

## ***Interactive comment on “Technical note: Evaluation of three machine learning models for surface ocean CO<sub>2</sub> mapping” by Jiye Zeng et al.***

**Jiye Zeng et al.**

zeng@nies.go.jp

Received and published: 2 March 2017

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.

Q1. This Technical note compares the results of three machine learning models for sea surface CO<sub>2</sub> mapping. Two of those, self-organizing-maps (SOM) and feedforward neural networks (FNN), have already been used and compared (in the Surface Ocean CO<sub>2</sub> 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 dis-

C1

tributions show features that cannot be explained by the CO<sub>2</sub> field data. For example there is a CO<sub>2</sub> hotspot east of the African coast near the equator where no observations (February) or low CO<sub>2</sub> 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 CO<sub>2</sub> 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 CO<sub>2</sub> hotspot east of the African coast near the equator and suggested a further discussion. The hotspot CO<sub>2</sub> 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 CO<sub>2</sub> 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. As the issue cannot be cleared without thorough numerical comparisons to other models's output of the same spatial and time resolutions and extend this technical note to a full research paper, it is beyond the scope of this manuscript, which is aimed at comparing the three machine learning models.

Q2. 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 data sets of all models as supplement.

Q3. Page 1, line 14: include (Goddijn-Murphy et al, 2015).

Reply: We included the reference.

Q4. Page 2, line 13: please explain “circular property” and why it can therefore not be

C2

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.”

Q5. Page 2, line 14: sine and cosine transformed components of LON and MON? How of MON?

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

Q6. 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.

Q7. 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

Q8. Page 3, line 8, refer to Table 1 here

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

Q9. 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. . .”

Q10. Page 3, line 12: “dependent of” should be “dependent on”.

Reply: We corrected the mistake.

### C3

Q11. 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)”.

Q12. Page 4, line 2: give references for preliminary studies.

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

Q13. Page 4, line 13: replace “to model” with “and modelled”.

Reply We revised the expression accordingly.

Q14. 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.

Q15. 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”.

Q16. Page 5, line 8: differences are expressed as mean difference standard deviation?

Reply: Yes.

Q17. Page 5, line 8: replace “respectively” with “for SOM”.

Reply: We corrected the mistake.

Q18. 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.

### C4

Q19. 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 CO<sub>2</sub>. 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 CO<sub>2</sub>” 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.

Q20. Page 8, line 8: “prediction” should be “predictions”.

Reply: We corrected the mistake.

Q21. Acknowledgements. Include, as suggested on SOCAT’s website: “The Surface Ocean CO<sub>2</sub> 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 CO<sub>2</sub> 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.

Q22. 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.

Q23. Table 1, line 9: ‘closet’ should be ‘closest’.

Reply: We corrected the mistake.

Q24. Figure 3: The labels in white font are too small to read.

C5

Reply: We enlarged the labels.

Please also note the supplement to this comment:

<http://www.ocean-sci-discuss.net/os-2016-73/os-2016-73-AC1-supplement.zip>

---

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