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

Guizhi Wang[1,2], Samuel S. P. Shen[3], Yao Chen[3], Yan Bai[4], Huan Qin[3], Baoshan Chen[1], Xianghui Guo[1],
5 Minhan Dai[1]

[1]State Key Laboratory of Marine Environmental Science and College of Ocean and Earth Sciences, Xiamen University, Xiamen, 361102, China
[2]Fujian Provincial Key Laboratory for Coastal Ecology and Environmental Studies, Xiamen University, Xiamen, 361102, China
10 [3]Department of Mathematics and Statistics, San Diego State University, San Diego, CA 92182, USA
[4]State Key Laboratory of Satellite Ocean Environment Dynamics, Second Institute of Oceanography, State Oceanic Administration, Hangzhou, 310012, China

Correspondence to: Samuel S. P. Shen (sshen@sdsu.edu)

Abstract. Sea surface partial pressure of CO_2 (pCO_2) data with high spatial-temporal resolution are important in studying the
15 global carbon cycle and assessing the oceanic carbon uptake capacity. However, the observed sea surface pCO_2 data are usually limited in spatial and temporal coverage, especially in marginal seas. This study provides an approach to reconstruct the complete sea surface pCO_2 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 pCO_2 and observed pCO_2. Empirical orthogonal functions (EOFs) were computed from the remote sensing derived pCO_2. Then, a multilinear regression was applied to the observed pCO_2 as the response
20 variable with the EOFs as the explanatory variables. EOF1 explains the general spatial pattern of pCO_2 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 north coast of the SCS. The reconstructions always agree with observations. When pCO_2 observations cover a sufficiently large area, the reconstructed fields successfully display a pattern of relatively high pCO_2 in the mid-and-southern basin. The rate of sea surface pCO_2 increase in the SCS is 2.383 µatm per year based on the spatial
25 average of the reconstructed pCO_2 over the period of 2000–2017. All the data for this paper are openly and freely available at PANGAEA under the link https://doi.pangaea.de/10.1594/PANGAEA.921210 (Wang et al., 2020).

1 Introduction

Ocean plays an important role in absorbing atmospheric CO_2 and consequently helps slow down the Earth's 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
30 increasing rate from 1.0±0.5 Gt C yr^{-1} in the 1960s to 2.4±0.5 Gt C yr^{-1} during 2008–2017 (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; Landschutzer et al., 2014; Le Quere et al., 2010; Rodenbeck et al., 2015; Turi et al., 2014). Thus, it is necessary to improve the spatial-temporal resolution and accuracy of the data in the evaluation of oceanic carbon
35   uptake capacity to better understand the global carbon cycle and to better project the future climate.

The sea–air CO_2 flux helps quantify the oceanic carbon uptake capacity and is primarily determined by the difference in the atmospheric and sea surface partial pressure of CO_2 (pCO_2). Although the measurement records of sea surface pCO_2 have been increasing to 14.7 million, are available in almost all ocean basins in 2014, and continue to receive more data for compilation (Rodenbeck et al., 2015; Sheu et al., 2010), the observations are still severely limited in the spatial and temporal pCO_2 field of
40   the global ocean surface, especially in marginal seas. 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 two 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).
45   These studies usually present their pCO_2 fields in a monthly time scale and at a $1° \times 1°$ 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 have 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
50   SCS is mostly blank (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
55   in situ observations in the SCS with a grid resolution of $0.5° \times 0.5°$, 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 data to calculate empirical orthogonal functions (EOFs) for spatial patterns of pCO_2. The remote sensing data are 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 data to compute the EOFs. These EOFs form an
60   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 $0.5° \times 0.5°$ over the SCS.

Section 2 will describe the datasets and methods, Section 3 includes results and discussion, and the conclusions are in Section
65   4.

