Response to referee #1 ---------------------------

We thank referee#1 for the thoughtful and constructive feedback on the paper. We have addressed major concerns in the revised manuscript and documented our responses to the referee's comments point-by-point as follows.

1. This Technical note compares the results of three machine learning models for sea surface CO2 mapping. Two of those, self-organizing-maps (SOM) and feedforward neural networks (FNN), have already been used and compared (in the Surface Ocean CO2 Mapping inter comparison initiative, SOCOM) and a new one, the support vector machine (SVM), is introduced in this paper. The SVM performs best but requires big computer memory. This is valuable work as with ever increasing computer power SVM will become available to more users. I have one concern: the resulting model distributions show features that cannot be explained by the CO2 field data. For example there is a CO2 hotspot east of the African coast near the equator where no observations (February) or low CO2 observations (July) are shown in the top panel. In July there is an unexplained hotspot in the Southern Ocean west of South America where there are no observations. I presume these features are produced by the correlation of sea surface CO2 with proxy variables such as SST, SSS CHL and MLD? Are these hotspots known / expected from previous publications? The authors should discuss this further in the discussion of Figure 3.

   Reply: Referee #1 pointed out the CO2 hotspot east of the African coast near the equator and suggested a further discussion. As the issue cannot be cleared without thorough numerical comparisons with outputs of other models, it would go beyond the scope of this manuscript, which is aimed at comparing the three machine learning models. The hotspot CO2 seems quite high comparing to the nearby measurement made by the one cruise in July 1995. It is very difficult to judge whether this is an issue because the observed CO2 shown in the figure is trend removed with a universal rate, which could be another source of uncertainty. Takahashi et al. (2009) shown that the trend could be quite different in different areas. However, it difficult to use multiple rates for global mapping.

2. My final question is: is the dataset produced by SVM available for download somewhere or can it be retrieved from the authors? Could this be added as a supplement possibly?

   Reply: We uploaded the dataset as supplement.

3. Page 1, line 14: include (Goddijn-Murphy et al, 2015).

   Reply: We included the reference.

4. Page 2, line 13: please explain "circular property" and why it can therefore not be used.

Reply: In the manuscript we added “For instance, longitude -180 degree is geographically connected to longitude 180 degree, but numerically they appear to be two extreme longitude values to the models.”

5. Page 2, line 14: sine and cosine transformed components of LON and MON? How of MON?

   Reply: The transforms are cos(MON*2*pi/12), sin(MON*2*pi/12), cos(Lon*2*pi/360) and sin(LON*2*pi/360). See Zeng et al. (2015). We didn’t give the transform here because MON and LON were not used.

6. Page 2, line 14: “The approach” is meaning “Our approach” or “Zeng et al.’s approach”?

   Reply: We revised “The approach …” to “Their approach …” to explicitly mean Zeng et al.’s approach.

7. Page 3, line 4: which two CHL products, calculated from OC3 and OCI algorithms?

   Reply: The footnote indicates that the products used the OCI algorithm

8. Page 3, line 8, refer to Table 1 here

   Reply: We revised “The Appendix summarizes…” to “The Appendix and Table 1 summarize…”.

9. Page 3, line 11: 10% of the measurements randomly chosen?

   Reply: Yes, they are randomly chosen. We revised “we used 10% of…” to “we randomly chose 10% of…”

10. Page 3, line 12: “dependent of” should be “dependent on”.

    Reply: We corrected the mistake.

11. Page 3, line 17: insert “” in “all variables ”; explain all variables (SST, SSS, CHL, MLD, dSST?).

    Reply: We revised the sentence to “we scaled all input variables LAT, SST, SSS, CHL, MLD, and dSST by their minimum and maximum to confine them in the range (0, 1)”.

12. Page 4, line 2: give references for preliminary studies.

Reply: We revised “Based on preliminary studies” to “Based on our preliminary correlation analysis”.

13. Page4, line 13: replace “to model” with “and modelled”.

    Reply We revised the expression accordingly.

14. Page 4, line 18: modeled and observed CO2 of “all / selected/ non-selected” data points?

    Reply: We added “of the selected data points” to the end of the sentence.

15. Page 5, line 6: random 10%?

    Reply: Yes, they were randomly selected. We revised “with 10% of the data” to “with 10% of randomly selected data points”.

16. Page 5, line 8: differences are expressed as mean difference standard deviation?

    Reply: Yes.

17. Page 5, line 8: replace “respectively” with “for SOM”.

    Reply: We corrected the mistake.

18. Page 5, line 9: give range of measurement uncertainties, how small is small?

    Reply: In the revision we add information for the standard deviation of gridded data for the discussion.

19. Page 5, line 15-17, Fig. 3: The panels for both February and July show features in all three model distributions that are not seen in the field CO2. For example there is a hotspot on the eastern African coast in the western Indian Ocean that is not seen in the observations (top panel). Likewise in July there is an unexplained hotspot west of South America in the Southern Ocean. So, “the models captured the major features of spatial distribution of observed CO2” plus quite a bit more. Can the authors discuss this further in page 5, line 30 - page 6, line 2?

    Reply: See the reply to question 1.

20. Page 8, line 8: “prediction” should be “predictions “.

    Reply: We corrected the mistake.

21. Acknowledgements. Include, as suggested on SOCAT’s website: “The Surface Ocean CO2 Atlas (SOCAT) is an international effort, endorsed by the International Ocean Carbon Coordination Project (IOCCP), the Surface Ocean Lower Atmosphere Study (SOLAS) and the Integrated Marine Biogeochemistry and Ecosystem Research program (IMBER), to deliver a uniformly quality-controlled surface ocean CO2 database. The many researchers and funding agencies responsible for the collection of data and quality control are thanked for their contributions to SOCAT.

    Reply: We revised the acknowledgements as suggested.

22. Table 1: Add a first column ‘Feature’, e.g., 1-input space mapping, 2-prediction by, 3-problems, 4-data scaling, 5-results affected by. Then revise the SVM, FNN, SOM columns accordingly.

    Reply: We revised the table as suggested.

23. Table 1, line 9: ‘closet’ should be ‘closest’.

    Reply: We corrected the mistake.

24. Figure 3: The labels in white font are too small to read.

    Reply: We enlarged the labels.

Response to referee #2 --------------------------

We thank referee#2 for the thoughtful comments, especially on the appendix. We have revised the appendix substantially to address the reviser's constructive opinions. Here we documented our responses to the reviewer's comments point-by-point.

1. This "technical note" discusses the formation of global maps of surface ocean CO2 from limited measurements using inferred dependence on (latitude, surface temperature SST, salinity, chlorophyll concentration, mixed-layer depth, difference between monthly- and annual-mean SST). The dependence is inferred by three methods: selforganisation map (SOM), feedforward neural network (FNN) and a new method (support vector machine; SVM). The results of these three methods, "trained" on a fraction of the data, are compared with the remaining data. The correlations are not particularly good for any (best at R2 = 0.715 for SVM) considering there are 6 independent variables aiding the fit. However, the results of all three methods for global air-sea CO2 flux are very close and the CO2 maps are visually similar. This similarity extends to a band of high CO2 concentration in February 2005 extending west from Chile where there are apparently no CO2 measurements. This extrapolation from CO2 observations is presumably via a similar feature in (at least) one of the 6 independent variables. There should be more discussion: (i) of the quality of the fit to observed data, especially in relation to the estimates of air-sea flux and the danger that the methods agree with each other more than with reality; (ii) of the extrapolation feature west of Chile (in particular ‐ perhaps also a careful examination for whether there are others) and whether it can be believed in terms of the values of the independent variables – is this set of six values closely approximated somewhere else where there are CO2 measurements constraining the CO2 estimate?

   Reply: First on the quality of fit. As the three model are unbiased, i,e, the mean difference between modeled CO2 and observation is statistically zero, we take the quality here means the correlation and the standard deviation between modeled CO2 and observation. Then quality is not only determined by the capability of the models, but also the variability of the data (we added this information in the revised manuscript).

   Second on the relation to the estimates of air-sea flux. The models may produce differently higher than observed CO2 in some areas and lower CO2 in others. The flux could be largely affect by this and by different wind. That all three models produced similar global fluxes indicates that the effect is small.

   Third on the danger that the methods agree with each other more than with reality. They are quite different issues. The answer to the quality of fit indicates the models cannot agree with reality than the variability of the reality. Whereas, the agreement between models is determined mainly by their similarity. For example, SVM is considered to be a one layer FNN in some articles; so FNN agrees better with SVM than with SOM.

Fourth on the extrapolation. Yes, the extrapolation approximate unmeasured area with somewhere that has a similar biogeochemical property and CO2 measurement.

2. Although the organisation and English are generally good, I think some sections and especially the Appendix are unclear/obscure, mainly due to inconsistent or missing explanations, definitions or notation. Most of the following detailed comments are about this aspect.

Reply: We have revise the appendix substantially to address to issue.

3. Page 2, lines 12 and 18. "dSST denotes the difference between the monthly and annual means of SST" implies 12 discrete values of dSST; how does this "improve expressing the seasonal variable continuously"?

Reply: Let's consider three measurements taken on January 1, January 31, and February 1. Using month as the seasonal variable, the variable values of the first two measurements are 1 and the last is 2. However, the seasonally the last two are nearly the same. dSST reflects better the actual change of seawater property caused by season change.

4. Page 3, Line 13. I think you mean ". . to the range (0, 1) for the SVM . . ."

Reply: We revised the expression accordingly.

5. Page 3,. Line 21 (i.e. line after (5)). Why between 0.1 and 0.9 not between 0 and 1? "better" compared with what? Why should scaling the output help?

