Mund, J., Ott, L., Collatz, G.J., Aalto, T., 2020. CarbonTracker documentation CT2019 release. In: Global Monitoring Laboratory-carbon Cycle Greenhouse Gases.
Jiang, L.Q., Cai, W.J., Wanninkhof, R., Wang, Y., Lüger, H., 2008. Air-sea CO_2 fluxes on the US South Atlantic bight: spatial and seasonal variability. J. Geophys. Res. Oceans 113 (C7).
Joshi, A., Kumar, V., Warrior, H., 2022. Modeling the sea-surface $p$$CO_2$ of the central Bay of Bengal region using machine learning algorithms. Ocean Model. 178, 102094.
Kim, K., Kim, K.-R., Rhee, T.S., Rho .K., H, Limeburner, R., Beardsley, R.C., 1991. Identification of Water Masses in the Yellow Sea and the East China Sea by Cluster Analysis, 54. Elsevier Oceanography Series, pp. 253–267.
Kim, D., Choi, S.-H., Shim, J., Kim, K.-H., Kim, C.-H., 2013. Revisiting the seasonal variations of sea-air CO_2 fluxes in the Northern East China Sea. Terr. Atmos. Ocean. Sci. 24 (3).
Landschützer, P., Gruber, N., Bakker, D.C., Schuster, U., 2014. Recent variability of the global ocean carbon sink. Global Biogeochem. Cycles 28 (9), 927–949.
Laruelle, G.G., Dürr, H.H., Slomp, C.P., Borges, A.V., 2010. Evaluation of sinks and sources of CO_2 in the global coastal ocean using a spatially-explicit typology of estuaries and continental shelves. Geophys. Res. Lett. 37 (15).
Laruelle, G.G., Lauerwald, R., Pfeil, B., Regnier, P., 2014. Regionalized global budget of the CO_2 exchange at the air-water interface in continental shelf seas. Global Biogeochem. Cycles 28 (11), 1199–1214.
Laruelle, G.G., Landschützer, P., Gruber, N., Tison, J.-L., Delille, B., Regnier, P., 2017. Global high-resolution monthly $p$$CO_2$ climatology for the coastal ocean derived from neural network interpolation. Biogeosciences 14 (19), 4545–4561.
Laruelle, G.G., Cai, W.-J., Hu, X., Gruber, N., Mackenzie, F.T., Regnier, P., 2018. Continental shelves as a variable but increasing global sink for atmospheric carbon dioxide. Nat. Commun. 9 (1), 454.
Le, C., Gao, Y., Cai, W.-J., Lehrter, J.C., Bai, Y., Jiang, Z.-P., 2019. Estimating summer sea surface $p$$CO_2$ on a river-dominated continental shelf using a satellite-based semi-mechanistic model. Remote Sens. Environ. 225, 115–126.
Liu, Q., Guo, X., Yin, Z., Zhou, K., Roberts, E.G., Dai, M., 2018. Carbon fluxes in the China seas: an overview and perspective. Sci. China Earth Sci. 61 (11), 1564–1582.
Liu, J., Bellerby, R., Li, X., Yang, A., 2022. Seasonal Variability of the Carbonate System and Air–Sea CO_2 Flux in the Outer Changjiang Estuary, East China Sea.
Lv, H., Bai, Y., Li, Q., Jiang, H., 2018. Satellite remote sensing retrieval of aquatic $p$$CO_2$ in summer in the Pearl River Estuary (in Chinese). J. Mar. Sci. 36 (2), 1–11.
Olsen, A., Triñanes, J.A., Wanninkhof, R., 2004. Sea–air flux of CO_2 in the Caribbean Sea estimated using in situ and remote sensing data. Remote Sens. Environ. 89 (3), 309–325.
Peng, B.T.H., Hung, J.J., Wanninkhof, R., Millero, F.J., 1999. Carbon budget in the East China Sea in spring. Tellus B 51 (2), 531–540.
Peng, L., Tu, Y., Huang, L., Li, Y., Fu, X., Chen, X., 2022. DAESTB: inferring associations of small molecule–miRNA via a scalable tree boosting model based on deep autoencoder. Brief. Bioinform. 23 (6), bbac478.
Qiu, Y., Zhou, J., Khandelwal, M., Yang, H., Yang, P., Li, C., 2021. Performance evaluation of hybrid WOA-XGBoost, GWO-XGBoost and BO-XGBoost models to predict blast-induced ground vibration. Eng. Comput. 1–18.
