Biogeosciences Discuss.,
https://doi.org/10.5194/bg-2018-276-RC1, 2018

Biogeosciences
Discussions
Open Access

Interactive comment on “Variation of Summer Oceanic $p\mathrm{CO}_2$ and Carbon Sink in the Prydz Bay Using SOM Analysis Approach” *by* Suqing Xu et al.

Anonymous Referee #1

Received and published: 6 August 2018

Summary:

Xu and colleagues investigate the regional sea surface pCO2 and air-sea flux in Prydz Bay Antarctica using observations from the CHINARE cruise in February 2015. The authors divide the study regions into 3 sub-regions, based on the physical and biogeochemical controls of these sub-regions. Using a self-organizing map approach, the authors extrapolate the cruise data to the entire study region in order to estimate the carbon exchange of Prydz Bay.

The Southern Ocean is still among the least observed and certainly least well understood ocean basins, hence I found this process study – investigating carbon variability

and air-sea exchange in Prydz Bay - to be very interesting and certainly relevant for the BG readership. More details on the strengths and weaknesses are listed below.

Strengths:

I found the manuscript and particularly the discussion of the processes comprehensive and logically built-up. The authors further make use of an appropriate and previously applied method based on machine learning (i.e. the SOM method) to extrapolate the cruise information to the full region of interest. They use independent validation data to test how well their approach reproduces observations from the SOCAT dataset and use this information to estimate the uncertainty of their integrated air-sea flux.

Weaknesses:

Up-front, I would like to note that there are several language issues – too many to be all named here (just one examples: line 232: “In Pacific Ocean” should be “In the Pacific Ocean”) – hence I do recommend English language editing.

During my review, I have encountered a few things that need clarification or some more information from the authors. They are listed from the most to least concerning. Additional comments (not of major concern) with line-numbers can be found at the end of this document:

.)Method section: At the moment, it is impossible for a reader who has not worked with the SOM approach to understand the methods section. Sentences like:” The SOM is trained using unsupervised learning to project the input space of training samples to a feature space (Kohonen, 1984), which is usually represented by grid points in two-dimension space.” Imagine a BG reader who is interested in the carbon exchange of Prydz Bay but has never worked with a SOM. How is that person supposed to understand wording like “unsupervised, feature space, weight vector, training data, labelling data, etc.” without reading several other papers first? As a SOM user I had no issues to follow this section but in my view, it has to be simplified for the more general BG audience. Furthermore, the authors miss to mention what distance function the SOM uses to detect the “winner neuron” (Euclidean distance maybe?). Furthermore, I don’t think the phrase “resolve nonlinear relationships” (see abstract) is appropriate, since a SOM is a clustering algorithm that clusters based on similarities, but does not explicitly “resolve” a relationship.

.)Training data: This links a bit to my point above but goes a bit more in-depth: I am not sure how data have been handled. On line 177 the authors state that the data have been “the four proxy parameters were logarithmically normalized” but table 1 suggests otherwise. In table 1 all values are absolute values. Besides that, I am not convinced that it makes sense to logarithmically normalize all 4 proxies. It makes sense for the skewed MLD and CHL-a but not really for salinity and temperature. Besides, I wonder how the normalization effects the distance function (which is not mentioned). Euklidian distances depend on the data-value range of each proxy. Also, what I am missing is a discussion why exactly the 4 proxies have been chosen? Why not sea surface height, wind speed, sea level pressure? What makes the 4 proxies so unique? I know they have been used by other authors, but the reader of THIS study needs this information.

.)Uncertainty: line 389 states: “increased from week-1 (2.13 TgC) to week-2 (2.24 TgC) due to increased wind speed.” I was a bit disappointed here. First there is the effort to calculate uncertainties, then it is neglected in the text. Given the final uncertainty estimate, it is very unlikely that this regional difference of 0.1 TgC is significant. In general I suggest to add uncertainties wherever possible to avoid such misinterpretations.

