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.

Report #1

The paper is much improved and the authors should be commended on their efforts. The new analyses based on the SEATS data are very helpful, and much work has been done to better quantify the uncertainty. It is also excellent that the paper is confirming that it is presenting new data, and this will elevate the impact of the work.

Response: Thank you very much for your positive confirmation of the improvement and value of our paper.

I yet feel the uncertainty assessment would benefit from another revision. It is not enough to omit one 0.5x0.5 degree grid cell at a time and recompute estimates in that grid cell. That is just a test of how well the method works when the pCO2 estimate has proximal (in both space and time) measurements. The test should omit entire cruises or entire swathes of data in the tests (i.e., the western or easternmost halves of all data in each year... though even that would provide somewhat of an underestimate since the estimates would always benefit from temporally-proximal measurements).

Response: Following your suggestion, we have performed another cross-validation analysis: Leaving out the western half and eastern halves of the data in a year. The analysis was done for the years with better spatial coverage: 2007, 2009, and 2012. The resulted RMSE is within the range of the RMSEs of the leave-one-out cross validation. This is another confirmation of the reliability of our reconstruction. The details of this cross-validation calculation are as follows.

The western halves (longitude < 115.5° E) of data in 2007, 2009, and 2012 were removed, respectively. The reconstructions were made using the remaining half data. The resulted RMSE between the removed data and the reconstructed data were 2.77, 4.46, and 3.82 μatm for 2007, 2009, and 2012, respectively. Similarly, when the eastern halves (longitude >115.5° E) of data in these years have been removed, the RMSEs are 4.32, 3.66, and 3.55 in 2007, 2009, and 2012, respectively. These values fall in the range of the RMSEs of the leave-one-out cross validation. In the revision, this leave-half-out cross validation has been included in the revised paper (See Lines 257-266).

This type of cross-validation can have infinitely many kinds of variations, depending on the size of the region where the in situ data are withheld. It is also obvious that the regional leave-out would not work if the remaining half has no in situ data. In this revision, rather than exhausting all possibilities, we have only included the regional leave-out cross-validation for 2007, 2009, and 2012 according to the 115.5ºE meridional line.

Also, I urge the authors to focus on estimate bias for each year instead of estimate RMSE, and to propagate the mean bias for each year's estimates into an estimate of the error in the average fluxes. Probably the most important quantity is RMSE of the annual average pCO2 (similar to average annual absolute bias).

Response: For the bias estimate, we have used the difference between the reconstructed pCO_2 and the observed underway pCO_2. Figure 9 in the paper shows the difference for each grid in each year. To show the statistics of this figure, we generated boxplots (Figure R1) of the difference data.

Although the bias propagation is an important question, it is not the focus of this data reconstruction paper. The flux of air-sea pCO_2 will be among our further studies. The bias of the reconstructed pCO_2 will then be transferred to the flux.

As we pointed out in Section 2.1 in our paper "This study focuses on the summer data since the greatest temporal coverage of the sampling occurs in summer." Data in other seasons are scarce. It is still under investigation whether they can be used for annual reconstruction. When the annual reconstruction is made, the corresponding RMSE will be provided.

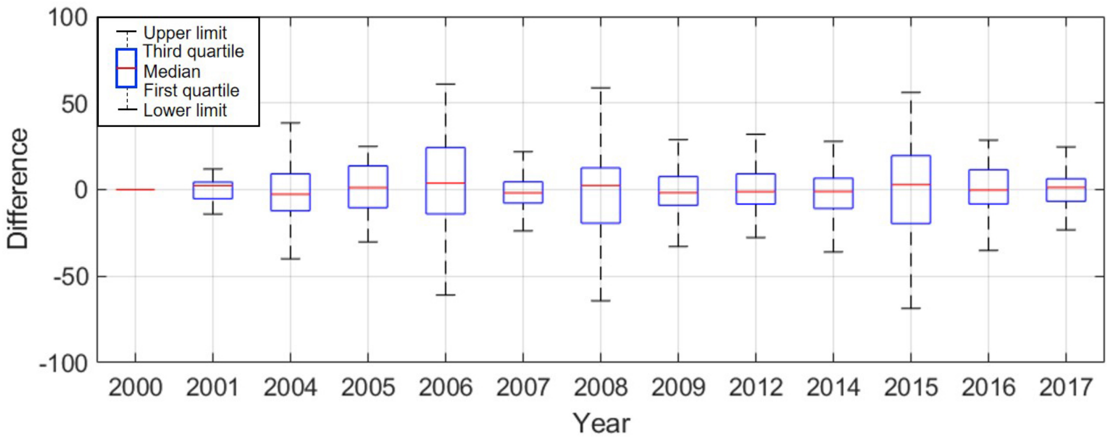

Figure R1: The boxplot of the difference between the reconstructed summer pCO_2 and the observed underway pCO_2 (unit: µatm) in 2000, 2001, 2004–2009, 2012, and 2014–2017. The upper limit is defined as Q3+1.5×(Q3-Q1), where Q3 is the third quartile and Q1 is the first quartile. The lower limit is defined as Q1-1.5×(Q3-Q1). The outliers outside the range determined by the upper and lower limits are not shown.

