Feasibility of reconstructing the summer basin-scale sea surface partial pressure of carbon dioxide from sparse in situ observations over the South China Sea

Guizhi Wang1,2,3, Samuel S. P. Shen4, Yao Chen4, Yan Bai5, Huan Qin4, Zhixuan Wang1,2, Baoshan Chen1,2, Xianghui Guo1,2, and Minhan Dai1,2

1State Key Laboratory of Marine Environmental Science, Xiamen University, Xiamen, 361102, China
2College of Ocean and Earth Sciences, Xiamen University, Xiamen, 361102, China
3Fujian Provincial Key Laboratory for Coastal Ecology and Environmental Studies, Xiamen University, Xiamen, 361102, China
4Department of Mathematics and Statistics, San Diego State University, San Diego, CA 92182, USA
5State Key Laboratory of Satellite Ocean Environment Dynamics, Second Institute of Oceanography, Ministry of Natural Resources, Hangzhou, 310012, China

Correspondence: Samuel S. P. Shen (sshen@sdsu.edu) and Guizhi Wang (gzhwang@xmu.edu.cn)

Received: 23 June 2020 – Discussion started: 11 September 2020
Revised: 16 February 2021 – Accepted: 17 February 2021 – Published: 1 April 2021

Abstract. Sea surface partial pressure of CO2 ($p_{CO2}$) data with a high spatiotemporal resolution are important in studying the global carbon cycle and assessing the oceanic carbon uptake. However, the observed sea surface $p_{CO2}$ data are usually limited in spatial and temporal coverage, especially in marginal seas. This study provides an approach to reconstruct the complete sea surface $p_{CO2}$ field in the South China Sea (SCS) with a grid resolution of 0.5° × 0.5° over the period of 2000–2017 using both remote-sensing-derived $p_{CO2}$ and observed underway $p_{CO2}$, among which the gridded underway $p_{CO2}$ data in 2004, 2005, and 2006 are presented for the first time. Empirical orthogonal functions (EOFs) were computed from the remote-sensing-derived $p_{CO2}$. Then, a multilinear regression was applied to the observed $p_{CO2}$ as the response variable with the EOFs as the explanatory variables. EOF1 explains the general spatial pattern of $p_{CO2}$ in the SCS. EOF2 shows the pattern influenced by the Pearl River plume on the northern shelf and slope. EOF3 is consistent with the pattern influenced by coastal upwelling along the northern coast of the SCS. When $p_{CO2}$ observations cover a sufficiently large area, the reconstructed fields successfully display a pattern of relatively high $p_{CO2}$ in the mid and southern basin. The rate of sea surface $p_{CO2}$ increase in the SCS is 2.4±0.8 μatm yr$^{-1}$ based on the spatial average of the reconstructed $p_{CO2}$ over the period of 2000–2017. This is consistent with the temporal trends at Station SEATS (SouthEast Asia Time-series Study; 18° N, 116° E) in the northern basin of the SCS and at Station ALOHA (A Long-Term Oligotrophic Habitat Assessment; 22°45′N, 158° W) in the North Pacific. We validated our reconstruction with a leave-one-out cross-validation approach, which yields the root-mean-square error (RMSE) in the range of 2.4–5.2 μatm, smaller than the spatial standard deviation of our reconstructed data and much smaller than the spatial standard deviation of the observed underway data. The RMSE between the reconstructed summer $p_{CO2}$ and the observed underway $p_{CO2}$ is no larger than 31.7 μatm, in contrast to (a) the RMSE from 12.8 to 89.0 μatm between the remote-sensing-derived $p_{CO2}$ and the underway data and (b) the RMSE from 32.6 to 44.5 μatm between the neural-network-produced $p_{CO2}$ and the underway data. The difference between the reconstructed $p_{CO2}$ and those calculated from observations at Station SEATS is in the range from −7 to 10 μatm. These comparison results indicate the reliability of our reconstruction method and output. All the data for this paper are openly and freely available at PANGAEA under the link https://doi.org/10.1594/PANGAEA.921210 (Wang et al., 2020).
1 Introduction

The ocean plays an important role in absorbing atmospheric CO\(_2\) and consequently helps slow down global warming (Le Quere et al., 2018a). Over the last half-century the ocean has taken up approximately 24 % of the total emitted CO\(_2\) at an increasing rate from 1.0±0.5 in the 1960s to 2.6±0.6 Gt C yr\(^{-1}\) in 2019 (Friedlingstein et al., 2020; Le Quere et al., 2018b). The ocean has been found to be responsible for up to 40 % of the decadal variability of CO\(_2\) accumulation in the atmosphere (DeVries et al., 2019). However, the regional and global patterns of the oceanic carbon sink vary both spatially and temporally (Doney et al., 2009; Fay and McKinley, 2013; Landschuter et al., 2014; Le Quere et al., 2010; Rodenbeck et al., 2015; Turi et al., 2014). Thus, it is necessary to improve the spatiotemporal coverage and accuracy of the data in the evaluation of oceanic carbon uptake capacity in order to better understand the global carbon cycle and to better project the future climate.

The sea–air CO\(_2\) flux is primarily determined by the difference in the atmospheric and sea surface partial pressure of CO\(_2\) (pCO\(_2\)). The measurements of sea surface fugacity of CO\(_2\) (fCO\(_2\), which is equal to pCO\(_2\) corrected for the non-ideal behavior of the gas; Pfeil et al., 2013) have increased to 28.2 million and are presently available in almost all ocean basins based on the Surface Ocean CO\(_2\) Atlas version 2020 (Bakker et al., 2020). However, for a given year the observations of sea surface pCO\(_2\) may still have sparse spatial coverage. Thus, interpolation and/or extrapolation methods are needed to obtain a complete pCO\(_2\) field in space and time over the concerned oceanic areas. Various methods have been applied for this purpose in the past 2 decades, including statistical interpolation (Chou et al., 2005) and empirical formulas between pCO\(_2\) and proxies such as sea surface temperature, salinity, chlorophyll a, sea surface height, and mixed-layer depth (Boutin et al., 1999; Denvil-Sommer et al., 2019; Jo et al., 2012; Laruelle et al., 2017; Lefevre and Taylor, 2002; Ono et al., 2004; Zhai et al., 2005a). These studies usually present their pCO\(_2\) fields in a monthly timescale and at a 1°×1° or even coarser grid. In marginal seas a finer grid resolution is needed to discern influences posed by local forces such as plumes and upwelling.