Reply: For fCO_2 close to 0 and 1, and a small change in fCO_2 requires very large adjustment of model parameters, which slows down the convergence of training. We added this in the revised manuscript.

6. Page 4, Lines 1-2. "We used Eq. (4) to scale . . SOM". There is no mention of this in Appendix A.1, indeed after (A1) it is stated that the diagonal factors of the scale matrix f are equal to 1.

Reply: In the appendix section for SOM, we added "In our application, the data for each input variable were scaled to be unitless by its mean and standard deviation”.

7. Page 4, Lines 2-3. “Based on preliminary studies, we applied a factor of 2 to . . SST and CHL . .”. What preliminary studies? Is this subjective, i.e. why should SST and CHL be emphasised?

   Reply: SOM is indeed subjective. In our knowledge, other applications scaled the data with different subjective factors to change the impact of independent variables on the distance defined by Eq.(A1). In our application, we scaled the data non-subjectively and uses the scale factors to change the impact, which in our opinion is easier to understand. Because CO_2 shows a much higher correlation with SST and CHL than with others, we subjectively used a factor of 2 for them. There is no theoretical basis for this choice. We revised the manuscript to address the issue and revised “Based on preliminary studies” to “Based on our preliminary correlation analysis”

8. Page 4, Line 7. “prediction for an input” needs explaining. Inputs are supposed to be known, not “predicted”.

   Reply: We revised “Making prediction for an input” to “Making a CO_2 prediction for an input”

9. Page 4, Lines 8, 9. “map size”. In normal language the map size is the earth’s surface area. Do you mean resolution, equivalent to the number of CO_2 output locations? Please explain / use correct word.

   Reply: We revised “the feature map size” to “the number of neuron cells”

10. Page 5, Line 8. “respectively” should be “for SOM”

    Reply: We corrected the mistake.

11. Page 5, Lines 11, 15. Please explain “normalized”/“normalization”

    Reply: We revised the manuscript and explained “normalized”/“normalization”.

12. Page 6. To have value, this needs to be understood in its own terms; the reader should not have to refer to cited references to understand the words used and the overall meaning. Too many words are not defined or explained. Also, it is too abstract. This is a manuscript about

"output" CO_2, depending on "inputs" LAT, SST, SSS, CHL, MLD, dSST. Presumably this applies to A.1, A.2 and A.3 – say so and do not use vague terms like "feature space" – at present the reader has to guess what you mean.

Reply: We have revised the appendix substantially according to address the concern.

13. (A.1 . .) Page 6 Line 23. What is "feature space" in oceanographic terms?

    Reply: We removed the jargon.

14. Lines 23-24. "usually represented by grid points in two dimensional space". Never mind about "usually"; describe in terms of the problem here.

    Reply: We revised the expression to be specific.

15. Line 24. "weight vector w". This name is confusing.

    Reply: We changed the symbol and revised the descriptions.

16. page 7, lines 7-8 weights (weight factors) h are defined by (A3). "w" is the result of applying the weights "h" to combine values of "v" at various locations [presumably to represent "v" at grid locations rather than original locations, but this is not clear to the reader. If this the case, then "w" is "gridded v" or "interpolated v"]. See also the comment on page 7 line 21. Line 25. Not "a data vector" which might refer to any vector at all, but "an input data vector" (I guess).

    Rely: We changed the symbol and revised the descriptions.

17. page 7, Line 30. "best matching cell (BMC)" needs explaining.

    Reply: We revised the descriptions.

18. page 7, Line 30. “minimizing the distance”. What is varied to do this?

    Reply: We revised the descriptions.

19. Page 7 Line 4. “matched”. Either this is the wrong word or it needs explaining.

    Reply: We revised the descriptions.

20. Line 17. “vector x of input data”. In A.1 the input data were “v”. Use consistent names for variables.

    Reply: We revised the descriptions.

21. Lines 20-22. You have input data, hidden neurons and output. There should be distinct variable names for each of these, e.g. v, x, y respectively. Here you have y for the hidden neurons and for the output, which is confusing.

    Reply: We changed the symbols and revised the descriptions.

22. Line 21. “w is the weight vector”. Indeed this seems correct for its use in (A4) but that is very different from its use in (A1). Use different terms for different quantities (c.f. comment on page 6 line 24).

    Reply: We changed the symbols and revised the descriptions.

23. Line 22. “The training updates the offset and weight parameters”. What are the starting values before updating? Do you mean “weight vector” as in line 21?

    Reply: We revised the descriptions. The parameters are initialized randomly between -1 and 1. We added this information in the revised manuscript.

24. Line 23. What is “e” or is it defined by (A5)? Please make this clear.

    Reply: Yes. It is the “e” defined b (A5).

25. Line 24. “modelled . . y” is unclear (especially because you use “y” for hidden neurons and output). Why are two “y” in this line in bold type but not the third or “y” in (A4)?

    Reply: We revised the description. Bold font indicate vector or matrix.

26. Line 24. “w includes both . .” This seems to be defining a vector with more components; it should have a new name.

    Reply: We used a new name.

27. Line 28. “ is the learning rate”. How is its value decided?

    Reply: The initial value is about 0.25. It is determined by try-and-error. A small value make training slow. A large value make a training diverge. We added the information in the revised manuscript.

28. Line 30. “derivatives of e by w”. Do you mean “derivatives of e with respect to w”.

    Reply: We revised the description.

29. Page 8, Lines 6-10. “The SVM . . SVM parameters.” Is this relevant?

    Reply: We removed this part.

30. Line 14. “independent variables”, “high dimensional space”, “target variable”. Please define these in terms of the oceanographic problem in question.

Reply: We removed these jargons.

31. Line 16. “minimizes” – what is varied to do this?

Reply: We revised the description.

32. Lines 18-19. “subjecting to the constraint”. (A11) looks like a definition of “e” and is not a constraint unless “e” is defined in some other way which needs to be stated.

Reply: We revised the description and re-arranged the equation.

33. Line 27. Can there be an explicit expression for '? Where has “b” in (A9) gone to? Table 1. SOM column half way down. “closest” not “closet”!

Reply: We corrected the mistake.

34. Figure 3 caption. Please explain “normalized to 2005”.

Reply: We revised “normalized to 2005” and in the section 4 added that fCO2 means trend-removed fCO2 unless specified otherwise.

Response to referee #3 --------------------------

We thank referee#3 for many valuable comments. As not all questions could be answered satisfactorily without extending the short technical note to a full research paper, the following responses address the referee's opinions in the scope of the technical note.

1. General points: More detail of the exact data application steps are required: Did the application of the methods follow the biogeochemical province-by-province approach of SOCOM, or was all global data combined together?

   Reply: All global data were combined together to trained the models. We did not model the biogeochemical provinces of SOCOM for the reason that not all the provinces have sufficient data for training the models. Dealing with the discontinuity near the borders of provinces are also problematic in global mapping. Although SOCOM compared models by the province-by-province approach, most of the models did not follows the province approach. One of the defined the provinces subjectively, another used SOM to define the provinces, but not of them discussed the border problem in detail.

2. General points: A comment regarding the use of a single trend normalization rate would be welcome. It is known that this is not globally uniform (e.g. Takahashi et al., 2014) and so it would be good to understand the impact of this choice.

   Reply: It's would be interesting to see the impact of using different rates for different areas. However, the approach is a challenge itself as it is difficult to determine the applicable areas for different rates without introducing subjective factors; therefore, it is not realistic for this study that focus on comparing machine learning models.

3. Why are the correlations so much poorer than that achieved by the application of the SOM-FFN approach of Landschutzer et al, 2014)?

   Reply: No model can fit data better than the variability of the data. When CO2 data are subdivided by region or by biogeochemical province, the variability becomes smaller and the data can be fitted better. Landschutzer et al (2014) subdivided the data, so it's not a surprise that their fitting showed better correlations.

4. Within the model validation section, was the random selection of 50% data carried out only once or multiple times? What is the effect of this random selection compared to say, using data clustered around 2005, or only data from regions where pCO2 varies the most, or only

using the most recent data? I would imagine this would be useful information for other researchers looking to apply the methods themselves, whether to map sea surface pCO2 or indeed other biogeochemical parameters. As mentioned above, the study would benefit with comparison with independent dataset e.g. time series at BATS / HOTS. There is very little coverage on uncertainties. More detail on how these are calculated, especially for regions where there are no observational data with which to compare (e.g. South Pacific / Southern Ocean) would be very welcome. This could useful be useful in explaining the anomolous flux feature currently prevalent in Figure 3 in the South Pacific, which is not mentioned in the text and does not appear to be supported by observations or previous studies (e.g. the Takahashi climatology). They are substantial

Reply: The random selection of data was determined by a random number seed. We tested that using different random seeds did change the results significantly. Regarding selecting data clustering around 2005 or recent years, we would like to point out that this may be carried out regionally, but not globally because of scarce measurements. In each month of a year, there might be one or two cruises or none at all doing measurements for the whole globe. Applying the machine learning models to BATS/HOTS should be an independent subject as more data become available the model equation and inputs should be different. For example, the LAT variable should be removed from the model and the measured SST, SSS, and CHL should be used.

5. Figures: - Figure 2 - unity line is not easily seen. Possibly changing the color of data points to gray could remedy this? - Figure 3 - needs larger labelling as to what they are showing. A column title would be useful, and a more color-blind friendly colorscale.

   Reply: We used gray for data points. This improves the figures' appearance.

6. p5 l7 - what do the uncertainties represent? Are these the standard error of the fit, standard deviation of the mean difference between predicted and observed values? How do these compare to other non neural network methods applied during SOCOM?

