First of all, we would like to thank the reviewer for a review of this paper. We have addressed all comments and suggestions. Please find our responses below

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 CaCO3 saturation. It also require indepth 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 farfetched and highly speculative conclusions on the causes for decreased saturation in the area.

--Please see below our response to each specific comment.

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.

--We are sorry that we have submitted the manuscript with many typographical errors and redundant text. We will ask an English Language Service to edit our revised manuscript. Many of the word "hotspot" has been removed. We have add a section describing the study area in the revised manuscript. Thank you for your suggestion.

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!

--Our ship-based observation in 2013 was performed in high saturation season when waters were under an influence of photosynthesis as suggested by moored sensors of DO and chlorophyll a. We believe that this does not conflict with "increasing undersaturation" with increasing anthropogenic CO_2.

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!.

--We agree. The title has been changed to:
"Seasonal variation of $CaCO_3$ saturation state in bottom water of a biological hotspot in the Chukchi Sea, Arctic Ocean"

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).

--We have revised the manuscript to describe results in a good honest manner. Sentences have been changed to be more exact and fair, for example, "bottom water was kept at aragonite undersaturation for most of the winter" was changed to "Ω in bottom water was kept low during winter" and "intermittent undersaturation was found" was changed to "intermittent undersaturation was suggested".
We hope the revised manuscript will meet the requirements for publication.
Information of DO sensor has been included (see below).

The regression analysis is lacking information

--Please see below

and includes large errors (RMSE) which are referred as "well correlated".

-- The word "well" was removed.

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).

--We found that the concern of the reviewer is based on misunderstanding. We have validated the algorithms by comparing independent data of shipboard DIC and TA in 2000, 2002, 2006, 2009 and 2010 with estimated DIC and TA using equations obtained for 2012/2013 data. Comparison showed that the algorism could reproduce Ω for multiple cruises even the area has high variability. Therefore we concluded that the algorism is applicable to the study area. We have modified the text to explain this clearly:

"In order to evaluate regression equations obtained from 2012-2013 cruises, we have applied the same equations to independent data from R/V Mirai cruises in the Chukchi Sea in 2000, 2002, 2006, 2009 and 2010..... We have calculated DICest and TAest from T, S and AOU data of these cruises using the equations obtained for 2012-2013. AOU was calculated from bottle DO data. DICest and TAest agreed with independent shipboard observations of DIC and TA with r^2=0.96 and 0.83, respectively. Ω calculated from DICest and TAest (Ωest) was…"

Major issues summary:

1) Methodology lacks important information on analytical methods and sensor information.

--See our response above.

Show also DIC and TA data as well as AT vs S relationship. Also, the manuscript would benefit from showing fCO2 (µatm) in relation to DO since that could provide information on the causes of the difference between the two years.

--Figures suggested will be added to the revised manuscript. fCO2 and AOU correlated well as described below. Please see figures of S-TA and fCO2-AOU submitted with this reply.

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.

--As mentioned below, errors in Ω resulted from errors in analytical methods, select of constants, difference between observed and estimated have been estimated. Cumulative errors in estimated values will be indicated in Figures in the revised manuscript.

3) The uncertainties in the calculations of the anthropogenic CO2 impact is not clearly described and emphasized in the abstract without a solid ground.

-- We will indicate uncertainties in Figure 7. The sentence was removed from the abstract. Abstract has been modified as shown below.

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).

--The title has been changed to:

"Seasonal variation of $CaCO_3$ saturation state in bottom water of a biological hotspot in the Chukchi Sea, Arctic Ocean"

Abstract:

Include range of saturation instead of only mentioning undersaturation

Row 16-: Highly speculative statement that it is anthropogenic CO2 uptake that drives the duration of CaCO3 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.