The South China Sea (SCS) is the largest marginal sea in the western Pacific. Measurements of sea surface pCO\(_2\) in the SCS started as early as 2000 (Zhai et al., 2005b). Seasonal and spatial variations are present in different domains of the SCS (Li et al., 2020; Zhai et al., 2013). However, the data coverage is still so sparse each year that on global compilation maps the SCS has large blank areas (Bakker et al., 2016; Fay and McKinley, 2013; Takahashi et al., 2009). For example, the summer observations of 2017 cover 7 % of the SCS, and those of 2001 cover only 1 %. Consequently, the observational data themselves cannot quantitatively depict the pCO\(_2\) field over the entire SCS basin. Thus, it is necessary to reconstruct a space–time complete pCO\(_2\) field in the SCS in order to better assess the CO\(_2\) source and sink features in the SCS and to supplement the global pCO\(_2\) map.

The purpose of this paper is to demonstrate the feasibility of reconstructing the pCO\(_2\) field over the SCS basin from the sparse in situ observations in the SCS with a grid resolution of 0.5°×0.5°, using a method illustrated in the flowchart of Fig. 1. This paper focuses on the pCO\(_2\) reconstruction for the summer season. As indicated in Fig. 1, we need to use an auxiliary dataset, the remote-sensing-derived pCO\(_2\) estimates, e.g., from Bai et al. (2015a), to calculate empirical orthogonal functions (EOFs) for spatial patterns of pCO\(_2\). The remote sensing pCO\(_2\) estimates are relatively complete in the space–time grid but less accurate, compared with in situ observations. The singular-value decomposition (SVD) method is applied to the remote sensing estimates to compute the EOFs. These EOFs form an orthogonal basis for the spectral optimal gridding (SOG) method (Shen et al., 2014, 2017; Gao et al., 2015; Lammlein and Shen, 2018). The method uses a multilinear regression to blend the in situ data (treated as the data of the response variable in the regression) and the EOFs (treated as the explanatory variables) together to reconstruct the complete summer pCO\(_2\) field at 0.5°×0.5° over the SCS.

![Figure 1. Reconstruction procedure of the sea surface pCO2 in the SCS. Here, RS pCO2 means the original remote-sensing-derived pCO2; Obs. pCO2 represents the original observed in situ pCO2. RS estimates are the grid-aggregated remote-sensing-derived pCO2, and Obs. data are the grid-aggregated observed pCO2. The standard deviations are the temporal standard deviation of the RS pCO2 estimates on each grid box.](https://doi.org/10.5194/essd-13-1403-2021)
Section 2 will describe the datasets and methods; Sect. 3 includes results and discussion; and the conclusions are in Sect. 4.

2 Datasets and methods

2.1 Observed data in the SCS

In the SCS, the underway sea surface \( p\text{CO}_2 \) data are hardly available for every month of each year, so we decided to compile the data seasonally. This study focuses on the summer data, since the greatest temporal coverage of the sampling occurs in summer. The available underway summer \( p\text{CO}_2 \) data from 2000 to 2017 are compiled in this study and shown in Table 1, in which the \( p\text{CO}_2 \) data in August 2004, July 2005, and June 2006 are new and obtained continuously with a non-dispersive infrared gas analyzer (LI-COR 7000). The summer data are the June–August mean for each year in this period excluding 2002, 2003, 2010, 2011, and 2013 (Li et al., 2020). Thus, we have observed underway \( p\text{CO}_2 \) data for 13 summers during 2000–2017. Figure 2 shows that these data are distributed mainly on the northern shelf and slope and in the northern and mid basin of the SCS with a different frequency of summer observations on different grid boxes. The observational data were aggregated onto 0.5° x 0.5° grid boxes in the 5–25° N, 109–122° E region of the SCS. The aggregation used a simple space–time pattern of the observed underway data in a grid box. The aggregated data for the 13 summers are shown in Fig. 3, which shows the distribution pattern of the observed underway \( p\text{CO}_2 \) data of each year. The aggregated \( p\text{CO}_2 \) in general falls in the range of 160–480 µatm with relatively larger spatial variation nearshore and smaller spatial variability in the basin. In addition, the large differences are apparent in the spatial coverage from year to year. For example, in the summer of 2007 the observed underway \( p\text{CO}_2 \) data cover a spatial range of 12° in latitude and 13° in longitude, with 231 grid boxes with data that cover 22% of the SCS. The data fall in the range of 281–480 µatm. In the summer of 2017 the observed data cover a spatial range of 13° in latitude and 6° in longitude, with 77 grid boxes with data that cover 7% of the SCS. The data are in the range of 279–440 µatm. The summer of 2000 has only five grid boxes (covering 0.5% of the SCS) with data in the range of 400–425 µatm. The lowest observational \( p\text{CO}_2 \) values appear on the northern SCS shelf due to the influence of the Pearl River plume (see Fig. 2), where nutrient-stimulated phytoplankton uptake consumes CO\(_2\). The relatively high sea surface \( p\text{CO}_2 \) values occur mainly in the basin, which are often higher than the atmospheric \( p\text{CO}_2 \) (Li et al., 2020; Zhai et al., 2005a). The high \( p\text{CO}_2 \) values off the northeastern coast of the SCS and the southern coast of Hainan Island in the summer of 2007 are consistent with local upwelling occurrences, which bring CO\(_2\)-enriched water from the subsurface (Li et al., 2020). In the summer of 2012, the spatial coverage is 7° in latitude and 9.5° in longitude. The \( p\text{CO}_2 \) data are in the range of 191–480 µatm with the lowest value appearing on the northwestern shelf of the SCS due to the Jianjiang River plume and the highest values occurring on the northeastern shelf and off the eastern coast of Hainan Island due to upwelling (Gan et al., 2015; Jing et al., 2015). Some other data, for example, in the summer of 2000, however, are relatively localized so that no certain spatial pattern is shown before the reconstruction. Our reconstruction results will help display the spatial patterns of the complete sea surface \( p\text{CO}_2 \) field.

As a dataset for our reconstruction validation, we calculated the sea surface \( p\text{CO}_2 \) from the observed temperature, salinity, total alkalinity, dissolved inorganic carbon, phosphate, and silicate at Station SEATS (SouthEast Asia Time-series Study; 18° N, 116° E) in the northern basin of the SCS in the summer of 2007, 2009, 2012, 2014, and 2017. The nutrient sample collection and measurement are described in Du et al. (2013, 2017). The samples of total alkalinity and dissolved inorganic carbon were collected and measured following the same procedure in Guo et al. (2015). The calculation of \( p\text{CO}_2 \) was made using the program of Lewis and Wallace (1998), in which the apparent dissociation constants for carbonic acid of Mehrbach et al. (1973) refit by Dickson and Millero (1987) and the dissociation constant for bisulfate ion from Dickson (1990) were employed. Another sea surface \( p\text{CO}_2 \) dataset calculated in the same way at Station SEATS in the summer of 2000, 2001, 2004, and 2006 was compiled from Lui et al. (2020).

