Dear Editor,

We are happy to submit a revised version of our manuscript Nordic Seas Acidification. Following the comments of you and the reviewers, the manuscript has undergone a major revision.

Important changes that have been done include:

- Inclusion of all individual references to the cruises that we use from the GLODAP database
- inclusion of a larger set of ESMs from CMIP5 (Figure S5 and Section 4.2)
- Addition of uncertainty analysis with error propagation (section 2.2.3)
- Addition of an analysis where we quantify the drivers of the spatial pH-variations in the Nordic Seas (section 2.2.1 and 3.1)
- Corrections of our calculations, i). past and future pH are now calculated with the modelled changes in the drivers, and not the modelled changes in pH, ii) the calculation of perfectly tracked pH has been revised after the comment of reviewer 2
- Inclusion of modelled drivers in the main manuscript (Figs. 12 and 13)
- Inclusion of a subsection called “Theoretical Background” (Sect. 1.1), where we better describe the chemistry behind ocean acidification, including the connection between pH and H^+ ions
- Revision of the writing
- We found the uncertainty in the sea surface pCO_2 trends to be too large to be able to state that they are increasing faster than the atmospheric one, and we therefore now put less focus on the pCO_2.

Please find below point-by-point answers to the comments of the reviewers (pages 2-30, 21-39, and 39-40, respectively), and a more detailed description of what changes that have been made. Note that the line numbers are slightly different from the ones in the comments posted online, since we added some more references in the manuscript after this was posted.

We are truly grateful to you and the three reviewers for your work and comments, which we think has led to major improvement of our manuscript!

Best regards,
Filippa Fransner and co-authors

Responses to Reviewer 1:

We want to thank reviewer 1 for the extensive work and very detailed review. We enjoyed working through the comments, and we think that they have led to a major improvement of the manuscript. Please find below our responses (in blue italics) to all the comments (in black.) References to line numbers, figures and tables refer to the revised manuscript, unless stated otherwise.

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.

We agree that the discussion on underlying processes of observed changes can be improved, and we therefore address this more fully in the revised version.
We have found that the most important driver in all regions and over all time scales is the increase in DIC from anthropogenic CO_2 *uptake. The second most important driver is a general increase in alkalinity. Because changes in alkalinity related to salinity changes are cancelled out by the respective changes in DIC, the processes of alkalinity-driven pH changes are due to biogeochemical processes. The third most important driver of pH change is increases in temperature. We have added references that discuss the reasons behind these changes, and the relative role of different water masses in Section 4.1.*

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.

We have carefully gone through the manuscript to make sure that we provide quantitative assessments. In particular, we have made an additional analysis of the present distribution of pH and saturation states to quantify the underlying drivers of spatial variations (sections 2.2.1 and 3.1).
However, we prefer not being too quantitative/detailed in the abstract; it would make it too exhaustive. This is why we have not followed the suggestions of the reviewer (that are found further down) in this part of 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 pCO 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.

It is a good idea to do some merging of these sections. After some thinking, we ended up dividing the introduction into two subsections; one general introduction on ocean acidification and the Nordic Seas, and one on “theoretical background”, where we give an introduction to how ocean acidification works with some chemcical background. We also added some text on current modelling of the region.

Undersaturation of pCO^c2: 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 pCO_2. At other times, I am very confused when you compare the pCO 2 in deep waters (2000 m) to pCO 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 pCO 2 decreases faster as their solubility increases than they can take up 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 ant . With increasing atmospheric CO 2 , the differences in surface ocean C T should become smaller, which would also lead to a smaller difference in marine and atmospheric pCO 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 ant is larger in the northward flowing waters than expected from atmospheric CO 2 concentrations. This larger C 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.

