

## **Review: Nordic Seas Acidification by Fransner et al.**

The authors have put together an overview of past, present, and future ocean acidification in the Nordic Seas using measurements, gridded assembled data, and model projections. The work thus presents an exhaustive overview of ocean acidification in the Nordic Seas that resembles most available information and will as such certainly prove to be very useful for other scientists and to people outside of Science, such as policy makers. Personally, I very much enjoyed reading the manuscript and want to congratulate the authors.

The manuscript would, however, significantly improve if underlying processes would be discussed further to explain the drivers between the different reaction of the regional seas to ocean acidification. What drives the changes in alkalinity, dissolved inorganic carbon, temperature and salinity? Is it changes in water masses, circulation, deep water formation, freshening from the Arctic Ocean or land ice melt from Greenland? A better explanation would help to understand past and future changes and also help to understand how well the one model used in this study performs regarding the expected changes. Furthermore, I have a number of questions regarding the Methods that would need to be addressed before publication.

### **Major comments:**

General: Please be quantitative as often as possible and avoid formulations like “it agrees well. This should be carefully addressed throughout the manuscript.

Introduction: I suggest merging sections 1,2, and 3. Some parts are repeated, and it would be helpful to have some information earlier. For example, I think a better explanation of how ocean acidification works chemically is needed and what the saturation state is and why it matters (or why not?). Why do the Nordic Seas have naturally low saturation states, etc. It is absolutely right, that the Nordic Seas are of high importance and very interesting, so strive on it. Why does it have short residence times and what is the link between the residence time and the  $p\text{CO}_2$  undersaturation? Which water masses meet? It all makes more sense after reading the sections 2 and 3, but section 1 is not understandable for oceanographers who do not know the Nordic Seas. Overall, I think most of the information is in the manuscript, but restructuring could strongly improve the Introduction. And as you discuss future projections, you could also highlight the difficulties that models have in these regions.

Undersaturation of  $p\text{CO}_2$ : Please introduce the concept of undersaturation carefully. Presented like this in the abstract without explanation leads to large confusion. Most of the time, I think I understand what you mean with the concept of undersaturation of  $p\text{CO}_2$ . At other times, I am very confused when you compare the  $p\text{CO}_2$  in deep waters (2000 m) to  $p\text{CO}_2$  in the atmosphere at the same time. Given that the age of the water masses at 2000 m is likely much older, it has “seen” the atmosphere a long time ago. How does it make sense to speak of undersaturation or a perfectly tracked pH?

I think a lot of people are not familiar with this concept, including. How do you calculate/define undersaturation? Why does the undersaturation exists in pre-industrial conditions? I suppose that northward flowing waters cool down and their  $p\text{CO}_2$  decreases

faster as their solubility increases than they can take up  $\text{CO}_2$  from the atmosphere (Broecker et al., 1974). If that is the dominating process, the weakening of the undersaturation is somehow expected, as cold waters hold less  $C_{\text{ant}}$ . With increasing atmospheric  $\text{CO}_2$ , the differences in surface ocean  $C_T$  should become smaller, which would also lead to a smaller difference in marine and atmospheric  $p\text{CO}_2$  in the northward flowing waters in the Nordic Seas. It should be very similar to the processes in the Barents Sea, where Terhaar et al. (2020) showed that  $C_{\text{ant}}$  is larger in the northward flowing waters than expected from atmospheric  $\text{CO}_2$  concentrations. This larger  $C_{\text{ant}}$  is equivalent to a weakening of the undersaturation of  $C_T$ . With a changing climate, other drivers could potentially also become important, such as a change in circulation, i.e. is the quantity of northward flowing waters larger or smaller, are they flowing faster or slower, do they extend further into the Nordic Seas? all these factors could come into play regarding the undersaturation and need to be understood to properly discuss this effect.

Overall, I think this concept needs to be introduced in more detailed with the respective literature if it is a core concept of the paper. In fact, much is explained in the Discussion (Lines 314-328). This should be moved to the Introduction. In the Discussion, it would be really great if you could pin down the most important mechanisms for each regional sea. As it is now, the manuscript is more an atlas to look up numbers, explaining the mechanisms would put it on a whole different level. I strongly encourage you to do this.

Line 9: You point out that the Barents Sea Opening is different to all other regions because of an increase in  $A_T$ . This sounds very interesting and it is thus even more disappointing that no explanation exists why the  $A_T$  increases? Is it via biological processes such as an increased production of organic matter via primary production or is it a change in circulation, i.e. are Atlantic water masses with higher  $A_T$  expanding further and displacing Arctic Ocean waters? I really would like to have an explanation here.

