Interactive comment on "Feasibility of reconstructing the basin–scale sea surface partial pressure of carbon dioxide from sparse in situ observations over the South China Sea" by Guizhi Wang et al.

Our response is in blue with line numbers listed where revisions are made.

Anonymous Referee #1
Received and published: 25 September 2020

The subject is important but the manuscript suffers from two major flaws. Fortunately, both could be amended.

My major concern is the reliability of the result. The manuscript gave a rate of pCO2 increase of 2.38 uatm/yr, which is very high. Unfortunately, no uncertainty was given. Judged by the large scatter of the data (Fig. 8a) the standard deviation of the rate must be very large. Note other studies, for instance, that of Lui et al. (2020, Transient carbonate chemistry in the expanded Kuroshio region, in Changing Asia Pacific Marginal Seas, pp 307-320) gave a much lower increasing rate of only 0.8 uatm/yr at the SEATS station. I fully recognize that different sampling locations, sampling periods, and sampling frequency could contribute to large differences in the results. Yet, exactly because of this the result must be qualified and compared with other studies. In addition, there ought to be other independent checks of the pCO2 data generated by the satellite chlorophyll data. There is an abundance of alkalinity, DIC, and pH data in various parts of the South China Sea, especially at SEATS. It would be relatively easy to generate pCO2 from these data to check model-derived pCO2.

Response: Thank you for your comments and input.

First, we have quantified the uncertainty of our rate of increase as 2.4±0.8 μtam/yr, where 0.8 μtam/yr is the standard error of the rate, in the revision (Line 287 and Figure 11).

Second, the rate in Lui et al. (2020) of 0.8 μatm/yr is for the period of 1998-2006 and was calculated based on the data of spring, summer, fall, and winter. The reason for their much slower rate might be due to the peak pCO_2 value approximately 405 μatm in 1999, as shown in Fig. 16.7 in their paper. This peak value at an earlier time forced a lower rate. We re-calculated the rate using their summer-only data from the year of 2000, which is the beginning year of our data, to the year of 2006, we obtained a rate of 2.5±1.0 μatm/yr, which is almost the same as our rate of 2.4±0.8 μatm/yr. We have added their rate, as well as the reasons for the differences between their rate and ours, and the recalculated rate in the revision (Line 288-294).

Third, regarding the question about independent checks for the remote-sensing derived pCO_2 estimates, we have compared our observed underway pCO_2 with the remote-sensing derived estimates as a quality check. In general, most of the remote-sensing derived pCO_2 overestimate the sea surface pCO_2 by no more than 50 μatm (Figure 5 in the revision). The root-mean-square-errors (RMSE) between the remote-sensing derived pCO_2 and the observed underway pCO_2 fall in the range of 12.8-89.0 μatm with a median of 33.8 μatm (Table 2 in the revision). The RMSE values are high in the years when the underway data covered only the shelf regions. These texts (Line

135-142), one Figure (Figure 5), and one table (Table 2) are added in the revision.

Fourth, we have used the pCO_2 data calculated from alkalinity and DIC data observed at Station SEATS from Lui et al. (2020) and our own database to check the reconstructed pCO_2 (Figure 10 in the revision). The difference between the reconstructed pCO_2 and the observed data at Station SEATS ranges from -7 to 10 μatm with the relative difference from -1.5 to 2.1 %. This comparison indicates that our reconstruction is reliable. We have added these texts (Line 273-276) and one figure (Figure 10) in the revision.

We have included a description of the calculation and the observed data used in the calculation in the revised data section of the paper (Line 119-127).

My second major concern is the coverage of the data. The manuscript covers data from only 13 years, and in most years the region covered was very small. In fact, none of the cruise tracks covers the southern South China Sea. It seems that the authors used only their own data but why not include other people's data as well? For instance, the open-access SOCAT database covers tracks in the southern South China Sea. One minor issue is that the title does not reflect correctly that only summer data were covered.

Response: The word 'summer' have been added before 'basin-scale' in the title in the revision. Our method allows us to include any group's data if they are available.

As for the data temporal coverage, to our knowledge the underway summer pCO_2 data are available by this date only for 13 years. A pCO_2 dataset calculated from observed total alkalinity and DIC is present for June 2010 in the northern South China sea in Guo et al. (2015). We have considered this dataset. However, we are still examining an outlier of the dataset. When the outlier and consistency issues of this dataset are resolved, we may include their data in our future reconstruction. Of course, we will also include any group's data when they become available. Observed data from other groups are published either at one or two buoys or stations, e.g., Lui et al. (2020) and Yu et al. (2020) or in other seasons, e.g., Xu et al. (2016). However, the data at one or two buoys or stations cannot provide a spatial coverage needed for the reconstruction. Although the SOCAT database has tracks in the southern South China Sea, it does not have summer data for the region. We plan to make reconstruction in other seasons in the future and we will then include these data.

As for the spatial coverage, the poor coverage of observed data in the South China Sea, especially in the southern South China Sea, is exactly the main reason for us to make the reconstruction. The purpose of our work is to reconstruct a complete pCO_2 field in the South China Sea.

References

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.

Guo, X. H., and Wong, G. T. F.: Carbonate chemistry in the Northern South China Sea Shelf-sea in

June 2010, Deep-Sea Res. Part II, 117, 119-130, doi:10.1016/j.dsr2.2015.02.024, 2015.
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: Chen, C.-T., and Guo, X. (eds) Changing Asia-Pacific Marginal Seas. Atmosphere, Earth, Ocean & Space. Springer, Singapore, doi:10.1007/978-981-15-4886-4_16, 2020.
Xu, X., Yu, P., Cai, X. Pan, J., Hu, J., and Zhang, H.: Distributions of the partial pressure of carbon dioxide and sea-air CO_2 flux in the western South China Sea in autumn, J. Tropical Oceanogr., 35(3), 55-64, doi:10.11978/2015035, 2016.
Yu, P., Wang, Z. A., Churchill, J., Zheng, M., Pan, J., Bai, Y., and Liang, C.: Effects of typhoons on surface seawater pCO_2 and air-sea CO_2 fluxes in the northern South China Sea, J. Geophys. Res., 125, doi:10.1029/2020JC016258, 2020.

Anonymous Referee #2
Received and published: 7 November 2020

The authors have used a remote-sensing-based pCO_2 field to derive EOFs, fit those EOFs to in situ pCO_2 observations collected over almost 2 decades, and then used the scaled EOFs to estimate the full surface pCO_2 record in the South China Sea.
It is a very interesting paper and many parts of it are clearly communicated, but it is also incomplete. The method validation and uncertainty quantification are missing. These should be an entire section of the paper and not just an added sentence or two, so the paper should be returned to the authors for major revisions.
Response: Thank you for bringing up the issues of validation and uncertainty quantification. We first addressed the cross validation issue. The regression-based reconstruction is often valid when outliers are not present. Our reconstruction follows this approach. Nonetheless, we agree with you and have conducted a cross-validation check of our reconstruction. The maximum RMSE of our cross-validation is 5.2 μatm, which occurred in 2006 when there were only 25 grid boxes with in situ pCO_2 data and which had the largest spatial standard deviation, 49.4 μatm, among the 13 years under consideration. This accuracy is very good compared to the spatial standard deviation of the in situ data in the same year. The temporal 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 cross-validation result. We have included our cross-validation method and result in the revised paper (Line 247-258, Table 2).

Second, we addressed the uncertainty issue. We have made grid-by-grid comparisons between the observed pCO_2 and reconstructed pCO_2 in two ways. One is comparison with observed underway data (Figure 9 and Table 2 in the revision) and the other is comparison with pCO_2 calculated from observed DIC and total alkalinity at Station SEATS (18° N, 116° E) (Figure 10 in the revision). The difference between the reconstructed data and the observed underway data mostly falls within the range from -30 to 30 μatm. 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) at least 29 %. The difference between the reconstructed data and the observed data at Station SEATS ranges from -7 to 10 μatm with the relative error within 2.1 %. Both comparisons have been provided as other ways of validation in the revised paper (Line 264-276, Figures 9 and 10, Table 2).

It is a bit unclear whether this paper is presenting new data along with a new method or just a new method. There are two cruises in 2005 and 2006 with a reference given as "this paper" and, if these data sets are truly being published for the first time in this paper, then the manuscript should highlight that there are new data in the abstract. This would raise the value of this paper if there are indeed new data being made available along with the analysis. I might have missed the text that explained this.

Response: This paper is presenting new data along with a new method. In the revised data the new data in 2004, 2005 and 2006 have been highlighted in the abstract (Line 18-19) and in the main text (Line 86-88). In addition, pCO_2 data from literature and calculated from our alkalinity and DIC at Station SEATS (18° N, 116° E) have been presented to compare with the reconstructed pCO_2 as a way of validation in the revised paper (Line 119-127 for the data description and Line 273-276 for comparison with the reconstructed estimates).

The part of the paper that deals with the pCO_2 mapping approach is not yet complete because the authors have not assessed the uncertainties of their approach. I recommend one or two exercises. First, the approach should be repeated after removing some of the in situ pCO_2 measurements. Each cruise should be removed, one at a time. After removing a cruise, the analysis should be conducted using only the remaining data. Then the withheld cruise can be used to quantify how good of a job the mapping procedure does at reconstructing the withheld cruise. This should be repeated for every cruise in the dataset to get bulk statistics. If there is only one cruise worth of data in each year, then (I believe this reconstruction wouldn't work and instead) large swaths of latitude/longitude should be removed from the cruises and the remaining data should be used to reconstruct the missing data. This will allow the errors in the approach to be quantified. Second, if a model is available for the South China Sea that has pCO_2, then the model can also have the Bai et al. 2015 approach applied, be subsampled where the cruise measurements are, and be analyzed in the same way proposed here. This will reveal both the point-by-point reconstruction errors and allow the uncertainties for the overall pCO_2 average estimate, for example, to be quantified. Currently, the validation is left as an unsupported statement that the results look about right, which is insufficient for publication of a paper describing a quantitative method.