Their response to "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" did not convince me. The method is only valid when its estimates can be trusted. With no validation measurements in the South China Sea, the results cannot be trusted in that region. The paper would be stronger if the results from that region were omitted from at least one regional trend, or held aside and given an appropriate caution.

Response: The reconstruction is a spatial prediction of the pCO_2 field for the region without in situ observations. We could do so because we have used the EOFs patterns from remote-sensing derived estimates, anchored on the sparse in situ data. This has been described in detail in the method section 2.3 in the paper.

Ideally, the in situ data should be evenly distributed in the entire South China Sea. Hence, the reconstruction can be validated for the entire region. However, the real situation is not the case. The in situ data are limited to regions as shown in Figure 3 in the paper. Nonetheless, our SOG reconstruction can still be made using the EOF patterns.

I didn't understand the authors' response to "If the Bai et al. approach gives a different average pCO2 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." The response seemed to explain how EOFs work using jargon. Perhaps I just didn't understand, but the reply didn't seem to address the concern. The concern is that a big error in the mean state in the underlying climatology would have to be compensated by spurious scaling of EOFs that are intended to capture anomalies from the climatology (not errors in the mean state). The new analysis suggests that the RS pCO2 is indeed overestimating average pCO2, and sometimes by up to 50 uatm (separately, this should be quantified with an average bias statistic across all years as well). This means the climatological values might indeed be a poor representation of the true state of the region.

Response: Yes, indeed. Our EOF reconstruction step is for the anomalies. The full field is recovered by adding the climatology and then multiplied by standard deviation, as illustrated in Figure 1 and described in Section 2.3 in the paper. We understand that the remote-sensing derived pCO_2 estimates have overestimated the field as shown in Figure 5 in the paper. When the accuracy of the remote-sensing data is improved, it is certainly helpful to improve our reconstruction. However, this RS dataset is the best so far available for computing the spatial pattern. We actually tested ocean model-derived pCO_2. The result was inferior. In addition, our reconstructed results may help correct the bias from the remote-sensing estimates, which needs further studies.

For the RS pCO_2 bias estimate, we have used the difference between the remote-sensing derived pCO_2 and the observed underway pCO_2 for each grid in each year as shown in Figure 5 in the paper. Figure R2 below shows the statistics of Figure 5's data using boxplots.

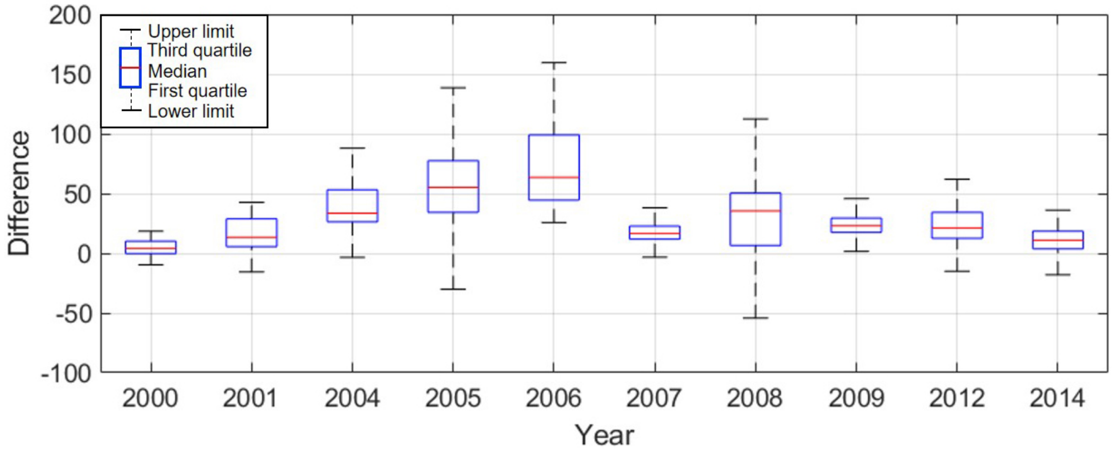

Figure R2: The boxplot of the difference between the remote-sensing derived summer pCO_2 and the observed underway pCO_2 (unit: µatm) in 2000, 2001, 2004–2009, 2012, and 2014. The upper limit is defined as Q3+1.5×(Q3-Q1), where Q3 is the third quartile and Q1 is the first quartile. The lower limit is defined as Q1-1.5×(Q3-Q1). The outliers outside the range determined by the upper and lower limits are not shown.