Line 16-17/ lines 119-122: No need to request further model studies. The output from CMIP5 and CMIP6 models are online available. While differences in pH and saturation among models are generally small between models, they can be large below the surface in the North Atlantic (Goris et al., 2018) and in the Arctic Ocean (Terhaar et al., 2020b). I suspect thus that the differences in the Nordic Seas would be large, too. Moreover, the NorESM-ME model seems to be at the high end of carbon uptake in both, the North Atlantic and the Arctic Ocean and thus likely at the high end in the Nordic Seas as well. This is likely caused by an overall strong AMOC in the model (Wang et al., 2014).

Given that the NorESM-ME model is at the high-end for the North Atlantic anthropogenic carbon sink, I doubt that it is representative for future ocean acidification. The cited papers only treat the surface (mainly driven by atmospheric  $p\text{CO}_2$ ) and the bottom ocean where changes in all variables are small and models “automatically” agree. I am not aware that these papers analyze model differences in waters from 100 to 2000 m.

Given the detailed analysis of the historical data, I strongly believe that it is necessary to analyze a larger number of models with annual mean data. I feel that it is absolutely necessary in order to make strong statements about the future acidification in the Nordic Seas, such as that undersaturation can only be avoided for RCP2.6 and RCP4.5.

Lines 95-97: Annual means/seasonal bias: Although not enough data exist to de-seasonalize the data, the model output gives you several possibilities to assess the possible uncertainties.

You could subsample the model at the same time and place and compare these to the model annual mean values. Or even easier, you could compare the seasons within the model output to understand the seasonal variations. And are the under sampled months crucial in these regions, i.e. is winter under sampled and is deep-water formation occurring in winter? If that is the case, you might even expect a higher  $p\text{CO}_2$  in winter. Given the strong seasonality, I feel that an analysis of the potential bias is necessary.

pH uncertainties: I have trouble in seeing the standard deviation over the entire Nordic Seas as an uncertainty. This is rather a measure of heterogeneity. The large gradients in T, S,  $C_T$ , and  $A_T$  over the Nordic Seas thus increase artificially the uncertainties in pH when compared to observations. I have no good proposition but I would consider using uncertainties from the dissolution constants using mocsy2.0 (Orr et al., 2018) instead of the heterogeneity. Whatever you do in the end, it would be great if you could find a solution that permits to compare the model to the observations on the same basis.

Trends and significance: In many figures you show trends for each subsea and depth. I am still wondering how you calculate if a trend is significant (maybe I missed the explanation). Especially confusing are for example Figure 5f,g. Measurements are only available for the 2<sup>nd</sup> half of the period and uncertainties become enormous for 1981 as no measurement exists there. How is it possible to speak of a significant trend given the large gap in observations? I am not strong in statistics and would need a better explanation here.

## Minor notes:

General: I would suggest using  $C_T$  and  $A_T$  instead of DIC and ALK as recommended by the guide for best practice (Dickson et al., 2007). Along the same lines, please write  $pCO_2$  instead of  $pCO_2$ .

Line 2: Why do the high-latitude location lead to a higher sensitivity to acidification? People who are not ocean biogeochemists might not directly understand that high-latitude oceans have a higher solubility, hold more dissolved inorganic carbon ( $C_T$ ) and have thus a naturally lower pH and lower saturation states. I feel that it would be helpful if you lead the reader a bit more here.

Moreover, aren't projected changes in the saturation states smaller in the high latitudes than in the tropics because the absolute numbers are already lower. It would probably be similar for pH? So strictly speaking, the Nordic Seas are more vulnerable towards ocean acidification because they start with more  $C_T$  and lower saturation states and pH, but they are not more sensitive.

Line 4: Could you be more exact instead of writing pre-industrial? The definition of preindustrial is crucial for the historical anthropogenic carbon uptake (Bronselaer et al., 2017) and in consequence also for historical acidification.

Line 6: Please be quantitative. How much larger are they?

Line 10: What exactly is the "acidification signal"? Which regions are you speaking, please be more specific? And 1000-2000 m is not very precise for the saturation horizon at present-day conditions. It would also be helpful to know where it was in pre-industrial conditions? Did it increase by 50 m or by 1000 m?

Line 11: What do you mean by significant?

Line 12: The switch between two sentences from present day saturation states to future pH makes it hard to read and to compare.

Line 13: Again, a switch from surface pH to the saturation state over the entire water column. You are losing me here.

Line 13: I suggest to write "high-emission" instead of "worst case".

Line 15: At what depths exactly? Everywhere or online over a depth of 10 m?

Line 15: What is the majority? Literally all or 51%?

Line 20: Are the 25% exact? Or around 25%?

Line 22: "More serious downside" is not necessary. Just delete that part of the sentence to make it shorter.

Line 23-24: What threat are they imposing? Can you be more precise please?

Line 27-29: I think this sentence is wrong. I am not aware of well constrained ocean acidification at global scale. This might hold for the surface ocean but definitely not for subsurface to deep ocean basins. Differences in models lead to large differences (see bottom ocean pH in Kwiatkowski et al. (2020) in regions where waters reach the bottom. As soon as the waters leave the surface, I believe that historic acidification and its projections are rather ill constrained.