Response: Again, you have suggested a cross-validation procedure. As aforementioned, we have 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 is done for every grid box with in situ data for each year. The final cross-validation result is output as RMSE. The maximum RMSE is 5.2 μatm, which occurred in 2006, and the minimum is 2.4 μatm, which occurred in 2017. The year 2017 has 77 in situ data grid boxes. The spatial standard deviation of the data in 2017 is 17.6 μatm. Compared to the 2006 data described earlier, 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. We have included our cross-validation method and result in the revised paper (Line 247-258, Table 2).

Your second suggestion can be mathematically proven, because the cross-validation RMSE of reconstruction from the sub-sample of the Bai et al. (2015) complete data is only the truncation error, which is equal to zero or very close to be zero. The reason is that the EOFs computed from Bai et al. (2015) data form a complete basis for the same data. Thus, the original data field can be exactly represented by a linear span of the EOFs.

There are other smaller problems that should also be addressed if a revised version of the paper is submitted:

1. The model should not be used in any region where there is no fitting data. This includes most of the South China Sea south of ~12.5 N.

Response: This is exactly the point that shows the power of the spectral optimal gridding (SOG) method using EOFs, in contrast to the traditional optimal interpolation method, such as kriging and inverse distance weighting. EOFs are a diagonalized representation of the covariance of climate dynamics, and thus are providing a consistency constraint of the pCO_2 field. This allows us to use a small number of grid boxes with in situ data to interpolate and extrapolate to the entire region. Of course, the reconstruction accuracy is better when more observed data are available.

2. There should be an assessment of how good of a job the Bai et al. approach does at reproducing the in situ observations in a RMSE and bias sense. The estimates from this approach should be compared to the measurements from the data sets that are used here (and that Bai et al. did not use to design their routine). If the Bai et al. approach gives a different average pCO_2 than the in situ measurements, then the climatology created from the remote sensing product should not be used to generate the Standardized Anomalies of Obs. Data (as indicated in Figure 1). I believe an independent climatology would then be needed. Otherwise, a significant average bias would have to be compensated by a large average value for one or more EOFs. In a best-case scenario, that would be EOF 1, but if, for example, the observations were mostly found in the dark blue patch of Figure 6c then the resulting reconstruction would be problematic. It seems likely that a large average value of EOF 3, which is highly variable spatially, would then be fit to the measurements to fix a homogenous bias between the in situ and remote-sensing records. This is just one example of the kinds of problems that could occur if the Bai reconstruction doesn't adequately resolve the mean or the variability. If nothing else, the Bai et al. validation should be discussed in this paper.

It would also be interesting to see how this approach compares to competing approaches, for example a neural network that relates the in situ pCO_2 measurements to seawater property values that can be measured using remote sensing. This approach is more commonly used in global reconstructions. The Bai et al. approach is another clear competing approach.

Response: We have made a grid-by-grid assessment of the remote-sensing derived pCO_2 using the observed underway pCO_2 in the revision. In general, most of the remote-sensing derived pCO_2 overestimate the sea surface pCO_2 by no more than 50 μatm (Figure 5 in the revision). The root-mean-square-errors (RMSE) between the remote-sensing derived pCO_2 and the observed underway pCO_2 fall in the range of 12.8-89.0 μatm with a median of 33.8 μatm (Table 2 in the revision). The RMSE values are high in the years when the underway data covered only the shelf

regions. These texts (Line 135-142), one Figure (Figure 5), and one table (Table 2) are added in the revision. In addition, a comparison between the observed underway pCO_2 and reconstructed pCO_2 has been provided in the revision (Line 264-273, Figure 9). Both comparisons have been provided as validations in the revised paper. The RMSEs of the two comparisons have been shown in Table 2 in the revision. Furthermore, pCO_2 data from literature and calculated from our alkalinity and DIC at Station SEATS (18° N, 116° E) were compared with the reconstructed pCO_2 as another way of validation in the revised paper (Line 119-127 for the data description and Line 273-276 for comparison with the reconstructed estimates).

With regard to EOFs and independent climatology, the mathematical theory is like Fourier expansion of orthogonal polynomials, which can be sine functions, Legendre polynomials, and any set of eigenfunctions of a self-adjoint operator. Thus, EOFs form a complete basis for a data field although they may be different when using different anomalies computed from different climatologies and standard deviations. With different anomalies, variances may be re-distributed to different EOFs due to the different anomaly calculation methods. EOF rotation may help to reorganize certain variances into some specific EOF modes, and hence to provide an explanation of climate dynamics. However, this EOF rotation is not needed for the purpose of reconstruction as long as our EOFs form a complete basis. This completeness is guaranteed by the SVD algorithm for computing our EOFs here.

As for reconstruction using a neural network approach, the data produced by Jo et al. (2012) show an overall RMSE of 32.6-44.5 in summer of years 2004-2007, which is twice as much as the median RMSE between our reconstructed pCO_2 and the underway pCO_2. We have added this comparison in the revision (Line 270-273).

Specific comments:
15: consider deleting “capacity”
Response: The suggestion has been taken in the revised paper.

23: “The reconstructions always agree with observations.” Delete or quantify this statement. The agreement is not absolute.
Response: This statement has been changed in the revised paper to “The RMSE between the reconstructed summer pCO_2 and the observed underway pCO_2 is no larger than 31.7 μatm, in contrast to (a) the RMSE from 12.8–89.0 μatm between the remote-sensing derived pCO_2 and the underway data, and (b) the RMSE from 32.6–44.5 μatm between the neural network produced pCO_2 and the underway data. The difference between the reconstructed pCO_2 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.” (Line 29-34).

28: The ocean
Response: The suggestion has been taken in the revised paper (Line 37).

36: The sea-air CO_2 flux is the negative of the ocean carbon uptake, so this sentence is partially tautological.

Response: In the revised paper "helps quantify the oceanic carbon uptake capacity" has been deleted.

37: This sentence has several language errors. It also needs to be better-quantified or referenced. What is the decorrelation length scale for pCO_2 generally? How much of the ocean is constrained by those measurements alone without the newly proposed spatio-temporal mapping techniques? Mostly, I think a reference should be added to this sentence that supports this statement.
Response: The language errors of the sentence have been eliminated in the revised paper. The paper Bakker et al. (2020) has been added as a reference here and the content has been updated based on this paper in the revision (Line 47-49). This paper is the most recent published compilation of measured pCO_2.

57: References needed for RS pCO_2 here.
Response: The suggestion has been taken in the revision. Bai et al. (2015) has been added here in the revised paper (Line 68).

Figure 1. What is meant by "standard deviations"? Standard deviations of grid values, or deviations of values within each grid cell?
Response: It is the temporal standard deviation of the RS pCO_2 values on each grid box. In the revision, this explanation has been added (Line 80).

Figure 2. Consider changing this map to a 2 dimensional histogram showing number summers with measurements (probably with colored bins).
Response: This map has been changed in the revision showing the number of summers with underway data (Figure 2).

109: These estimates were... change "data" to "estimates" in this section since pCO_2 is not measured. In figure 1 as well.
Response: The changes have been made here (Line 133, 135) and in Figure 1 in the revised paper.

138: Where is this symbol used?
Response: The symbol $\langle\cdot\rangle$ is only used once for expected value in Eq. (2) in the paper. In the revised version, we have replaced the symbol $\langle\cdot\rangle$ by E[ ] (Line 171), which is more commonly used in statistics and science, while $\langle\cdot\rangle$ is a symbol commonly used in the field of theoretical physics.

141: how many EOFs were used? Say here.
Response: The suggestion has been taken in the revised paper (Line 175-176). Eight EOFs were used for every year except 2000, which had only four EOFs because the year had only five grid boxes with observed underway data.

182: This only shows the fields. One must compare this field to other figures to get an indication of how well the reconstruction performs. A plot showing differences between observed and reconstructed values is required.
Response: We have taken your suggestion and have produced two figures showing differences

between observed and reconstructed values in the revision (Figures 9 and 10).

187: It is not enough to say "we fit the model to the data, so it fits the data." Statistics of goodness-of-fit should be presented. Furthermore, demonstrating that the method works requires withholding several cruises worth of data from the training data set and then using those cruises to verify that the method reconstructs the withheld data. Statistics and plots are required to quantify how well the reconstruction does.
Response: Again, this is a cross-validation issue discussed earlier. We have included our results of cross-validation and uncertainty quantification in the revised paper (Line 246-283, Figures 9 and 10, Table 2) as shown in our response to the previous comments.

195: What is meant by reasonable?
Response: In the revised paper the RMSE, 7.3 μatm, has been provided here (Line 227), which indicates that the reconstruction appears reasonable.

205: it is unclear what is meant by "the large spatial gradient of in situ data."
Response: Here it means the large spatial variation of in situ data. The "gradient" is changed to "variation" in the revision (Line 239).

214: 2.383 is given to excessive precision. An attempt should be made to quantify the uncertainty and the data should be reported to the appropriate precision.
Response: The suggestion is taken. The rate has been given as 2.4±0.8 μatm/yr in the revised paper (Line 287, Figure 11).

229: why would a higher rate be expected in a marginal sea? I would argue that 2.4 uatm/year is completely within expectations of the atmospheric pCO_2 trend over this time period given the large uncertainties in this approach and the likely natural variability in surface ocean pCO_2.
Response: Considering the uncertainty in the rate is 0.8 μatm/yr, we agree with the reviewer that our rate is consistent with the trend shown at Station HOT in the Pacific. In the revised paper, this statement has been included (Line 296) and the sentence about "a higher rate be expected in a marginal sea" has been deleted.

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, 2331–2349, doi:10.1002/2014JC010632, 2015.