2.2 Remote-sensing-derived sea surface \( p\text{CO}_2 \) data

The satellite remote-sensing-derived sea surface \( p\text{CO}_2 \) in the SCS was estimated for the years of 2000–2014 using a “mechanistic semi-analytical algorithm” (MeSAA) developed by Bai et al. (2015a). In the summer in the SCS, the thermodynamic, mixing, and biological effects on the

| Year | Cruise time      | Data source                  |
|------|------------------|------------------------------|
| 2000 | July 2000        | Zhai et al. (2005a)          |
| 2001 | June 2001        | Zhai et al. (2005a)          |
| 2004 | July–August 2004 | Zhai et al. (2013), this study|
| 2005 | July 2005        | This study                   |
| 2006 | June 2006        | This study                   |
| 2007 | July–August 2007 | Zhai et al. (2013)           |
| 2008 | July–August 2008 | Li et al. (2020)             |
| 2009 | August 2009      | Li et al. (2020)             |
| 2012 | July–August 2012 | Li et al. (2020)             |
| 2014 | June 2014        | Li et al. (2020)             |
| 2015 | July–August 2015 | Li et al. (2020)             |
| 2016 | June 2016        | Li et al. (2020)             |
| 2017 | June 2017        | Li et al. (2020)             |
sea surface $pCO_2$ were parameterized in the MeSAA algorithm as a function of major controlling factors derived from multiple-satellite-derived sea surface temperature, colored dissolved organic matter, and chlorophyll. The spatial resolution of the remote-sensing-derived $pCO_2$ estimates is $1' \times 1'$. These estimates were aggregated into $0.5^\circ \times 0.5^\circ$ grid boxes in our study region (5–25° N, 109–122° E). As shown in Fig. 4, the gridded remote-sensing-derived $pCO_2$ data cover almost all the areas of the SCS (see the boxes of RS $pCO_2$ and RS estimates of full coverage in Fig. 1). We made a validation study for the remote-sensing-derived $pCO_2$ by comparison with the observed underway $pCO_2$ (Fig. 5). In general, most of the remote-sensing-derived $pCO_2$ overestimate the sea surface $pCO_2$ but not by more than 50 µatm. The root-mean-square error (RMSE) falls in the range from 12.8 to 89.0 µatm with a median of 33.8 µatm (Table 2). The RMSE values are high in the years when the underway data covered only the shelf regions. With the MeSAA algorithm, the derived $pCO_2$ dataset represents the major $CO_2$ variation in large scales. However, variations shown by these remote-sensing-derived $pCO_2$ data are much less than those shown by the observed $pCO_2$ data because the current MeSAA algorithm does not consider some local processes, such as eddies and potentially different carbonate system patterns in coastal areas. Larger spatial variations are expected especially in areas influenced by river plumes. This makes it necessary to reconstruct a $pCO_2$ field not only from the remote-sensing-derived $pCO_2$ but also constrained by the observed in situ $pCO_2$ data from cruises.

### 2.3 Reconstruction method

Figure 1 is a flowchart of our method. We used the remote-sensing-derived data to compute the EOFs for the SOG reconstruction. The grid with $0.5^\circ \times 0.5^\circ$ resolution covered 5–25° N and 109–122° E with 1040 grid boxes in total. The land area data were marked with NaN. The data were arranged in a $1040 \times 15$ space–time matrix with rows for grid boxes and columns for time. Then, we removed the 143 land grid boxes from the data and computed the climatology and standard deviation for the remaining 897 non-NaN grid boxes from the 15 years of remote-sensing-derived data from 2000 to 2014. The standardized anomalies were computed for each grid box using the remote-sensing-derived data minus the climatology and subsequently dividing the difference by the standard deviation. The singular-value decomposition (SVD) method was applied to the standardized anomalies in the space–time matrix to compute the EOFs. The results are shown in Sect. 3. The climatology and standard deviation calculated from the remote-sensing-derived data were also used to compute the standardized anomalies of the observed data, which were used as the response variable in the SOG regression reconstruction. Following the reconstruction of the standardized anomalies, the remote-sensing-derived climatology and standard deviation were then used to produce the full field as the final reconstruction result.

The SOG reconstruction method is basically a multilinear regression model for the space–time field at grid box $x$ and time $t$, expressed as follows:

$$P(x, t) = \beta_0(t) + \sum_{m \in M} \beta_m(t) E_m(x) / \sqrt{a(x)} + e(x, t),$$

where $P(x, t)$ is the response variable whose data are the standardized anomalies of the observed data, $\beta_0(t)$ is the regression intercept, $\beta_m(t)$ is the regression coefficient for the $m$th EOF $E_m(x)$, the least-square estimator of $\beta_m(t)$ is denoted by $b_m(t)$, $a(x) = \cos(\phi_x)$ is the area factor, $\phi_x$ is the centroid’s latitude (expressed in radians) of the grid box $x$, and $e(x, t)$ is the regression error. The error is assumed to be normally distributed with a zero mean and has an independent error variance of

$$\sigma^2(x, t) = E[e^2(x, t)],$$

where $E$ denotes the mathematical operation of the expected value. The explanatory variables in the above multilinear regression are $E_m(x)$, computed from the area-weighted standardized anomalies of the remote-sensing-derived data. The
anomalies were written as an $897 \times 15$ space–time data matrix. The SVD method was applied to this matrix to compute the spatial patterns, which are EOFs; the temporal patterns, which are principal components (PCs); and their corresponding variances. $\mathbf{M}$ is the set of EOFs selected for our regression reconstruction. It contained eight EOFs for every year except 2000, which had only four EOFs because the year had only five grid boxes with observed underway data.

For a given year, the grid boxes with observed data are known. Then, the linear regression model can be computed based on the observed data $P(x_d, t)$ and the EOFs in the grid boxes $x_d$ with the observed data $E_m(x_d)$. For example, the year 2002 had only 17 grid boxes with the observational data: $x_1, x_2, \ldots, x_{17}$. The data in these 17 boxes were used to estimate the intercept $b_0(t)$ and coefficients $b_m(t)$ of the regression. With the estimates $b_0(t)$ and $b_m(t)$, $m \in \mathbf{M}$, the reconstructed standardized anomalies are expressed as

$$\hat{P}(x, t) = b_0(t) + \sum_{m \in \mathbf{M}} b_m(t) E_m(x) / \sqrt{a(x)},$$

where $x$ runs through the entire 893 grid boxes over our study region in the SCS. These anomalies were converted to the full field by adding the climatology and multiplying the standard deviation computed from the remote-sensing-derived data for each of the 893 grid boxes. In this way, the full reconstructed field was produced and is presented in Sect. 3.