We agree that the concept of undersaturation of pCO$_2$ needs a better introduction and discussion, and be used more carefully, such as avoiding this term when discussing deep waters. We have therefore added some text about the air-sea pCO$_2$ difference in the introduction, both in the paragraph about the Nordic Seas, and also in the subsection theoretical background. Here, we now explain what we mean by pCO$_2$ undersaturation, and the processes underlying the pCO$_2$ undersaturation in the Nordic Seas (lines 37-40). We have further removed the comparison with atmospheric CO$_2$ change in the subplots of current trends below surface waters. However, because changes in the degree of undersaturation is not the main topic of the paper, we don't want to go too much into details in the introduction. To this end, to actually pin down the underlying processes we would need an extensive analysis far beyond the scope of this paper, for example one would need to take into account upstream changes in the seawater chemistry, which would require a larger set of observational data. This is why we have left the discussion on the possible mechanisms behind a change in the pCO$_2$ undersaturation as it is. Here, we are also discuss the potential role of an increasing import of anthropogenic carbon as suggested by Olsen et al. (2006) and Anderson and Olsen (2002), which is similar to the process that you mention was described in Terhaar et al., 2020.
Also, please note that after the analysis of the impact of seasonal undersampling, we found that the uncertainty is too large for us to state that the ocean pCO$_2$ is increasing faster than the atmospheric one. We therefore removed the statement on pCO$_2$ from the abstract. We added a discussion on this in Section 4.1.

Anderson, L. G. and Olsen, A.: Air–sea flux of anthropogenic carbon dioxide in the North Atlantic, Geophysical Research Letters, 29, 16–1–16–4, https://doi.org/10.1029/2002GL014820, https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/2002GL014820, 2002.

Olsen, A., Omar, A. M., Bellerby, R. G. J., Johannessen, T., Ninnemann, U., Brown, K. R., Olsson, K. A., Olafsson, J., Nondal, G., Kivimäe,C., Kringstad, S., Neill, C., and Olafsdottir, S.: Magnitude and origin of the anthropogenic CO2 increase and 13C Suess effect in the Nordic seas since 1981, Global Biogeochemical Cycles, 20, https://doi.org/10.1029/2005GB002669, https://agupubs.onlinelibrary.wiley. com/doi/abs/10.1029/2005GB002669, 2006.

Line 9: You point out that the Barents Sea Opening is different to all other regions because of an increase in AT . 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 AT expanding further and displacing Arctic Ocean waters? I really would like to have an explanation here.

We agree that the strong increase in AT is very interesting. Unfortunately, it is not possible to make any conclusions on the underlying mechanisms from the little data we have, and we can only discuss the possible reasons behind. The decomposition into a freshwater-driven component and a biogeochemical - driven component (Fig. S19) suggest that it is not a result of changing salinity, but of biogeochemical processes, or possibly sampling issues. One biogeochemical process that could impact AT is the recurrent blooms of coccolithophores in the Barents Sea, which have shown a tendency to increase. We are now discussing this in section 4.1 at lines 495-507.
Please note that, after investigating the effect of seasonal undersampling, we found that it is the apparent cooling that lies behind the weak trend in pH, and not the strong increase in alkalinity (there is also an exceptionally strong increase in CT, counteracting the effect of alkalinity).

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 pCO 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.

Thank you for this comment. We decided to analyze the output of several CMIP5 models for the section on cold-water corals (section 4.2), where the results (i.e. the future location of the saturation horizon), is sensitive to model uncertainties. For the other sections, which mostly focus on drivers and thus process understanding, we decided to stick with NorESM1-ME only. We additionally made a time-series plot showing the simulated pH at three different depths of the Nordic Seas from NorESM1 and several CMIP5 models, that we put in the supplementary material (Fig. S5) and that we mention in Sections 3 and 4.2. Unfortunately, for emission driven runs, only output for the high emission scenario (which we used for the manuscript) are available on the CMIP5 data portals, which is why we could not make a sensitivity analysis for the other scenarios.

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 pCO_2 in winter. Given the strong seasonality, I feel that an analysis of the potential bias is necessary.

