General comments

The manuscript 'variation of Summer Oceanic pCO_2 and Carbon Sink in the Prydz Bay Using SOM Analysis Approach' by Suqing Xu et al. presents their cruise data plus its analysis regarding oceanic and atmospheric pCO_2 and the related air-sea pCO_2 flux. The results can potentially be of interest to readers interested in the Southern Ocean carbon cycling, and its variability in time and space. It also provides an opportunity to the authors to show a practical example of the application of SOM in biogeochemistry. In order for the manuscript to be appreciated by the biogeochemical community, the authors should provide a better description of its relevance and importance for the greater Southern Ocean. As I am not an expert on SOM or neural networks, I cannot judge the methodology on that method in detail. I should however be able to understand what is presented in section 2.2, and I find this difficult at times. Several times mention is made of methods (like 'a linear method' or 'Linear regression extrapolation method') without further information on what is done: This makes reproducibility of the work without consulting the authors impossible. Besides that, I unfortunately often find the language to be confusing/imprecise, and therefore recommend professional English language checking before resubmitting. The language made it more difficult for me to judge the value of the manuscript, and I expect I can provide a more in-depth review after the language is improved. The manuscript would also improve if it were shortened as compared to the current version, as there is enough space to increase the information density in the manuscript in my opinion.

Specific comments

1. The introduction

   The introduction thoroughly describes the geographic setting of the Prydz Bay. I appreciate this, but it makes the introduction unbalanced as the questions 'why is this study of relevance' and 'what is new' are only covered by a few sentences. The authors describe the issue that the manuscript wants to address, namely the sparse spatiotemporal coverage of the Southern Ocean (SO) carbon cycle. They also tell the reader that they address the issue using the SOM approach. However, to what extent does research on the Prydz Bay support our understanding of the SO carbon cycle? On page 2, line 38-39 it is mentioned that the Prydz Bay is the third largest embayment in the Antarctic continent. No other reasons are given for the study of in specific this bay: What makes this bay (potentially) important for the SO carbon cycle even though it is small as compared to the total surface area of the SO? To what extent is this Bay representative for the SO as a whole (or just other parts of the SO), i.e. do the authors think their approach or data are useful for and representative of other areas in the SO? Why was the month February chosen to do the cruise?

   In the first sentence, it is mentioned that the SO is important for anthropogenic CO_2 uptake. The authors cannot distinguish between natural and anthropogenic carbon fluxes based on their measurements: Some sentences should be added to describe that the SO is a natural source of carbon to the atmosphere, but a sink for anthropogenic carbon – and that both are highly variable but creating a net sink for total carbon over the past decades. Here an argument could be made for their own study and cruise, which aims to reduce the spatiotemporal sparsity of the data and get a better understanding of the variability of the contemporary ΔpCO_2 and its driving mechanisms. The authors call the Bay a sink at several instances (for example P4, l101 and P5, l125): Some numbers from previous studies should

be given to support the statement that the Bay as a whole is a sink for carbon before presenting your own results.

In Figure 1, an inset could be added to visualise the location of Fig. 1 on the Antarctic continent.

P3, l64-66: How does a marine ecosystem interact with the physical environment to make it complicated to study pCO2? Clarify your statement, as it currently is imprecise.

When describing the methods, clarify that in situ data from the cruise are combined with remotely sensed data to arrive at a gridded product.

2.1 In situ data

Here the authors present how they took their underway measurements and present them in Fig. 2. The first time I read this section, I missed a good structure: The section starts with an explanation of the cruise and instruments used (until line 115). Then, the following paragraphs came to me as a bit of a surprise. One could help the reader find a better flow through the text by explaining that there are several processes/water characteristics that can influence the pCO2 flux (which is the topic of this study). Then, the sea ice paragraph (lines 116-120), the information on the SSS and SST collection (lines 120-124), the biology/CHL paragraph (lines 125-131), and the MLD paragraph (lines 132-end of section) come more naturally. It is important to defend why specifically these proxies/data are used to do your study (create a gridded pCO2 map). Don't forget to start the title with a capital letter i. It is unclear to me whether the results presented in Fig. 2 are 4-week mean results or how they are calculated from the 4 cruise legs: Add more information to both the caption and the text.