   Reply: The uncertainty is the standard deviation of the difference between predicted and observed values. We added this to the manuscript. In our opinion, comparing the uncertainties of different models is not meaning full. For example, a model in SOCOM used spline fitting. As we know that spline fitting can fit data perfectly well, but a perfect spline fitting may lead to over interpolation. Another example is SOM. Given a very large number of neuron cells, SOM can also produce perfect fittings, but then the prediction for the spatial distribution of CO2 would be uninterpretable.

7. p5 l9 - what are the measurement uncertainties?

   Reply: The gridded SOCAT includes standard deviation varying from 0.1 μatm to 71.2 μatm. We added this information in the revised manuscript.

8. p5 10 - what is this uncertainty from temperature?

   Reply: Yes, it is. This is not relevant anymore. We used measurements uncertainty for the discussion.

9. p5 11 - what is the average standard deviation of repeat measurements (should also reference)

   Reply: About 12.5 uatm. We added this to the manuscript.

10. p5 13 - why is only july looked at, what is the uncertainty for the full year? How much of this is due to the normalization method?

    Reply: We thought that the manuscript only showed CO2 maps in July and February, so using July as an example was sufficient. Now we included the standard deviation for all months. The effect on the STD by normalization is small. The STD of normalized fCO2 range from 0.1 uatm to 103.1 uatm and the mean is 12.5 uatm; whereas the STD non-normalized fCO2 range from 0.1 uatm to 107.5 uatm and the mean is 14.6

11. p5 25 - there seems some agreement with other studies for 2000 but substantial disagreement with other estimates (Wanninkhof et al., 2013, Rodenbeck et al., 2015) for 2010. This is surprising given that this is when there are most observational data and so it could be assumed that this era would be best modelled. Equally it is rather worrying that the same models as used in the SOCOM study are showing substantially higher estimates for the air-sea CO2 flux for the same input dataset. Is this related to the choice of wind field or how the mapped pCO2 fields are built? How do the mapped pCO2 fields compare with other methods? Some comment on this discrepancy would be greatly appreciated. In particular, comment on how fluxes for years other then 2000 are calculated would be useful as this is not currently explained. Is the systematic trend of 1.5uatm/year simply reintroduced.

    Reply: Yes, the flux estimate is highly dependent on wind products as shown by Wanninkhof et al. (2013) and Zeng et al. (2014). We added a short comment to the manuscript.

12. p5 l27 - the within-model differences are smaller, but this would be expected as they are essentially iterations of a similar technique. More disconcerting is the substantial offset of this group of models with other independent approaches. As mentioned above, more comment/discussion on this aspect would be useful.

    Reply: SOCOM shows that FNN agree well with other models in general. Inter-comparison of model by different authors is important but beyond the scope of this manuscript.

Technical note: Evaluation of three machine learning models for surface ocean CO2 mapping

Jiye Zeng[1], Tsuneo Matsunaga[1], Nobuko Saigusa[1], Tomoko Shirai[1], Shin-ichiro Nakaoka[1], Zheng-Hong Tan[2]

[1]National Institute for Environmental Studies, Tsukuba, Japan
[2]Department of Environmental Science, Hainan University, China

Correspondence to: Jiye Zeng (zeng@nies.go.jp)

Abstract. Reconstructing surface ocean CO_2 from scarce measurements plays an important role in estimating oceanic CO_2 uptake. There are varying degrees of differences among the 14 models included in the Surface Ocean CO_2 Mapping (SOCOM) inter-comparison initiative, in which five models used neural networks. This investigation evaluates two neural networks used in SOCOM, self-organization map and feedforward neural network, and introduces a machine learning model called support vector machine for ocean CO_2 mapping. The technique note provides a practical guide to selecting the models.

1 Introduction

The global ocean is a major sink for anthropogenic carbon and therefore an important contributor for slowing down the human-induced global warming (Stocker et al., 2013). For calculating the oceanic CO_2 uptake, various models have been used to interpolate scarce CO_2 measurements in the surface ocean spatially and temporarily to obtain basin-wide (e.g. Zeng et al., 2002; Lefevre et al., 2005; Chierici et al., 2006; Sarma et al., 2006; Jamet et al., 2007; Friedrich and Oschlies, 2009; Telszewski et al., 2009; Takamura et al., 2010; Landschützer et al., 2013; Nakaoka et al., 2013; Iida et al., 2015; Goddijn-Murphy et al, 2015) and global ocean CO_2 maps (Takahashi et al., 2002, 2009 and 2014; Park et al., 2010. Rödenbeck et al., 2013; Sasse et al., 2013; Jones et al., 2015; Zeng et al., 2015). The Surface Ocean CO_2 Mapping (SOCOM) inter-comparison initiative revealed varying degrees of differences among 14 models (Rödenbeck et al., 2015), of which 5 used neural networks. They include self-organizing maps (SOM) and feedforward neural networks (FNN). The SOM has a long history in CO_2 mapping (Lefevre et al., 2005; Friedrich and Oschlies, 2009; Telszewski et al., 2009; Nakaoka et al., 2013). Recently, the FNN is gaining popularity in this field (Landschützer et al., 2015; Zeng et al., 2014 and 2015). In this investigation we introduce a machine learning model called support vector machine (SVM) for ocean CO_2 mapping and compare the SVM with the SOM and FNN. We intend to provide a practical guide for using these machine learning models.

2 Model Equations

The machine learning models included in this study cannot directly model the long term trend of CO_2. Therefore, we express the dependence of CO_2 fugacity (fCO_2) on year (*YR*), month (*MON*), latitude (*LAT*), and longitude (*LON*) as the sum of a nonlinear static component and a linear trend component:

$$fCO_2 = F_{static}(LAT, LON, MON) + F_{trend}(YR). \quad (1)$$

As available observations are scarce with respect to the biogeochemical properties of the surface ocean, we used sea surface temperature (*SST*), sea surface salinity (*SSS*), chlorophyll-a concentration (*CHL*), and mixed layer depth (*MLD*) as the proxy variables of space and time. These proxy variables were commonly used by models included in the SOCOM. The model equation becomes

$$fCO_2 = F_{static}(LAT, SST, SSS, CHL, MLD, dSST) + F_{trend}(YR), \quad (2)$$

where d*SST* denotes the difference between the monthly and annual means of *SST*. Here we excluded *LON* and *MON*. They have a circular property and therefore cannot be used directly. For instance, longitude -180 degree is geographically connected to longitude 180 degree, but numerically they appear to be two extreme longitude values to the models. Zeng et al. (2014 and 2015) circumvented this problem by using sine and cosine transformed components. Their approach could unintentionally enhance the influence of *LON* and *MON* on fCO_2 as one more derived variable from each of them were added to the model. We excluded *LON* for the belief that the combination of *SST*, *SSS*, *CHL*, and *MLD* contains sufficient spatial information, but retained *LAT* for its different seasonal and geophysical meanings in the northern and southern hemispheres. Replacing *MON* by d*SST* also improves expressing the seasonal variable continuously, especially for those measurements taken near the start or end of a month.

3 Data

We extracted monthly fCO_2 from the track-gridded database of the Surface Ocean CO_2 Atlas (SOCAT) version 3.0[1] (Pfeil et al., 2013; Sabine et al., 2013; Bakker et al. 2013). The database has a 1°×1° spatial resolution and includes global measurements from 1970 to 2014. Similar to Zeng et al. (2014), we excluded some data points by these criteria: (i) fCO_2 values smaller than 250 µatm or larger than 550 µatm, (ii) ocean depth smaller than 500 m, (iii) salinity smaller than 25.0, and (iv) year before 1990. A total of 158,052 data points were extracted with these conditions.

[1] http://www.socat.info/

The monthly SST data of 1990 to 2015 were extracted from the Optimum Interpolation (OI) V2 product[2] of NOAA (Reynolds et al., 2002). The monthly SSS climatology was extracted from the World Ocean Atlas 2013 (WOA13) product[3] (Boyer et al., 2013), which contains the monthly mean SSS from June 27, 1896 to December 25, 2012. The monthly CHL climatology was calculated using the MODIS Aqua and SeaWiFS climatology[4], which covers the period of 2012 to 2015. The mean of the two CHLs was used as the CHL climatology. The mixed layer data were derived from the Monthly Isopycnal and Mixed-layer Ocean Climatology[5] of NOAA (Schmidtko et al., 2013), which includes the period of 1955 to 2012.

4 Machine Learning Models

The Appendix and Table 1 summarize the algorithms of the three models. Here we focus on discussing their usage in CO_2 mapping.

Zeng et. al. (2014) presented a method to model the linear component in Eq.(2). Instead of repeating the process, we used their annual rate of 1.5 μatm to remove trend from fCO_2 to normalize it to the reference year 2005, i.e.,

$$fCO_2^{normalized} = fCO_2 - 1.5 * (YR - 2005) \qquad (3)$$

Although Takahashi et al (2014) obtained a global mean rate of 1.9 μatm yr^{-1}, we used 1.5 μatm yr^{-1} as this rate was obtained by using the gridded fCO_2 of SOCAT version 2. The normalized fCO_2 was used to model the nonlinear component in Eq.(2). In later discussions, fCO_2 means the normalized fCO_2 unless explicitly stated. Similarly, we applied the log transform of Zeng et. al. (2014) to *CHL* prior to data scaling discussed below, i.e.,

$$CHL = \log_{10}(1.0 + CHL). \qquad (4)$$

For a given dataset, the SVM requires a prior step to find the optimal value for the parameter γ in Eq.(A10) and the parameter σ in Eq.(A15). To shorten the training time, we randomly chose 10% of the measurement data in this step and obtained 10 for γ and 0.06 for σ. Note that these values are dependent on data scaling, which is necessary in this case to avoid overflow problem in solving Eq.(A12). We scaled all input variables LAT, SST, SSS, CHL, MLD, and dSST by their minimum and maximum to confine them in the range (0, 1), i.e.,

$$v = \frac{v - v_{min}}{v_{max} - v_{min}}. \qquad (5)$$

---

[2] http://www.esrl.noaa.gov/psd/data/gridded/data.noaa.oisst.v2.html

[3] https://www.nodc.noaa.gov/OC5/woa13/

[4] http://oceancolor.gsfc.nasa.gov/cgi/l3

[5] http://www.pmel.noaa.gov/mimoc/

Data scaling is not necessary for the FNN, but can improve its performance. Following Zeng et al. (2014), we scaled the input variables by their mean and standard deviation as

$$v = \frac{v - \bar{v}}{s}. \tag{6}$$

The output variable fCO_2 is scaled by

$$v = 0.1 + 0.8 \frac{v - v_{min}}{v_{max} - v_{min}}. \tag{7}$$

This confines the scaled fCO_2 between 0.1 and 0.9 for better response to changes of input variables. The kernel function Eq.(A4) has the property that for any input in (-∞,+∞), the output varies between 0 and 1. For fCO_2 close to 0 or 1, a small change in fCO_2 requires very large adjustment of model parameters, which slows down the convergence of training.

