Review of Yamamoto-Kawai et al Biogeosciences Discussion doi:10.5194/bg-2016-74

The manuscript describes an important issue regarding challenges to monitor ocean acidification in a highly dynamic area with large seasonal and interannual variability. The described method uses measurements and data using shipboard sampling and conventional (relatively, although the TA method using spectrophotometry is not entirely established) methods during 2 weeks each in 2012 and 2013 (both in summer time) to create algorithms. These algorithms are developed as to provide a method to achieve data on the carbonate system and $CaCO_3$ saturation using proxies from sensors attached to moorings in the Chukchi Sea/(Beaufort shelf). The method has been relatively successful in a few other areas (Alin et al..). However, there are several precautions to consider for the Arctic shelves which probably show larger seasonal and interannual variability, in a way also shown in the large difference in the oxygen concentration levels in the bottom water between the two years as presented here. The ship-board data is used in regression analysis using proxies to provide algorithms to determine the variability in the carbonate system at other times of the same year and even for other years. This requires knowledge on the study area and its major drivers to discern the appropriate proxies to describe the carbonate system and $CaCO_3$ saturation. It also require in-depth analysis of the uncertainties and the cumulative errors in the methodology. Unfortunately, this type of error analysis is largely lacking and also it is not entirely clear how the proxies were chosen. The authors also "press on" to use the algorithms, not only for several years where there is very little ship measurements to compare with, but also make far-fetched and highly speculative conclusions on the causes for decreased saturation in the area. The manuscript also seems to be written in haste (check spelling of co-authors and references) and needs a thorough language check. There are also many parts with redundant or repeating text, sometimes the same sentences appear a few sentences apart, also, the word "hotspot" is used 14 times in the text and most of them could be removed. Part of the redundancy could be arranged easily if the authors include a section describing the study area, including the physical oceanography, explain causes for this area to be a hotspot for biology as well as the ice conditions for the two shipboard measurement years. This would also facilitate the interpretation of the differences between the two years later in the results section. The authors also refer to trends of wide spread and increasing undersaturation but show clearly that the undersaturation they found in 2012 was not present in 2013! Moreover, the large uncertainty in the analysis (shown as range of RMSE as shown in Figure 6), clearly shows that it is not possible to state prolonged undersaturation as the title does. This is biased statement, the positive RMSE do not show undersaturation, neither to ship board data for 2013!. Unfortunately, I cannot recommend this manuscript for publication as it is now and there are substantial revisions and clarifications to be made to make it possible for publication. I suggest that the authors rewrite the manuscript focusing on the methodology and describe it in a "good-honest" manner, including sensor and instrument information (resolution, response time, brand, latest calibration). The regression analysis is lacking information and includes large errors (RMSE) which are referred as "well correlated". One example of limitations in the manuscript is that the algorithm is not validated using independent data on DIC and TA, and the authors use it on several years prior to the data used for the development of algorithms where sensor data was collected. However, during this >10 year of time, was the sensors the

same? How frequently were they calibrated? The year that is selected needs to be possible to validate with independent data not only the advantages. Moreover, the limitations to use this method as well as a true and honest error analysis rather than try to push observational and predictive powers that is clearly not within reach for this method in this highly dynamic area, (part of the reason why this is a hotspot).

Major issues summary:

1) Methodology lacks important information on analytical methods and sensor information. Show also DIC and TA data as well as AT vs S relationship. Also, the manuscript would benefit from showing fCO_2 (µatm) in relation to DO since that could provide information on the causes of the difference between the two years.
2) The regression analysis does not fulfill a proper error progression and is not open about the limitations and the true predictive capacity. The year 2012 is described as an anomalous year. How can it then be used for extrapolation in both time and space in the regression analysis and further in the text? The analysis should also use published data for validation and check the spatial validity of the algorithms. Probably they are very limited and cannot be extrapolated to any other region. RMSE are large for DIC and TA, what consequences do that have for the calculated Ω? Explain.
3) The uncertainties in the calculations of the anthropogenic CO_2 impact is not clearly described and emphasized in the abstract without a solid ground.

More specifically:

Title: Since 2013 did not show undersaturation, also the large range of RMSE also show "prolonged" oversaturation! The title is biased towards showing undersaturation and overstates the "negative" results and does not mirror the large uncertainty. Suggest to add "possible" or focus in the 2012 event with largely undersaturated waters due to biological processes (likely).