Many computer software packages are available to compute the EOFs using SVD and to compute multilinear regres-
sions. This paper chose to use R, a computer program language that has become a popular data science tool in the last few years, for this purpose.

SOG usually uses the first few EOFs or the first $M$ EOFs that account for more than 80% of the total variance or that are determined by response data via a correlation test (Smith et al., 1998). Here, $M$ is the size of set $M$. The current paper used eight EOFs that explain 87% of the total variances (Fig. 6).

3 Results and discussion

3.1 EOFs and PCs

EOF1 demonstrates the mode of an average level of $p$CO$_2$ with lower or higher values near the coastal regions of mainland China (Fig. 7a). This mode accounts for 49% of the variance, which indicates the dominance of the average field and hence a small overall spatial variation, except in the coastal regions. The remote-sensing-derived $p$CO$_2$ data support this mode well. EOF2 shows a north–south dipole (Fig. 7b), which is supported by the observed data shown in Fig. 3, particularly in the summer of 2017, showing lower values in the north on the shelf and slope and higher values in the south in the ocean basin. The minimum values in the north occur where the Pearl River plume dominates (Li et al., 2020; Zhai et al., 2013). EOF3 shows an east–west pattern (Fig. 7c), in addition to the north–south dipole in EOF2. EOF3 thus reflects a spatial variation of a smaller scale. This pattern is consistent with that influenced by coastal upwelling along the northeastern China coast and off eastern Hainan Island (Gan et al., 2015; Jing et al., 2015).

The PCs are a temporal stamp of the occurrence of the spatial patterns. PC1 basically shows the temporal trend (Fig. 8d). It has been concluded that surface SCS $p$CO$_2$ has an increasing trend with time (Tseng et al., 2007). PC2 indicates the strength of the north–south dipole. This strength seems to be related to the strength and extent of the Pearl River plume on the northern shelf and slope (Bai et al., 2015b; Li et al., 2020; Zhai et al., 2013). PC3 shows the temporal variation corresponding to the east–west spatial pattern of EOF3.

3.2 Reconstruction results in the SCS

Figure 8 shows that the reconstructed $p$CO$_2$ fields in the SCS have successfully displayed the spatial patterns of the observed $p$CO$_2$ and in general are consistent with previous studies (Li et al., 2020; Zhai et al., 2013). Relatively low values appear in the northern coastal region where the Pearl River plume is dominant in summer, and generally high values occur in the mid and southern basins.

The reconstructions have taken the advantages of both the observed underway data for retaining spatial and temporal variations and the remote-sensing-derived data for EOF patterns. By default, the reconstructed field has fidelity to the in situ data because the SOG reconstruction method is a fit of EOFs to the in situ data. The reconstruction is, thus, consistent with the in situ observations. When the in situ data cover a sufficiently large area and hence provide a proper constraint to the EOF fitting through the SOG procedure, the reconstruction result is more faithful to the reality. For example, the reconstructions of the summers of 2004, 2007, 2009, 2012, and 2014–2017 nicely demonstrate the spatial $p$CO$_2$ patterns (Fig. 8c, f, h, and i–m) that are consistent with observations (Li et al., 2020; Zhai et al., 2013) and ocean dynamics (Gan et al., 2015; Jing et al., 2015).

When the observational data are scarce, as long as the in situ data provide a proper constraint to the EOFs, the reconstruction can still yield reasonable results. For example, the summer of 2001 has few in situ data, but its reconstruction, with an RMSE of 7.3 µatm between the reconstructed data and the observed data, appears reasonable (Fig. 8b).

In cases of extreme data scarcity, the reconstruction may not be reliable. For example, the reconstructed data in the summer of 2000 appear to be of poor quality (Fig. 8a), since the relatively low values in the mid-SCS basin may not be realistic. These poorly reconstructed data may be due to the poor spatial coverage of the in situ $p$CO$_2$ data in the summer of 2000, which had only five grid boxes with data (Fig. 3a). These five boxes are all located together and cover only 0.5% of the SCS. Similarly, the reconstructed $p$CO$_2$ data for the summers of 2005, 2006, and 2008 are not well constrained by
the in situ $pCO_2$ data that cover only the northern shelf and slope of the SCS so that the reconstructed $pCO_2$ in the mid basin are less than 350 µatm (Figs. 8d, e, and g). These small values are unlikely, since the sea surface $pCO_2$ in the basin is generally higher than the atmospheric $pCO_2$ (380–420 µatm) (Li et al., 2020; Zhai et al., 2013). Another cause of the less ideal reconstruction results for the summers of 2005, 2006, and 2008 may be the large spatial variations in the in situ data. These variations, such as those for the summer of 2008 (Fig. 3g), in the in situ data can cause a large deviation of the regression coefficients because the linear regression is not robust.

The reconstruction results have demonstrated the feasibility of the SOG reconstruction of the sea surface $pCO_2$ over the SCS, as long as the in situ data provide a proper constraint to the EOFs. The percentage of the in situ data coverage needs not necessarily be large. However, large spatial variations in the in situ data can distort the reconstruction and lower the quality of reconstruction because the linear regression method is not robust.

3.3 Reconstruction validation and uncertainty quantification

To quantitatively validate our reconstruction, we conducted a leave-one-out cross-validation study: withholding a grid box datum, making the reconstruction using the remaining in situ data, and computing the difference between the withheld datum and the reconstructed datum at the same grid box. This
was repeated for every grid box with in situ data for each year. The final cross-validation result is output as RMSE (Table 2). The maximum RMSE is 5.2 µatm, which occurred in 2006 when there were only 25 grid boxes with in situ \( pCO_2 \) data, and the in situ data had the largest spatial standard deviation, 49.4 µatm, among the 13 years under consideration. The minimum RMSE is 2.4 µatm, which occurred in 2017 with 77 in situ data grid boxes and a spatial standard deviation of 17.6 µatm for the in situ data. This accuracy is very good compared to the spatial standard deviation of the in situ data in the same year. Compared to the 2006 data, a more accurate reconstruction for 2017 is expected because of more grid boxes with in situ data and smaller spatial variability. This is supported by the cross-validation result. The spatial standard deviation of the reconstructed data is in the range of 2.1–6.6 µatm. The cross-validation RMSEs are in the range of 2.4–5.2 µatm. We thus conclude that the reliability of our reconstruction is well supported by the leave-one-out cross-validation result.