We used 64 hidden neurons for the FNN as Zeng et al. (2014) did. The learning rate in Eq.(A6) was set to 0.25 by try-and-error. A small value makes training slow; whereas a large value may make a training diverge. The constant in Eq.(A8) was determine dynamically in each iterative training loop. It was taken as 10 times the mean of absolute differences between modelled and observed fCO_2. We experienced that this method improves the performance of training.

Data scaling is critical for the SOM, as the distance defined by Eq.(A1) would be affected by variable units. We used Eq.(6) to scale input variables in training the SOM. Based on our preliminary correlation analysis, we applied a factor of 2 to enhance the influence of *SST* and *CHL* on the distance. Using such a subjective factor is the only way to make the correlations between the output and the input variables more in line with observed correlations.

From the labelling procedure of SOM described in the Appendix, it is not difficult to see that the number of neuron cells in SOM affects the labelling and hence the prediction. Unfortunately, there is no guideline for choosing the size. Based on previous studies (Telszewski et al., 2009 and Nakaoka et al. 2013), we used 20,000 neuron cells, roughly one neuron cell for one 1x1 grid cell of sampled areas.

5 Model Validation

We examined the goodness of fitting by randomly selecting 10% to 50% of the data points to train the FNN and SVM, and to label the SOM; and then calculated the correlation coefficient between modelled and observed CO_2 of the selected data points.

The SOM yields the best correlation in the case of 10% of randomly selected data points and the correlation decreases with the number of data points (Fig.1). The reason is that for a given number of neuron cells, the fewer the data points, the less possible a neuron cell will be labelled by multiple measurements and the more likely that the prediction will find the same CO_2 value used for labelling. Therefore, the goodness of fit does not necessary mean good SOM modelling.

The correlations obtained by the SVM and FNN do not vary much with the number of data points. While the SVM's correlation decreases monotonically, even though by only a little, with the number of data points, the FNN's correlation obtained with 75000 data points is larger than that with 60000 data points. The FNN is known for not being able to find the global optimum in training. This case could be an indication of an imperfect training. The FNN appears inferior to SVM in all case. However, imperfect training does not account for all the differences. If we use the number of model parameters to be determined by the training as the indicator of the dimension of the model space, the FNN's dimension is far smaller than that of the SVM. The former is determined by the number of hidden neurons and input variables, whereas the latter is determined by the number of training data. For 6 input variables, 15000 training data, and 64 hidden neurons, the number of parameters is 509 for the FNN and 15001 for the SVM.

A better indicator for the performance of the models would be the goodness of prediction. To emulate the situation that the sampled area was only a small portion of the global ocean, we evaluated the goodness of prediction by training FNN and SVM and labelling SOM with 10% of randomly selected data to make prediction for the rest of the data. Fig.2 shows that the SVM yielded the best correlation (R^2=0.72), the FNN fell behind (R^2=0.67), and the SOM performed the worst (R^2=0.54). The differences between predicted and observed fCO_2 are 0.1±17.4 µatm for SVM, 0.1±18.9 µatm for FNN, and 0.2±23.3 µatm for SOM. Comparing to the variation of fCO_2 measurements, these differences are small and their uncertainties are in the same order of magnitude as the variation of measurements. Let's examine the standard deviation (STD) of fCO_2 in those grids having at least 3 data points. The track-gridded fCO_2 in SOCAT version 3.0 includes STD ranging from 0.1 µatm to 71.2 µatm and the mean is 5.2 µatm. Calculating the STD of normalized fCO_2 in the same grids and in the same months of all years yielded a mean of 12.5 µatm in the range of 0.1 µatm to 103.1 µatm. The normalization had little effect on the STD as the calculation for none normalized fCO_2 gives a mean STD of 14.6 µatm in the range of 0.1 µatm to 107.5 µatm.

6 Differences

Figure 3 shows fCO_2 maps in February and July, 2005, which is the reference year for normalization. In the mapping, we randomly selected 50% of the data to train the FNN and SVM and to label the SOM. All models captured the major features of observed fCO_2 distribution. The SOM exhibits obvious discontinuity because of its discrete characteristics of picking up fCO_2 values from the labelled SOM. For year 2005, the mean fCO_2 difference is -0.05±12.73 µatm for FNN-SVM and -0.6±18.80 for SOM-SVM. The uncertainty is the standard deviation of the mean difference between predicted and observed values. The statistics indicates that FNN agrees better with SVM than SOM does.

Although the differences among models might be on the order of 10 to 20 μatm, the effect on the global ocean CO_2 flux estimate is small (Fig.4). The fluxes are calculated using the wind speed from ECMWF's interim product (Deea et al., 2011). Our estimate for the oceanic uptake is on the higher end among those in Wanninkhof et al. (2013) and Le Quéré et al. (2015). For example, Wanninkhof et al. (2013) reported that the median sea–air anthropogenic CO_2 fluxes centered on year 2000 ranged from 1.9 to 2.5 PgC yr^{-1} among the seven models. In comparison, our estimates by the three models are about 2.4 PgC yr^{-1}. The mean difference of CO_2 flux is 0.02 PgC yr^{-1} between the FNN and the SVM (FNN-SVM) and 0.06 PgC yr^{-1} between the SOM and the SVM (SOM-SVM). They are small in comparison with those differences among the models in Wanninkhof et al. (2013) and Le Quéré et al. (2015). Note that the flux estimate is highly dependent on wind products as shown by Wanninkhof et al. (2013) and Zeng et al. (2014).

On the spatial scale of tens of degrees, the three models show good mutual agreement for modelled fCO_2 distributions among them. However, each model shows distinguished fine structures, which are determined by the biogeochemical processes in the ocean, by model parameters obtained from training, and by the characteristics of the models. We believe that the modelled monthly fCO_2 distributions are true to the degree given by the model validations.

7 Summary

The main features of the three machine models are listed in Table 1. The SVM is recommended when the computer has enough memory to store the matrix in Eq.(A12), which is proportional to the square of the number of training data. The SVM performs the best, but the training time could become very long when the number of training data is too large to be handled by a computer without using virtual memory. For any given dataset, using the SVM requires a prior step to find the optimal value for the parameter γ in Eq.(A10) and the parameter σ in Eq.(A15).

The FNN model does not perform as well as the SVM, but the number of training data does not affect its training as much as the SVM's. The training time can become long when a large number of hidden neurons are used and many iterations are needed to achieve convergence. It takes longer time to train the FNN than the SVM for a small number of data points. However, the FNN is simpler to use as it requires no prior step.

The SOM is recommended only when the other two models have over fitting or over interpolation problems. The SOM performs the worst and is not as straightforward as the others as its result depends too much on data scaling and the number of neurons. An advantage of the SOM is that once trained, re-labelling the SOM with new CO_2 measurements and making a new prediction is fast. Although the SOM does not have the over interpolation problem of the other two, it may produce nonsense predictions due to its strong dependence on data scaling.

Appendix

A.1 Self-Organization Map

A self-organizing map (SOM) is a type of artificial neural network that is trained using unsupervised learning (Kohonen, 1984). The SOM in our application comprises grid points on a two dimensional plane. Each grid point, also called neuron cell, has the same number of parameters as the input variables, which include LAT, SST, SSS, CHL, MLD, and dSST in our case. Training the SOM is to use samples of input variables to adjust the parameters to make neighbourhood neuron cells having similar parameter values that reflect certain biogeochemical features of the surface ocean.

We used the batch learning algorithm (Abe et al., 2002) to train the SOM as the result does not depend on the sequential order of training samples. The parameters were initialized randomly in the range (-1,1). In each iterative training loop, each training sample is associated with a neuron cell to which the distance defined as follow is the smallest than to other neuron cells:

$$d = |\mathbf{f}(\mathbf{p} - \mathbf{x})|, \quad \text{(A1)}$$

where **p** denotes the vector of neuron cell parameters, **x** the vector of input variables, and **f** the scale matrix that we introduced to change the influence of certain variables on the distance. The components of **f** are all zero except for those on the diagonal, which are set to 1 by default. In our application, the data for each input variable were scaled to be unitless by its mean and standard deviation to eliminate the effect of units on the distance.