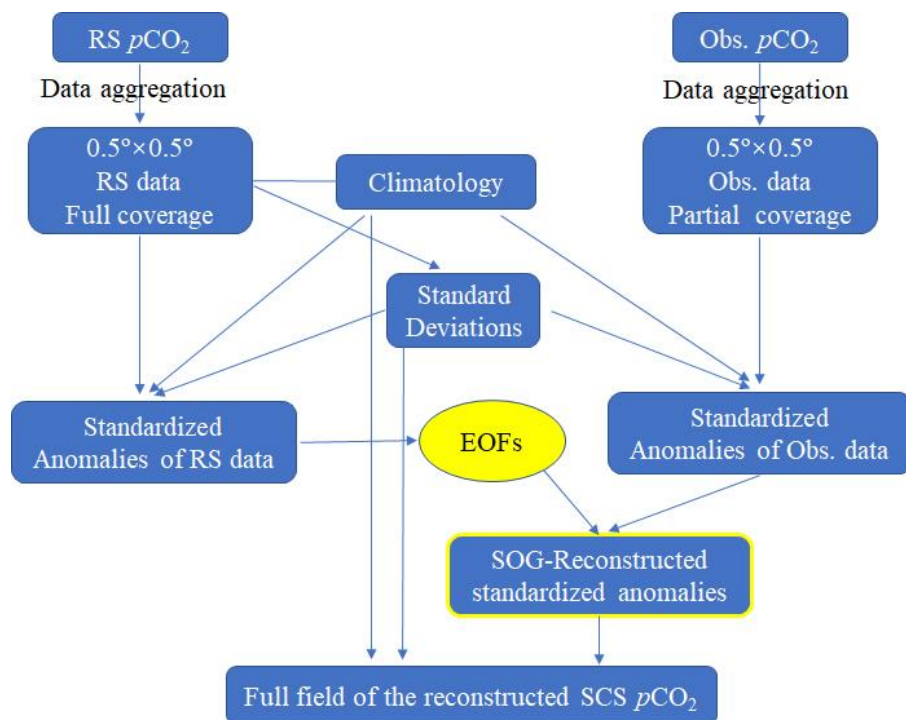



Figure 1: Reconstruction procedure of the sea surface pCO_2 in the SCS. Here, RS pCO_2 means the original remote-sensing derived pCO_2, Obs. pCO_2 represents the original observed in situ pCO_2, RS data are the grid-aggregated remote-sensing derived pCO_2, and Obs. data are the grid-aggregated observed pCO_2.

2 Datasets and methods

2.1 Observed data in the SCS

In the SCS, the underway sea surface pCO_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 observed summer pCO_2 data from 2000 to 2017 are compiled in this study and shown in Table 1. 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; Zhai et al., 2005a). Thus, we have observed pCO_2 data for 13 summers during 2000–2017. The blue cruise tracks of Fig. 2 indicate all the sea surface pCO_2 observations in the 13 summers. The tracks indicate that these data are distributed mainly on the northern shelf and slope, and in the northern-and-mid basin of the SCS. The coverage of an individual summer is only a subset of the blue tracks. See Fig. 3 for the subset of each year. These observational data were aggregated onto 0.5°×0.5° grid boxes in the (5–25º N, 109–122º E) region that covers most of the SCS. The aggregation used a simple space-time average of the data in a grid box. The aggregated data for 13 summers are shown in Fig. 3. The aggregated pCO_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

pCO$_2$ data cover a spatial range of 12° in latitude and 13° in longitude, 231 grid boxes with data that cover 22 % of SCS. The

85   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, 77 grid boxes with data that cover 7 % of SCS. The data are in the range of 279–440 µatm. The summer of 2000

has only 5 grid boxes (covering 0.5 % SCS) with data in the range of 400–425 µatm. The lowest observational pCO$_2$ 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 pCO$_2$ values occur mainly in the basin, which are often

90   higher than the atmospheric pCO$_2$ (Li et al., 2020; Zhai et al., 2013). The high pCO$_2$ values off the northeastern coast of 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 pCO$_2$ 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 northeast shelf and off the eastern coast of

95   the 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 pCO$_2$ field.

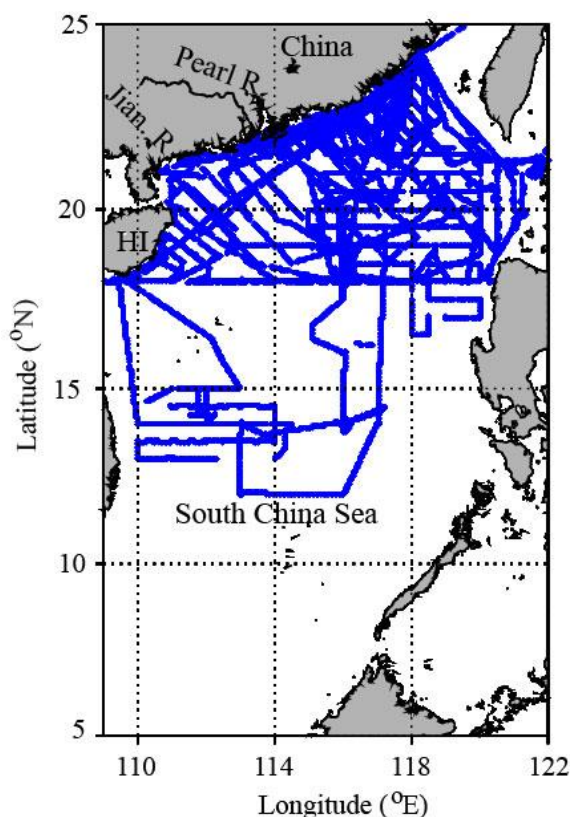



Figure 2: Cruise tracks of sea surface pCO$_2$ observations in the SCS in the summers from 2000-2017. HI represents Hainan Island,
100   **Jian. R. is the Jianjiang River, and Pearl R. represents the Pearl River.**

Table 1. Underway sea surface pCO₂ data in summer in the SCS compiled in this study.