Line 34: Here, you could introduce your acidification signals.

Line 37-38: I like this last sentence, maybe you could make it a topic sentence of the paragraph. Please also write that and why the Nordic Seas have low saturation states.

Line 48: I am confused by the word “anthropogenic” here. Above, you used it in combination with anthropogenic carbon. I suppose now it is not to be read in combinations with carbon because changes in T, S, and  $A_T$  are also anthropogenic, right?

Table 1: I am not sure if this table is needed. If you keep it, please define “minor”.

Line 60: I think you can delete large parts of this sentence, such as “by the strongest base in seawater” and all after “which has been supplied ...”. It does not seem to be such an important information in this context.

Figure 1: Please add a colorbar. Could you change the color for the continents, it is hard to distinguish continents and shelf seas? It is also very difficult to see the dashed lines, could you somehow make them stand out better? A lot of points are outside your defined region, did you use these points? If not, I would suggest deleting them to make the figure less busy. Could you also highlight the two time series stations that you mention in 4.1?

Section 4.1: The section is highly important, but it is difficult to read and to extract the information. For example, the topic sentence speaks about DIC, ALK, temperature and salinity. Later on, uncertainties are also reported for oxygen and nutrients. Did you also use oxygen and nutrients or only the 4 variables in the topic sentence? If you did not use nutrients and oxygen, no need to mention them.

Could you also please provide an assessment of how much volume of the Nordic Seas is covered by the GLODAPv2 data? Is it mainly surface data, how deep does it go, are some regions under sampled? Given, that you look at changes over the data collection time period, it would also be useful to see the measurement on a timeline to see if the measurements are biased towards the end of the period.

Please use paragraphs to divide between GLODAPv2, the two time series stations and the framework of the Norwegian ocean acidification monitoring program. This will likely help to read the section. By the way, what exactly is the Norwegian ocean acidification monitoring program? Could you introduce it with one sentence?

Line 86: Please be more precise than “approximately 4 times”. Are they covering all seasons? Can you be more precise than 10-20 depth levels? Are these depth levels concentrated at the surface or evenly spaced?

Line 109: Is there a reason why you did not include RCP6.0? Could you give the citations for the scenarios (Meinshausen et al., 2011; van Vuuren et al., 2011)?

Lines 110: What is similar? Please be quantitative. And please make the comparison for the Nordic Seas. If the NorESM2 model is doing a good job in the Pacific and Southern Ocean, what does it tell me with respect to the Nordic Seas?

Line 112: What is comparable? Please be quantitative here. The reader can then judge if that is good enough or not.

Line 113: This is definitely a red flag to me. What is broadly? How much? You do not even show me what happens under the surface. What simulated acidification? Over which period? The relatively close agreement at the surface is somehow expected as pH follows mainly the atmospheric  $p\text{CO}_2$ . Below the surface, I do not believe this statement without seeing it. Please revise the model evaluation and really try to convince me that the model works. As it is written now, I am absolutely not convinced.

Lines 113-115: Do you add relative changes in pH and saturation states to the absolute saturation states? I think you should not do this with non-linear variables (Fassbender et al., 2020). Please add simulated changes in  $C_T$ ,  $A_T$ ,  $T$ , and  $S$  to the respective variables in GLODAPv2 and use the projected  $C_T$ ,  $A_T$ ,  $T$ , and  $S$  to calculate projected pH and saturation states. And please precise how saturation states are calculated within the model world, using CO2SYS or mocsy2.0? Are you using simulated nutrients as you do for the observations?

Lines 124/125: Could you assess how good the pre-industrial  $C_T$  values from GLODAPv2 are? If I understand right, a very basic version of the TTD method was used to calculate these numbers, although the TTD method is uncertain in the Nordic Seas and the Arctic Ocean? The time-dependence of saturation of CFCs and SF6 (Tanhua et al., 2008) as well as the change in saturation of surface ocean  $C_T$  that you document here and that were documented in the Arctic Ocean (Terhaar et al., 2020). It would be good to let the reader know here that this pre-industrial  $C_T$  has these uncertainties.

Line 130: I am a little bit confused here. Before you talked not really about nutrients when you discussed measurements and now, they are here. Did you have nutrient measurements everywhere? What did you do when nutrients measurements were not available?

Lines 134/135: Why do you need this sentence? It is confusing after the sentences above and it does not really add something. I suggest deleting it.

Lines 130-135: I suggest using the mocsy uncertainty propagation for the observations (Orr et al., 2018). That would help to know how reliable trends are.

Lines 139-142: This reads nicely but fits better into the Introduction. Here in the Methods, it would be helpful to motivate why you chose exactly these regions and not others. What makes them special? So far, I am still more or less in the dark about the special aspects of each region, although this is a good start.