The associated neuron cell is called the best matching cell (BMC). After the BMC for all training samples are found, the parameters are updated by

$$p_i = \frac{\sum_k h_{ik} x_k}{\sum_k h_{ik}}, \quad \text{(A2)}$$

where i and k denote indexes of neuron cells and training samples, respectively. The neighbourhood function that determines the weight factor h is defined as

$$h_{ik} = \exp\left(-\frac{|\mathbf{r}_{ik}|}{q}\right), \quad \text{(A3)}$$

where $|\mathbf{r}_{ik}|$ denotes the geographic distance between the ith neuron cell and the BMC of the kth training sample and q is a factor that decreases linearly with iteration loop. In another words, the procedure adjusts the parameters of neuron cells toward those training samples whose BMC are close to them and the amount of adjustment decreases exponentially with the geographic distance between neuron cells and linearly with the training loop.

The trained SOM needs to be labelled by $f\mathrm{CO_2}$ for making prediction. The values of $f\mathrm{CO_2}$ measurements are assigned to their BMC. Predicting $f\mathrm{CO_2}$ for a set of input variables is realized by finding the BMC labelled with fCO2 and extract its mean $f\mathrm{CO_2}$ value.

A.2 Feedforward Neural Network

A feedforward neural network (FNN) is an artificial neural network that is trained using supervised learning. Our FNN comprises three layers (Zeng et al., 2014): An input layer, a hidden layer, and an output layer. The number of neurons in the input layers is determined by the number of input variables, i.e., LAT, SST, SSS, CHL, MLD, and dSST in our case. The output layer has only one neuron for $f\mathrm{CO_2}$. Each neuron in the hidden layer uses the following kernel function to transform all input variables:

$$y_h = \frac{1}{1+\exp(-(b+\mathbf{w}\cdot\mathbf{x}))}, \qquad \text{(A4)}$$

where **w** denotes the vector of weight parameters and b the offset parameter. The y_h of all hidden neurons become the inputs of the output neuron, which uses the same kernel function to transform y_h to produce $f\mathrm{CO_2}$.

The training updates the offset and weight parameters, which are initialized randomly in the range (-1,1), by minimizing the cost function

$$f(\mathbf{w}') = \frac{1}{2}\mathbf{e}\cdot\mathbf{e} = \frac{1}{2}|\mathbf{y}_m - \mathbf{y}_o|. \qquad \text{(A5)}$$

where $\mathbf{w}'$ is the extended vector that include b and **w**; $\mathbf{y}_\mathrm{m}$ and $\mathbf{y}_\mathrm{o}$ stand for the vectors of modelled and observed fCO2, respectively. In the gradient descent training algorithm, updating $\mathbf{w}'$ at the training iteration t can be expressed as

$$\mathbf{w}'(\mathrm{t}) = \mathbf{w}'(t-1) - \alpha\mathbf{g} \qquad \text{(A6)}$$

where α is the learning rate (a positive number smaller than 1), and **g** the first-order derivative of the cost function:

$$\mathbf{g} = \nabla f(\mathbf{w}') = \mathbf{J}^\mathrm{T}\mathbf{e}, \qquad \text{(A7)}$$

where **J** is the Jacobian matrix whose components are derivatives of **e** with respect to **w'** using back propagation method. We used the efficient Levenberg-Marquardt algorithm (Wilamowski et al., 2010), which derives the gradient as

$$\mathbf{g} = (\mathbf{J}^T\cdot\mathbf{J} + \mu\mathbf{I})^{-1}\mathbf{J}^T\,\mathbf{e}, \qquad \text{(A8)}$$

where μ is a constant.

A.3 Support Vector Machine

A support vector machine (SVM) is a supervised learning model that was conceptualized in the in 1960s for classification problems and later extended to regression analysis (Basak et al., 2007). We used the so called least-square support vector machine for regression (Pelckmans et al., 2002) which, similar to FNN, seeks to minimize the error between model outputs and measurements. The SVM models the dependence of $f\mathrm{CO_2}$ on LAT, SST, SSS, CHL, MLD, and dSST as

$$\mathbf{y}(\mathbf{x}) = \mathbf{c}^T\varphi(\mathbf{x}) + b \qquad \text{(A9)}$$

where **y** is the vector of outputs, **x** the vector of inputs, **c** the vector of coefficients, b the offset parameter, and φ the kernel function. The goal of training SVM is to minimizes the cost function

$$F(\mathbf{c}) = \frac{1}{2}(\mathbf{c}^T\mathbf{c} + \gamma|\boldsymbol{e}|) \qquad \text{(A10)}$$

where

$$\boldsymbol{e} = \mathbf{y}(\mathbf{x}) - \mathbf{c}^T\varphi(\mathbf{x}) - b \quad \text{(A11)}$$

and γ is a constant whose optimal value depends on the data used for training. By applying the Lagrangian multiplier, the optimization problem eventually becomes solving the linear equation of

$$\begin{bmatrix} 0 & \mathbf{u}^T \\ \mathbf{u} & \mathbf{\Omega} \end{bmatrix} \begin{bmatrix} b \\ \boldsymbol{\alpha} \end{bmatrix} = \begin{bmatrix} 0 \\ \mathbf{y} \end{bmatrix}, \quad \text{(A12)}$$

where $\mathbf{u}$ is a vector with all components being 1, and the components of $\mathbf{\Omega}$ are

$$\Omega_{ij} = \varphi(\mathbf{x}_i)^T\varphi(\mathbf{x}_j). \quad \text{(A13)}$$

Once Eq.(a12) is solved, making a prediction is done by

$$y_i = \sum_j^n \alpha_j \varphi(\mathbf{x}_i)^T\varphi(\mathbf{x}_j) \quad \text{(A14)}$$

In this investigation, we used the radial basis kernel function, i.e.,

$$\varphi(\mathbf{x}_i)^T\varphi(\mathbf{x}_j) = \exp\left(-\frac{\|\mathbf{x}_i - \mathbf{x}_j\|^2}{2\sigma^2}\right), \quad \text{(A15)}$$

where σ is a parameter whose optimal value depends on the data used for training.

Acknowledgements

The Surface Ocean CO2 Atlas (SOCAT) is an international effort, endorsed by the International Ocean Carbon Coordination Project (IOCCP), the Surface Ocean Lower Atmosphere Study (SOLAS) and the Integrated Marine Biogeochemistry and Ecosystem Research program (IMBER), to deliver a uniformly quality-controlled surface ocean CO2 database. The many researchers and funding agencies responsible for the collection of data and quality control are thanked for their contributions to SOCAT.

References

Abe, T., S. Kanaya, M. Kinouchi, Y. Ichiba, T. Kozuki, T. Ikemura, 2002. A Novel Bioinformatic Strategy for Unveiling Hidden Genome Signatures of Eukaryotes: Self-Organizing Map of Oligonucleotide Frequency. Genome Informatics 13: 12–20.

Bakker, D. C. E., B. Pfeil, K. Smith, S. Hankin, A. Olsen, S. R. Alin, C. Cosca, S. Harasawa, A. Kozyr, Y. Nojiri, K. M. O'Brien, U. Schuster, M. Telszewski, B. Tilbrook, C. Wada, J. Akl, L. Barbero, N. R. Bates, J. Boutin, Y. Bozec, W.-J. Cai, R. D. Castle, F. P. Chavez, L. Chen, M. Chierici, K. Currie, H. J. W. de Baar, W. Evans, R. A. Feely, A. Fransson, Z. Gao, B. Hales, N. J. Hardman-Mountford, M. Hoppema, W.-J. Huang, C. W. Hunt, B. Huss, T. Ichikawa, T. Johannessen, E. M. Jones, S. D. Jones, S. Jutterström, V. Kitidis, A. Körtzinger, P. Landschützer, S. K. Lauvset, N. Lefèvre, A. B. Manke, J. T. Mathis, L. Merlivat, N. Metzl, A. Murata, T. Newberger, A. M. Omar, T. Ono, G.-H. Park, K. Paterson, D. Pierrot, A. F. Ríos, C. L. Sabine, S. Saito, J. Salisbury, V. V. S. S. Sarma, R. Schlitzer, R. Sieger, I. Skjelvan, T. Steinhoff, K. F. Sullivan, H. Sun, A. J. Sutton, T. Suzuki, C. Sweeney, T. Takahashi, J. Tjiputra, N. Tsurushima, S. M. A. C. van Heuven, D. Vandemark, P. Vlahos, D. W. R. Wallace, R. Wanninkhof, and A. J. Watson, 2014. An update to the Surface Ocean CO2 Atlas (SOCAT version 2). Earth Syst. Sci. Data, 6: 69–90, doi:10.5194/essd-6-69-2014

Basak D, S, Pal S, and D. D. Patranabis. 2007. Support vector regression. Neural Information Processing-Letters and Reviews 11: 203–224.

Boyer, T.P., J. I. Antonov, O. K. Baranova, C. Coleman, H. E. Garcia, A. Grodsky, D. R. Johnson, R. A. Locarnini, A. V. Mishonov, T.D. O'Brien, C.R. Paver, J.R. Reagan, D. Seidov, I. V. Smolyar, and M. M. Zweng, 2013. World Ocean Database 2013, NOAA Atlas NESDIS 72, S. Levitus, Ed., A. Mishonov, Technical Ed.; Silver Spring, MD, 209 pp.

Chierici, M., A. Fransson, and Y. Nojiri, 2006. Biogeochemical processes as drivers of surface fCO2 in contrasting provinces in the subarctic North Pacific Ocean. Glob. Biogeochem. Cycle, 20, GB1009, doi:10.1029/2004GB002356.

Deea, D. P., S. M. Uppalaa, A. J. Simmonsa, P. Berrisforda, P. Polia, S. Kobayashib, U. Andraec, M. A. Balmasedaa, G. Balsamoa, P. Bauera, P. Bechtolda, A. C. M. Beljaarsa, L. van de Bergd, J. Bidlota, N. Bormanna, C. Delsola, R. Dragania, M. Fuentesa, A. J. Geera, L. Haimbergere, S. B. Healya, H. Hersbacha, E. V. H´olma, L. Isaksena, P. K°allbergc, M. K¨ohlera, M. Matricardia, A. P. McNallya, B. M. Monge-Sanzf , J.-J. Morcrettea, B.-K. Parkg, C. Peubeya, P. de Rosnaya, C. Tavolatoe, J.-N. Th´epauta and F. Vitarta. 2011. The ERA-Interim reanalysis: configuration and performance of the data assimilation system. Quarterly Journal of the Royal Meteorological Society, 137: 553–597.