Year	Cruise time	Data source
2000	July 2000	Zhai et al., 2005a
2001	June 2001	Zhai et al., 2005a
2004	July–Aug. 2004	Zhai et al., 2013; This study
2005	July 2005	This study
2006	June 2006	This study
2007	July–Aug. 2007	Zhai et al., 2013
2008	July–Aug. 2008	Li et al., 2020
2009	Aug. 2009	Li et al., 2020
2012	July–Aug. 2012	Li et al., 2020
2014	June 2014	Li et al., 2020
2015	July–Aug. 2015	Li et al., 2020
2016	June 2016	Li et al., 2020
2017	June 2017	Li et al., 2020

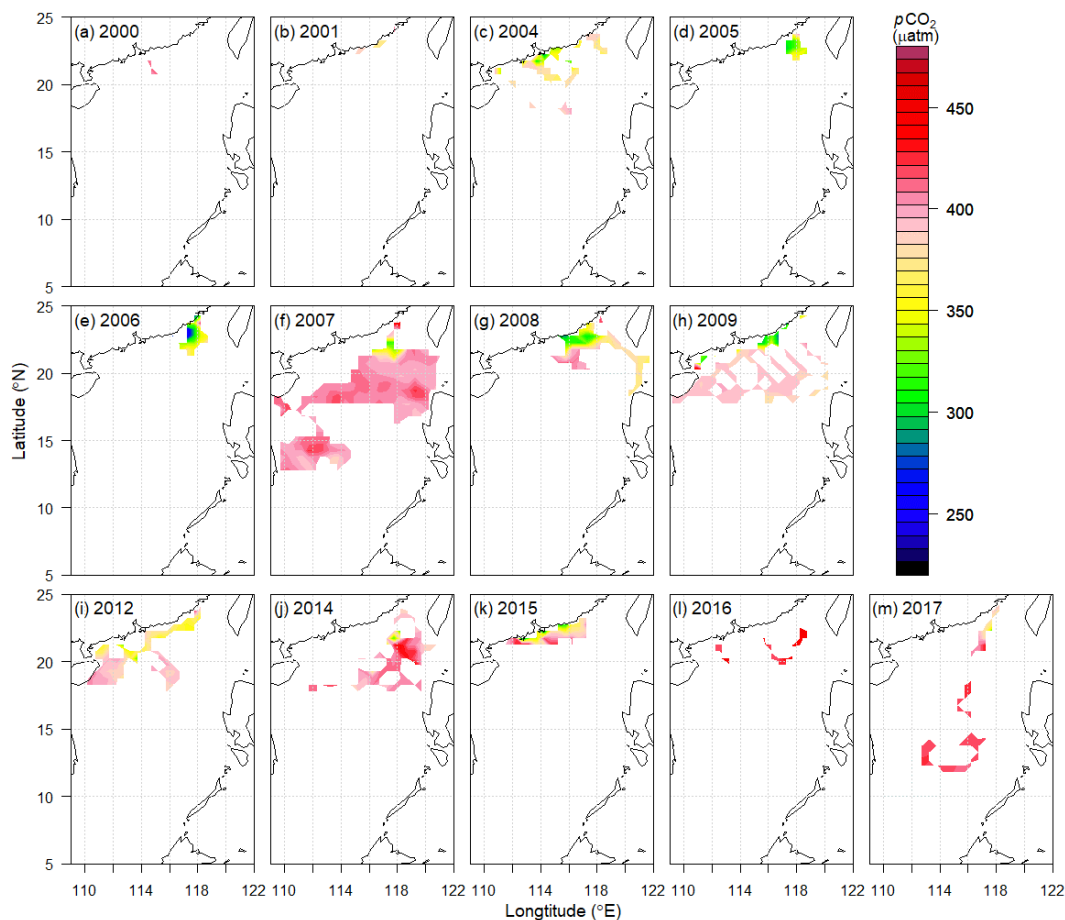



Figure 3: The aggregated in situ observational pCO₂ data in 0.5º×0.5º grid boxes in the SCS in the 13 summers during 2000–2017.

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

The satellite remote-sensing derived sea surface pCO$_2$ in the SCS were estimated for the years of 2000–2014 using a semi-analytical algorithm developed by Bai et al. (2015a). The algorithm treats pCO$_2$ as a function of major controlling factors derived from multiple satellite remote sensing products, including sea surface temperature, chlorophyll a, and salinity. The spatial resolution of the remote-sensing derived pCO$_2$ data is 1´×1´. These data were aggregated into 0.5°×0.5° 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 data full coverage in Fig. 1). However, variations shown by these remote-sensing derived pCO$_2$ are much less than those shown by the observed pCO$_2$ data. 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 the cruise samplings.