Line 144-145: If you reduce the Fram Strait, can you still call it the Fram Strait? To me, the Fram Strait is characterized by northward flowing Atlantic water and southward flowing Arctic water. If you chose the region in order to focus on the Atlantic waters and exclude almost 2 thirds of the width (looks like in Fig 1), you can hardly call it Fram Strait, can you? Maybe call it Eastern Fram Strait? Similar for the other regions, why do you choose such small subsamples. The argument of aliasing effects might be a good one, but at the moment it is not convincing to me. I think you have to elaborate a little bit on it.

Lines 146-147: Does 200 m correspond the maximum mixed layer? Or does a strong gradient exist between 0-100 m and 100-200 m?

Line 149: How did you define significant trends? Did you look at  $r^2$  or p values? An explanation is necessary.

Line 150: For pH trends and uncertainties, wouldn't it be better to look at [H+]? Otherwise it is not a linear trend and the uncertainties are not linear either? (Fassbender et al., 2020).

Line 151: Didn't you introduce T and S already earlier?

Line 171: To be clear, you used only GLODAPv2 pre-industrial for past changes (section 4.3), but you used model output for the driver of these past changes? I think that I am not understanding something here.

Line 172: You do not consider changes in nutrients as potential driver, so I take it that they are negligible? Or are they somehow included in the alkalinity?

Line 173-175: As explained above, this needs much more introduction. At this point of the manuscript, I still have no idea why this should matter.

Line 176/177: If possible, use a topic sentence here to tell the reader what this first paragraph is about. And please be quantitative here ("Agrees well" does not tell me anything). Do you mean that it agrees within the uncertainties?

Line 179: What does 'it' refer to at the beginning of the sentence?

Line 179: Can you compare the two uncertainties? I suppose that you used the entire region simulated in the mode whereas the GLODAPv2 does not cover the entire region (or did you use the gridded version here from Lauvset et al.? In any way, please be more precise. In case that the model and the data product do not cover the same regions, I think you cannot compare the uncertainties as they are a measure of the homogeneity of the sample and not the real uncertainty. As the observations are likely concentrated in some regions, this might reduce the homogeneity and thus the uncertainty. Moreover, how confident are you with respect to the pre-industrial uncertainties given that you did not account for changes in T, S, and  $A_T$ ?

Line 180: I do not buy this sentence. The different pH largely stems from the temperature, doesn't it? Globally, waters are warmer than in the Nordic Seas and thus the global surface average pH is smaller, right?

Line 181: Do you show this somewhere? You could simply illustrate that you are right, if you calculate the average of surface pH values in the observations. Otherwise this is just a speculative guess.

Line 182: Can you be more precise than "about 0.05"? Uncertainties would be good and regional differences as well. I also would make a single paragraph about the comparison to the global ocean to make it easier to read. As it is presented now, the comparison to the global ocean makes it hard to grab the information about the Nordic Seas.

Line 183: See comment above. As observations and model are not sampled at the same places, I think you compare measurement uncertainties with the spatial heterogeneity (which is large in the Nordic Seas). I do not think that this is valid here. Moreover, the trend in the observations seems to be much larger than the simulated trend. It would be helpful to subsample the model at the same places as the observations to make a proper comparison.

Line 184-187: The trends are significantly different. Slightly stronger is misleading when the difference is 29% and more than two times the standard deviation. Could you discuss this please and try to explain why this is the case? You mention the variability at the beginning of the time period, but I only see one measurement at the beginning of the time period (first 5 years). And if I only look at the years from 1994-2020 (excluding the beginning, the trend seems to be even stronger). I think you cannot wave this away by mentioning variability.

Line 188/189: Please be quantitative: How much more than 0.4? How much below 7.7? What are the uncertainties? You have the numbers, so just give them here. Please be consistent with the significant digits in earlier reported values, either always give two numbers behind the comma or one.

Line 189/190: As written above, the shrinking difference between global pH and high-latitude ocean pH is expected as cold waters can hold less  $C_{\text{ant}}$ . Please tell the reader here how much closer they get. It would be better to make a paragraph for the global-Nordic Sea comparison.

Lines 190/191: This hypothesis is very speculative and comes out of nowhere. This is your results section. If you mention it here, you have to show results that support your claim. Can you somehow use the model output to show what drives the shrinking difference between global and Nordic Sea surface ocean pH?

Lines 193-195: Again, please be exact and give uncertainties. Please also restructure the paragraph. First you mention RCP8.5, then you switch to the concept of reduced undersaturation with possible explanations and then you go back to the projections. It would be more readable if you first present the results for the scenarios and then present other mechanisms. Furthermore, the information that the pH decrease under RCP4.5 is 0.2 and half of RCP8.5 is redundant. Everyone is capable of comparing 0.2 and 0.4.

Line 195: Can you give a reference that supports that the RCP2.6 is necessary to limit warming to 2°C. I think, depended on the model other scenarios might also be good enough?