Bakker, D. C. E., Alin, S. R., Bates, N., Becker, M., Castaño-Primo, R., Cosca, C. E., Cronin, M., Kadono, K., Kozyr, A., Lauvset, S. K., Metzl, N., Munro, D. R., Nakaoka, S., O'Brien, K. M., Ólafsson, J., Olsen, A., Pfeil, B., Pierrot, D., Smith, K., Sutton, A. J., Takahashi, T., Tilbrook, B., Wanninkhof, R., Andersson, A., Atamanchuk, D., Benoit-Cattin, A., Bott, R., Burger, E. F., Cai, W.-J., Cantoni, C., Collins, A., Corredor, J. E., Cronin, M. F., Cross, J. N, Currie, K. I., De Carlo, E. H., DeGrandpre, M. D., Dietrich, C., Emerson, S., Enright, M. P., Evans, W., Feely, R. A., García-Ibáñez, M. I., Gkritzalis, T., Glockzin, M., Hales, B., Hartman, S. E., Hashida, G., Herndon, J., Howden, S. D., Humphreys, M. P., Hunt, C. W., Jones, S. D., Kim, S., Kitidis, V., Landa, C. S,

Landschützer, P., Lebon, G. T., Lefèvre, N., Lo Monaco, C., Luchetta, A., Maenner Jones, S., Manke, A. B., Manzello, D., Mears, P., Mickett, J., Monacci, N. M., Morell, J. M., Musielewicz, S., Newberger, T., Newton, J., Noakes, S., Noh, J.-H., Nojiri, Y., Ohman, M., Ólafsdóttir, S., Omar, A. M., Ono, T., Osborne, J., Plueddemann, A. J., Rehder, G., Sabine, C. L, Salisbury, J. E., Schlitzer, R., Send, U., Skjelvan, I., Sparnocchia, S., Steinhoff, T., Sullivan, K. F., Sutherland, S. C., Sweeney, C., Tadokoro, K., Tanhua, T., Telszewski, M., Tomlinson, M., Tribollet, A., Trull, T., Vandemark, D., Wada, C., Wallace, D. W. R., Weller, R. A., and Woosley, R. J.: Surface Ocean CO_2 Atlas Database Version 2020 (SOCATv2020) (NCEI Accession 0210711), NOAA National Centers for Environmental Information, https://doi.org/10.25921/4xkx-ss49. 2020.

Jo, Y. H., Dai, M. H., Zhai, W. D., Yan, X. H., and Shang, S. L.: 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.

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

[1]State Key Laboratory of Marine Environmental Science and College of Ocean and Earth Sciences, Xiamen University, Xiamen, 361102, China
[2]Fujian Provincial Key Laboratory for Coastal Ecology and Environmental Studies, Xiamen University, Xiamen, 361102, China
[3]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: Guizhi Wang (gzhwang@xmu.edu.cn), Samuel S. P. Shen (sshen@sdsu.edu)

Abstract. Sea surface partial pressure of CO_2 (pCO_2) data with a high spatial-temporal resolution are important in studying the 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 0.5º×0.5º over the period of 2000–2017 using both remote-sensing derived pCO_2 and observed underway pCO_2, among which ~~,~~the gridded underway pCO_2 data in 2004, 2005, and 2006 are presented for the first time. 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 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. 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 2.4±0.8~~383~~ μatm per year based on the spatial average of the reconstructed pCO_2 over the period of 2000–2017. This is consistent with the temporal trends at Station SEATS (18º N, 116º E) in the northern basin of the SCS and at Station HOT (22º45′ N, 158º 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 pCO_2 and the observed underway pCO_2 is no larger than 31.7 μatm, in contrast to (a) the RMSE from 12.8–89.0 μatm between the remote-sensing derived pCO_2 and the underway data, and (b) the RMSE from 32.6–44.5 μatm between the neural network produced pCO_2 and the underway data. The difference between the reconstructed pCO_2 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.pangaea.de/10.1594/PANGAEA.921210 (Wang et al., 2020).

1 Introduction

The oOcean 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 increasing rate from 1.0±0.5 Gt C yr^{-1} in the 1960s to 2.64±0.65 Gt C yr^{-1} during 2008–2017in 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; 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 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.

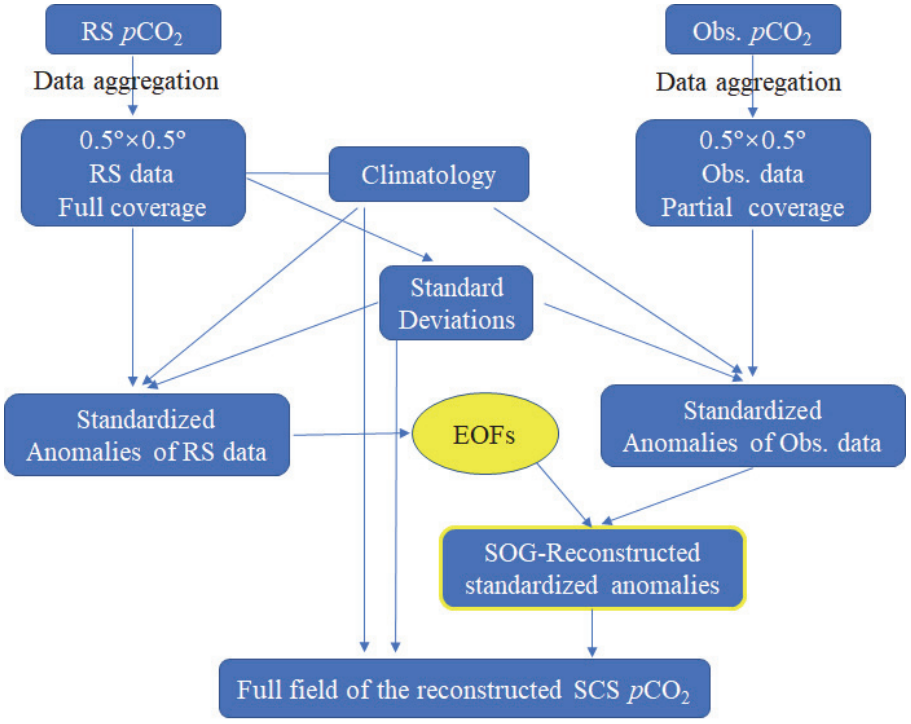
The sea–air CO_2 flux helps quantify the oceanic carbon uptake capacity and is primarily determined by the difference in the atmospheric and sea surface partial pressure of CO_2 (pCO_2). Although tThe measurement values records of sea surface fugacity of CO_2 (fCO_2, which is equal to pCO_2 corrected for the non-ideal behavior of the gas (Pfeil et al., 2013)) have been increasingincreased to 14.728.2 million and, are presently available in almost all ocean basins in 2014 based on the Surface Ocean CO_2 Atlas Version 2020 (Bakker et al., 2020), and continue to receive more data for compilation (Rodenbeck et al., 2015; Sheu et al., 2010). However, for a given year the observations of sea surface pCO_2 are still severelymay still have limited sparse in the spatial and temporalcoverage pCO_2 field of 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). These studies usually present their pCO_2 fields in a monthly time scale and at a $1^o \times 1^o$ 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 SCS is mostly blank (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 0.5º×0.5º, 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~~data~~, e.g., from Bai et al. (2015), to calculate empirical orthogonal functions (EOFs) for spatial patterns of pCO_2. The remote sensing pCO_2~~data~~ 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 ~~data~~ 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 0.5º×0.5º over the SCS.

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



80

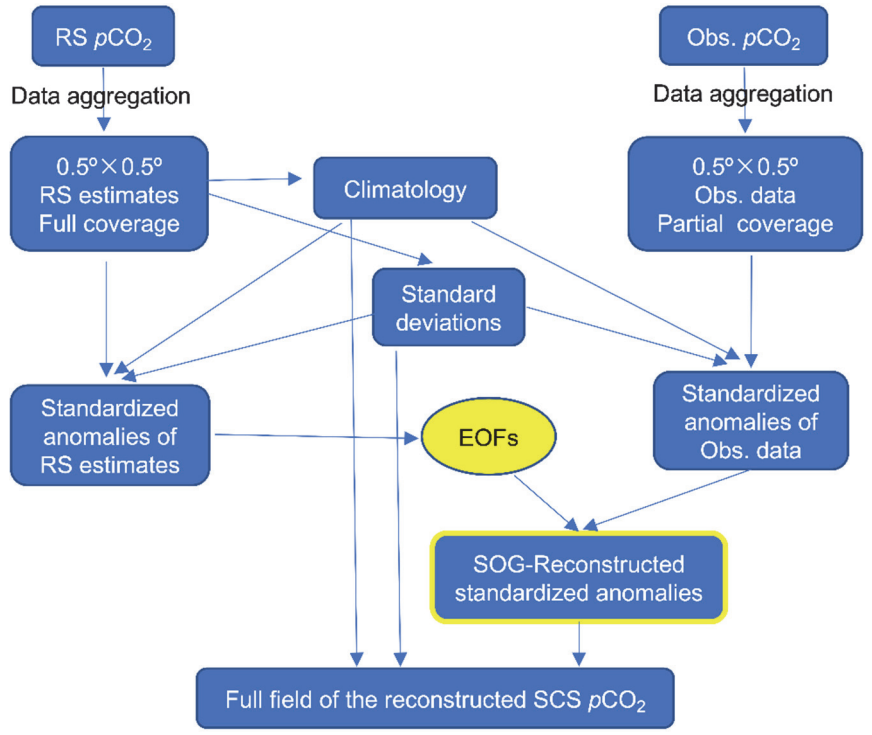


Figure 1: Reconstruction procedure of the sea surface pCO_2 in the SCS. Here, RS pCO_2 means the original remote-sensing derived pCO_2, Obs. pCO_2 represents the original observed in situ pCO_2, RS estimates~~data~~ are the grid-aggregated remote-sensing derived

pCO_2, and Obs. data are the grid ~~–~~aggregated observed pCO_2. The standard deviations are the temporal standard deviation of the RS pCO_2 estimates on each grid box.

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~~ underway summer pCO_2 data from 2000 to 2017 are compiled in this study and shown in Table 1~~.~~, in which the pCO_2 data in August 2004, July 2005 and June 2006 are new, 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; Zhai et al., 2005a). Thus, we have observed underway 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~~Figure 2 ~~indicate~~ 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 coverage of an individual summer is only a subset of the blue tracks. See Fig. 3 for the subset of each year.~~ The~~se~~ observational data were aggregated onto 0.5º×0.5º 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 the 13 summers are shown in Fig. 3, which shows the distribution pattern of the observed underway pCO_2 data of each year. 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 underway pCO_2 data cover a spatial range of 12º in latitude and 13º in longitude, 231 grid boxes with data that cover 22 % of SCS. The data fall in the range of 281–480 µatm. In the summer of 2017 the observed data cover a spatial range of 13º in latitude and 6º 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 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 7º in latitude and 9.5º 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 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.