We have also considered other types of cross-validations, such as leaving out data in half of the study region. A numerical test was made for the following situation: leaving out the western or eastern half of the data in a year, making the reconstruction using the remaining half of data, and computing the RMSE between the removed data and the reconstructed data. The analysis was done for the years with better spatial coverage: 2007, 2009, and 2012. When the western halves (longitude \(<115.5^\circ\ E\) of data in these years were removed, the resulted RMSEs were 2.77, 4.46, and 3.82 µatm for 2007, 2009, and 2012, respectively. When the eastern halves (longitude \(>115.5^\circ\ E\) of data were removed, the RMSEs were 4.32, 3.66, and 3.55 µatm in 2007, 2009, and 2012, respectively. These RMSEs fall within the range of the RMSEs of the leave-one-out cross-validation. These numerical results are another confirmation of the reliability of our reconstruction.

The uncertainty in our reconstruction was quantified by grid-by-grid comparisons of the reconstructed \( pCO_2 \) with the observed \( pCO_2 \) in two ways. One is the comparison with the observed underway data (Fig. 9). The difference between the reconstructed data and the observed underway data mostly falls within the range from −30 to 30 µatm (Fig. 9). The greatest deviation from the underway data appears near the coast, likely due to the lack of some typical patterns in coastal areas transferred via EOFs from the remote sensing estimates. The RMSE between the reconstructed data and the observed underway data is no larger than 31.7 µatm with a median of 16.5 µatm, which is smaller than the RMSE between the remote-sensing-derived \( pCO_2 \) and the underway data with the relative difference between the two RMSEs (rows 1 and 3 in Table 2) by at least 29 %. When comparing the \( pCO_2 \) data produced by Jo et al. (2012) in the northern
SCS by a neural network approach in the summer of the years 2004–2007 with the underway \( p\text{CO}_2 \), the resultant RMSE falls in the range from 32.6 to 44.5 \( \mu \text{atm} \) and is twice as much as the median RMSE between our reconstructed \( p\text{CO}_2 \) and the underway \( p\text{CO}_2 \) (Table 2). Another uncertainty quantification for our reconstruction is to compare it with the \( p\text{CO}_2 \) calculated from long-term observations at Station SEATS (Fig. 10). The difference between the reconstructed \( p\text{CO}_2 \) and the observed data at Station SEATS ranges from \(-7\) to \(10 \mu \text{atm} \) with the relative difference from \(-1.5\%\) to \(2.1\%\). Both comparisons confirm that our reconstruction results are reliable.

As an application of our reconstruction and a validation, we examine the temporal trend of sea surface \( p\text{CO}_2 \) over the SCS. The rate based on the linear temporal trend of the spatial average of the reconstructed sea surface \( p\text{CO}_2 \) over the SCS from 2000 to 2017 is \(2.4 \pm 0.8 \mu \text{atm yr}^{-1}\) (see Fig. 11a). It is lower than the rate of \( f\text{CO}_2 \) increase (4 \( \mu \text{atm yr}^{-1}\)) for the period of 1999–2003 (Tseng et al., 2007), while it is higher than the rate of \( p\text{CO}_2 \) increase for the period of 1998–2006 (0.8 \( \mu \text{atm yr}^{-1}\)) at Station SEATS (Lui et al., 2020). The differences between their rates and ours exist because (a) our rate is a spatial average in summer and their rates are based on data collected in spring, summer, fall, and winter at a basin station and (b) the period to derive our rate is much longer than theirs. Using the summer data in Lui et al. (2020), the rate we estimated from the year 2000, which is the beginning year of our data, to the year 2006 at Station SEATS is \(2.5 \pm 1.0 \mu \text{atm yr}^{-1}\). Although the period of 2000–2006 is much shorter than our period of 2000–2017, the summer rate at Station SEATS is almost the same as our rate based on the reconstructed data over the SCS. When compared with the summer rate of observed \( p\text{CO}_2 \) at Station ALOHA (A Long-Term Oligotrophic Habitat Assessment station) (22°45′N, 158°W) in the North Pacific, which is \(2.0 \pm 0.2 \mu \text{atm yr}^{-1}\) over 2000–2017 (Dore et al., 2009) (see Fig. 11b), our rate is consistent with the rate in the North Pacific. The consistency of the trend in our reconstructed sea surface \( p\text{CO}_2 \) over the SCS with the local trend at Station SEATS and the North Pacific trend at Station ALOHA confirms that our reconstruction is reasonable.

3.4 Outliers of the observed data in the reconstruction

The SOG method is basically a linear regression method, which is known to be sensitive to the outliers of the response
Figure 9. The difference between the reconstructed summer pCO$_2$ and the observed underway pCO$_2$ (unit: µatm) in 2000, 2001, 2004–2009, 2012, and 2014–2017.

Figure 10. The comparison between the summer sea surface pCO$_2$ calculated from the observations and those from our reconstruction at Station SEATS (18° N, 116° E). The pCO$_2$ data calculated from the observations in the years of 2000, 2001, 2004, and 2006 are from Lui et al. (2020).
Figure 11. (a) Time series and linear trend of the spatial averages of the reconstructed summer \( p\text{CO}_2 \) data in the SCS in the period of 2000–2017; (b) summer sea surface \( p\text{CO}_2 \) at Station ALOHA in the years 2000–2017 adapted from Dore et al. (2009).

data. Some outliers, whether due to observational biases or extreme events, can cause a large change in the regression coefficients and hence the regression results and can even make the regression results outside the physically valid domain, such as negative \( p\text{CO}_2 \) values in the reconstructed data. Although we cannot conclude that the outliers of 3\( \sigma \) away from the mean in the observed data are due to data biases, we have decided not to use them in our reconstruction to avoid the unphysical reconstruction results. Table 3 shows the 14 outlier entries excluded from our response data for regression. These outliers are located in the region of 21.25–23.25° N, 113.25–116.75° E. This region is near the Pearl River estuary. Thus, these extremely low \( p\text{CO}_2 \) values may result from the Pearl River plume where the observed \( p\text{CO}_2 \) can be very low. These very low values, such as those at least 3\( \sigma \) away from the mean, may cause a very large gradient in the observed \( p\text{CO}_2 \). Our reconstruction has excluded these extremely low values influenced by the river plumes. Our reconstructed data may therefore overestimate the \( p\text{CO}_2 \) values in the Pearl River estuary and its nearby region.

### 4 Code availability

The R computer codes and their required files for this paper are freely available at the following link: https://doi.org/10.5281/zenodo.4567859 (Wang et al., 2021).