Qu, B., Song, J., Li, X., Yuan, H., Li, N., Ma, Q., 2013. $p$$CO_2$ distribution and CO_2 flux on the inner continental shelf of the East China Sea during summer 2011. Chinese J. Oceanol. Limnol. 31 (5), 1088–1097.
Qu, B., Song, J., Yuan, H., Li, X., Li, N., 2014. Air-sea CO_2 exchange process in the southern Yellow Sea in April of 2011, and June, July, October of 2012. Cont. Shelf Res. 80, 8–19.
Qu, B., Song, J., Yuan, H., Li, X., Li, N., Duan, L., Chen, X., Lu, X., 2015. Summer carbonate chemistry dynamics in the Southern Yellow Sea and the East China Sea: regional variations and controls. Cont. Shelf Res. 111, 250–261.
Qu, B., Song, J., Yuan, H., Li, X., Li, N., Duan, L., 2017. Comparison of carbonate parameters and air–sea CO_2 flux in the southern Yellow Sea and East China Sea during spring and summer of 2011. J. Oceanogr. 73 (3), 365–382.
Ramraj, S., Uzir, N., Sunil, R., Banerjee, S., 2016. Experimenting XGBoost algorithm for prediction and classification of different datasets. Int. J. Control Theory Applic. 9 (40).
Roobaert, A., Laruelle, G.G., Landschützer, P., Gruber, N., Chou, L., Regnier, P., 2019. The spatiotemporal dynamics of the sources and sinks of CO_2 in the global coastal ocean. Global Biogeochem. Cycles 33 (12), 1693–1714.
Sarma, V., 2003. Monthly variability in surface $p$$CO_2$ and net air-sea CO_2 flux in the Arabian Sea. J. Geophys. Res. Oceans 108 (C8).
Sarma, V., Saino, T., Sasaoka, K., Nojiri, Y., Ono, T., Ishii, M., Inoue, H., Matsumoto, K., 2006. Basin-scale $p$$CO_2$ distribution using satellite sea surface temperature, Chl a, and climatological salinity in the North Pacific in spring and summer. Global Biogeochem. Cycles 20 (3).
Shim, J., Kim, D., Kang, Y.C., Lee, J.H., Jang, S.-T., Kim, C.-H., 2007. Seasonal variations in $p$$CO_2$ and its controlling factors in surface seawater of the northern East China Sea. Cont. Shelf Res. 27 (20), 2623–2636.
Smirnov, A., Berrendorf, M., Shprits, Y., Kronberg, E.A., Allison, H.J., Aseev, N.A., Zhelavskaya, I.S., Morley, S.K., Reeves, G.D., Carver, M.R., 2020. Medium energy electron flux in earth's outer radiation belt (MERLIN): a machine learning model. *Space*. Weather 18 (11), e2020SW002532.
Song, N.-q., Wang, N., Lu, Y., Zhang, J.-r., 2016a. Temporal and spatial characteristics of harmful algal blooms in the Bohai Sea during 1952–2014. Cont. Shelf Res. 122, 77–84.
Song, X., Bai, Y., Cai, W.-J., Chen, C.-T.A., Pan, D., He, X., Zhu, Q., 2016b. Remote sensing of sea surface $p$$CO_2$ in the Bering Sea in summer based on a mechanistic semi-analytical algorithm (MeSAA). Remote Sens. (Basel) 8 (7), 558.
State Oceanic Administration, C, 2013. Bulletin of marine environmental status of China in 2022 (in Chinese). https://nmdis.org.cn/hygb/zghyhjzlgb/2012nzghyhjzkgb/ (accessed in Dec. 2012).
Stephens, M.P., Samuels, G., Olson, D.B., Fine, R.A., Takahashi, T., 1995. Sea-air flux of CO_2 in the North Pacific using shipboard and satellite data. J. Geophys. Res. Oceans 100 (C7), 13571–13583.
Su, J.l., 1998. Circulation dynamics of the China Seas north of 18° N. The sea 11, 483–505.
Sweeney, C., Gloor, E., Jacobson, A.R., Key, R.M., McKinley, G., Sarmiento, J.L., Wanninkhof, R., 2007. Constraining global air-sea gas exchange for CO_2 with recent bomb 14C measurements. Global Biogeochem. Cycles 21 (2).