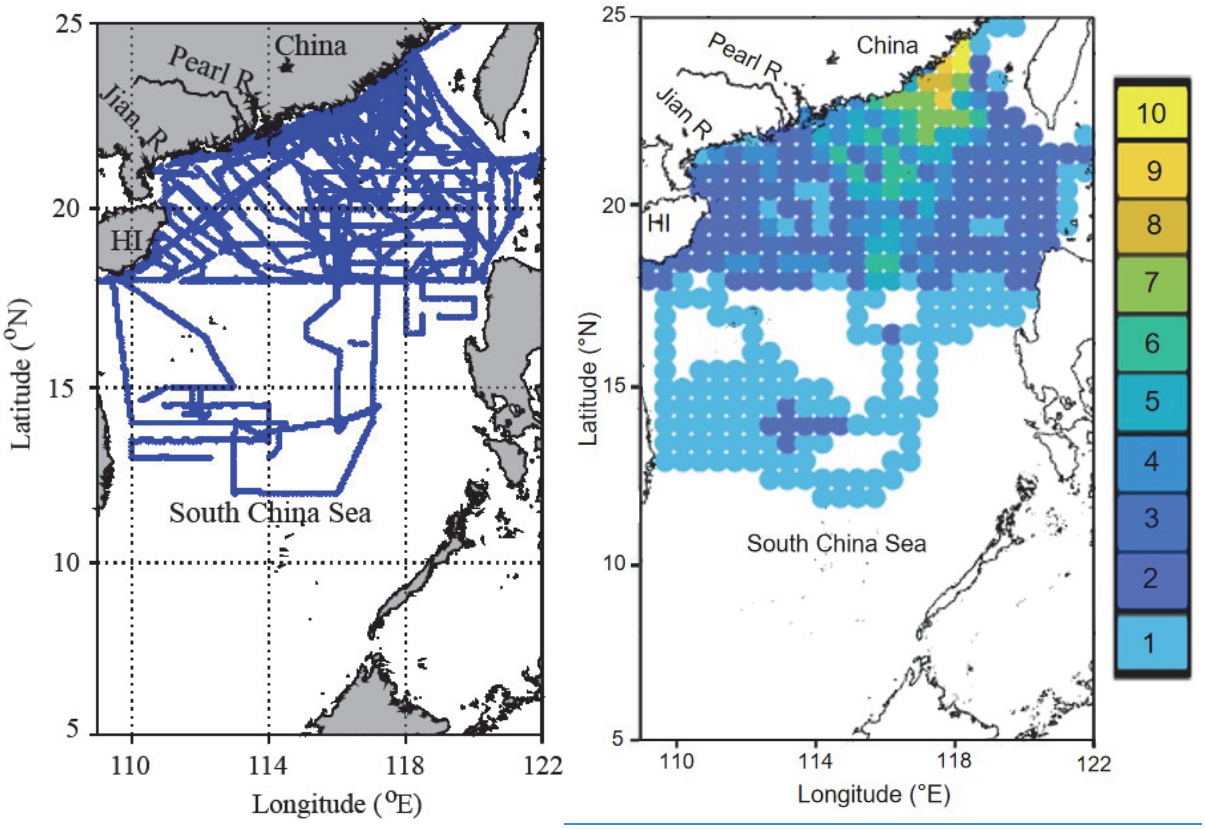


Figure 2: The number of summers with~~Cruise tracks of~~ underway sea surface pCO_2 observations in the SCS ~~in the summers from~~ in years 2000–2017. HI represents Hainan Island, Jian. R. is the Jianjiang River, and Pearl R. represents the Pearl River.

Table 1. Underway sea surface pCO_2 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

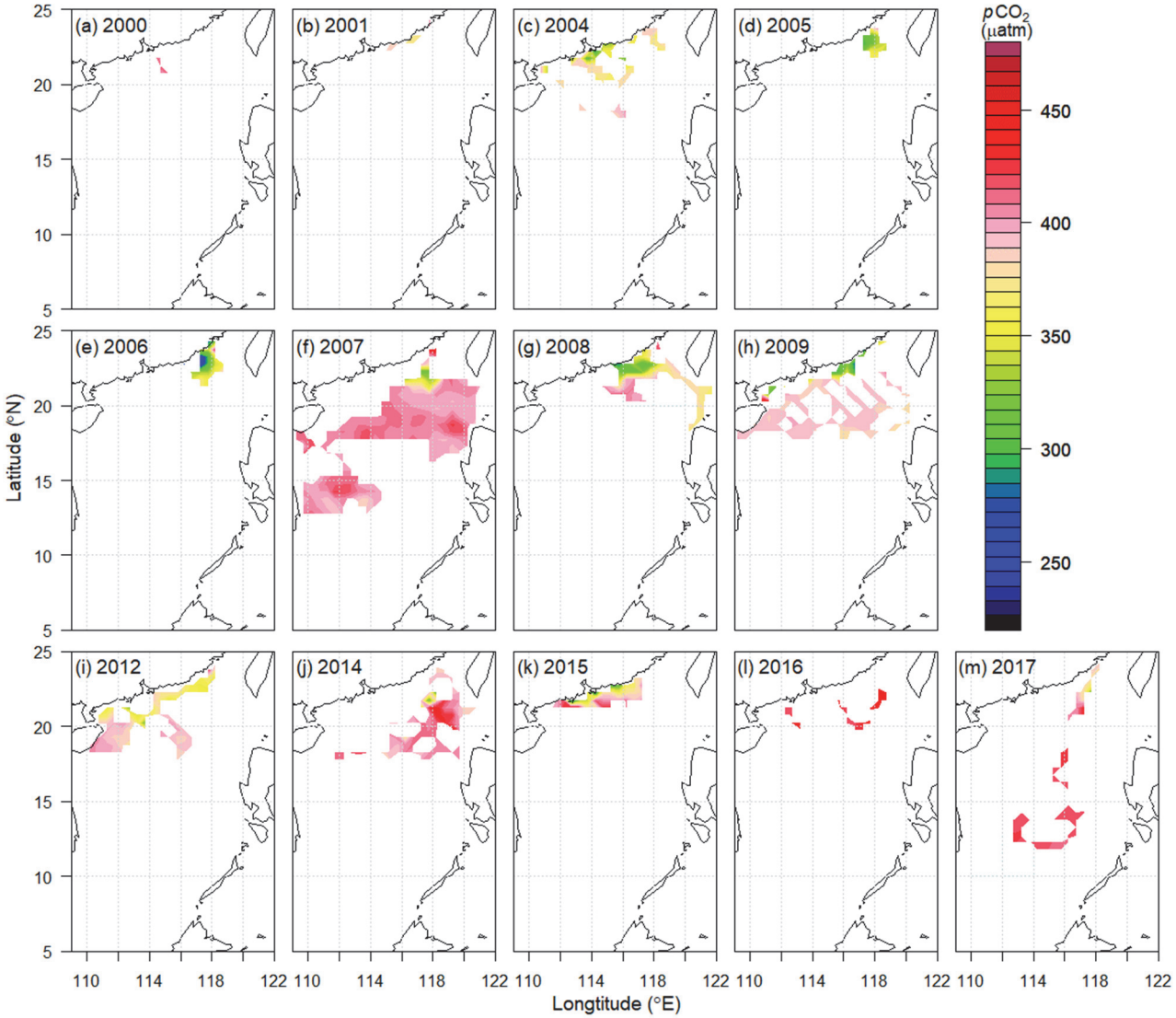


Figure 3: The aggregated in situ observational pCO_2 data in 0.5°×0.5° grid boxes in the SCS in the 13 summers during years 2000–2017.

As a dataset for our reconstruction validation, we calculated the sea surface pCO_2 from the observed temperature, salinity, total alkalinity, dissolved inorganic carbon, phosphate, and silicate at Station SEATS (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 pCO_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 pCO_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 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 "mechanic semi-analytical algorithm" (MeSAA) developed by Bai et al. (2015a). ~~The algorithm treats pCO_2~~ In the summer of SCS, the thermodynamic, mixing and biological effects on the sea surface pCO_2 were parameterized in the MeSAA algorithm as a function of major controlling factors derived from multiple satellite-derived ~~remote sensing products, including~~ sea surface temperature, colored dissolved organic matter, and chlorophyll ~~*a*, and salinity~~. The spatial resolution of the remote-sensing derived pCO_2 ~~data~~ estimates is 1′×1′. These ~~data~~ estimates were aggregated into 0.5º×0.5º grid boxes in our study region (5–25º N, 109–122º 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~~ estimates 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 more than 50 µatm. The root-mean-square-error (RMSE) falls in the range from 12.8–89.0 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 are much less than those shown by the observed pCO_2 data because the current MeSSA 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 ~~the~~ cruises ~~samplings~~.