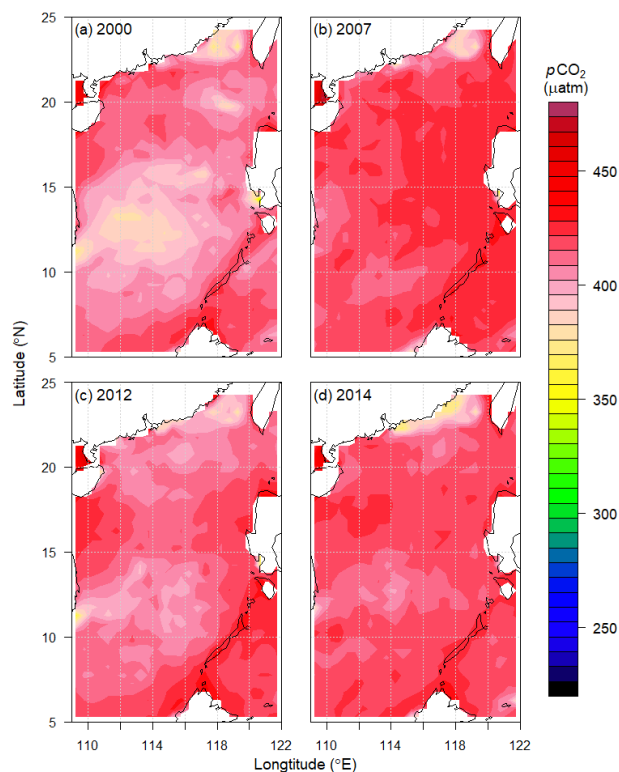



Figure 4: Remote-sensing derived sea surface pCO$_2$ in summer in selected years from 2000 to 2014.

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°×0.5° resolution covered from 5° to 25° N and from 109° to 122° E with 1040 grid boxes in total. The land area data were marked with NaN. The data were arranged in a 1040×15 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)

125 method was applied to the standardized anomalies in the space–time matrix to compute the EOFs. The results are shown in Section 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.

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

$$P(x,t) = \beta_0(t) + \sum_{m \in \mathcal{M}} \beta_m(t) E_m(x)/\sqrt{a(x)} + e(x,t), \tag{1}$$

Here, $P(x,t)$ 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 mth EOF $E_m(x)$, the least square estimator of $\beta_m(t)$ is denoted by $b_m(t)$,

135 $a(x) = \cos(\phi_x)$ is the area-factor, ϕ_x is the centroid's latitude, expressed in radian, of the grid box x, and $e(x,t)$ is the regression error. The error is assumed to be normally distributed with zero mean and has an independent error variance

$$\varepsilon^2(x,t) = \langle e^2(x,t) \rangle, \tag{2}$$

where $\langle \cdot \rangle$ denotes the mathematical operation of expected value. The explanatory variables in the above multivariate regression are $E_m(x)$, computed from the area-weighted standardized anomalies of the remote-sensing derived data. The anomalies were

140 written as an 897×15 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. $\mathcal{M}$ is the set of EOFs selected for our regression reconstruction.

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

145 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 $\beta_0(t)$ and coefficients $\beta_m(t)$ of the regression. With the estimates $b_0(t)$ and $b_m(t)$, $m \in \mathcal{M}$, the reconstructed standardized anomalies are expressed as

$$\hat{P}(x,t) = b_0(t) + \sum_{m \in \mathcal{M}} b_m(t) E_m(x)/\sqrt{a(x)}, \tag{3}$$

where x runs through the entire 893 grid boxes over our study region in the SCS. These anomalies were converted to the full

150 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 Section 3.

Many computer software packages are available to compute the EOFs using SVD and to compute multilinear regressions. 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. The R computer codes and their required files for this paper are freely available at
155 https://github.com/Hqin2019/pCO2-reconstruction

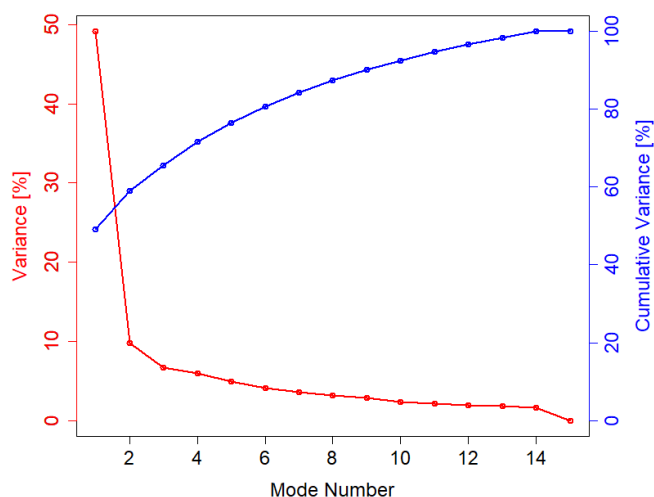
SOG usually uses the first few EOFs, or the first *M* EOFs that account for more than 80 % of the total variance, or determined by response data via a correlation test (Smith et al., 1998). The current paper used eight EOFs that explain 87 % of the total variances (Fig. 5). However, the year 2000 was an exception and used only four EOFs, because the year has only five grid boxes with the observed data.