### 5 Data availability

The gridded underway sea surface \( p\text{CO}_2 \) data, the remote-sensing-derived sea surface \( p\text{CO}_2 \) estimates, and the reconstruction result data are openly and freely available at PANGAEA at the following link: https://doi.org/10.1594/PANGAEA.921210 (Wang et al., 2020).

| Year | Grid ID | Latitude (N) | Longitude (E) | \( p\text{CO}_2 \) (µatm) |
|------|---------|--------------|---------------|---------------------|
| 2006 | 926     | 22.75°       | 116.75°       | 208                 |
| 2006 | 952     | 23.25°       | 116.75°       | 197                 |
| 2009 | 896     | 22.25°       | 114.75°       | 212                 |
| 2009 | 923     | 22.75°       | 115.25°       | 217                 |
| 2012 | 836     | 21.25°       | 110.75°       | 248                 |
| 2014 | 873     | 21.75°       | 116.25°       | 191                 |
| 2014 | 874     | 21.75°       | 116.75°       | 219                 |
| 2016 | 841     | 21.25°       | 113.25°       | 265                 |
| 2016 | 842     | 21.25°       | 113.75°       | 272                 |
| 2016 | 868     | 21.75°       | 113.75°       | 239                 |
| 2016 | 869     | 21.75°       | 114.25°       | 205                 |
| 2016 | 870     | 21.75°       | 114.75°       | 216                 |
| 2016 | 896     | 22.25°       | 114.75°       | 210                 |
| 2016 | 897     | 22.25°       | 115.25°       | 274                 |
6 Conclusions

This study has demonstrated the feasibility of using the SOG method to reconstruct the sea surface $p$CO$_2$ data into regular grid boxes. We compiled the observed underway and remote-sensing-derived sea surface $p$CO$_2$ data in the SCS in summer over the period of 2000–2017 and aggregated these data with a grid resolution of 0.5° × 0.5° for reconstruction. The SOG method based on the multilinear regression was applied to reconstruct the space–time complete $p$CO$_2$ field in the SCS. The method took the EOFs calculated from the remote-sensing-derived $p$CO$_2$ as the explanatory variables and treated the observed $p$CO$_2$ as the response variable. The EOFs reflect reasonably well the general spatial pattern of the sea surface $p$CO$_2$ in the SCS and reveal features affected by regional physical forcing such as the river plume and coastal upwelling in the northern SCS. As long as the in situ data provide a proper constraint to the EOFs, the reconstructed $p$CO$_2$ fields are, in general, consistent with the patterns of the observed $p$CO$_2$ and demonstrate relatively low values along the northern coast affected by the Pearl River plume and consistently high values in the ocean basin of the SCS. The leave-one-out cross-validation result validates our reconstruction with an RMSE smaller than the spatial standard deviation of the observed underway data in the same year. The grid-by-grid comparison of the reconstructed summer $p$CO$_2$ with the observed underway $p$CO$_2$ has an RMSE smaller than that of the remote-sensing-derived $p$CO$_2$, as well as that of the neural-network-produced $p$CO$_2$ in the same year. Moreover, our reconstructed $p$CO$_2$ compares well with the $p$CO$_2$ calculated from observations around Station SEATS in the northern basin of the SCS. These comparisons confirm that our reconstruction is reliable. The temporal rate of our reconstructed sea surface $p$CO$_2$ over the SCS is consistent with the local rate at Station SEATS and the North Pacific rate at Station ALOHA, which further validates our reconstruction. These reconstructed $p$CO$_2$ fields provide full spatial coverage of the sea surface $p$CO$_2$ of the SCS in summer over a temporal scale of almost 2 decades and therefore help fill the long-lasting blanks on the global sea surface $p$CO$_2$ map. Thus, the reconstruction products will help improve the accuracy of the estimate of the oceanic CO$_2$ flux of the largest marginal sea of the western Pacific so as to better constrain the global oceanic carbon uptake capacity.

Although the SOG method can optimize the information from both the in situ data and the remote-sensing-derived data, the reliability of the reconstructed results is still limited by the observed data. When the observed data are limited to only a few grid boxes in a small region, the reconstruction results may not be realistic. Additional constraints have to be considered.

Financial support. This research has been supported by the Ministry of Science and Technology of China (grant nos. 2015CB954001 and 2009CB9421200) and the National Natural Science Foundation of China (grant no. 40521003).

Competing interests. The authors declare that they have no conflict of interest.

Author contributions. MD conceptualized and directed the field program of the in situ observations. BC and XG participated in the in situ data collection. YB provided the remote-sensing-derived data. GW, YC, and SSPS developed the reconstruction method, wrote the MATLAB and R codes, analyzed the data, and plotted the figures. ZW participated in the uncertainty analysis of the reconstruction and plotted some of the figures. HQ developed the data repository and revised and tested the R codes. GW and SSPS wrote the paper. All the authors contributed to the original writing, editing, and revisions of the paper.

Acknowledgements. We thank Hon-Kit Lui for the communication about the data at Station SEATS. We are thankful to Weidong Zhai for providing the underway $p$CO$_2$ data in his publications and to Young-Heon Jo for providing the neural-network-produced $p$CO$_2$ estimates in his published paper. We appreciate the constructive comments from the two anonymous reviewers that have helped improve this paper. Aiqin Han, Tao Huang, and Lifang Wang helped in nutrient sample collection and measurements, and Liguo Guo, Wenping Jing, Yi Xu, Wei Yang, and Nan Zheng helped in the sample collection and measurements of total alkalinity and dissolved inorganic carbon at Station SEATS. The work by Guizhi Wang, Yan Bai, Xianhui Guo, and Minhan Dai was supported by grants from the Ministry of Science and Technology of China (nos. 2015CB954001 and 2009CB9421200). The observed underway $p$CO$_2$ data in 2004, 2005, and 2006 were collected under the support of the National Natural Science Foundation of China (grant no. 40521003).

Review statement. This paper was edited by David Carlson and reviewed by two anonymous referees.