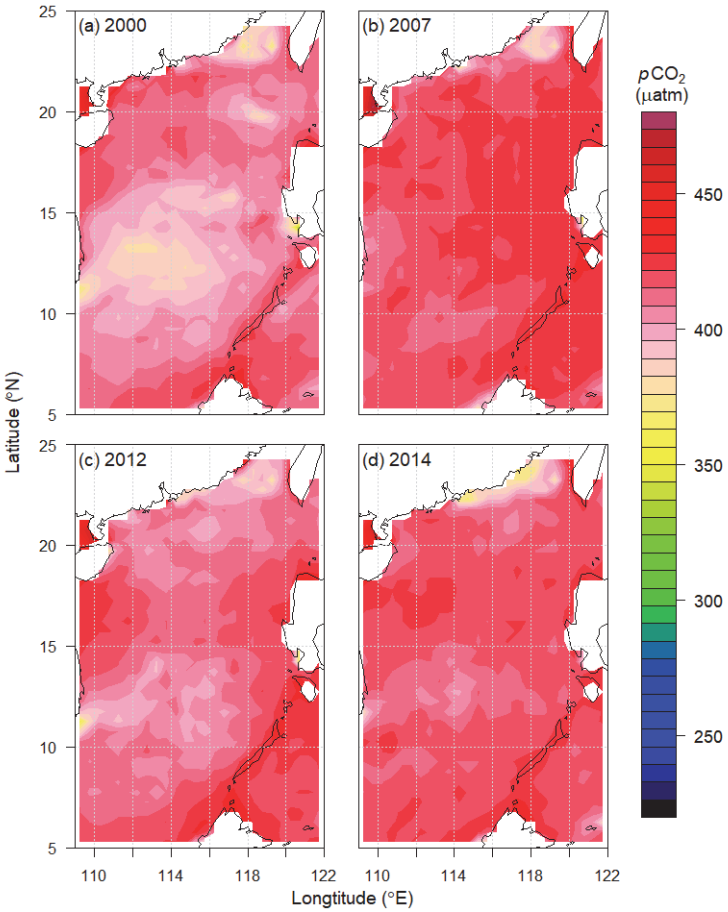


Figure 4: Remote-sensing derived sea surface pCO_2 in summer in selected years from 2000 to 2014.

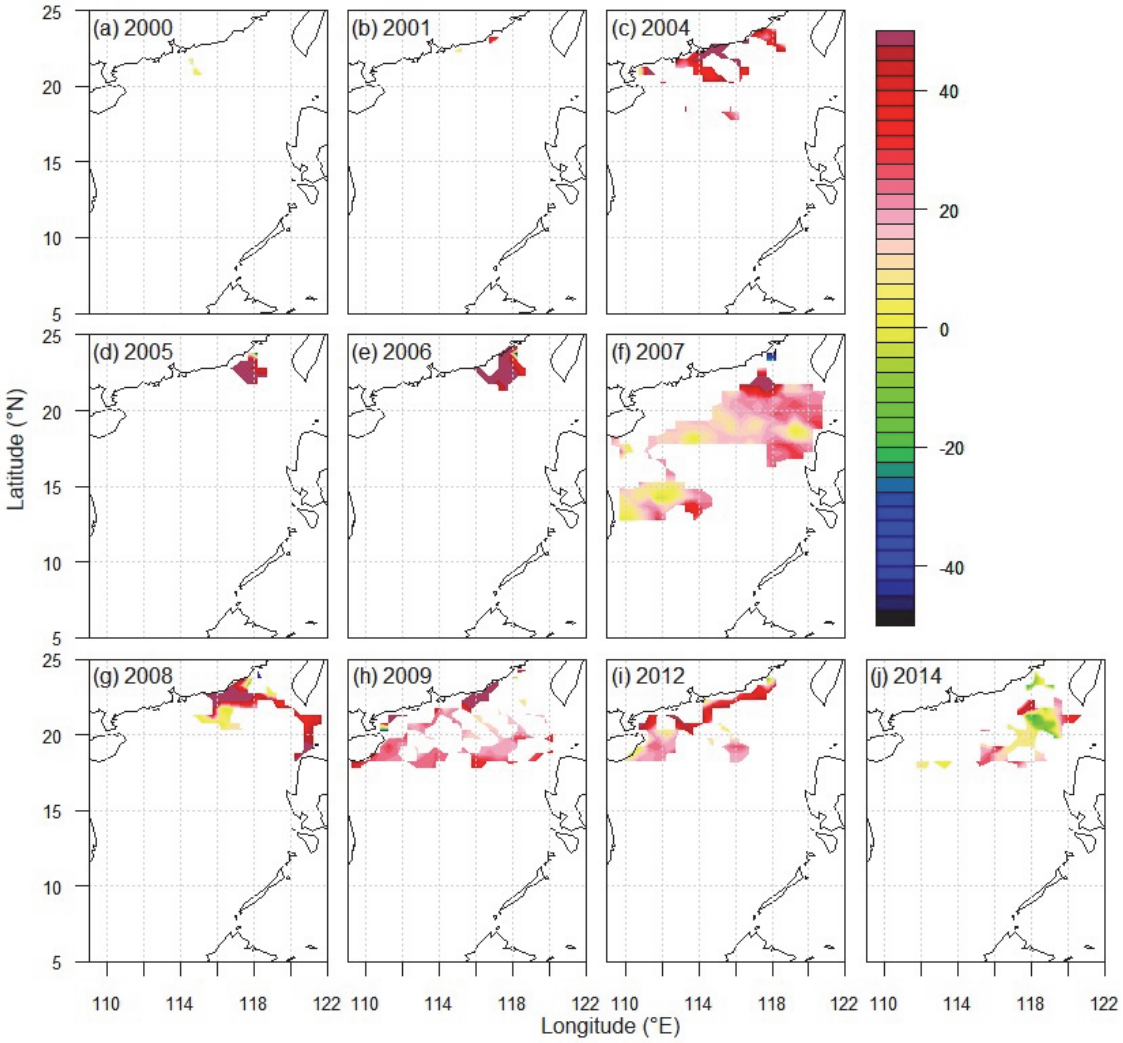


Figure 5: The difference between the remote-sensing derived summer pCO_2 estimates and the observed underway pCO_2 (unit: µatm) in 2000, 2001, 2004–2009, 2012, and 2014.

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 0.5º×0.5º 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) 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.

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

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

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 $b_m(t)$, $a(x) = \cos(\phi_x)$ 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

$$\varepsilon^2(x,t) = E[e^2(x,t)]~~(e^2(x,t))~~, \quad (2)$$

where ~~(.)~~E denotes the mathematical operation of expected value. The explanatory variables in the above multilinear~~ivariate~~ regression are $E_m(x)$, computed from the area-weighted standardized anomalies of the remote-sensing derived data. The anomalies were written as an 897×15 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. 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, …, 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

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

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 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 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). Here, M is the size of set $\mathcal{M}$. The current paper used eight EOFs that explain 87 % of the total variances (Fig. 65). ~~However, the year 2000 was an exception and used only four EOFs, because the year has only five grid boxes with the observed data.~~

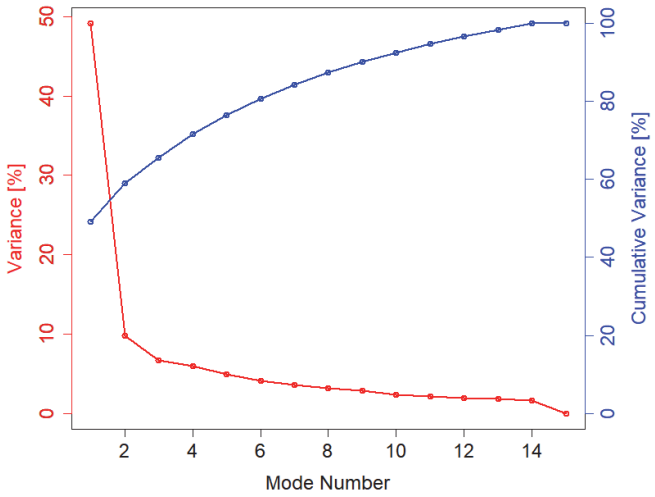


Figure 65: The percentage variances and cumulative variances based on the summer remote-sensing derived pCO_2 data for the period of years 2000–2014.

3 Results and discussion

3.1 EOFs and PCs

EOF1 demonstrates the mode of an average level of pCO_2 with lower or higher values near the coastal regions of China mainland (Fig. 76a). 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. 76b), 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. 76c), 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. 86d). It has 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~~ 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.

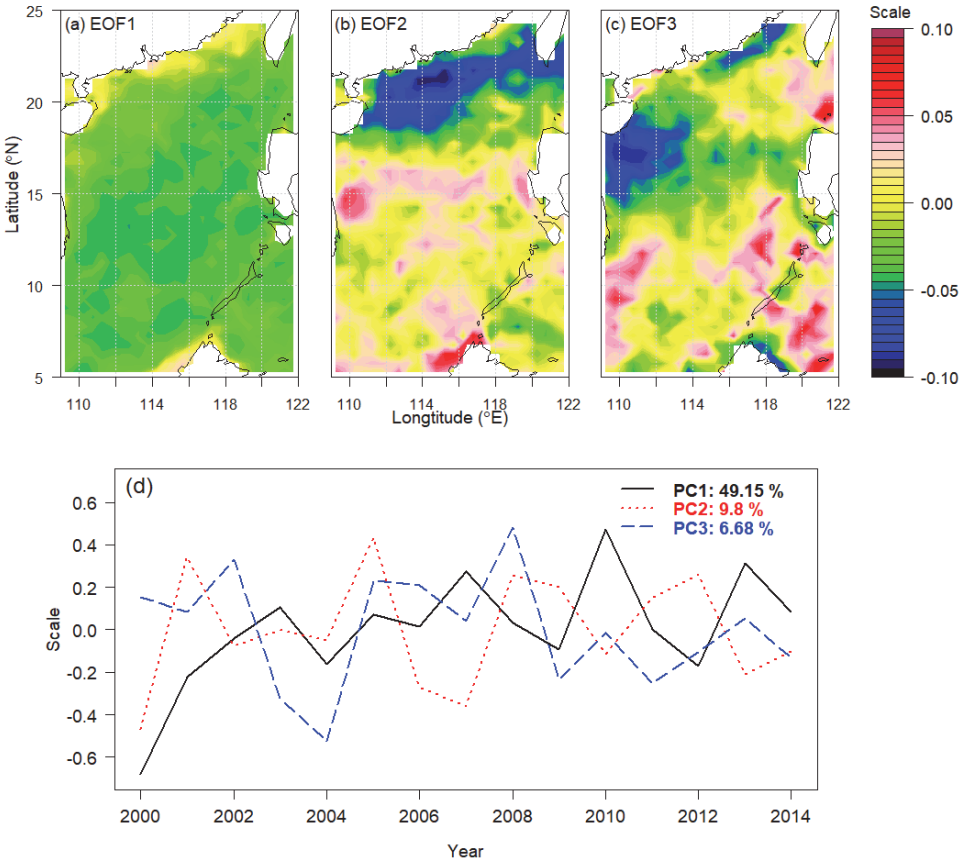


Figure ~~6~~7: EOFs and PCs of the remote-sensing derived pCO_2 estimates~~data~~. (a)–(c) EOFs, and (d) PCs.