160

Figure 5: The percentage variances and cumulative variances based on the summer remote-sensing derived pCO$_2$ data for the period of 2000–2014.

3 Results and discussion

3.1 EOFs and PCs

165 EOF1 demonstrates the mode of average level of pCO$_2$ with lower or higher values near the coastal regions of China mainland (Fig. 6a). 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 pCO$_2$ data support this mode well. EOF2 shows a north-south dipole (Fig. 6b), 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

170 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. 6c), 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 northeast China coast and off eastern Hainan Island (Gan et al., 2015; Jing et al., 2015).

The PCs are temporal stamp of the occurrence of the spatial patterns. PC1 basically shows the temporal trend (Fig. 6d). It has

175 been concluded that surface SCS pCO$_2$ 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 with 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.

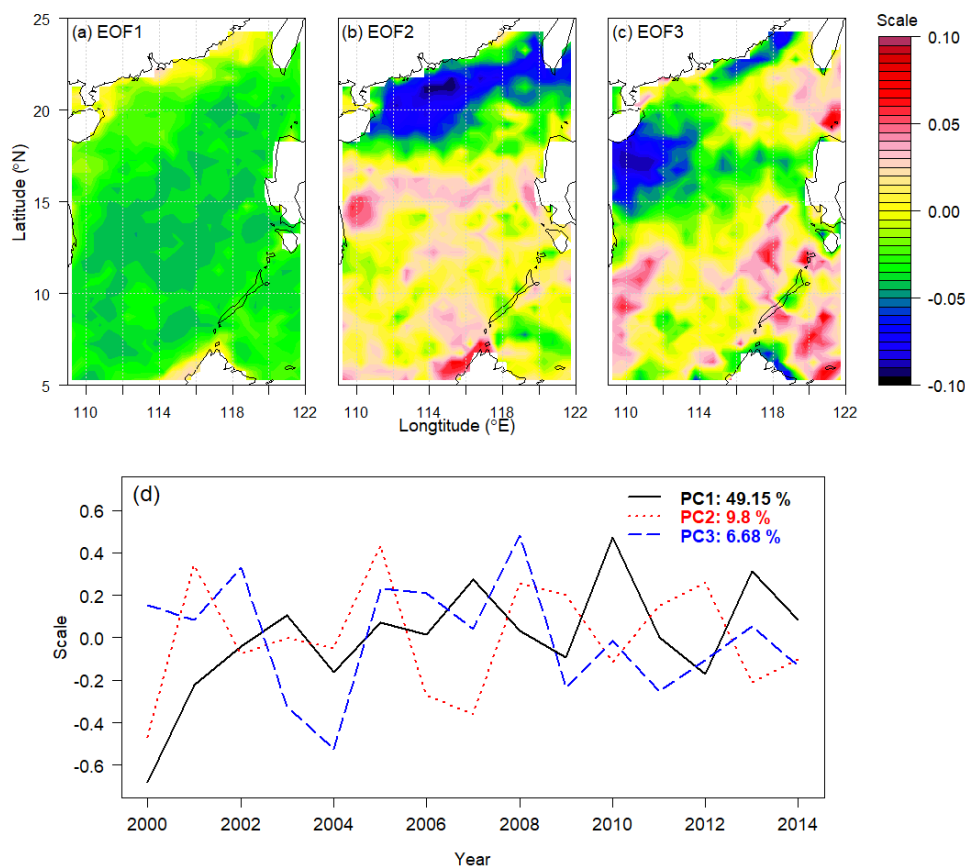



Figure 6: EOFs and PCs of the remote-sensing derived pCO$_2$ data. (a)–(c) EOFs, and (d) PCs.

3.2 Reconstruction results in the SCS

Figure 7 shows that the reconstructed pCO$_2$ fields in the SCS have successfully displayed the spatial patterns of the observed pCO$_2$ 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 basin.

The reconstructions have taken the advantages of both the in situ 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, 2014–2017 nicely demonstrate the spatial pCO_2 patterns (Figs. 7c, f, h, 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 appears

195   reasonable (Fig. 7b).

Figure 7: Reconstructed summer pCO_2 fields for the years of 2000–2017 in the SCS.

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 in poor quality (Fig. 7a) since the relatively low values in the mid SCS basin may not be realistic. These

200   poorly reconstructed data may be due to the poor spatial coverage of the in situ pCO_2 data in the summer of 2000, which had only 5 grid boxes with data (Fig. 3a). These 5 boxes are all located together and cover only 0.5 % of the SCS. Similarly, the reconstructed pCO_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. 7d, e, g). These small values are unlikely since the sea surface pCO_2 in the basin is generally higher than the atmospheric pCO_2

205 (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 gradient of in situ data. These gradients, 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

210 necessarily be large. However, large spatial gradients of the situ data can distort the reconstruction and lower the quality of reconstruction, because the linear regression method is not robust.