Friedrich, T., and A. Oschlies, 2009. Neural network-based estimates of North Atlantic surface pCO2 from satellite data: A methodological study. J. Geophys. Res. 114, C03020, doi:10.1029/2007JC004646.

Goddijn-Murphy, L. M., D. K. Woolf, P. E. Land, J. D. Shutler, and C. Donlon, 2015. The OceanFlux Greenhouse Gases methodology for deriving a sea surface climatology of CO2 fugacity in support of air–sea gas flux studies. Ocean Science 11: 519-541. doi:10.5194/os-11-519-2015.

Iida, Y., A. Kojima, Y. Takatani, T. Nakano, H. Sugimoto, T. Midorikawa and M. Ishii, 2015. Trends in pCO_2 and sea-air CO_2 flux over the global open oceans for the last two decades, J. Oceanogr.1-25.

Jamet, C., C. Moulin, and N. Lefevre, 2007. Estimation of the oceanic pCO2 in the North Atlantic from VOS lines in-situ measurements: parameters needed to generate seasonally mean maps. Ann. Geophys., 25: 2247-2257.

Jones, S. D., C. L. Quéré, C. Rödenbeck, A. C. Manning, and A. Olsen, 2015. A statistical gap-filling method to interpolate global monthly surface ocean carbon dioxide data, J. Adv. Model. Earth Syst., 7: 1942–2466.

Kohonen, T., Self-Organization and Associative Memory, Springer, Berlin, 1984.

Landschützer, P., N. Gruber, D. C. E. Bakker, U. Schuster, S. Nakaoka, M. R. Payne, T. Sasse, and J. Zeng, 2013. A Neural Network-based Estimate of the Seasonal to Inter-annual Variability of the Atlantic Ocean Carbon Sink. Biogeosciences Discuss., 10: 8799-8849.

Landschützer, P., N. Gruber, F. Haumann, C. Rödenbeck, D. Bakker, S. van Heuven, M. Hoppema, N. Metzl, C. Sweeney, T. Takahashi, B. Tilbrook, and R. Wanninkhof, 2015. The reinvigoration of the Southern Ocean carbon sink, Science, 349: 1221–1224.

Le Quéré1, C., R. Moriarty, R. M. Andrew, J. G. Canadell, S. Sitch, J. I. Korsbakken, P. Friedlingstein, G. P. Peters, R. J. Andres, T. A. Boden, R. A. Houghton, J. I. House, R. F. Keeling, P. Tans, A. Arneth, D. C. E. Bakker, L. Barbero, L. Bopp, J. Chang, F. Chevallier, L. P. Chini, P. Ciais, M. Fader, R. A. Feely, T. Gkritzalis, I. Harris, J. Hauck, T. Ilyina, A. K. Jain, E. Kato, V. Kitidis, K. Klein Goldewijk, C. Koven, P. Landschützer, S. K. Lauvset, N. Lefèvre, A. Lenton, I. D. Lima, N. Metzl, F. Millero, D. R. Munro, A. Murata, J. E. M. S. Nabel, S. Nakaoka, Y. Nojiri, K. O'Brien, A. Olsen, T. Ono, F. F. Pérez, B. Pfeil, D. Pierrot, B. Poulter, G. Rehder, C. Rödenbeck, S. Saito, U. Schuster, J. Schwinger, R. Séférian, T. Steinhoff, B. D. Stocker, A. J. Sutton, T. Takahashi, B. Tilbrook, I. T. van der Laan-Luijkx, G. R. van der Werf, S. van Heuven, D. Vandemark,

N. Viovy, A. Wiltshire, S. Zaehle, and N. Zeng, 2015. Global Carbon Budget 2015. Earth Syst. Sci. Data, 7: 349–396, doi:10.5194/essd-7-349-2015.

Lefevre, N., A. J. Watson, and A. R. Watson, 2005. A comparison of multiple regression and neural network techniques for mapping in situ pCO(2) data. Tellus Ser. B-Chem. Phys. Meteorol., 57: 375-384.

Nakaoka, S., M. Telszewski, Y. Nojiri, S. Yasunaka, C. Miyazaki, H. Mukai, and N. Usui, 2013. Estimating temporal and spatial variation of ocean surface pCO2 in the North Pacific using a Self Organizing Map neural network technique. Biogeosciences, 10: 6093-6106.

Park, G.-H., R. Wanninkhof, S.C. Doney, T. Takahashi, K. Lee, R.A. Feely, C.L. Sabine, J. Triñanes, and I.D. Lima, 2010. Variability of global net sea-air CO2 fluxes over the last three decades using empirical relationships, Tellus B, 62: 352-368. 10.1111/j.1600-0889.2010.00498.x.

Pelckmans, K., J.A.K. Suykens, T.V. Gestel, J.D. Brabanter, B. Hamers, D. Moor, and J. Vandewalle, 2002. LS-SVMlab: a MATLAB/C toolbox for Least Squares Support Vector Machines, http://www.esat.kuleuven.ac.be/sista/lssvmlab, presented at Neural Information Processing Systems (NIPS 2002).

Pfeil, B., A. Olsen, D. C. E. Bakker, S. Hankin, H. Koyuk, A. Kozyr, J. Malczyk, A. Manke, N. Metzl, C. L. Sabine, J. Akl, S. R. Alin, N. Bates, R. G. J. Bellerby, A. Borges, J. Boutin, P. J. Brown, W.-J. Cai, F. P. Chavez, A. Chen, C. Cosca, A. J. Fassbender, R. A. Feely, M. González-Dávila, C. Goyet, B. Hales, N. Hardman-Mountford, C. Heinze, M. Hood, M. Hoppema, C. W. Hunt, D. Hydes, M. Ishii, T. Johannessen, S. D. Jones, R. M. Key, A. Körtzinger, P. Landschützer, S. K. Lauvset, N. Lefèvre, A. Lenton, A. Lourantou, L. Merlivat, T. Midorikawa, L. Mintrop, C. Miyazaki, A. Murata, A. Nakadate, Y. Nakano, S. Nakaoka, Y. Nojiri, A. M. Omar, X. A. Padin, G.-H. Park, K. Paterson, F. F. Perez, D. Pierrot, A. Poisson, A. F. Ríos, J. M. Santana-Casiano, J. Salisbury, V. V. S. S. Sarma, R. Schlitzer, B. Schneider, U. Schuster, R. Sieger, I. Skjelvan, T. Steinhoff, T. Suzuki, T. Takahashi, K. Tedesco, M. Telszewski, H. Thomas, B. Tilbrook, J. Tjiputra, D. Vandemark, T. Veness, R. Wanninkhof, A. J. Watson, R. Weiss, C. S. Wong, and H. Yoshikawa-Inoue, 2013. A uniform, quality controlled Surface Ocean CO2 Atlas (SOCAT). Earth Syst. Sci. Data, 5: 125-143, doi:10.5194/essd-5-125-2013

Reynolds, R. W., N.A. Rayner, T. M. Smith, D. C. Stokes, and W. Wang, 2002. An Improved In Situ and Satellite SST Analysis for Climate. J. CLIMATE, 15: 1609-1625.

Rödenbeck, C., R.F., Keeling, D.C.E. Bakker, N. Metzl, A. Olsen, C. Sabine, and M. Heimann, 2013. Global surface-ocean pCO2 and sea–air CO2 flux variability from an observation-driven ocean mixed-layer scheme, Ocean Sci., 9: 193–216.

Rödenbeck, C., D. C. E. Bakker, N. Gruber, Y. Iida, A. R. Jacobson, S. Jones, P. Landschützer, N. Metzl, S. Nakaoka, A. Olsen, G.-H. Park, P. Peylin, K. B. Rodgers, T. P. Sasse, U. Schuster, J. D. Shutler, V. Valsala, R. Wanninkhof, and J. Zeng, 2015. Data-based estimates of the ocean carbon sink variability – first results of the Surface Ocean pCO2 Mapping intercomparison (SOCOM). Biogeosciences, 12: 7251–7278.

Sabine, C. L., S. Hankin, H. Koyuk, D. C. E. Bakker, B. Pfeil, A. Olsen, N. Metzl, A. Kozyr, A. Fassbender, A. Manke, J. Malczyk, J. Akl, S. R. Alin, R. G. J. Bellerby, A. Borges, J. Boutin, P. J. Brown, W.-J. Cai, F. P. Chavez, A. Chen, C. Cosca, R. A. Feely, M. Gonz´alez-D´avila, C. Goyet, N. Hardman-Mountford, C. Heinze, M. Hoppema, C. W. Hunt, D. Hydes, M. Ishii, T. Johannessen, R. M. Key, A. Körtzinger, P. Landschützer, S. K. Lauvset, N. Lefèvre, A. Lenton1, A. Lourantou, L. Merlivat, T. Midorikawa, L. Mintrop, C. Miyazaki, A. Murata, A. Nakadate, Y. Nakano, S. Nakaoka, Y. Nojiri, A. M. Omar, X. A. Padin, G.-H. Park, K. Paterson, F. F. Perez, D. Pierrot, A. Poisson, A. F. Ríos, J. Salisbury, J. M. Santana-Casiano, V. V. S. S. Sarma, R. Schlitzer, B. Schneider, U. Schuster, R. Sieger, I. Skjelvan, T. Steinhoff, T. Suzuki, T. Takahashi, K. Tedesco, M. Telszewski, H. Thomas, B. Tilbrook, D. Vandemark, T. Veness, A. J. Watson, R. Weiss, C. S.Wong, and H. Yoshikawa-Inoue, 2013. Surface Ocean CO2 Atlas (SOCAT) gridded data products. Earth Syst. Sci. Data, 5: 145-153, doi:10.5194/essd-5-145-2013

Sasse, T. P., B.I., McNeil, and G. Abramowitz, 2013. A new constraint on global air-sea CO2 fluxes using bottle carbon data, Geophys. Res. Lett., 40: 1594-1599.

