Ocean Sci. Discuss.,
doi:10.5194/os-2016-73-AC3, 2017

Ocean Science
Discussions
Open Access

Interactive comment on "Technical note: Evaluation of three machine learning models for surface ocean CO_2 mapping" *by* Jiye Zeng et al.

Jiye Zeng et al.

zeng@nies.go.jp

Received and published: 2 March 2017

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.

Q1. 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.

Q2. 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.

Q3. 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.

Q4. 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.

Q5. 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.

Q6. 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.

Q7. p5 l9 - what are the measurement uncertainties?

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

Q8. p5 10 - what is this uncertainty from temperature?

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

Q9. 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.

Q10. 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

Q11. 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.

Q12. 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. Intercomparison of model by different authors is important but beyond the scope of this manuscript.

Interactive comment on Ocean Sci. Discuss., doi:10.5194/os-2016-73, 2016.