Ocean Sci. Discuss.,
doi:10.5194/os-2016-73-RC1, 2016

Ocean Science Discussions

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

Anonymous Referee #1

Received and published: 29 November 2016

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 observa-

tions. 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.

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?

Detailed comments

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

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

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

Page 2, line 14: "The approach. . ." is meaning "Our approach. . ." or "Zeng et al.'s approach. . ."?

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

Page 3, line 8, refer to Table 1 here

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

Page 3, line 12: "dependent of" should be "dependent on"

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

Page 4, line 2: give references for preliminary studies

Page 4, line 13: replace "to model" with "and modelled"

Page 4, line 18: modeled and observed CO2 of "all / selected/ non-selected" data

points?

Page 5, line 6: random 10%?

Page 5, line 8: differences are expressed as mean difference $\pm$ standard deviation?

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

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

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?

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

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.”

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.

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

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

References Goddijn-Murphy, L. M., Woolf, D. K., Land, P. E., Shutler, J. D., Donlon, C. (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.

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