Sarma, V. V. S. S., T. Saino, K. Sasaoka, Y. Nojiri, T. Ono, M. Ishii, H. Y. Inoue, and K. Matsumoto, 2006. Basin-scale pCO2 distribution using satellite sea surface temperature, Chla, and climatological salinity in the North Pacific in spring and summer. Glob. Biogeochem. Cycle, 20, doi:10.1029/2005GB002594, 2006.

Schmidtko, S., G. C. Johnson, and J. M. Lyman, 2013. MIMOC: A global monthly isopycnal upper-ocean climatology with mixed layers. Journal of Geophysical Research, 118: 1658–1672, doi: 10.1002/jgrc.20122.

Stocker, T., Qin, D., and Platner, G.-K., 2013. Climate Change 2013 The Physical Science Basis, Cambridge University Press, Cambridge, United Kingdom.

Takahashi, T., S.C. Sutherland, C. Sweeney, A. Poisson, N. Metzl, B. Tilbrook, N. Bates, R. Wanninkhof, R.A. Feely, C. Sabine, J. Olafsson, Y. Nojiri, 2002. Global sea-air CO2 flux based on climatological surface ocean pCO2, and seasonal biological and temperature effects. Deep-Sea Research Part II, 49, 1601-1622.

Takahashi, T., S.C. Sutherland, R. Wanninkhof, C. Sweeney, R.A. Feely, D.W. Chipman, B. Hales, G. Friederich, F. Chavez, C. Sabine, A. Watson, D.C.E. Bakker, U. Schuster, N. Metzl, H. Yoshikawa-Inoue, M. Ishii, T. Midorikawa, Y. Nojiri, A. Körtzinger, T. Steinhoff, M. Hoppema, J. Olafsson, T.S. Arnarson, B. Tilbrook, T. Johannessen, A. Olsen, R. Bellerby, and C.S. Wong, 2009. Climatological mean and decadal change in surface ocean pCO2, and net sea-air CO2 flux over the global oceans. Deep-Sea Research Part II, 56: 554-577.

Takahashi, T., S.C. Sutherland, D.W. Chipman, J.G. Goddard, Cheng Ho, Timothy Newberger, Colm Sweeney, and D.R. Munro, 2014. Climatological distributions of pH, pCO2, total CO2, alkalinity, and CaCO3 saturation in the global surface ocean, and temporal changes at selected locations. Marine Chemistry 164: 95-145.

Takamura, T. R., H. Y. Inoue, T. Midorikawa, M. Ishii, and Y. Nojiri, 2010. Seasonal and Inter-Annual Variations in pCO2sea and Air-Sea CO2 Fluxes in Mid-Latitudes of the Western and Eastern North Pacific during 1999-2006: Recent Results Utilizing Voluntary Observation Ships. Journal of the Meteorological Society of Japan, 88: 883-898.

Telszewski, M., A. Chazottes, U. Schuster, A.J. Watson, C. Moulin, D.C.E. Bakker, M. González-Dávila, T. Johannessen, A. Körtzinger, H. Lüger, A. Olsen, A. Omar, X. A. Padin, A. F. Ríos, T. Steinhoff, M. Santana-Casiano, D. W. R. Wallace, and R. Wanninkhof, 2009. Estimating the monthly pCO2 distribution in the North Atlantic using a self-organizing neural network. Biogeosciences, 6: 1405-1421, doi:10.5194/bg-6-1405-2009.

Wanninkhof, R., G.-H. Park, T. Takahashi, C. Sweeney, R. Feely, Y. Nojiri, N. Gruber, S.C. Doney, G.A. McKinley, A. Lenton, C. Le Quéré, C. Heinze, J. Schwinger, H. Graven, and S. Khatiwala, 2013. Global ocean carbon uptake: magnitude, variability and trends. Biogeosciences, 10: 1983-2000, doi:10.5194/bg-10-1983-2013.

Wilamowski, B. M., and H. Yu, 2010: Improved Computation for Levenberg-Marquardt Training. Ieee Transactions on Neural Networks, 21: 930-937.

Zeng, J. Y., Y. Nojiri, P. P. Murphy, C. S. Wong, and Y. Fujinuma, 2002. A comparison of Delta pCO(2) distributions in the northern North Pacific using results from a commercial vessel in 1995-1999. Deep-Sea Research Part II, 49: 5303-5315.

Zeng, J., Y. Nojiri, P. Landschützer, M. Telszewski, and S. Nakaoka, 2014. A global surface ocean fCO2 climatology based on a feedforward neural network, J. Atmos. Ocean Technol., 31: 1838–1849.

Zeng, J., Y. Nojiri, S. Nakaoka, H. Nakajima, and T. Shirai, 2015. Surface ocean CO2 in 1990-2011 modelled using a feed-forward neural network, Geoscience Data Journal, 2: 47–51

Table 1 Feature comparison of the three machine learning models.

Feature	SVM	FNN	SOM
Input space projection	Projects the input variable space to a high dimensional space that is proportional to the number training samples.	Projects the input space to a high dimensional space that is proportional to the number of hidden neurons and input variables.	Projects the input space to a feature space whose size is determined by the number of neurons.
Prediction by	Continuous interpolation.	Continuous interpolation.	Picking up labelling samples that have the closest feature to the input.
Problems	May over fit and over interpolate.	May over fit and over interpolate.	Discrete interpolation leads to spatial discontinuity.
Data scaling	Helps solving the linear equation, but has no effect on the result.	Helps the convergence of training, but has insignificant effect on the result.	Significant effect on the result.
Results affected by	The parameter values for regularization and kernel function.	The number of hidden neurons.	The number of neurons and data scaling.

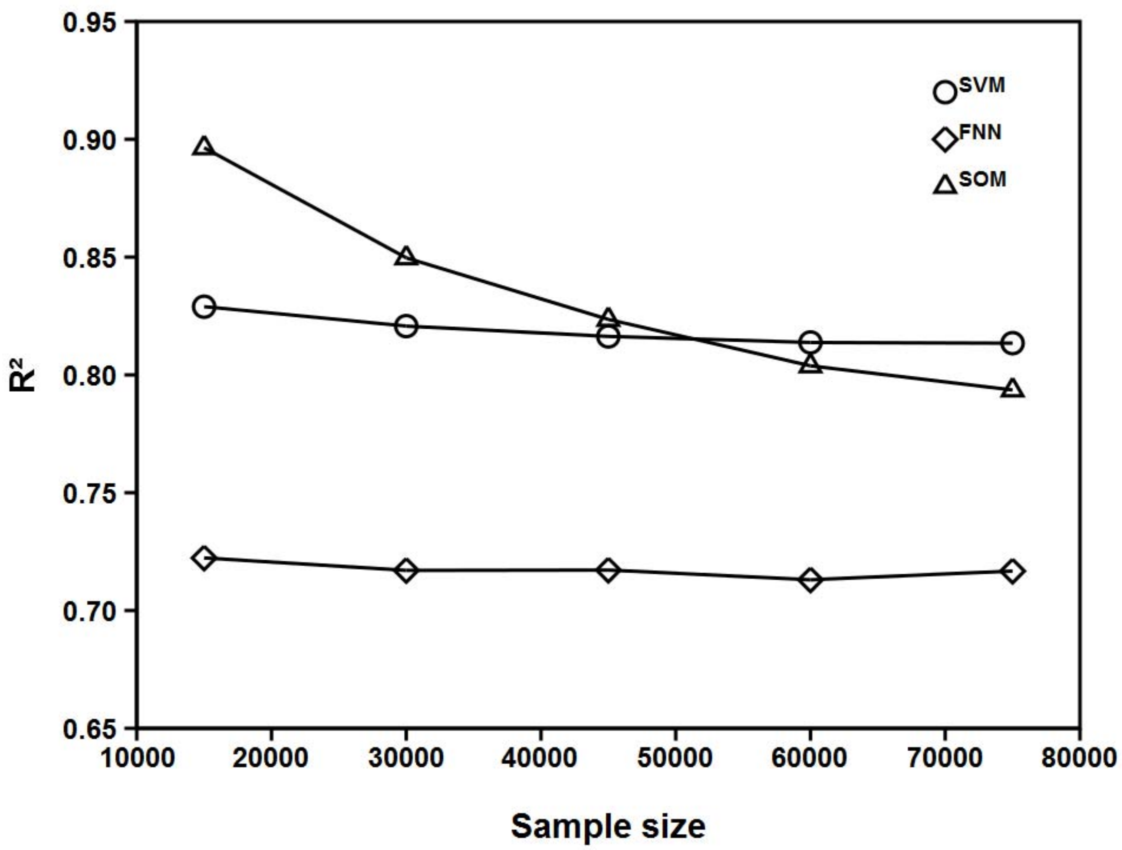
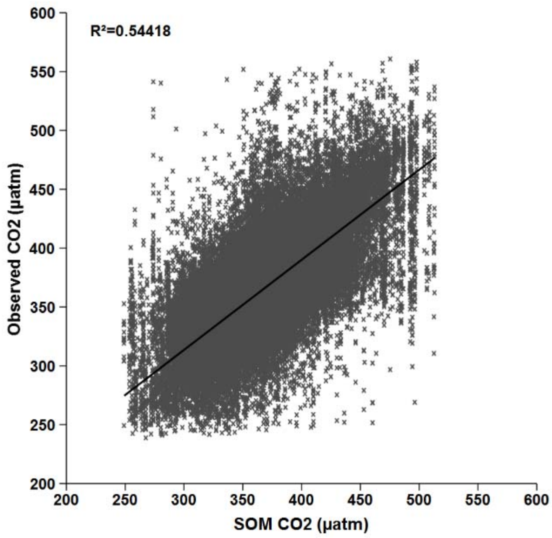
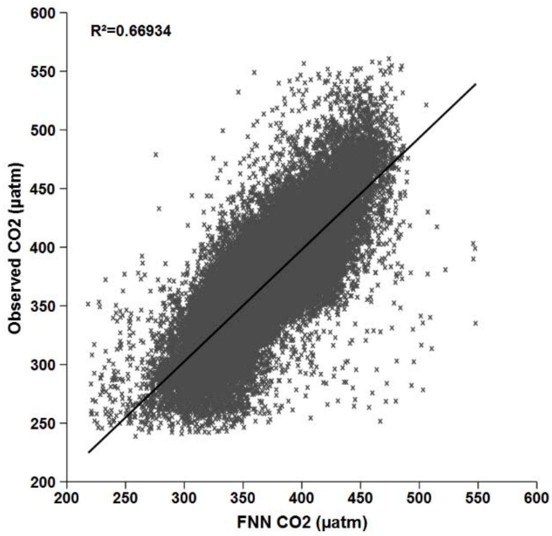