References

Bai, Y., Cai, W.-J., He, X., Zhai, W. D., Pan, D., Dai, M., and Yu, P.: A mechanistic semi-analytical method for remotely sensing sea surface $p$CO$_2$ in river-dominated coastal oceans: a case study from the East China Sea, J. Geophys. Res., 120, 2331–2349, https://doi.org/10.1002/2014JC010632, 2015a.
Bai, Y., Huang, T. H., He, X. Q., Wang, S. L., Hsin, Y. C., Wu, C. R., Zhai, W. D., Lui, H. K., and Chen, C. T. A.: Intrusion of the Pearl River plume into the main channel of the Taiwan Strait in summer, J. Sea Res., 95, 1–15, https://doi.org/10.1016/j.seares.2014.10.003, 2015b.
Bakker, D. C. E., Pfeil, B., Landa, C. S., Metzl, N., O’Brien, K. M., Olsen, A., Smith, K., Cosca, C., Harasawa, S., Jones, S. D., Nakaoka, S., Nojiri, Y., Schuster, U., Steinhoff, T., Sweeney, C., Takahashi, T., Tilbrook, B., Wada, C., Wanninkhof, R., Alin, S., Bakker, D. C. E., Pfeil, B., Landa, C. S., Metzl, N., O’Brien, K. M., Olsen, A., Smith, K., Cosca, C., Harasawa, S., Jones, S. D., Nakaoka, S., Nojiri, Y., Schuster, U., Steinhoff, T., Sweeney, C., Takahashi, T., Tilbrook, B., Wada, C., Wanninkhof, R., Alin, S., and. https://doi.org/10.5194/essd-13-1403-2021

Earth Syst. Sci. Data, 13, 1403–1417, 2021
Gao, J., Shen, S. S. P., Yao, T., Tafolla, N., Risi, C., and He, Y.: Reconstruction of precipitation δ18O over the Tibetan Plateau since 1910, J. Geophys. Res.-Atmos., 120, 4878–4888, https://doi.org/10.1002/2015JD023233, 2015.

Guo, X. H. and Wong, G. T. F.: Carbonate chemistry in the Northern South China Sea Shelf-sea in June 2010, Deep-Sea Res. II, 117, 119–130, https://doi.org/10.1016/j.dsr2.2015.02.024, 2015.

Jing, Z., Qi, Y., Du, Y., Zhang, S. W., and Xie, L. L.: Summer upwelling and thermal fronts in the northwestern South China Sea: observational analysis of two mesoscale mapping surveys, J. Geophys. Res., 120, 1993–2006, https://doi.org/10.1002/2014jc010601, 2015.

Jo, Y.-H., Dai, M., Zhai, W., Yan, X.-H., and Shang, S.: On the variations of sea surface pCO2 in the northern South China Sea: A remote sensing based neural network approach, J. Geophys. Res., 117, C08022, https://doi.org/10.1029/2011JC007745, 2012.

Lammllein, L. J. and Shen, S. S. P.: A multivariate regression reconstruction of the quasi-global annual precipitation on 1-degree grid from 1900 to 2015, Advances in Data Science and Adaptive Analysis, 10, 185008, https://doi.org/10.1142/S2424922X18500080, 2018.

Landschützer, P., Gruber, N., Bakker, D. C. E., and Schuster, U.: Recent variability of the global ocean carbon sink, Global Biogeochem. Cy., 28, 927–949, https://doi.org/10.1002/2014gb004853, 2014.

Laruelle, G. G., Landschützer, P., Gruber, N., Tison, J.-L., Delille, B., and Regnier, P.: Global high-resolution monthly pCO2 climatology for the coastal ocean derived from neural network interpolation, Biogeosciences, 14, 4545–4561, https://doi.org/10.5194/bg-14-4545-2017, 2017.

Lefèvre, N. and Taylor, A.: Estimating pCO2 from sea surface temperatures in the Atlantic gyres, Deep-Sea Res. Pt. I, 49, 539–554, https://doi.org/10.1016/s0967-0637(01)00064-4, 2002.

Le Quéré, C., Takahashi, T., Buitenhuis, E. T., Rodenbeck, C., and Sutherland, S. C.: Impact of climate change and variability on the global oceanic sink of CO2, Global Biogeochem. Cy., 24, GB4007, https://doi.org/10.1029/2009GB003599, 2010.

Le Quéré, C., Andrew, R. M., Friedlingstein, P., Sitch, S., Pongratz, J., Manning, A. C., Korsbakken, J. I., Peters, G. P., Canadell, J. G., Jackson, R. B., Boden, T. A., Tans, P. P., Andrews, O. D., Arora, V. K., Bakker, D. C. E., Barbero, L., Becker, M., Betts, R. A., Bopp, L., Chevallier, F., Chini, L. P., Ciais, P., Cosca, C. E., Cross, J., Currie, K., Gasser, T., Harris, I., Hauck, J., Haverd, V., Houghton, R. A., Hunt, C. W., Hurr, G., Ilyina, T., Jain, A. K., Kato, E., Kautz, M., Keeling, R. F., Klein Goldewijk, K., Körtzinger, A., Landschützer, P., Lefèvre, N., Lentot, A., Lierett, S., Larnzdorozzi, D., Metzl, N., Millero, F., Monteiro, P. M. S., Munro, D. R., Nabel, J. E. M. S., Nakaoka, S., Nojiri, Y., Padin, X. A., Peregon, A., Pfeil, B., Piorrot, D., Poulter, B., Rehder, G., Reimer, J., Rödenbeck, C., Schwinger, J., Séférian, R., Skjelvan, I., Stocker, B. D., Tian, H., Tilbrook, B., Tubiello, F. N., van der Laan-Luijx, I. T., van der Werf, G. R., van Heuven, S., Viovy, N., Vuichard, N., Walker, A. P., Watson, A. J., Wilshire, A. J., Zaehle, S., and Zhu, D.: Global Carbon Budget 2017, Earth Syst. Sci. Data, 10, 405–448, https://doi.org/10.5194/essd-10-405-2018, 2018b.

Lewis, E. R. and Wallace, D. W. R.: Program Developed for CO2 System Calculations, United States: N. p., Web, https://doi.org/10.15485/1464255, 1998.

Li, Q., Guo, X., Zhai, W., Xu, Y., and Dai, M.: Partial pressure of CO2 and air–sea CO2 fluxes in the South China Sea: synthesis of an 18 year dataset, Prog. Oceanogr., 182, 102272, https://doi.org/10.1016/j.pocean.2020.102272, 2020.

Lui, H.-K., Chen, C.-T. A., Hou, W.-P., Yu, S., Chan, J.-W., Bai, Y., and He, X.: Transient Carbonate Chemistry in the Expanded Kuroshio Region, in: Changing Asia-Pacific Marginal Seas, eds: Chen, C.-T. and Guo, X., Atmosphere, Earth, Ocean & Space, Springer, Singapore, https://doi.org/10.1007/978-981-15-4886-4_16, 2020.

Mehrbach, C., Culiberso, C. H., Hawley, J. E., and Pytkowicz, R. M.: Measurement of partial dissociation constants of carbonic acid in seawater at atmospheric pressure, Limnol. Oceanogr., 18, 897–907, https://doi.org/10.4319/lo.1973.18.6.0897, 1973.

Ono, T., Saito, T., Kurita, N., and Sasaki, K.: Basin-scale extrapolation of shipboard pCO2 data by using satellite SST and ChlA, Int. J. Remote Sens., 25, 3803–3815, https://doi.org/10.1080/01431160310001657515, 2004.