As an application of our reconstruction and a validation, we examine the temporal trend of sea surface pCO_2 over the SCS. The rate based on the linear temporal trend of the spatial average of the reconstructed sea surface pCO_2 over the SCS from 2000 to 2017 is 2.383 μatm per year (See Fig. 8a). It is lower than the rate of increase with time in the fugacity of CO_2

215 calculated for the mixed layer for the period of 1999-2003 at the South East Asia Time-Series Station (Station SEATS) (18º N, 116º E) in the northern basin of the SCS, which is 4 μatm per year (Tseng et al., 2007). This makes sense since our rate is a spatial average in summer. When compared with the summer rate at the Hawaii Ocean Time-Series Station (Station HOT) (22º45´ N, 158º W) in the North Pacific, which is 1.976 μatm per year over 2000–2017 (Dore et al., 2009) (See Fig. 8b), our rate is about 0.4 μatm per year higher. This is reasonable for a marginal sea where a higher rate of increase in pCO_2 would be

220 expected.

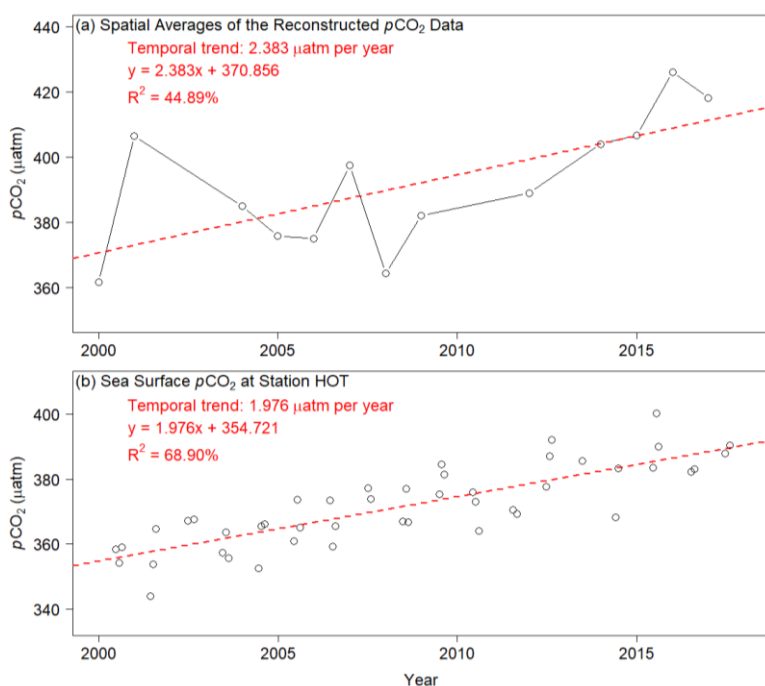



Figure 8: (a) Time series and linear trend of the spatial averages of the reconstructed summer pCO_2 data in the period of 2000–2017; (b) Summer sea surface pCO_2 at Station HOT in 2000–2017 adapted from Dore et al. (2009).

3.3 Outliers of the observed data in the reconstruction

225 The SOG method is basically a linear regression method, which is known to be sensitive to the outliers of the response 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 pCO_2 values in the reconstructed data. Although we cannot conclude that the outliers of 3σ 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.

230 Table 2 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 pCO_2 values may result from the Pearl River plume where the observed pCO_2 can be very low. These very low values, such as at least 3σ away from the mean, may cause a very large gradient in the observed pCO_2. Our reconstruction has excluded these extremely low values influenced by the river plumes. Our reconstructed data may therefore overestimate the pCO_2 values in the Pearl River

235 Estuary and its nearby region.

Table 2. Outliers excluded from the SOG reconstruction.

Year	Grid ID	Latitude (N)	Longitude (E)	pCO_2 (µatm)
2006	926	22.75°	116.75°	208
	952	23.25°	116.75°	197
2009	896	22.25°	114.75°	212
	923	22.75°	115.25°	217
2012	836	21.25°	110.75°	248
2014	873	21.75°	116.25°	191
	874	21.75°	116.75°	219
2016	841	21.25°	113.25°	265
	842	21.25°	113.75°	272
	868	21.75°	113.75°	239
	869	21.75°	114.25°	205
	870	21.75°	114.75°	216
	896	22.25°	114.75°	210
	897	22.25°	115.25°	274

4 Data availability

The gridded in situ data, the remote-sensing derived sea surface pCO_2 data, and the reconstruction result data are openly and

240 freely available at PANGAEA under the link https://doi.pangaea.de/10.1594/PANGAEA.921210 (Wang et al., 2020).