Line 196: What is “slightly” above 8? As mentioned above, please use the same number of significant digits throughout the manuscript. Please also give uncertainties for the end-of-century pH under RCP2.6

Figure 2: I do not know if the global line is needed, it makes it more complicated. Please make the shading transparent so that the other scenarios can also be seen in the first half of the 21<sup>st</sup> century. In the label you can delete the phrase “including those outside our regional boxes”, it is redundant as you say before that you use all observations in the Nordic Seas.

I suppose your observations for every year come from a limited number of cruises that are in a well-defined part of the Nordic Seas and does not represent the entire Nordic Seas. Does it this make sense to make a global ‘Nordic Sea’ trend? If a cruise takes place north of Iceland or in the Fram Strait in that year, the pH should be relatively large, but if it takes place close to Norway, it would probably be small as suggested by Figure 3b. The fact that the mapped product (Lauvset et al., 2016) is in good agreement with the model suggest that the stronger trend from 1981-2019 in the observations might just be due to the sampling locations.

Could you also try to explain why the model uncertainty in 1850 is almost twice as large as the one from the gridded GLODAPv2 product? Is the model heterogeneity too large, or does the gridded product overlook the real heterogeneity because of the extrapolation of a limited number of measurements?

Overall, I would propose a very different model-data comparison as mentioned above. I suggest sampling the model at the same time and place as the obs. I would then make a scatter plot with the observations of the x axis and the model data on the y axis. You could then plot one point per year and the points should be on one line. If the points are constantly above or below, a constant bias could be seen. Maybe color the dots depending on the year to get the time dimension in the plot.

Section 5.1: I like that you discuss this with the water masses. I would make this even more prominent. Could you for example show the different water masses approximately with lines in Figure 3? How are these water masses defined? This would certainly help to understand changes in each water mass or changes of the extend of water masses. Instead of writing warm waters from the South, I suggest for example writing warm Atlantic waters.

Moreover, this section has a lot of qualitative assessments that are not supported by evidence (see comments below). Please go carefully through it and see how you can support all the claims with evidence.

Lines 205/206: You write that  $C_T$  relates to salinity. I think you want to speak here about freshening from Greenland ice melt. If that is what you are writing about, please make it clear and do not leave the author left in the unknown about the relationship between  $C_T$  and salinity.

Line 206/207: I would present it here as southward flowing Arctic waters along the Greenland coast with high pH and northward flowing Atlantic waters with high pH or something similar. And how do you define quantitatively that temperature is the main driver? The alkalinity

distribution looks also similar to the pH distribution. If you make a statement about the main driver, can you give evidence?

Lines 206/207: I do not think that  $A_T$  and  $C_T$  effects dominate along the Greenland coast. It seems to be rather salinity/freshening. I also suggest deleting that a southeast northwest gradient exist. As you mention just afterwards, that gradient is not stable over the entire Nordic Seas.

Line 210: This information is needed in the Methods (see comments above). How well do the observations cover the Nordic Seas in space and time? Without this information, it is almost impossible to put the results in context.

Line 211: Here you speak about contrast of polar waters and Atlantic waters. That is great! I would present the results all around this idea. Polar waters, Atlantic waters, Greenland coastal waters. I think that would make it much easier to understand all these mechanisms and great findings that you present. However, at the moment I do not really know what I would expect as the paragraph above does not really use the words polar waters and Atlantic waters.

Line 214: Are the two citations the best ones here? Furthermore, the relationship between carbonate ion and temperature holds globally, but in regions with strong salinity and  $A_T$  gradients, I would not think that it holds well enough. Do you have evidence that the relationship holds in these conditions? I really miss the effect of freshening, which can be large in high-latitude oceans.

Lines 216-218: This is speculative and no result. I would suggest deleting it.

Line 222: Please be quantitative, what do you mean by “rather uniform”? Isn’t there a gradient from west to east? Especially along the Norwegian coast, the changes seem to be larger. Can you explain this?

Line 223-224/227-2289: Isn’t the change in the saturation states just larger in the Atlantic waters because the saturation states are larger and therefore the same changes in the drivers ( $C_T$ ,  $A_T$ , T, and S) lead to different changes in the saturation state due to the non-linear scale of the saturation state? Wouldn’t it be better to look at  $[H^+]$ ?

Line 225: The  $pCO_2$  undersaturation comes out of the blue here. Do you have a figure to show this? Do you show somewhere the larger  $CO_2$  uptake in Atlantic waters? In any way, wouldn’t it be expected that Atlantic waters take up more  $C_T$  given the higher Revelle factor in Arctic waters compared to Atlantic waters? And what is the role of sea ice changes over the historical period? I find this a little bit fast given the complexity of the system.

Line 229: I do not agree that the impact of acidification changes with depth. It is the acidification that changes with depth or the impact of changes in  $C_T$  on acidification. This is a difference. Furthermore, I would not say “limited connection” but rather refer to the time it takes to ventilate the deeper ocean and the age of the deep waters.