Pfeil, B., Olsen, A., Bakker, D. C. E., Hankin, S., Koyuk, H., Kozyr, A., Malczyn, J., Manke, A., Metzl, N., Sabine, C. L., Akl, J., Alin, S. R., Bates, N., Bellerby, R. G. J., Borges, A., Boutin, J., Brown, P. J., Cai, W.-J., Chavez, F. P., Chen, A., Cosca, C., Fassbender, A. J. J., Feely, R. A., González-Dávila, M., Goyet, C., Hales, B., Hardman-Mountford, N., Heinze, C., Hood, M., Hoppema, M., Hunt, C. W., Hydes, D., Ishii, M., Johannesson, T., Jones, S. D., Key, R. M., Körtzinger, A., Landschützer, P., Lauvset, S. K., Lefèvre, N., Lentot, A., Lourantou, A., Merlivat, L., Midorikawa, T., Minotrop, L., Miyazaki, C., Murata, A., Nakada, A., Nakano, Y., Nakaoka, S., Nojiri, Y., Omar, A. M., Padin, X. A., Park, G.-H., Paterson, K., Perez, F. F., Piorrot, D., Poisson, A., Rios, A. F., Santana-Casiano, J. M., Salisbury, J., Sarma, V. V. A., Park, G.-H., Paterson, K., Perez, F. F., Piorrot, D., Poisson, A., Rios, A. F., Santana-Casiano, J. M., Salisbury, J., Sarma, V. V. A., Park, G.-H., Paterson, K., Perez, F. F., Piorrot, D., Poisson, A., Rios, A. F., Santana-Casiano, J. M., Salisbury, J., Sarma, V. V.
Rödenbeck, C., Bakker, D. C. E., Gruber, N., Iida, Y., Jacobson, A. R., Jones, S., Landschützer, P., Metzl, N., Nakaoka, S., Olsen, A., Park, G.-H., Peylin, P., Rodgers, K. B., Sasse, T. P., Schuster, U., Shuter, J. D., Valsala, V., Wanninkhof, R., and Zeng, J.: Data-based estimates of the ocean carbon sink variability – first results of the Surface Ocean \( p\text{CO}_2 \) Mapping intercomparison (SOCOM), Biogeosciences, 12, 7251–7278, https://doi.org/10.5194/bg-12-7251-2015, 2015.

Shen, S. S. P., Tafolla, N., Smith, T. M., and Arkin, P. A.: Multivariate regression reconstruction and its sampling error for the quasi-global annual precipitation from 1900-2011, J. Atmos. Sci., 71, 3250–3268, https://doi.org/10.1175/JAS-D-13-0301.1, 2014.

Shen, S. S. P., Behm, G., Song, T. Y., and Qu, T. D.: A dynamically consistent reconstruction of ocean temperature, J. Atmos. Ocean. Tech., 34, 1061–1082, https://doi.org/10.1175/JTECH-D-16-0133.1, 2017.

Smith, T. M., Livezey, R. E., and Shen, S. S. P.: An improved method for interpolating sparse and irregularly distributed data onto a regular grid, J. Clim., 11, 1717–1729, https://doi.org/10.1175/1520-0442(1998)011<1717:AIMFAS>2.0.CO;2, 1998.

Takahashi, T., Sutherland, S. C., Wanninkhof, R., Sweeney, C., Feely, R. A., Chipman, D. W., Hales, B., Friederich, G., Chavez, F., Sabine, C., Watson, A., Bakker, D. C. E., Schuster, U., Metzl, N., Yoshikawa-Inoue, H., Ishii, M., Midorikawa, T., Nohira, Y., Kortzinger, A., Steinhoff, T., Hoppema, M., Olaofsson, J., Armaison, T. S., Tilbrook, B., Johannessen, T., Olsen, A., Bellerby, R., Wong, C. S., Delille, B., Bates, N. R., and de Baar, H. J. W.: Climatological mean and decadal change in surface ocean \( p\text{CO}_2 \), and net sea-air \( \text{CO}_2 \) flux over the global ocean, Deep-Sea Res. Pt. II, 56, 554–577, https://doi.org/10.1016/j.dsr2.2008.12.009, 2009.

Tseng, C. M., Wong, G. T. F., Chou, W. C., Lee, B. S., Sheu, D. D., and Liu, K. K.: Temporal variations in the carbonate system in the upper layer at the SEATS station, Deep-Sea Res. Pt. II, 54, 1448–1468, https://doi.org/10.1016/j.dsr2.2007.05.003, 2007.

Turi, G., Lachkar, Z., and Gruber, N.: Spatiotemporal variability and drivers of \( p\text{CO}_2 \) and air-sea \( \text{CO}_2 \) fluxes in the California Current System: an eddy-resolving modeling study, Biogeosciences, 11, 671–690, https://doi.org/10.5194/bg-11-671-2014, 2014.

Wang, G. Z., Shen, S. S. P., Chen, Y., Qin, H., Bai, Y., Chen, B. S., Guo, X. H., and Dai, M. H.: Summer partial pressure of carbon dioxide from the South China Sea from 2000 to 2017, PANGAEA, https://doi.org/10.1594/PANGAEA.921210, 2020.

Wang, G., Shen, S. P. S., Chen, Y., and Qin, H.: Hqin2019/pCO2-reconstruction: \( p\text{CO}_2 \)-reconstruction (Version v1.0.2), Zenodo, https://doi.org/10.5281/zenodo.4567859, 2021.

Zhai, W. D., Dai, M. H., Cai, W. J., Wang, Y. C., and Hong, H. S.: The partial pressure of carbon dioxide and air-sea \( \text{CO}_2 \) fluxes in the northern South China Sea in spring, summer and autumn, Mar. Chem., 96, 87–97, https://doi.org/10.1016/j.marchem.2004.12.002, 2005a.

Zhai, W. D., Dai, M. H., Cai, W. J., Wang, Y. C., and Wang, Z. H.: High partial pressure of \( \text{CO}_2 \) and its maintaining mechanism in a subtropical estuary: the Pearl River estuary, China, Mar. Chem., 93, 21–32, https://doi.org/10.1016/j.marchem.2004.07.003, 2005b.

Zhai, W.-D., Dai, M.-H., Chen, B.-S., Guo, X.-H., Li, Q., Shang, S.-L., Zhang, C.-Y., Cai, W.-J., and Wang, D.-X.: Seasonal variations of sea-air \( \text{CO}_2 \) fluxes in the largest tropical marginal sea (South China Sea) based on multiple-year underway measurements, Biogeosciences, 10, 7775–7791, https://doi.org/10.5194/bg-10-7775-2013, 2013.