5 Conclusions

This study has demonstrated the feasibility of using the SOG method to reconstruct the sea surface pCO_2 data into regular grid boxes. We compiled the observed and remote-sensing derived sea surface pCO_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

245 multilinear regression was applied to reconstruct the space–time complete pCO_2 field in the SCS. The method took the EOFs calculated from the remote-sensing derived pCO_2 as the explanatory variables and treated the observed pCO_2 as the response variable. The EOFs reflect reasonably well the general spatial pattern of the sea surface pCO_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 pCO_2 fields are, in general, consistent with the patterns of the

250 observed pCO_2 and demonstrate relatively low values along the north coast affected by the Pearl River plume and consistently high values in the ocean basin of the SCS. These reconstructed pCO_2 fields provide full spatial coverage of the sea surface pCO_2 of the SCS in summer over a temporal scale of almost two decades and therefore fill the long-lasting blanks in the global sea surface pCO_2 mapping. 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.

255 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.

Author contribution

Minhan Dai conceptualized and directed the field program of the in situ observations. Baoshan Chen and Xianghui Guo
260 participated in the in situ data collection. Yan Bai provided the remote-sensing derived data. Guizhi Wang, Yao Chen and Samuel S. P. Shen developed the reconstruction method, wrote the Matlab and R codes, analyzed the data, and plotted the figures. Huan Qin developed the data repository, and revised and tested the R codes. Guizhi Wang and Samuel S. P. Shen wrote the manuscript. All the authors contributed to the original writing, editing and revisions of the manuscript.

Competing interests

265 The authors declare that they have no conflict of interest.

Acknowledgements

The work by Guizhi Wang, Yan Bai, Xianghui Guo, and Minhan Dai was supported by a grant from the Ministry of Science and Technology of China (2015CB954001). The observed pCO_2 data in 2004, 2005 and 2006 were collected under the support

of the National Natural Science Foundation of China (40521003). Acknowledgement is for the data support from "National
270   Earth System Science Data Sharing Infrastructure, National Science & Technology Infrastructure of China
(http://www.geodata.cn)".

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 pCO2 in river-dominated coastal oceans: a case study from the East China Sea, J. Geophys. Res. 120,
275   2331–2349, doi: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,
doi:10.1016/j.seares.2014.10.003, 2015b.

Boutin, J., Etcheto, J., Dandonneau, Y., Bakker, D. C. E., Feely, R. A., Inoue, H. Y., Ishii, M., Ling, R. D., Nightingale, P. D.,
280   Metzl, N., and Wanninkhof, R.: Satellite sea surface temperature: a powerful tool for interpreting in situ pCO_2 measurements
in the equatorial Pacific Ocean, Tellus B., 51(2), 490–508, doi:10.1034/j.1600-0889.1999.00025.x, 1999.

Chou, W. C., Sheu, D. D. D., Chen, C. T. A., Wang, S. L., and Tseng, C. M.: Seasonal variability of carbon chemistry at the
SEATS time-series site, northern South China Sea between 2002 and 2003, Terr. Atmos. Ocean. Sci., 16(2), 445–465,
doi:10.3319/TAO.2005.16.2.445(O), 2005.

285   Denvil-Sommer, A., Gehlen, M., Vrac, M., and Mejia, C.: LSCE-FFNN-v1: a two-step neural network model for the
reconstruction of surface ocean pCO_2 over the global ocean, Geosci. Model Dev., 12(5), 2091–2105, doi:10.5194/gmd-12-
2091-2019, 2019.

DeVries, T., Le Quere, C., Andrews, O., Berthet, S., Hauck, J., Ilyina, T., Landschutzer, P., Lenton, A., Lima, I. D., Nowicki,
M., Schwinger, J., and Seferian, R.: Decadal trends in the ocean carbon sink, P. Natl. A. Sci. USA, 116(24), 11646–11651,
290   doi:10.1073/pnas.1900371116, 2019.

Doney, S. C., Tilbrook, B., Roy, S., Metzl, N., Le Quere, C., Hood, M., Feely, R. A., and Bakker, D.: Surface-ocean CO_2
variability and vulnerability, Deep–Sea Res. Pt. II, 56(8–10), 504–511, doi:10.1016/j.dsr2.2008.12.016, 2009.

Dore, J. E., Lukas, R., Sadler, D. W., Church, M. J., and Karl, D. M.: Physical and biogeochemical modulation of ocean
acidification in the central North Pacific, P. Natl. A. Sci. USA, 106, 12235–12240, doi:10.1073/pnas.0906044106, 2009.

295   Fay, A. R., and McKinley, G. A.: Global trends in surface ocean pCO_2 from in situ data, Global Biogeochem. Cy., 27(2), 541–
557, doi:10.1002/gbc.20051, 2013.

Gan, J., Wang, J., Liang, L., Li, L., and Guo, X.: A modeling study of the formation, maintenance, and relaxation of upwelling
circulation on the Northeastern South China Sea shelf, Deep–Sea Res. Pt. II, 117, 41–52, doi:10.1016/j.dsr2.2013.12.009, 2015.

Gao, J., Shen, S.S.P., Yao, T., Tafolla, N., Risi, C., and He, Y.: Reconstruction of precipitation δ18O over the Tibetan Plateau
300   since 1910, J. Geophys. Res.–Atmos., 120, 4878–4888, doi:10.1002/2015JD023233, 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(3), 1993–2006, doi:10.1002/2014jc010601, 2015.