Line 232: Is this change significant when accounting for uncertainties? See Terhaar et al. (2020) as an example of uncertainties related to the ASH.

Figure 3: Please adjust the colorbars. For your work, the Baltic Sea is not important, so you might not need to go down to pH values below 8.05. Along the same lines, it would be helpful to adjust the range of the colorbars of the saturation states. For calcite the shown range seems to be 2-5 while the colorbar goes from 0 to 5. Similarly, you might be able to better show changes (c,f,i) if you chose an adjusted range. Consider also to mask the Baltic Sea as ESMs are usually not made for these small basins.

Figure 4: Can you show the section on a map, maybe in figure 3? As mentioned for figure 3, could you try to adjust the range of the colorbar. The shown colors do not seem to go far below the 8, but the range of the colorbar goes down to 7.8. Moreover, I find it confusing to see pH and the saturation horizon on the same plot. Wouldn't it be better to show the filled contours of the saturation state?

Section 5.3: In line with the comment above, I suggest changing the term "Present trends". I would furthermore suggest to restructure the section and to explain regions that are similar together. Going from region type to region type might be easier to follow than going from the surface ocean to the deep ocean. I struggled to read this.

Line 237: What do you mean by decreasing order?

Line 237-238: It is normal that largest decreases are seen at the surface, right? As the increase in atmospheric CO<sub>2</sub> is exponential, the decrease in pH should be larger in years that were recently in contact with the surface.

Line 238: Can you explain what this uncertainty refers to, how it is calculated?

Line 239: If you only have observations for a short time, does it make sense to calculate a trend over the entire period? You make an extrapolation and assume that the trend holds. I really do not think that this is a good idea.

Line 243: Can you provide uncertainties for the estimates by Skjelvan et al. (2014)?

Line 244: If Skjelvan et al. (2014) used different observations (sampling period, different region, and different seasons), can you compare these two estimates? If you provide this as an explanation for the difference, I would like a prove. Could you use the same region as Skjelvan et al. (2014) and the same time period (1981-2013) to demonstrate that your hypothesis is right? If not, I do not see how these two estimates are comparable. And if they are comparable, I would like to know why you find a significantly reduced trend in pH. I would be really interested in knowing if a shift exists in the reaction of pH in the Norwegian basin after 2013. Alternatively, the change in the trend might also be due to the non-linear scale of pH referring to the question if a linear trend in pH (a non-linear unit) makes sense.

Line 247: I would not write relatively strong and just give the units.

Lines 253-259: Most of this is material for the Discussion and I suggest moving it there.

Lines 269-270: What exactly is “close”? Can you be quantitative?

Figure 5/6: Please zoom into each subplot. The way it is represented at the moment, it is hard or impossible to read the data and the uncertainties (see 6t as an example). I see that you want to have the same limits for all depth levels for better comparison but that makes most subplots unreadable.

Could you precise how you calculated the uncertainties in the aragonite saturation state and pH? Are they calculated in GLODAPv2? Did you calculate them using uncertainty propagation? Did you account for uncertainties in the equilibrium constants?

Line 278/279: Why do you present these numbers twice in the Results? I think I have already seen them earlier. Please add the uncertainties.

Line 283: What exactly is a “Small region”

Line 284: Please show the changes in  $p\text{CO}_2$ . You prominently highlight  $p\text{CO}_2$  in the abstract, so I would like to see changes.

Line 295: Wasn't the seafloor already undersaturated? And isn't it the change in the saturation state that leads to undersaturation?

Line 295/296: This is speculative. Can you add evidence for this hypothesis?

Figures 7/9: Please adapt the colorbar range according to the colors shown in the plots. I strongly suggest changing the colors around saturation states around 1. Having the same color for saturation states from 0.5 to 1.5 does not clearly indicate where waters change from undersaturated to oversaturated, although this is the most interesting region.

Figures 8/10: I suggest merging both figures. I would also show the aragonite saturation state instead of pH and I would add uncertainties for the saturation horizon (See comment above). More importantly, how representative is this section for the Nordic Seas. As you show, the acidification trends in deeper waters strongly depend on the regional sea. Thus, the rise in the saturation horizon should also depend on the regional sea. Is this section at the higher end or lower end of Nordic Sea acidification?

Lines 299-300: See comments above about the idea of changes in ocean pH expected from atmospheric  $\text{CO}_2$  increase. It would be quite naïve to think that the ocean pH depends only on the atmospheric  $\text{CO}_2$  increase. Although it might be a useful comparison, I would not state it this prominently at the beginning of the Discussion.

Line 301: When did you show something about climate variability? Do you mean variability in the atmospheric climate? I really do not know what you are referring to here.

Line 307: Can you extend the buffer capacity explanation a little bit more?

Line 312: In the Fram Strait and the Greenland Sea, the slope of atmospheric CO<sub>2</sub> and marine pCO<sub>2</sub> is almost the same. The difference between atmosphere and ocean is only valid in a subset of the Nordic Seas, right? And if that is true, can you explain these differences?