Abstract:

Include range of saturation instead of only mentioning undersaturation

Row 16-: Highly speculative statement that it is anthropogenic CO_2 uptake that drives the duration of $CaCO_3$ saturation in the bottom waters. See also later comment. The authors mention the limitations of the method they used in the actual chapter which is not the case in the abstract. Rephrase this part to give possible rates of change per decade at a moderate scenario. RC8.5 is considered unrealistic.

Introduction:

Row 16 to 17: This is not correct. Just because it is a shallow shelf it does not mean that the Chukchi Sea has large content of anthropogenic CO_2. One of the largest content of anthropogenic CO_2 is found in the North Atlantic, which is several thousand meters deep. Use a solid reference if that is the case, if not, delete sentence.

2. Observation and analysis

Generally, lack of detailed information on methodology

Row: 5 to 7. Two different methods are used to determine TA, spectrophotomery and potentiometric titration. How did these two methods differ? Was this assessed? The quality of

the TA data plays a large role for the determination of the $CaCO_3$ saturation. Is there any comparison performed between these two methods? Or was another parameter measured for example pH, as to perform internal consistency check on the quality of the TA and DIC? This information would greatly support some of the findings later in the manuscript and also strengthen the algorithm development.

Row 5: How was dissolved oxygen measured on the ship? What was the method for nutrients? If they were sensor data should also include information on what sensors and how and when they were calibrated. Perhaps the difference on DO in bottom water (Figure 2) between 2012 and 2013 only due to sensor differences? That should be explained.

Row 8 to 10: The precision for DIC is quite high, was there a reason for that? How much will that error result in calculated Ω?
Row 12: Why was Lueker constants chosen? It is preferable if the authors would present results from several other determinations of constants, estimate the mean for all and deviation from the mean for each constant as to assess the range in uncertainty/error by using different set of constants.

Row 17: Describe the sensors (brand), resolution and how the accuracy was established.
Row 17: Was the sensor measuring chlorophyll- a or fluorescence? What sensor was used? Was it calibrated?

Results and Discussion
3.1 Ship based observations
Since the Figure does not refer to any of the sampling locations such as (Pt Barrow or Hope valley) it is difficult to follow. Please add this information in the figure referred to in the text.

The authors should also show the TA and DIC not only $CaCO_3$ saturation. The discussing that follows on the differences between the two years would greatly benefit from analysis of the TA and DIC data.

Row 18 to 20: Section is repetitive and is almost the same as stated in introduction, redundant.

Row 24: Interesting that undersaturation was found in the Bering Strait. This was not found in 7-years earlier (water column data summer 2005 by Chierici and Fransson (2009). Could that support the statement that undersaturation has progressed in the area? Interesting comparison to add. They also discuss different constants and the result in Ω (see previous comment).

Row 32-end of section: The authors explain the difference between the two years of shipboard measurements to be caused by differences in organic matter accumulation. Was there evidence for larger primary production in 2012 than in 2013? Perhaps sediment trap data? Or satellite data showing differences in primary production? Later in the manuscript the authors refer to Grebmeier et al 2015 for chlorophyll a data. This could be developed further.

3.2: Mooring observations

Please point out where the mooring was located in Figure 1.

Row 3: Display that sensor data agreed well, how well? Show some numbers?

Row 6: what does the author mean by" T, S, DO showed larger and high frequent variability"?

The changes in water masses limit the possibility to use algorithms to estimate aragonite saturation in this area, since it is highly variable. Should add information on the limitations of the spatial extent for the algorithms.

Row 11-12: What about the ice conditions between the two years? Could variability in sea ice formation/melting cause these changes? Did they differ between the years?

3.3 Regression analysis

Generally: This chapter is not open on its limitations, and does not give a ground-based insight into the development of algorithms or the predictive power of the proxies. This requires a much more thorough and detailed and unbiased description of the methodology that is used here and its limitations. For example: What is the cumulative error in the regression analysis including all error (analytical methods, CO2SYS calculation, difference between observed and estimated). The Figures where this data is used should all include error bars.

Row 15: DIC is also influenced by mixing of watermasses, not only TA.

Row 17: What about calcification by primary producers in the area? That could also cause changes in the TA.

Row 19: the authors should include the TA –S figure. This will add information on the scatter in the data.

Why was the chl a not used? Chl a is commonly used as proxy for primary production. AOU could be a proxy for respiration. Check also with fCO2 which generally correlates well with DO. If they do not correlate that could help in explaining other causes for differences. Explain the choices better.

Row 20: First the authors mention that the best predictions were obtained using all three proxies, then they end up using fewer and fewer, this is understandable since it is better to be able to estimate the whole carbonate system as simple as possible. But it is not convincing and do not explain the selection of proxies. Most published algorithm development use a step-wise method to go forward with the proxy selection.