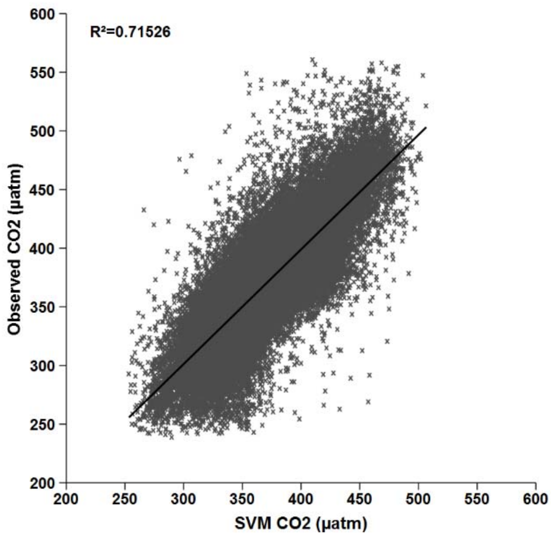
Figure 1: Correlation coefficient between modelled and observed *f*CO_2 (uatm). The sample size is the number of data points randomly selected to train FFN and SVM and to label SOM.



(A)



(B)



(C)

Figure 2: Predicted vs observed *f*CO_2 (µatm). Ten percent of data points was selected randomly to train FNN and SVM and to label SOM, and the rest was used for validation.

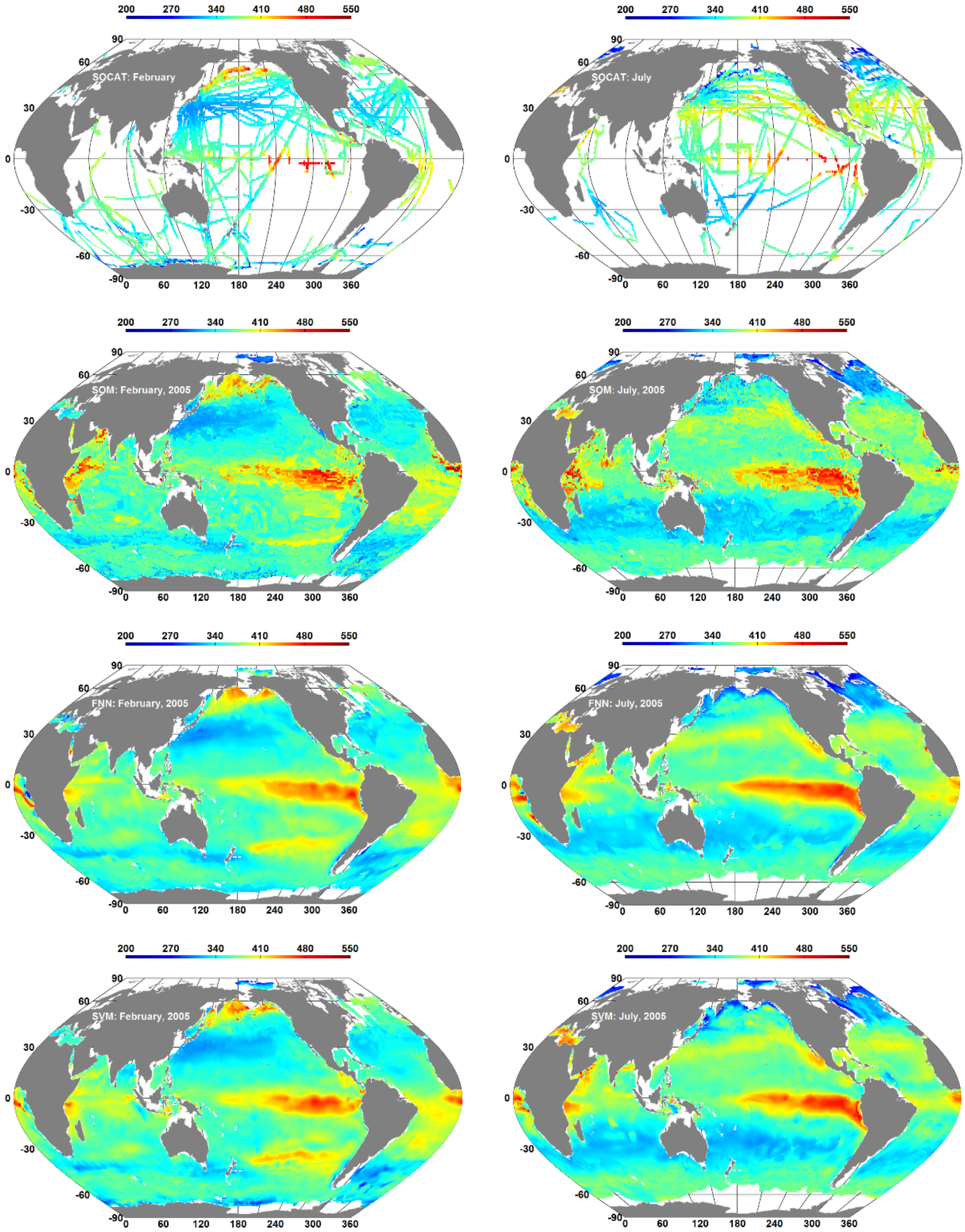


Figure 3: Distributions of modelled and observed fCO_2. The composite map for observations includes fCO_2 in 1990-2014. Half of randomly selected data points were used to train FNN and SVM and to label SOM to make prediction. The left panels show February and the right panels show July.

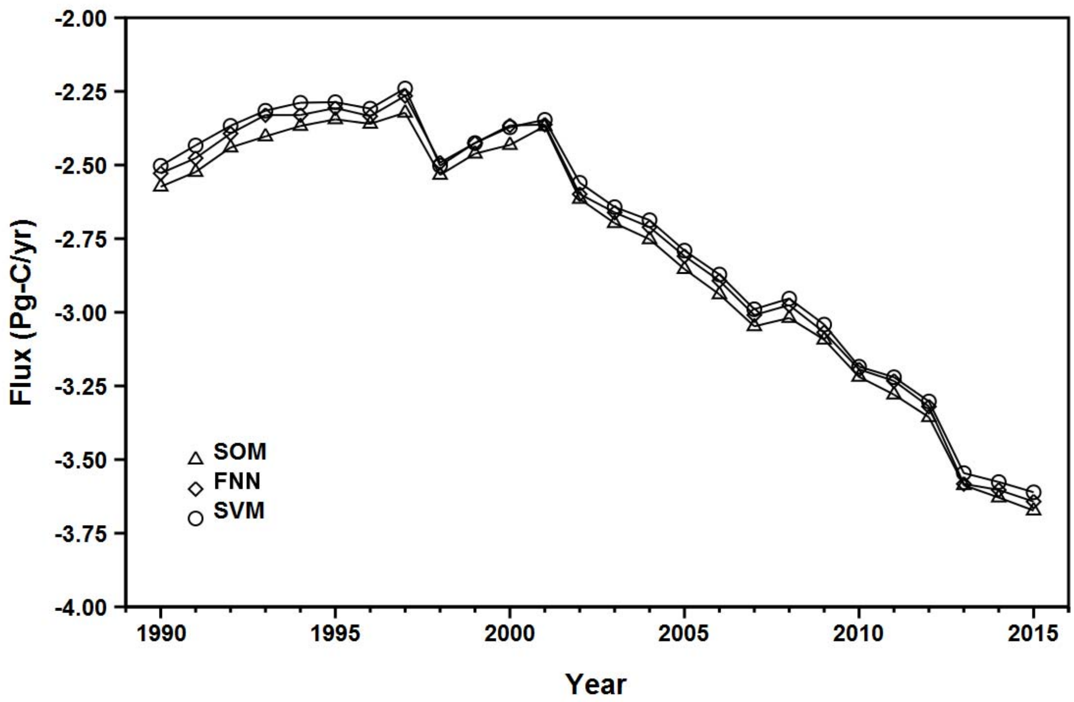


Figure 4: Modelled global CO_2 fluxes. A negative value indicates oceanic uptake.