Takahashi, T., Olafsson, J., Goddard, J.G., Chipman, D.W., Sutherland, S., 1993. Seasonal variation of CO_2 and nutrients in the high-latitude surface oceans: a comparative study. Global Biogeochem. Cycles 7 (4), 843–878.
Takahashi, T., Sutherland, S.C., Wanninkhof, R., Sweeney, C., Feely, R.A., Chipman, D. W., Hales, B., Friederich, G., Chavez, F., Sabine, C., 2009. Climatological mean and decadal change in surface ocean $p$$CO_2$, and net sea–air CO_2 flux over the global oceans. Deep-Sea Res. II Top. Stud. Oceanogr. 56 (8–10), 554–577.
Takahashi, T., Sutherland, S.C., Kozyr, A., 2017. Global ocean surface water partial pressure of CO_2 database: measurements performed during 1957–2016 (version 2016). In: ORNL/CDIAC-161, NDP-088 (V2015), Carbon Dioxide Information Analysis Center, Oak Ridge National Laboratory, US Department of Energy, Oak Ridge, Tennessee, Dataset. Document available at: https://www.nodc.noaa.gov/ocads/oceans/LDEO_Underway_Database/NDP-088_V2016.pdf (last access: 31 July 2020).
Tiyasha, T., Tung, T.M., Bhagat, S.K., Tan, M.L., Jawad, A.H., Mohtar, W.H.M.W., Yaseen, Z.M., 2021. 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. 170, 112639.
Tsao, S.-E., Shen, P.-Y., Tseng, C.-M., 2023. Rapid increase of $p$$CO_2$ and seawater acidification along Kuroshio at the east edge of the East China Sea. Mar. Pollut. Bull. 186, 114471.
Tseng, C.M., Liu, K.K., Gong, G.C., Shen, P.Y., Cai, W.J., 2011. CO_2 uptake in the East China Sea relying on Changjiang runoff is prone to change. Geophys. Res. Lett. 38 (24).
Tseng, C.-M., Shen, P.-Y., Liu, K.-K., 2014. Synthesis of observed air–sea CO_2 exchange fluxes in the river-dominated East China Sea and improved estimates of annual and seasonal net mean fluxes. Biogeosciences 11 (14), 3855–3870.
Tsunogai, S., Watanabe, S., Sato, T., 1999. Is there a "continental shelf pump" for the absorption of atmospheric CO_2? Tellus B 51 (3), 701–712.
Wang, S.-y., Zhai, W.-d., 2021. Regional differences in seasonal variation of air–sea CO_2 exchange in the Yellow Sea. Cont. Shelf Res. 218, 104393.
Wang, S.-L., Chen, C.-T.A., Hong, G.-H., Chung, C.-S., 2000. Carbon dioxide and related parameters in the East China Sea. Cont. Shelf Res. 20 (4–5), 525–544.
Wang, Z.A., Wanninkhof, R., Cai, W.-J., Byrne, R.H., Hu, X., Peng, T.-H., Huang, W.-J., 2013. The marine inorganic carbon system along the Gulf of Mexico and Atlantic coasts of the United States: insights from a transregional coastal carbon study. Limnol. Oceanogr. 58 (1), 325–342.
Wang, G., Dai, M., Shen, S.S., Bai, Y., Xu, Y., 2014. Quantifying uncertainty sources in the gridded data of sea surface CO_2 partial pressure. J. Geophys. Res. Oceans 119 (8), 5181–5189.
Wang, Z., Wang, G., Guo, X., Bai, Y., Xu, Y., Dai, M., 2022. Spatial reconstruction of long-term (2003−2020) sea surface $p$$CO_2$ in the South China Sea using a machine learning based regression method aided by empirical orthogonal function analysis. Earth Syst. Sci. Data Discuss. 1–30.
Wanninkhof, R., 1992. Relationship between wind speed and gas exchange over the ocean. J. Geophys. Res. Oceans 97 (C5), 7373–7382.
Wanninkhof, R., 2014. Relationship between wind speed and gas exchange over the ocean revisited. Limnol. Oceanogr. Methods 12 (6), 351–362.
Wanninkhof, R., Doney, S.C., Takahashi, T., Mcgillis, W.R., 2002. The effect of using time-averaged winds on regional Air-Sea CO_2 fluxes. Geophys. Monogr. Am. Geophys. Union 127, 351–356.

