

Interactive comment on “Technical note: Evaluation of three machine learning models for surface ocean CO₂ mapping” by Jiye Zeng et al.

Anonymous Referee #3

Received and published: 26 February 2017

The basis of the this technical note is a more detailed examination of the application of various configurations of the neural network approach to create monthly maps of sea surface pCO₂ from temporally and spatially sparse observations. These look to build predictive relationships from high resolution satellite sea surface temperature (SST), chlorophyll a, salinity, mixed layer depth and datasets and SST anomaly. Here, the authors focus on two previously explored methodologies (involving self-organising maps - SOM, and feed-forward neural networks - FNN) in addition to a third novel approach (support vector machine - SVM). This is a timely study, as neural network-based approaches have found increasing application and influence within the community. A thorough investigation of the advantages and weaknesses of individual configurations is an important next step.

Unfortunately, this study is not able to achieve this. It is quite a threadbare investigation,
C1

and spends more time comparing outputs from each configuration against each other, rather than the more important comparison against real observations, previous studies, and alternative independent methods. More attention should be made describing the exact application of the method to the data, the assumptions made and their impact, an assessment of the true uncertainty on pCO₂ mapping, and how and why the estimates diverge from observations / other methods. For instance, has a riverine fCO₂ flux been accounted for in the flux estimates presented here, as this needs to be done in order to be compare like for like

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

Structure: I feel it would be better to have the methodological description section currently situated within the appendix to be within the main body of the text. To a non-user of neural networks, it seems disingenuous to direct readers to the end of the manuscript in order to understand the details underpinning the outputs.

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.

Specific points: 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? p5 l9 - what are the measurement uncertainties? p5 l10 - what is this uncertainty from temperature? p5 l11 - what is the average standard deviation of repeat measurements (should also reference) p5 l13 - why is only july looked at, what is the uncertainty for the full year? How much of this is due to the normalization method? p5 l25 - 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 than 2000 are calculated would be useful as this is not currently explained. Is the systematic trend of 1.5uatm/year simply reintroduced. 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

C3

on this aspect would be useful

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

C4