-- The abstract has been changed to read:

"Distribution of calcium carbonate saturation state (Ω) was observed in bottom waters of the Chukchi Sea in autumn 2012 and early summer 2013. Aragonite and calcite undersaturation with Ω as low as 0.3 and 0.5, respectively, were found in high productivity regions in autumn 2012 but not in early summer 2013. Comparison with other parameters has indicated that biological processes -respiration and photosynthesis- are major factor controlling regional and temporal variability of Ω. From these ship-based observations, we have obtained empirical equations to reconstruct Ω

from temperature, salinity and apparent oxygen utilization. Using two-year-round mooring data and these equations, we have reconstructed seasonal variation of Ω in bottom water in Hope Valley, a biological hotspot in the southern Chukchi Sea. Estimated Ω was high in spring and early summer, decreased in later summer, and kept relatively low in winter. Calculations indicated a possibility that bottom water could have been undersaturated for aragonite on an intermittent basis even in the pre-industrial period, and that anthropogenic CO_2 has extended the period of aragonite undersaturation to two- or three-fold longer by now."

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 CO2. One of the largest content of anthropogenic CO2 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.

--We do not mean Chukchi Sea bottom water has a large content or higher content of anthropogenic CO_2 than in the North Atlantic but mean that anthropogenic CO_2 can penetrate easily to bottom water by mixing of surface water into bottom water induced by wind, tide, and atmospheric cooling.

We modified the sentence to:

"because of shallow bottom depth of ~50 m, vertical mixing induced by wind, tide or atmospheric cooling brings anthropogenic CO_2 into bottom water to which benthos are exposed."

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 CaCO3 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.

--We do not have pH or other measurements. Intercomparison of two method showed good agreement of 0.88±2.03 umol/kg (Li et al., 2013). Also, a comparison of S-TA relationship for our two cruises did not show any offset. These are now mentioned in the text.

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.

--Bottle DO was determined by Winkler titration following World Ocean Circulation Experiment Hydrographic Program operations and methods (Dickson, 1996). Nutrient samples were analyzed according to the GO-SHIP Repeat Hydrography Manual (Hydes et al., 2010). These are now mentioned in the text.

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?

-- We do not know the reason for relatively high precision for DIC in 2013. Uncertainty of 5.5 umol/kg in DIC results in error of ±0.05 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.

--This kind of analysis has been done elsewhere (Azetsu-Scott et al., 2010) and we do not think this is worth to repeat the similar analysis in our manuscript. We have used Lueker following our previous publications. When constants of K1, K2 from Mehrbach et al., 1973 (refit by Dickson and Millero, 1987) were used, Ω differs from original estimate by only < 0.01.

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?

-- We have added description about DO sensor as follows:

"DO sensor used for mooring observation was an AROW-USB phosphorescent DO sensor (JFE Advantech Co., Ltd., Kobe, Japan). The sensor was calibrated using oxygen-saturated and anoxic water to determine the linear relationship between them with 2 % accuracy (Nishino et al., 2016). .....difference between sensor and bottle DO data was 4 umol/kg."

Because we do not use chlorophyll a or turbidity data, we do not think detailed information of these sensor and its calibration is needed to be mentioned in our manuscript. They are described in Nishino et al. (2016) as referred in the text.

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 CaCO3 saturation. The discussing that follows on the differences between the two years would greatly benefit from analysis of the TA and DIC data.

--Revised as suggested.

3.1 Ship based observations

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

--The sentence has been deleted.

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).

--Their observations in Bering Strait was made in July/August. Therefore, we think that the difference between observations reflects seasonal variation of Ω. As shown above, use of different constant cannot produce a large difference in Ω as observed between cruises.

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.

--As mentioned in the text, we think that stronger stratification in 2012 caused an accumulation of more CO_2 in the bottom water because photosynthetic activity was lower in 2012 than in 2013 associated with stronger stratification and mixing with sea

ice meltwater (Nishion et al., 2016).

3.2: Mooring observations

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

--Differences between bottle data and sensor data for the day of bottle sampling will be indicated in a table as suggested by the reviewer. Difference between sensor DO and bottle DO of 4 umol/kg in July 2013 is mentioned in the text.

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

--We meant that these sensor data showed a lot of ups and downs. We have modified the sentence to:

"T, S and DO showed large fluctuation"

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.

--We believe that the obtained equations are applicable for waters in the eastern Chukchi Sea as confirmed by comparison of estimates with observations for multiple cruises. If processes affecting Ω as well as source of waters are the same, the same equation could be applicable in the western Chukchi Sea. However, we do not have any data to validate this.

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?

--Winter conditions are similar between two years. More sea ice meltwater was found in summer of 2012. This is stated in the text as the cause of stronger stratification and accumulation of more CO_2 in bottom water in 2012.