Lines 314-328: You mention a lot of possible mechanisms and explain them well. This is what I would like to see earlier (in the Introduction). Here, it becomes all more understandable why you insist on the change in undersaturation. You say that any further exploration is beyond the scope of this study. I find this very disappointing. Your manuscript concentrates on this, leads the reader to this point. You show all changes and then stop at explaining them. It would really make the manuscript much stronger if you could somehow at least conceptually find the most important mechanisms in each region.

Line 330: Could you add some Discussion about the alkalinity behavior in the Barents Sea Opening. In light of the recent study by Asbjørnsen et al. (2020), it seems that Atlantic water extends far more northward and would thus increase alkalinity. It is more a displacement of Arctic Water in the BSO than a change of the existing water mass. Could you somehow elaborate on this and discuss possible impacts? You somehow waver around it, without really explaining the driving mechanisms.

Lines 337-341: It looks surprising that all the cited studies find a warming of the Nordic Seas while you find a cooling at the surface. You only give a possible explanation without evidence. This is disappointing as differences to other studies are most of the time what helps understanding the underlying mechanisms. Could you somehow test, if the sampling is the reason for the opposite trends? If it is, how robust are your estimates overall? If it is not the sampling, why are your results different?

Lines 343-345: Is the increase in temperature in these depths caused by increased temperatures in Atlantic waters that flow northward at these depths towards the Arctic Ocean? Please use the Discussion to explain the mechanisms. What did Osterhus and Gammelsrod (1999) say about it.

Line 346: I think you already mentioned the compensation of C<sub>T</sub> and A<sub>T</sub>? Do you need to repeat it here?

Lines 346-349: You point out regions where changes of C<sub>T</sub> and A<sub>T</sub> cancel each other without mentioning the drivers. I find that unsatisfactory. Once, you show me something interesting, I would like to know why this happens. Can I expect that to continue in the future, or is A<sub>T</sub> only temporarily slowing down acidification and once the A<sub>T</sub> stops, acidification will be even stronger than before? A lot of open questions that I would like to see discussed.

Lines 349-352: If the C<sub>T</sub> signal goes deeper in the Lofoten basin, this looks like stronger deep-water formation in this region. Is this the case? Could you test it in the model or with observations of transient tracers?

Lines 353-360: Can you explain why alkalinity changes are very important in the present day trends, but in the past and the future C<sub>T</sub> is by far the most important driver in all regions? I would suspect that it might be due to the model. If I understand everything right, Figure 11 is

based on observations, but Figures S9-S11 are based on the model. Could you make Figure 11 only with model results. If the alkalinity contribution is not significant in the model but in the observations, I think that the model misses important processes. In this case, many of the conclusions in this paragraph do not hold ( $C_T$  being by far the main driver for the past and the future). If the model shows similar results as the observations, however, than it would be really interesting to know what is different in the present-day situation, maybe it is the beginning shift from constant to declining land and sea ice? And if the NorESM-ME model does not show the changes in alkalinity, it would be interesting to look at other models (something I think would be necessary anyway).

Line 360: This is speculative: If it is related to freshwater export from the Arctic Ocean, why would it only be seen in the Iceland Sea.

Lines 363/364: What do you mean by “We relate this ... excess  $CO_2$ ”? Please extend your explanation here. I think the undersaturation is just a “normal” part of the Nordic Seas and only in the RCP2.6 it is gone, because the atmospheric  $CO_2$  trend reverses.

Line 374: This is speculative. You need to show the AOU and try to quantify the effect on  $C_T$  if you want to make this point.

Line 376: You continue to speculate, including changes in deep-water formation. This is really interesting and makes it even more frustrating not to know why this happens. Can you try to elaborate on this? A reader does not want to be kept in suspense.

Line 389: Could you quantify this instead of saying most of? It is 321 out of 324, right? So you could just say: Out of 324 reefs in the Nordic Seas, 321 are at depths of 0-500 m for example.

Lines 393/395: I would not really spend so much time on one out of 324 reefs. It is not that important and somehow takes the wind out of the sails of your really strong message here.

Line 396: If I understand it right, under RCP2.6 and 4.5 only 3 out of 324 reefs will be exposed to undersaturated waters. That is good news. Can you try to make this clearer as it might be interesting to policy makers. At the moment you say the deepest reefs without saying how many reefs.

Figure 11: I really like the decomposition. Nice plot. However, as mentioned above, I do not think that the stars (expected trend) make sense for waters below the surface. As the deeper waters are not in contact with the atmosphere, they cannot see the accelerated trend in pH (see figure 2).

Conclusion: I find the Conclusion relatively weak in comparison to the findings of your paper. At the beginning it reads more like a summary of your results and it ends with a relatively complicated paragraph about the difference in partial pressure that is very technical. I would like to encourage you to really highlight your main findings and why they are important. You have many interesting messages in this manuscript!