Jo, Y.-H., Dai, M., Zhai, W., Yan, X.-H., and Shang, S.: On the variations of sea surface pCO_2 in the northern South China Sea: A remote sensing based neural network approach, J. Geophys. Res., 117, C08022, doi:10.1029/2011JC007745, 2012.

Lammlein, 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, doi:10.1142/S2424922X18500080, 2018.

Landschutzer, P., Gruber, N., Bakker, D. C. E., and Schuster, U.: Recent variability of the global ocean carbon sink, Global Biogeochem. Cy., 28(9), 927–949, doi:10.1002/2014gb004853, 2014.

Laruelle, G. G., Landschützer, P., Gruber, N., Tison, J.-L., Delille, B., and Regnier, P.: Global high-resolution monthly climatology for the coastal ocean derived from neural network interpolation, Biogeosciences, 14(19), 4545–4561, doi:10.5194/bg-14-4545-2017, 2017.

Le Quere, C., Andrew, R. M., Friedlingstein, P., Sitch, S., Hauck, J., Pongratz, J., et al.: Global Carbon Budget 2018, Earth Syst. Sci. Data, 10(4), 2141–2194, doi:10.5194/essd-10-2141-2018, 2018a.

Le Quéré, C., Andrew, R. M., Friedlingstein, P., Sitch, S., Pongratz, J., Manning, A. C., et al.: Global Carbon Budget 2017, Earth Syst. Sci. Data, 10(1), 405–448, doi:10.5194/essd-10-405-2018, 2018b.

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 CO_2, Global Biogeochemical Cycles, 24, GB4007, doi:10.1029/2009GB003599, 2010.

Lefevre, N., and Taylor, A.: Estimating pCO_2 from sea surface temperatures in the Atlantic gyres, Deep–Sea Res. Pt. I, 49(3), 539–554, doi:10.1016/s0967-0637(01)00064-4, 2002.

Li, Q., Guo, X., Zhai, W., Xu, Y., and Dai, M.: Partial pressure of CO_2 and air-sea CO_2 fluxes in the South China Sea: synthesis of an 18-year dataset, Prog. Oceanogr., 182, 102272, doi:10.1016/j.pocean.2020.102272, 2020.

Ono, T., Saino, T., Kurita, N., and Sasaki, K.: Basin-scale extrapolation of shipboard pCO_2 data by using satellite SST and Ch1a, Int. J. Remote Sens., 25(19), 3803–3815, doi:10.1080/01431160310001657515, 2004.

Rodenbeck, C., Bakker, D. C. E., Gruber, N., Iida, Y., Jacobson, A. R., Jones, S., Landschutzer, P., Metzl, N., Nakaoka, S., Olsen, A., Park, G. H., Peylin, P., Rodgers, K. B., Sasse, T. P., Schuster, U., Shutler, J. D., Valsala, V., Wanninkhof, R., and Zeng, J.: Data-based estimates of the ocean carbon sink variability first - results of the Surface Ocean pCO_2 Mapping intercomparison (SOCOM), Biogeosciences, 12(23), 7251–7278, doi:10.5194/bg-12-7251-2015, 2015.

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, doi:10.1175/JTECH-D-16-0133.1, 2017.

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, doi:10.1175/JAS-D-13-0301.1, 2014.

Sheu, D. D., Chou, W. C., Wei, C. L., Hou, W. P., Wong, G. T. F., and Hsu, C. W.: Influence of El Nino the sea-to-air CO_2 flux

335   at the SEATS time-series site, northern South China Sea, J. Geophys. Res.–Oceans, 115, C006013, doi:10.1029/2009JC006013, 2010.

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. Climate, 11, 1717–1729, doi: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., et al.: Climatological mean and

340   decadal change in surface ocean pCO$_2$, and net sea–air CO$_2$ flux over the global ocean,. Deep–Sea Res. Pt. II, 56(8–10), 554–577, doi: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(14–15), 1448–1468, doi:10.1016/j.dsr2.2007.05.003, 2007.

Turi, G., Lachkar, Z., and Gruber, N.: Spatiotemporal variability and drivers of and air–sea CO$_2$ fluxes in the California Current

345   System: an eddy-resolving modeling study, Biogeosciences, 11(3), 671–690, doi: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, doi:10.1594/PANGAEA.921210, 2020.

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 fluxes in the northern South China Sea in spring, summer and autumn, Mar. Chem., 96(1–2), 87–97,

350   doi: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 CO$_2$ and its maintaining mechanism in a subtropical estuary: the Pearl River estuary, China, Mar. Chem., 93(1), 21–32, doi: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 CO$_2$ fluxes in the largest tropical marginal sea (South China Sea) based on multiple-year underway

355   measurements, Biogeosciences, 10(11), 7775–7791, doi:10.5194/bg-10-7775-2013, 2013.