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.

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.

--As mentioned above, errors in Ω resulted from errors in analytical methods, select of constants, difference between observed and estimated have been estimated. Cumulative errors in estimated values will be indicated in Figures in the revised manuscript.

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

--This sentence has been deleted.

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

--This possibility is mentioned in the revised text:

"A bloom of calcifying primary producer can cause a drawdown of TA (Murata et al., 2002). However, neither bloom of calcifies nor TA drawdown in S-TA diagram was observed during our observations."

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

--Revised as suggested.

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.

--Chl a is a proxy for primary production at the moment of observation, while AOU represents the effect of primary production/respiration accumulated for some period. The latter is also the case for DIC. Therefore, AOU is a better proxy for effect of respiration/production on DIC. In fact, chl a and fCO2 does not show significant correlation ($r=0.12$, $p=0.08$) whereas AOU correlates well with fCO2 ($r=0.92$, $p<0.01$). A figure showing AOU-fCO2 correlation has been added to the revised manuscript.

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 stepwise method to go forward with the proxy selection.

--We have compared results of regression analysis and found the inclusion or replacement of proxies of biological processes, such as nutrients or chlorophyll a concentration did not improve the estimate of DIC and TA. This is now mentioned in the revised text.

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.

--Resulting error in Ω is indicated in the comparison between estimated and observed Ω in the text and in Figure 5a.

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.

--As mentioned above, errors in Ω resulted from errors in analytical methods, select of constants, difference between observed and estimated have been estimated. Cumulative errors in estimated values will be indicated in Figures in the revised manuscript.

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

--Large differences were found in surface waters with high temperature (>6℃). This might be due to rapid warming at the surface that could cause decoupling of oxygen and carbon due to a difference in temperature dependence of solubility or in gas exchange rate between two gases. We mention these possibilities in the revised text. We do not estimate RMSE for these data because scattering of these data is evident in Figure 5b and because they are out of range of bottom water in temperature and Ω.

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?

--We found that the concern of the reviewer is based on misunderstanding. We have validated the algorithms using independent data of shipboard DIC and TA with the estimated DIC and TA from the algorithm. Please see our reply above.

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.

--Because Ω is a ratio, relative standard deviation CV% does not make sense. However, we agree that we should be more objective in describing the results. We have removed the word "well" from the sentence.

3.4.

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.

--See our response above.

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 CO2. 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 CO2 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.

--Estimates of anthropogenic CO2 by Tanhua et al. (2009) are basically the same as estimated in our study for waters above the winter mixed layer, which is the bottom depth for the Chukchi Sea. TTD method is accounting for mixing of water with different ages, but such mixing does not significantly occur in our study area.

Assumptions that processes have remained the same since preindustrial period is not probable and therefore our estimate is rough.

This is now stated in the text as: "In order to roughly estimate the effect of anthropogenic CO_2…" and "Caveat here is that our calculation is based on an assumption that terms Δdiseq and Δbio have not changed since pre-industrial period and therefore provides only very rough estimates". Numerical number of counted days of undersaturation for preindustrial and future cases are not presented in the revised manuscript.

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.

--In the revised manuscript, we removed the expression of "50 years" for pCO2=650ppm case.

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 CO2 per decade?

--A rate of change in omega /decade differs significantly depend on revel factors that varies during a course of a year. This is shown in Figure 7, as the fact that difference between preindustrial and future cases varies with seasons. Describing a rate of change/decade for each season will beyond the limitation of our rough estimate.

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

--We did not combine them as 3 estimates with error bars make the figure too busy.

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

--We do not think this a redundant picture. It supports the statement in the text that many bivalves were found in Hope Valley hot-spot area, both well-grown adults and small young individuals. However, because a couple of Figures have been added in the revised manuscript as suggested by the reviewer, it should be better to reduce the number of figures. We therefore moved this picture to Figure 7 as an insert panel.

All of following comments were considered and manuscript was revised accordingly

Rewrite summary with regard to previous comments.

Biological hotspot is used many times.

Reference spelling! Kroeder!?

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

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

Figure 2: Figure 2c: add abbreviation DO which is used in the text and unit for DO.

Add plots showing DIC, TA and fCO2. 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 with different symbols or colors. Figure 5b (xaxis), 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.