3.2 Reconstruction results in the SCS

Figure ~~7~~8 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 basins.

The reconstructions have taken the advantages of both the ~~in situ~~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, 2014–2017 nicely demonstrate the spatial pCO_2 patterns (Figs. 8~~7~~c, 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, with an RMSE of 7.3 μatm between the reconstructed data and the observed data, appears reasonable (Fig. 8~~7~~b).

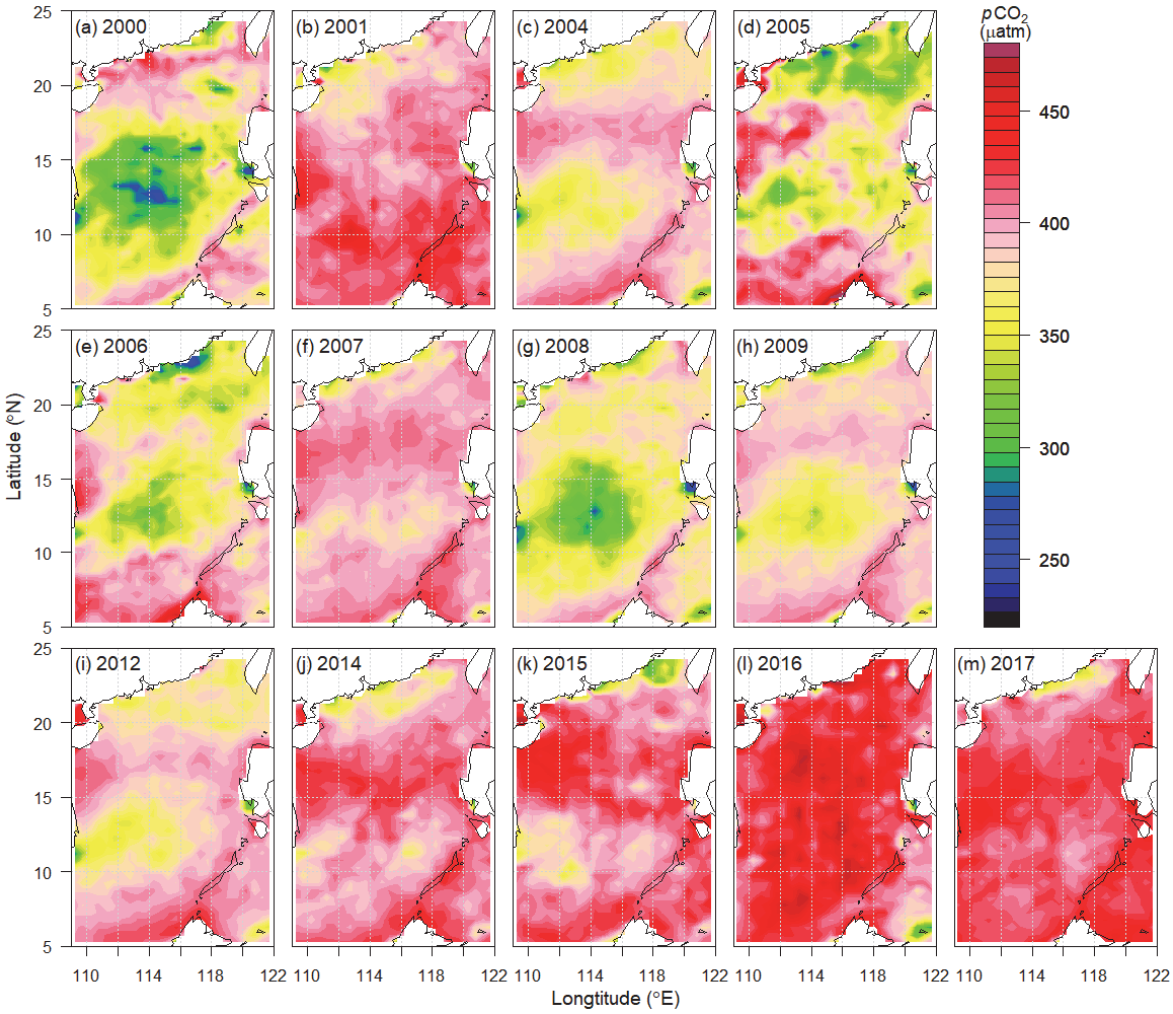


Figure 8~~7~~: Reconstructed summer pCO_2 fields for the years ~~of~~ 2000–2017 in the SCS.

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~~ of poor quality (Fig. 8~~7~~a) 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 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~~8d, e, 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~~gradient~~ in the~~of~~ in situ data. These ~~gradients~~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 ~~gradients~~ variations in~~of~~ the 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 cross-validation result.

Table 2: The RMSE between the remote-sensing derived pCO_2 estimates and the observed underway pCO_2 data ($RMSE_{RS}$), of the cross-validation ($RMSE_{CV}$), and between the reconstructed pCO_2 and the observed underway pCO_2 data ($RMSE_{RC}$) (unit: µatm).

Year	2000	2001	2004	2005	2006	2007	2008	2009	2012	2014	2015	2016	2017
$RMSE_{RS}$	12.8	20.2	47.9	65.7	89.0	25.1	43. 8	36.8	30.7	24.2	NaN	NaN	NaN
$RMSE_{CV}$	NaN	2.8	3.1	4.9	5.2	2.9	4.2	2.9	2.8	4.2	4.3	3.2	2.4
$RMSE_{RC}$	0.01	7.3	19.7	16.3	31.7	16.5	26.1	20.4	15.5	18.8	27.8	13.0	12.8

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) 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 years 2004–2007 with the underway pCO_2, the resultant RMSE falls in the range from 32.6–44.5 μatm and is twice as much as the median RMSE between our reconstructed pCO_2 and the underway pCO_2 (Table 2). Another uncertainty quantification for our reconstruction is to compare with the pCO_2 calculated from long-term observations at Station SEATS (Fig. 10). The difference between the reconstructed pCO_2 and the observed data at Station SEATS ranges from -7 to 10 μatm with the relative difference from -1.5 to 2.1 %. Both comparisons confirm that our reconstruction results are reliable.

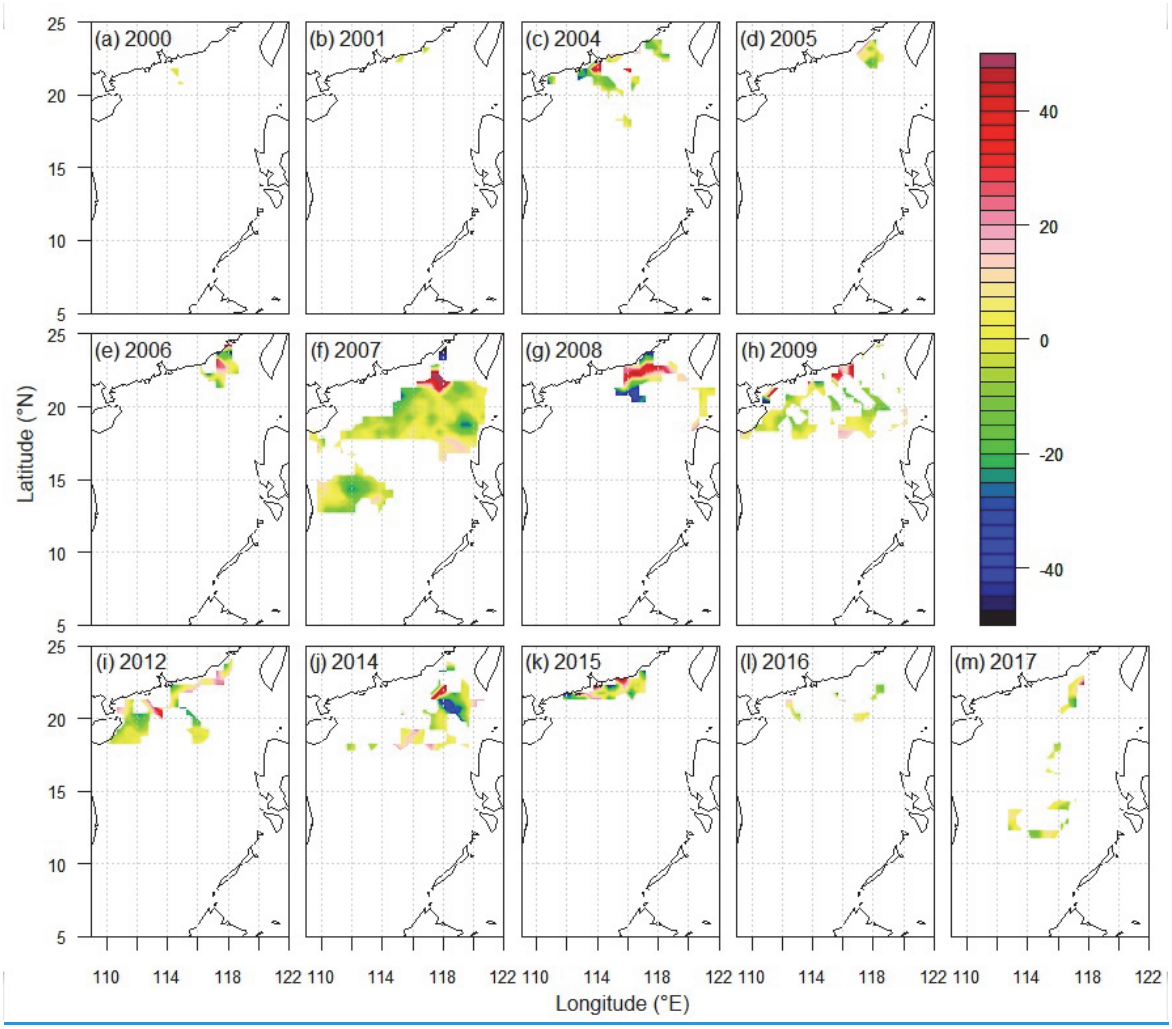


Figure 9: The difference between the reconstructed summer pCO_2 and the observed underway pCO_2 (unit: µatm) in 2000, 2001, 2004–2009, 2012, and 2014–2017.