.)Validation, comparison: I appreciate that the authors do a comparison with SOCAT data and include this in the overall flux uncertainty. I think that there need to be a bit more info in the text what cruise from SOCAT you are comparing to (this information is available on socat.info), or what the average spatial and temporal distance (which should be possible since a nearest grid method was used) between the cruises is. That certainly contributes to the mismatch as well. Otherwise, I was quite impressend by the relatively small ($\sim$22μatm) difference. It might not sound small at first but you are

comparing small special scale and high frequency temporal scale data based on the extrapolation of a single cruise. Therefore, 22μatm is impressive in my view. Furthermore, the RMSE tells the reader about the spread, but it would be valuable to add the mean (or absolute mean) difference between the SOM derived CO2 and the SOCAT cruise. This would give you an indication of the bias.

.)Methods section: On many occasions the authors re-grid data to the desired 0.1x0.1° resolution, but a bit more information on all data that were regridded and the algorithm would be appreciated. Ideally in form of a table. Additionally, I am missing the motivation why 0.1x0.1° was chosen. Why not 0.5x0.5 or even 0.05x0.05. Just to be clear, I don’t suggest changing the resolution, but the text needs some motivation/technical explanation on why the current resolution was chosen that justifies all the data handling (i.e. regridding of proxy data)

Recommendation:

I have found this study to be interesting and to be of value to the BG readership. While I have raised some (partly major) concerns above I think that they can be resolved by the authors. I therefor recommend major revisions of the manuscript.

Specific and minor comments to the text:

Abstract line 14: Please also add the temporal resolution to the spatial resolution

Abstract lines 27-29: This last sentence is out of context and is not something you can conclude from this study, hence it needs to be removed.

Lines 32-33 reads “The role of the ocean south of 60S in the transport of CO2 to or from the atmosphere is still uncertain despite of its importance of reducing anthropogenic CO2 in the atmosphere” – that is a conflicting statement as it currently reads. If we know the importance of reducing atmospheric CO2 how can its role be uncertain?

Lines 76-77: “Therefore, the direction of the sea-air CO2 transfer is mainly regulated by the oceanic pCO2” – this statement needs a reference

Line 84: “The SOM analysis, based on neural network (NN), a type of artificial neural network” – the second part (based on neural network) can be removed

Line 117: “Salinity records the physical processes” – When I read this sentence I also think of larger scale circulation and mixing in the context of physical processes, whereas this statement links to the follow-up discussion about brine rejection. Maybe a different term would be more appropriate.

Line 130: How was the interpolation done?

Lines 133-136: “The mixed layer links the atmosphere to the deep ocean and plays a critical role in climate variability. Very few studies have emphasized the importance of accounting for the vertical mixing through the mixed layer depth” – Firstly, I disagree. Several studies have emphasized the importance of vertical mixing of carbon (but also nutrients, etc) through the mixed layer. Secondly, I caution the authors to mention the role in climate variability here. Their study does not resolve the necessary timescales to discuss either seasonal or interannual or decadal (whatever variability the authors refer to) variability.

Lines 154-155 “SOM based multiple non-linear regression” – This must have been a mistake or typo here, since the SOM (unlike e.g. a backpropagation network) does not perform a regression (also not a non-linear one). Instead the SOM clusters data based on similar environmental conditions.

Lines 194-195: “until the neural network sufficiently represents the nonlinear interdependence of proxy parameters used in training.” – how is this judged? When do you know that its sufficient? I suppose this is judged by the number of SOM iterations, but how is set?

Line 215: “I could not figure out where the factor 30.8X10-4 comes from? Please explain in the text

Line 264: “robustly divided” – I caution the authors here: How can you be sure the

division is “robust”? Have you done any test that would proof robustness?

Lines 281-282: “region atmospheric pCO2 was stable from 374.6 μatm to 387.8 μatm” That is a difference of 13μatm – I would not call this stable at all! I suppose this difference is largely the result of sea level pressure variability and relative humidity in the surface layer, hence it would be interesting to see the molar fractions (in ppm) for comparison if available.

Line 285: “biological consume” – should be “biological uptake”

Line 318-319: “for a same period” – This would be important information. Furthermore, have you considered ARGO biogeochemistry floats from the SOCCOM array? They are deployed since 2013 and may add some additional independent estimate. This might however be beyond this manuscript.

Figure 4b: “It would be easier visible if x-axis an y-axis range would be the same.

Interactive comment on Biogeosciences Discuss., https://doi.org/10.5194/bg-2018-276, 2018.