2.2 SOM method and input variables

This section is generally hard to follow, maybe partly because I am not familiar with SOM. It should be improved so that also people new to SOM are able to understand and appreciate what you have done. Which 'environmental parameters' and which 'observational datasets' (Fig. 3) are used? Lines 205-220(or even up to 228) could be moved up in order to introduce the reader earlier to the datasets. Then the authors can explain what they are used for and how.

2.3 Validation of SOM derived oceanic pCO$_2$

This section raises a lot of questions from my side. To what extent is SOCAT comparable to your data? Are the data both summer data? Why do you talk about assimilating several years together, but then only take 2015 from SOCAT (line 239)? Could you maybe compare your data to a model estimate of pCO$_2$ for this region? Lines 232-235: How is the equilibrium used to assimilate a dataset over different years? There is generally no equilibrium between atmospheric and surface ocean pCO$_2$, do you mean pCO$_2$-disequilibrium? Why do you describe this if you did not apply this method after all?

2.4 Carbon uptake in the Prydz Bay

This section is quite clear to me: You have combined wind speed data and your pCO$_2$ measurements to arrive at a flux using Eq 2. However, you should clarify 1) where you used a 'scaling factor' (P 10, l247-248) (in Eq. 2?), and 2) that that you used your SOM-based pCO$_2$ product to calculate ΔpCO$_2$ in Eq. 2 (did you?). In addition, you write that the transfer velocity is a function of wind speed and temperature (line 245) and then you write about a

gas transfer rate (line 248) (=transfer velocity?) which you apply a scaling factor to. I am left with the question which gas transfer rate or velocity you have used / how you calculated it.

3.1 the distribution of underway measurements
Here you present your underway measurements for three areas. On what basis did you divide the Prydz bay in these subregions? You write the division is 'robust' (P11, l264): Did you test what effect the choice of your division has on your results? It would be helpful to the reader if you added a plot figure with the subdivision of the Prydz bay into its three regions. Add units to all numbers (especially salinity lacks the psu unit throughout this section). I assume you are describing the results that are visualized in Fig 2 in this section: you should make reference to it if this is the case. Throughout the text of this section, you should be more precise on whether the values are regional means, 4-week means, and how you calculated this (refer to the methods). When you say decrease or increase (like P12, l291), it is not always clear to me whether it decreases/increases in time or space or whether the mean is lower or higher than in the neighbouring sub-region. This causes for example confusion when SST's 'vary sharply' (l 293) but 'decreased slightly' just the sentence above (l 291). The readability of this section may improve by summarizing your main results in a table. A sentence should be added either here on the methods where the relationship between chlorophyll-a (as remotely observed) and biological productivity is stated.

3.2 Quality and maps of SOM-derived oceanic $p\text{CO}_2$
You compare your results to SOCAT and calculate the RMSE. Could you also provide the R^2 of the best-fit line (red line in Fig. 4b)? You say your RMSE is consistent but not as good as most of the neuron methods. Do you mean it is on the high side of the accuracies previously reported, or why is it not as good? Could you calculate/estimate how many extra data points you would need to gain an improved precision of your SOM approach? You could probably comment on the limited amount of data that retrieving more data is not realistic with the resources and time available. SOCAT is not perfect either: A comment on its limited overlap with your study area would be appropriate here. It is surprising that the SOM estimate is generally higher than the SOCAT one, as SOCAT does not cover the low-$p\text{CO}_2$ area towards the south. Did you sample your SOM-derived $p\text{CO}_2$ dataset on the SOCAT locations, or did you compare all SOCAT in the area to all your data points in Fig. 4b? The first would probably be a fairer comparison and provide a better outcome as well. Fig. 4a could be plotted in the same way as Fig. 2 to make it easier for the reader to compare the spatial coverage.

3.3 Spatial and temporal distributions of SOM-derived $p\text{CO}_2$
Here I expect the presentation of your main result: the $p\text{CO}_2$ maps of Figure 6. However, the text mostly describes the sea ice situation of the region: Why is this done here? Maybe a different title would be more appropriate? If sea ice is a main driving factor for $p\text{CO}_2$, this should be argued using the results. If the authors could add regional sub-division lines on the maps in Fig. 6, it might be easier to argue for the chosen sub-division (i.e. Shelf region, etc.).