Wanninkhof, R., Asher, W.E., Ho, D.T., Sweeney, C., McGillis, W.R., 2009. Advances in quantifying air-sea gas exchange and environmental forcing. Annu. Rev. Mar. Sci. 1, 213–244.

Weiss, R.F., 1974. Carbon dioxide in water and seawater: the solubility of a non-ideal gas. Mar. Chem. 2 (3), 203–215.

Weiss, R., Price, B., 1980. Nitrous oxide solubility in water and seawater. Mar. Chem. 8 (4), 347–359.

Xia, Y., Jiang, L., Wang, L., Chen, X., Ye, J., Hou, T., Wang, L., Zhang, Y., Li, M., Li, Z., 2022. Rapid assessments of light-duty gasoline vehicle emissions using on-road remote sensing and machine learning. Sci. Total Environ. 815, 152771.

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

Xue, L., Zhang, L., Cai, W.-J., Jiang, L.-Q., 2011. Air–sea CO_2 fluxes in the southern Yellow Sea: an examination of the continental shelf pump hypothesis. Cont. Shelf Res. 31 (18), 1904–1914.

Xue, L., Xue, M., Zhang, L., Sun, T., Guo, Z., Wang, J., 2012. Surface partial pressure of CO_2 and air–sea exchange in the northern Yellow Sea. J. Mar. Syst. 105, 194–206.

Yin, W., Qi, Y., Cao, Z., Zhang, Y., Tang, H., 2012. The environmental characteristics of the major greenhouse gases and seawater pCO_2 in the Bohai Sea (in Chinese). Trans. Oceanol. Limnol. 4, 189–193.

Zhai, W., Dai, M., 2009. On the seasonal variation of air–sea CO_2 fluxes in the outer Changjiang (Yangtze River) Estuary, East China Sea. Mar. Chem. 117 (1–4), 2–10.

Zhai, W.-D., Zheng, N., Huo, C., Xu, Y., Zhao, H.-D., Li, Y.-W., Zang, K.-P., Wang, J.-Y., Xu, X.-M., 2014. Subsurface pH and carbonate saturation state of aragonite on the Chinese side of the North Yellow Sea: seasonal variations and controls. Biogeosciences 11 (4), 1103–1123.

Zhai, W.d., Zhao, H.d., Su, J.l., Liu, P.f., Li, Y.w., Zheng, N., 2019. Emergence of summertime hypoxia and concurrent carbonate mineral suppression in the central Bohai Sea, China. J. Geophys. Res. Biogeosci. 124 (9), 2768–2785.

Zhang, Y., Shi, K., Sun, X., Zhang, Y., Li, N., Wang, W., Zhou, Y., Zhi, W., Liu, M., Li, Y., 2022. Improving remote sensing estimation of Secchi disk depth for global lakes and reservoirs using machine learning methods. GIScience Remote Sens. 59 (1), 1367–1383.

Zhang, S., Bai, Y., He, X., Yu, S., Song, Z., Gong, F., Zhu, Q., Pan, D., 2023. The carbon sink of the Coral Sea, the world's second largest marginal sea, weakened during 2006–2018. Sci. Total Environ. 162219.

Zheng, N., Xu, X.-m., Wei, Y.-w., Zhao, H.-d., 2021. Distribution and flux estimation of dissolved inorganic carbon in main rivers around Bohai Sea in summer. Chin. J. Mar. Environ. Sci. 40 (6), 908–914. https://doi.org/10.12111/j.mes.2021-x-0133.