Supplementary Figures: Please correct them in accordance with my comments on the figures in the main manuscripts.

## Technical notes:

Line 4: I think it should be “from...to” instead of “since...to”

Lines 5-6: It is difficult to compare “in the last 40 years” to “between 1850-1980”. Could you use the same format, for example “between 1980 and 2020” and “between 1850 and 1980”?

Line 13: I suggest to write “is projected to” instead of “will be”

Lines 20 and 24/25: “Since 1750” is repeated again, maybe you can look for another formulation?

Line 33: returns

Line 39: “Projected” instead of “expected”

Line 51: “are” instead of “is”

Line 51: Maybe write “qualitative effects” instead of “(direction only)”

Line 86: Is “visited” the right word?

Line 90: Are all these citations necessary? Can they be somehow grouped?

Line 136: I find it hard to understand the word “present trends”. I suggest writing the trends over the last 40 years or something similar.

Line 184: An uncertainty of 0.00 looks weird.

Lines 185/186: Please use the same number of significant digits. And is mpH a common unit? I suggest writing  $1e^{-3}$  pH, but it is your call.

Line 187: I would start a new paragraph here.

Line 212: To be consistent, I would not use the word saturation state but only  $\Omega$ .

Lines 247-250: Sometimes you give positive and negative trends. I see how that fits into your writing, but it is confusing. Could you stick to write about decreases and only give positive numbers or speak about trends and then give the negative numbers?

Line 275: Close to undersaturation or close to being undersaturated.

Line 273-277: I am not sure if this summary of the previous sections is needed.

Line 284: Delete “interestingly”.

General: I think it is undersaturated with respect to. Could you change this throughout the manuscript please?

## References

Asbjørnsen, H., Årthun, M., Skagseth, Ø., & Eldevik, T. (2020). Mechanisms underlying recent Arctic Atlantification. *Geophysical Research Letters*, 47, e2020GL088036. <https://doi.org/10.1029/2020GL088036>

Broecker, Wallace S., and T-H. Peng. "Gas exchange rates between air and sea." *Tellus* 26.1-2 (1974): 21-35.

Bronselaer, B., Winton, M., Russell, J., Sabine, C. L., & Khatiwala, S. (2017). Agreement of CMIP5 simulated and observed ocean anthropogenic CO<sub>2</sub> uptake. *Geophysical Research Letters*, 44, 12,298– 12,305. <https://doi.org/10.1002/2017GL074435>

Dickson, Andrew Gilmore, Christopher L. Sabine, and James Robert Christian. Guide to best practices for ocean CO<sub>2</sub> measurements. North Pacific Marine Science Organization, 2007.

Fassbender, A. J., Orr, J. C., and Dickson, A. G.: Technical note: Interpreting pH changes, *Biogeosciences Discuss.*, <https://doi.org/10.5194/bg-2020-348>, in review, 2020.

Goris, Nadine, et al. "Constraining projection-based estimates of the future North Atlantic carbon uptake." *Journal of Climate* 31.10 (2018): 3959-3978.

Kwiatkowski, L. et al. Twenty-first century ocean warming, acidification, deoxygenation, and upper-ocean nutrient and primary production decline from CMIP6 model projections, *Biogeosciences*, 17, 3439–3470, <https://doi.org/10.5194/bg-17-3439-2020>, 2020.

Meinshausen, M., Smith, S.J., Calvin, K. et al. The RCP greenhouse gas concentrations and their extensions from 1765 to 2300. *Climatic Change* 109, 213 (2011). <https://doi.org/10.1007/s10584-011-0156-z>

Orr, J. C., et al. "Routine uncertainty propagation for the marine carbon dioxide system." *Marine Chemistry* 207 (2018): 84-107.

Tanhua, T., D. W. Waugh, and D. W. R. Wallace (2008), Use of SF<sub>6</sub> to estimate anthropogenic carbon in the upper ocean, *J. Geophys. Res.*, 113, C04037, doi:10.1029/2007JC004416.

Terhaar, J., Tanhua, T., Stöven, T., Orr, J. C., & Bopp, L. (2020). Evaluation of data-based estimates of anthropogenic carbon in the Arctic Ocean. *Journal of Geophysical Research: Oceans*, 125, e2020JC016124. <https://doi.org/10.1029/2020JC016124>

Terhaar, Jens, Lester Kwiatkowski, and Laurent Bopp. "Emergent constraint on Arctic Ocean acidification in the twenty-first century." *Nature* 582.7812 (2020b): 379-383.

van Vuuren, D.P., Edmonds, J., Kainuma, M. et al. The representative concentration pathways: an overview. *Climatic Change* 109, 5 (2011). <https://doi.org/10.1007/s10584-011-0148-z>

Wang, Chunzai, et al. "A global perspective on CMIP5 climate model biases." *Nature Climate Change* 4.3 (2014): 201-205.