3.4 Carbon uptake in Prydz Bay
This section is quite clear, although it would be good to clarify when mean values are reported, and whether they are regional means or temporal means, or both. From the figure on page 17 (which has no number?) it is hard to read the $\Delta p\text{CO}_2$ changes: One could either

present it as a table, or adjust the y-axis range. Please make sure the figure is suitable for the colour blind (and check this throughout the manuscript): Use for example different shapes for the three different lines in the upper graph, and add shapes in the lower one.

Supplementary information
The text at the start of the SI is already used in the main text, I do not see the need to provide it twice, and would recommend to remove it from the SI.

Technical corrections

I made an effort to pick out the most important language issues. However, as recommended in the general comments, I would strongly advise the authors to revise their language throughout the manuscript and to have it checked before resubmitting.

1. Try to prevent the use of the word 'it' throughout the manuscript: replace by the actual subject of the sentence
2. Caption of Fig. 1: replace 'The circulations in the' by 'The ocean circulation in the'. Replace sentence 'The weekly sea ice extents for our study periods were overlapped on the cruise.' by 'During the 4-week cruise, the sea ice extent varied as indicated by the contoured white areas:' and replace 'the white shadow' by a fourth contoured area.
3. Check all figures on their suitability for colour-blind people
4. P2, l33: replace 'of reducing anthropogenic CO2 in the atmosphere' with 'in regulating atmospheric carbon and acting as a net sink for anthropogenic carbon' or similar.
5. P2, l35: replace 'this status derives' by ' This uncertainty comes'
6. P2, l36: replace 'for' with 'because of'
7. P2, l38: move 'lying in the Indian Ocean section' to the next sentence and replace 'lying' by 'situated'
8. P2, l39-40: move 'With Cape Darnley …. to the east' to the end of the sentence or rephrase whole sentence, try to use the main verb as early as possible in a sentence
9. P2, l41: replace 'varies' by 'increases' (or does it go up and down?)
10. P3, l51-52: Add 'the': 'The Fram Bank and the Four Ladies Bank'
11. P3, l52: a spatial barrier for
12. P3, l54: replace 'part of it' by 'partly'
13. P3, l63-64: rephrase sentence to clarify the sequence of events
14. P3, l67: the importance for what? Replace 'carbon cycle' by 'carbon cycling'. This relates to comment 1 as well: how does studying the Prydz Bay relate to the SO carbon cycle?
15. P3, l69: use present tense where possible: 'is'
16. P3, l72: remove first word 'the'
17. P3, l77: Add 'A' before 'linear'. Clarify that it was not you doing this by adding 'In earlier studies, …'
18. P4, l78: What is a big scale? The entire Prydz Bay, the SO?
19. P4, l79: Start a new sentence at 'however'. Simplicity can be a good thing: why is calculating pCO2 based on SST and CHL insufficient? How do you know what controlling factors to select?
20. P4, l83: remove 'the' before 'February'