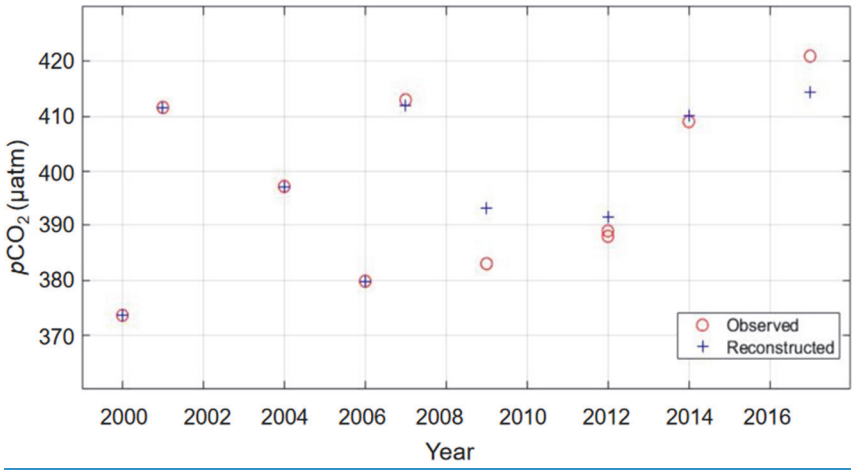
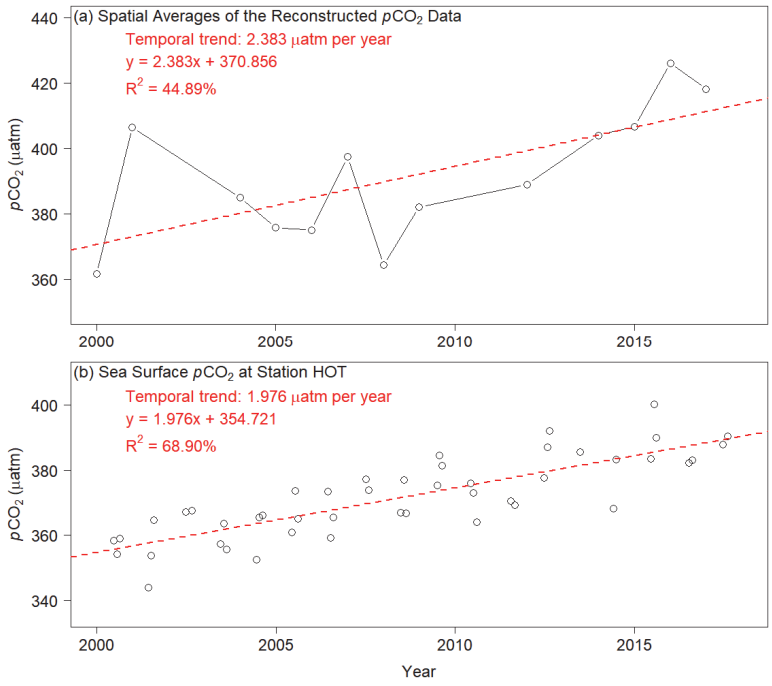


Figure 10: The comparison between the summer sea surface pCO2 calculated from the observations and those from our reconstruction at Station SEATS (18° N, 116° E). The pCO_2 data calculated from the observations in years of 2000, 2001, 2004, and 2006 are from Hui et al. (2020).

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 2.4±0.8~~383~~ µatm per year (See Fig. 11~~8~~a). It is lower than the rate of fCO_2 increase (4 µatm per year) ~~with time in the fugacity of CO_2 calculated for the mixed layer~~ for the period of 1999–2003 ~~at the South East Asia Time Series Station (Station SEATS) (18° N, 116° E) in the northern basin of the SCS, which is 4 µatm per year~~ (Tseng et al., 2007), while higher than the rate of pCO_2 increase for the period of 1998–2006 (0.8 µatm per year) at Station SEATS (Lui et al., 2020). ~~This~~ The differences between their rates and ours ~~makes sense since~~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 of 2000, which is the beginning year of our data, to the year of 2006 at Station SEATS is 2.5±1.0 µatm per year. 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 pCO_2 at the Hawaii Ocean Time-Series Station (Station HOT) (22°45′ N, 158° W) in the North Pacific, which is ~~1.976~~2.0±0.2 µatm per year over 2000–2017 (Dore et al., 2009) (See Fig. 11~~8~~b), our rate is ~~about 0.4 µatm per year higher~~consistent with the rate in the North Pacific. ~~This is reasonable for a marginal sea where a higher rate of increase in pCO_2 would be expected.~~ The consistency of the trend in our reconstructed sea surface pCO_2 over the SCS with the local trend at Station SEATS and the North Pacific trend at Station HOT confirms that our reconstruction is reasonable.



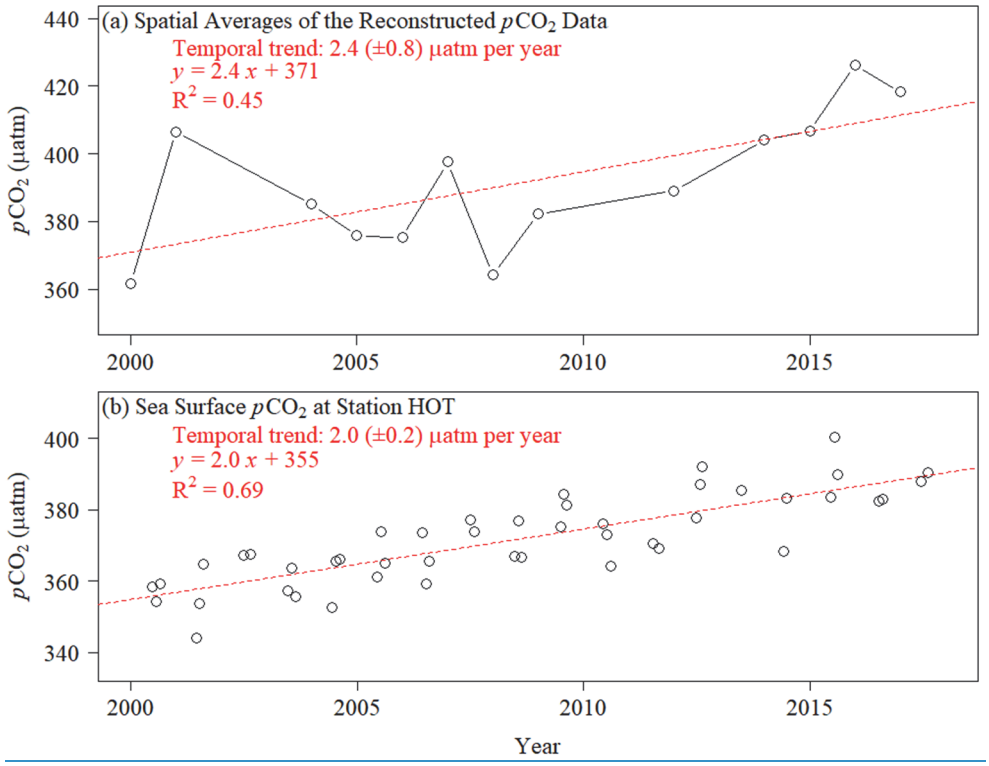


Figure 11~~8~~: (a) Time series and linear trend of the spatial averages of the reconstructed summer pCO_2 data in the SCS in the period of years 2000–2017; (b) Summer sea surface pCO_2 at Station HOT in years 2000–2017 adapted from Dore et al. (2009).

3.43 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 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. Table 32 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 Estuary and its nearby region.

Table 32. Outliers excluded from the SOG reconstruction.

Year	Grid ID	Latitude (N)	Longitude (E)	pCO_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

4 Data availability

The gridded ~~in situ~~underway sea surface pCO_2 data, the remote-sensing derived sea surface pCO_2 estimates~~data~~, and the reconstruction result data are openly and 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 underway 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 0.5º×0.5º for reconstruction. The SOG method based on the 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 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. 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 pCO_2 with the observed underway pCO_2 has an RMSE smaller than that of the remote-sensing derived pCO_2, as well as that of the neural network produced pCO_2 in the same year. Moreover, our reconstructed pCO_2 compares well with the pCO_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 pCO_2 over the SCS is consistent with the local rate at Station SEATS and the North Pacific rate at Station HOT, which further validates our reconstruction. 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 help fill the long-lasting blanks on~~in~~ the global sea surface pCO_2 map~~ping~~. 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.

Author contribution

Minhan Dai conceptualized and directed the field program of the in situ observations. Baoshan Chen and Xianghui Guo 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. Zhixuan Wang participated in the uncertainty analysis of the reconstruction and plotted some of 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

The authors declare that they have no conflict of interest.

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 pCO_2 data in his publications and to Young-Heon Jo for providing the neural-network produced pCO_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, Xianghui Guo, and Minhan Dai was supported by ~~a~~ grants from the Ministry of Science and Technology of China (2015CB954001, 2009CB421200). The observed underway 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 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 pCO_2 in river-dominated coastal oceans: a case study from the East China Sea, J. Geophys. Res. 120, 2331–2349, https://doi.org/~~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, https://doi.org/~~doi:~~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. R., Balestrini, C. F., Barbero, L., Bates, N. R., Bianchi, A. A., Bonou, F., Boutin, J., Bozec, Y., Burger, E. F., Cai, W.-J., Castle, R. D., Chen, L., Chierici, M., Currie, K., Evans, W., Featherstone, C., Feely, R. A., Fransson, A., Goyet, C., Greenwood, N., Gregor, L., Hankin, S., Hardman-Mountford, N. J., Harlay, J., Hauck, J., Hoppema, M., Humphreys, M. P., Hunt, C. W., Huss, B., Ibánhez, J. S. P., Johannessen, T., Keeling, R., Kitidis, V., Körtzinger, A., Kozyr, A., Krasakopoulou, E., Kuwata, A., Landschützer, P., Lauvset, S. K., Lefèvre, N., Lo Monaco, C., Manke, A., Mathis, J. T., Merlivat, L., Millero, F. J., Monteiro, P. M. S., Munro, D. R., Murata, A., Newberger, T., Omar, A. M., Ono, T., Paterson, K., Pearce, D., Pierrot, D., Robbins, L. L., Saito, S., Salisbury, J., Schlitzer, R., Schneider, B., Schweitzer, R., Sieger, R., Skjelvan, I., Sullivan, K. F., Sutherland, S. C., Sutton, A. J., Tadokoro, K., Telszewski, M., Tuma, M., van Heuven, S. M. A. C., Vandemark, D., Ward,

B., Watson, A. J., and Xu, S.: A multi-decade record of high-quality fCO_2 data in version 3 of the Surface Ocean CO_2 Atlas (SOCAT), Earth Syst. Sci. Data, 8, 383–413, https://doi.org/10.5194/essd-8-383-2016, 2016.