Moreover, the RMSE for TA is about 14 µmol/kg, which is quite substantial. What is the consequence for the calculated Ω? Need also to consider the RMSE i DIC which is even larger (24 µmol/kg). The authors need to address what that means for the calculated Ω values.

Row 30: the RMSE for aragonite and calcite are quite large and cannot be neglected. The authors should perform a proper error analysis and estimate the cumulative error and discuss the error in context of the seasonal and interannual differences they found. This is crucial for the discussions after on trends and future projections.

Row 29: What is the cause for the large difference in the surface waters? Explain
What is the RMSE for the fits mentioned here?

Row 32. It is repeatedly displayed in the results between the two years that this is a highly variable and dynamic area. This also means that the algorithms are likely to be highly uncertain and result in high bias for other times of the year and also between years. Thus, using the algorithms for 2000, 2002, 2006, 2009,and 2010 is highly speculative. The authors states that this is a validation of the method but to be a validation the authors need independent data on shipboard DIC and TA to compare with the estimated DIC and TA from the algorithms. This is a great weakness in the method and I think the authors should remove this analysis since it is highly speculative. Especially since the authors later on uses results from this analysis to make projections on future saturation states. There are likely other published DIC and TA data in this region which could be used for a proper validation. This area is one of the most studied in the Arctic Ocean with regard to the carbonate system.

Moreover, using sensor data from more than 10 years must give details on sensor types, were they the same for all those years? Calibrations? Precision? Drift?

Page 7: Row 1-4: the authors mention that the estimated and observed omega values correlate well, RMSE is 0.36 and 0.57!! this give a very large relative standard deviation (CV%) compared to the mean Ω, and is not what should be referred as "well correlated". Again, needs a proper error analysis and put into context of the seasonal and interannual variability in the omega values. Be more objective.

3.4.
Row 13-14: repeat biological hotspot again.

Page 8: Row10: If 2012 was such an unusual year (or autumn), how can the algorithm that is developed on this data give reliable estimates for other years (and other areas)? Sometimes 10 years back in time? Not solid and again shows that the data from this study area cannot be extrapolated to other times of the year nor to other years.

3.5 Anthropogenic impact
The used method assumes that the rate of uptake in the bottom water is the same as in the surface water uptake of anthropogenic CO_2. That should be added to the limitations of this method. The method also assumes that ocean mixing and all other processes have remained the same, which is partly commented by the authors regarding biological processes. It could be worthwhile using reported values of decadal uptake rates from for example Tanhua et al., 2007 Anthropogenic CO_2 in the Arctic Ocean) and use that rate of change to estimate the impact of anthropogenic uptake. That is more robust and also shows the rate of change at depth, which is likely not the same as in the surface waters. Regardless the authors should include more on the assumptions and limitations on the resulting omega values.

Also, after the last Paris (COP) meeting it was agreed that the RCP8.5 scenario is unrealistic and I suggest that the authors include a moderate scenario in addition.

Also, now it is difficult to follow the actual change in omega, it would be better to suggest a rate of change /decade. How much would omega change due to anthropogenic CO_2 per decade?

4. Summary and conclusions

Rewrite summary with regard to previous comments.
Biological hotspot is used many times. Reference spelling! Kroeder!?

Figures:

Figure 1: Please show the main study sites on the map Pt Barrow and Hope Valley, Bering Strait.

Figure 2: Figure 2c: add abbreviation DO which is used in the text and unit for DO.
Add plots showing DIC, TA and fCO_2. Include units. Moreover, there is no undersaturation in 2013 which is not the impression given in the text. Please, clarify in the text.

Figure 3: Describe what a negative and positive AOU refers to.

Figure 4: Difficult to see the bias between shipboard and sensor data in the plot, suggest to add a Table showing the differences. The Table should also include the number of samples/data points that were compared (N).

Figure 5a: Show the data from 2012 and 2013 i different symbols or colors. Figure 5b (x-axis), is the Ωobs based on 2012 and 2013 data? Clarify. Are the displayed data mean values for all years between 2000-2010 or separate years? If separate years, show the different years using different symbols for improved understanding of the bias to the observed for the different years.

Figure 6: The errors and biases are large as is shown in the RMSE (grey lines). This error "only" includes the error from reconstruction and do not include the cumulative error from the Ω calculation from DIC and TA. The figure text should include the full regression incl coefficient of correlation.

Figure 7: Perhaps combine Figure 6 and Figure 7 to show present, pre-ind and a future 50 years.

Figure 8: Redundant picture. Also, Figure caption text is wrong, should probably be "collected".