21. P4, l84: Is NN a type of neural network? The acronym NN is not used anywhere else in the manuscript – so not need to define it. What makes it artificial?
22. P4, l85: Remove 'been'
23. P4, l88: Add 'and' before 'chlorophyll'
24. P4, l92: Remove 'been' and replace 'a' before spatial-temporal by 'the'
25. P4, l97: Add the word 'cruise' after 'CHINARE'. Do the same on P4, l108.
26. P4, l98: replace 'to the early of March' with 'to early March'. Check general fluency of lines 97-99.
27. P4, l99: replace 'is show' by 'are shown'
28. P4, l101: here the authors suddenly discuss carbon absorption: the readers has not learned before that this area is considered to be a sink for carbon, so it would be could to introduce the reader to that earlier in the introduction
29. P4, l102: Replace 'followed' by 'follows'
30. P4, l104: Add ', and' and remove '.'
31. P4, l108: 'at the beginning of February 2015', did the cruise not extend into March? Why 'beginning'?
32. P5, l115: replace 'pCO_2 in atmosphere' by 'atmospheric pCO_2'. Check also that each time you use the word pCO_2, that you use an italicised letter p (also in captions, and axes titles)
33. P5, l116/117: Replace 'in polar region' by 'in polar regions'
34. P5, l117: Move sentence 'Salinity records the physical processes' to later in the paragraph, because you first need to explain what salinity has to do with sea ice. It would also fit to explain to the reader why this is all relevant for a study of pCO_2.
35. P5, l117-118: Replace 'During freezing, salt is excluded … [] … brine rejection' with 'During freezing, brine is rejected from ice, thereby increasing sea surface salinity'.
36. P5, l119: replace 'to dilute' with 'thereby diluting'
37. P5, l125: Remove 'clearly'
38. P5, l127-128: 'the active biological process': Do you mean photosynthesis?
39. P5, l128-129: Explain the relationship between chlorophyll-a and biological productivity before you directly connect them and the consecutive effect on pCO_2 in this sentence.
40. P5, l129: Clarify that you used remote sensing data, and provide the reader with uncertainties associated with this method. Be consistent writing Modis either as Modis or MODIS.
41. P5, l130: Replace link by appropriate reference.
42. P5, l138-139: This sentence seems to repeat lines 121-122 on this page.
43. P5/6, l139-141: Rephrase sentence to make clear to the reader that there are two main methods in use, and what the advantages are of the 'difference criterion' method in the SO.
44. P6, l141: Add 'therefore' between 'we' and 'calculated'
45. P6, l142: Replace 'the' with 'on'
46. P6, l142-143: 'of with …' Do you mean 'of which'? I do not understand this sentence, sorry.
47. P6, l143-144: Why where the data gridded? They were point data from the CTD taken along the track, so why where they not already on the right spatial and temporal 'resolution' (do you mean interval?)?
48. P6, l150-151: Start with a capital letter t. Some words have disappeared from the caption.
49. P7, l161: Replace 'dimension' by 'dimensional'

50. P7, l163: 'Input variables', how do these relate to the boxes in Fig. 3? 'as a vector' is more fluent than 'in a vector form'
51. P8, l173: did not all your underway measurements include measurement of pCO_2?
52. P8, l178: Why did you quantify skewness and what did you do with the results? Is taking the logarithm an accepted method to improve the N coverage? Why does the coverage increase when taking the log?
53. P8, l186: Why is this not done for SST and SSS?
54. P9, l198: Add 'part of the' between 'second' and 'process'. Also, it is either each neuron or all neurons (i.e. is it plural or singular here?)
55. P9, l213: What is meant with '8-d'? 8 dimensions, 8 days? If 8 days, why not 7 if used as weekly data?
56. P10, l243: Replace 'by two items' with 'using ΔpCO_2 and the transfer velocity across the air-sea interface' or something similar.
57. P10, l246: Replace 'delta' with 'Δ'
58. P10, l247: What scaling factor are you talking about here? Is it in Eq. 2?
59. P10, l251: Check that equation has one format/font and denote units in []-brackets.
60. P10, l252: Check superscripts of pCO_2_air and pCO_2_sea, also add 'and' before pCO_2_sea and end the sentence with 'respectively'
61. P10, l256: I am again confused by the use of the word regridding, you are working with sample data – why do you regrid? You mean you gridded the data from the point measurements you had of atmospheric pCO_2? What linear method did you use?
62. P10, l258-259: Do you mean you integrated the gridded flux over the area of Prydz Bay, taking into account the ice-free area only? How did you take ice into account?
63. P11, l267: No need to use the acronym AD if you only use it once
64. P12, l300: What is formed here? The subject of the sentence is the Shelf region, but a regions cannot be formed by modification of water.
65. P12, l305-306: If the region was ice-free, Fig. 5 cannot be correct?
66. P12, l314-315: When and where does the biological pump become the dominant factor setting the distribution of pCO_2? How do you know this is the main contributor to the pCO_2 variations?
67. P16, l371: What indicators did you use to conclude that the stability of the water was weak?
68. P16, l377: flew? Please rewrite this sentence.
69. P18, l395: 10^{12} gram=Tg
70. P18, l400: Pleas provide references to this statement and mention it earlier in the manuscript.
71. P18, l408-410: So does the region take up more carbon than on average in the ocean? I.e., is it a relatively large sink as compared to its area?