Bakker, D. C. E., Alin, S. R., Bates, N., Becker, M., Castaño-Primo, R., Cosca, C. E., Cronin, M., Kadono, K., Kozyr, A., Lauvset, S. K., Metzl, N., Munro, D. R., Nakaoka, S., O'Brien, K. M., Ólafsson, J., Olsen, A., Pfeil, B., Pierrot, D., Smith, K., Sutton, A. J., Takahashi, T., Tilbrook, B., Wanninkhof, R., Andersson, A., Atamanchuk, D., Benoit-Cattin, A., Bott, R., Burger, E. F., Cai, W.-J., Cantoni, C., Collins, A., Corredor, J. E., Cronin, M. F., Cross, J. N, Currie, K. I., De Carlo, E. H., DeGrandpre, M. D., Dietrich, C., Emerson, S., Enright, M. P., Evans, W., Feely, R. A., García-Ibáñez, M. I., Gkritzalis, T., Glockzin, M., Hales, B., Hartman, S. E., Hashida, G., Herndon, J., Howden, S. D., Humphreys, M. P., Hunt, C. W., Jones, S. D., Kim, S., Kitidis, V., Landa, C. S, Landschützer, P., Lebon, G. T., Lefèvre, N., Lo Monaco, C., Luchetta, A., Maenner Jones, S., Manke, A. B., Manzello, D., Mears, P., Mickett, J., Monacci, N. M., Morell, J. M., Musielewicz, S., Newberger, T., Newton, J., Noakes, S., Noh, J.-H., Nojiri, Y., Ohman, M., Ólafsdóttir, S., Omar, A. M., Ono, T., Osborne, J., Plueddemann, A. J., Rehder, G., Sabine, C. L, Salisbury, J. E., Schlitzer, R., Send, U., Skjelvan, I., Sparnocchia, S., Steinhoff, T., Sullivan, K. F., Sutherland, S. C., Sweeney, C., Tadokoro, K., Tanhua, T., Telszewski, M., Tomlinson, M., Tribollet, A., Trull, T., Vandemark, D., Wada, C., Wallace, D. W. R., Weller, R. A., and Woosley, R. J.: Surface Ocean CO_2 Atlas Database Version 2020 (SOCATv2020) (NCEI Accession 0210711), NOAA National Centers for Environmental Information, https://doi.org/10.25921/4xkx-ss49. 2020.

Boutin, J., Etcheto, J., Dandonneau, Y., Bakker, D. C. E., Feely, R. A., Inoue, H. Y., Ishii, M., Ling, R. D., Nightingale, P. D., 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, https://doi.org/~~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, https://doi.org/~~doi:~~10.3319/TAO.2005.16.2.445(O), 2005.

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, https://doi.org/~~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, https://doi.org/~~doi:~~10.1073/pnas.1900371116, 2019.

Dickson, A. G.: Standard potential of the reaction: $AgCl(s) + 12H_2(g) = Ag(s) + HCl(aq)$, and the standard acidity constant of the ion HSO_4^- in synthetic sea water from 273.15 to 318.15 K, J. Chem. Thermodyn., 22(2), 113–127, https://doi.org/10.1016/0021-9614(90)90074-Z, 1990.

Dickson, A.G., and Millero, F. J.: A comparison of the equilibrium constants for the dissociation of carbonic acid in seawater media, Deep-Sea Res. I, 34, 1733–1743, https://doi.org/10.1016/0198-0149(87)90021-5, 1987.

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, https://doi.org/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, https://doi.org/doi:10.1073/pnas.0906044106, 2009.

Du, C., Liu, Z., Dai, M., Kao, S. J., Cao, Z., Zhang, Y., Huang, T., Wang, L., and Li, Y.: Impact of the Kuroshio intrusion on the nutrient inventory in the upper northern South China Sea: insights from an isopycnal mixing model, Biogeosciences, 10(10), 6419–6432, https://doi.org/10.5194/bg-10-6419-2013, 2013.

Du, C. J., Liu, Z. Y., Kao, S. J., and Dai, M. H.: Diapycnal Fluxes of Nutrients in an Oligotrophic Oceanic Regime: The South China Sea. Geophys. Res. Lett., 44(22), 11510-11518, http://doi.org/10.1002/2017gl074921, 2017.

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, https://doi.org/doi:10.1002/gbc.20051, 2013.

Friedlingstein, P., O'Sullivan, M., Jones, M. W., Andrew, R. M., Hauck, J., Olsen, A., Peters, G. P., Peters, W., Pongratz, J., Sitch, S., Le Quéré, C., Canadell, J. G., Ciais, P., Jackson, R. B., Alin, S., Aragão, L. E. O. C., Arneth, A., Arora, V., Bates, N. R., Becker, M., Benoit-Cattin, A., Bittig, H. C., Bopp, L., Bultan, S., Chandra, N., Chevallier, F., Chini, L. P., Evans, W., Florentie, L., Forster, P. M., Gasser, T., Gehlen, M., Gilfillan, D., Gkritzalis, T., Gregor, L., Gruber, N., Harris, I., Hartung, K., Haverd, V., Houghton, R. A., Ilyina, T., Jain, A. K., Joetzjer, E., Kadono, K., Kato, E., Kitidis, V., Korsbakken, J. I., Landschützer, P., Lefèvre, N., Lenton, A., Lienert, S., Liu, Z., Lombardozzi, D., Marland, G., Metzl, N., Munro, D. R., Nabel, J. E. M. S., Nakaoka, S. I., Niwa, Y., O'Brien, K., Ono, T., Palmer, P. I., Pierrot, D., Poulter, B., Resplandy, L., Robertson, E., Rödenbeck, C., Schwinger, J., Séférian, R., Skjelvan, I., Smith, A. J. P., Sutton, A. J., Tanhua, T., Tans, P. P., Tian, H., Tilbrook, B., van der Werf, G., Vuichard, N., Walker, A. P., Wanninkhof, R., Watson, A. J., Willis, D., Wiltshire, A. J., Yuan, W., Yue, X., and Zaehle, S.: Global Carbon Budget 2020, Earth Syst. Sci. Data, 12(4), 3269–3340, https://doi.org/10.5194/essd-12-3269-2020, 2020.

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, https://doi.org/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 since 1910, J. Geophys. Res.–Atmos., 120, 4878–4888, https://doi.org/doi: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(3), 1993–2006, https://doi.org/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, https://doi.org/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, https://doi.org/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, https://doi.org/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, https://doi.org/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, https://doi.org/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, https://doi.org/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, https://doi.org/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, https://doi.org/doi:10.1016/s0967-0637(01)00064-4, 2002.

Lewis, E. R., and Wallace, D. W. R.: Program Developed for CO_2 System Calculations, United States: N. p., Web, http://doi.org/10.15485/1464255, 1998.

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, https://doi.org/doi: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: Chen, C.-T., and Guo, X. (eds) Changing Asia-Pacific Marginal Seas. Atmosphere, Earth, Ocean & Space. Springer, Singapore, https://doi.org/10.1007/978-981-15-4886-4_16, 2020.

Mehrbach, C., Culberso, C. H., Hawley, J. E., and Pytkowic, R. M.: Measurement of apparent dissociation constants of carbonic acid in seawater at atmospheric pressure, Limnol. Oceanogr., 18(6), 897–907, https://doi.org/10.4319/lo.1973.18.6.0897, 1973.

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, https://doi.org/doi:10.1080/01431160310001657515, 2004.

Pfeil, B., Olsen, A., Bakker, D. C. E., Hankin, S., Koyuk, H., Kozyr, A., Malczyk, 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., 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., Johannessen, T., Jones, S. D., Key, R. M., Körtzinger, A., Landschützer, P., Lauvset, S. K., Lefèvre, N., Lenton, A., Lourantou, A., Merlivat, L., Midorikawa, T., Mintrop, L., Miyazaki, C., Murata, A., Naka-date, A., Nakano, Y., Nakaoka, S., Nojiri, Y., Omar, A. M., Padin, X. A., Park, G.-H., Paterson, K., Perez, F. F., Pierrot, D., Poisson, A.,Ríos,A.F.,Santana-Casiano,J.M.,Salisbury,J.,Sarma,V.V. S. S., Schlitzer, R., Schneider, B., Schuster, U., Sieger, R., Skjelvan, I., Steinhoff, T., Suzuki, T., Takahashi, T., Tedesco, K., Telszewski, M., Thomas, H., Tilbrook, B., Tjiputra, J., Vandemark, D., Veness, T., Wanninkhof, R., Watson, A. J., Weiss, R., Wong, C. S., and Yoshikawa-Inoue, H.: A uniform, quality controlled Surface Ocean CO_2 Atlas (SOCAT), Earth Syst. Sci. Data, 5, 125–143, https://doi.org/10.5194/essd-5-125-2013, 2013.

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, https://doi.org/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, https://doi.org/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, https://doi.org/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 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, https://doi.org/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 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, https://doi.org/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 System: an eddy-resolving modeling study, Biogeosciences, 11(3), 671–690, https://doi.org/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, https://doi.org/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, https://doi.org/~~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, https://doi.org/~~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 measurements, Biogeosciences, 10(11), 7775–7791, https://doi.org/~~doi:~~10.5194/bg-10-7775-2013, 2013.