Thank you very much for pointing this out!
The idea of subsample the model to get an estimate of the uncertainty related to the seasonal undersampling is a good thought. However, there are also uncertainties related to such a model-observation comparison, which would make it questionable. These uncertainties are related to the model's capability to simulate the seasonal cycle (including the magnitude of the seasonal drawdown of DIC, and the timing of blooms), which have not yet been fully evaluated for the Nordic Seas. We believe that an evaluation of the seasonal cycle of NorESM1-ME lies beyond the scope of this paper, as we focus on longer time scales.

To address the issue of trends being biased due to an uneven sampling between winter and summer months, we instead calculated the trends for the productive season only (months March -September).

It did not result in any significant differences in the pH trends in the 0-200m layer, but it made us realise that the seasonal subsampling has quite an effect in the regions that we thought were sticking out. This includes the weak negative trend of pH in the Barents Sea Opening and the apparent cooling we see in the surface. Further, we found that our results with the sea-surface pCO_2 increasing at a faster rate than the atmospheric one were very sensitive to the choice of months, which makes these results very uncertain. This is discussed in section 4.1.

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.

The purpose of figure 2 is to give an overview of the different datasets and the acidification in the Nordic Seas. The idea behind the use of spatial STD for the model output is to show that the model and observations compare reasonably well. Because we use a spatial mean for the whole region from the model, which is not exactly comparable to a mean of the pointwise measurements from observations, we also want to put in the modelled spatial heterogeneity.
Considering that the objective of this paper is to analyse the observation and model data separately (present, past and future climate), we do not think that we need to make a more advanced comparison between the two datasets.

You are, however, right that it is not correct to call it an uncertainty, and that it should be heterogeneity. We have corrected this.

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 nd 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.

We now explain how the significance is calculated (lines 229-231 in the revised manuscript). We have additionally removed the 95% confidence intervals (the parabolic lines) from all figures. These lines showed the confidence intervals of y-values when using the equation obtained from the regression. We realise that this does not add any value to the paper, and that the significance of the slope, indicated in Table 5 and 6, is enough.

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 $\mathrm{pCO_2}$.

We have changed this according to your suggestions.

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.

The reviewer is correct that we want to refer to the lower saturation states induced by the low temperatures. It is therefore more appropriate to use the word vulnerable than the word sensitive. We have corrected this. However, projected changes in the high latitudes are smaller because of low initial carbonate ion concentration. It is the opposite for pH, we expect larger changes at high latitudes because lower carbonate concentration equates with lower buffer capacity. We now also give a more thorough introduction to this in the subsection Theoretical Background.

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.

In general we want to avoid too much detail in the abstract, and will therefore not clarify it here. However, we are now clarifying pre-industrial in the main manuscript in relation to each data-source.

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

We have removed the lines about the change in pCO_2 undersaturation from the abstract, as we found these results to be very sensitive to the choice of season (see explanation above).

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?

We have removed the use of "acidification signal" from the manuscript.
Regarding the saturation horizon: with the current analysis that we do in the manuscript we cannot be more precise than this. We are now mentioning the depth of the pre-industrial saturation horizon in the abstract.

Line 11: What do you mean by significant?

We have removed the word significant from the abstract. We now explain how we calculate significance in section 2.2.2 in the main manuscript.

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

We have made an attempt to make the abstract clearer by some rephrasing and restructuring.

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

We have made an attempt to make the abstract clearer by some rephrasing and restructuring.

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

We have changed this according to your suggestion.

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

After revision of the abstract this sentence has been removed.

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

After revision of the abstract this sentence has been removed.

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

You are right that it is not exactly 25%. We will change this to “about 25%”.

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

We have revised the sentence taking into account the reviewer’s comment.

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

We have added some text explaining this in the first paragraph of the introduction.

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.

The reviewer is correct that also the subsurface is ill-constrained, which is one of the reasons that we put some focus on that in this manuscript. We have reformulated the sentence.

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

We now avoid the term “ acidification signal” in the manuscript.

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.

We now give a deeper explanation of the connection between cold waters and low saturation states in the subsection Theoretical background. It is a good thought to make this sentence a topic sentence of the last paragraph. However, after all the revisions of the introduction, we think that it would not fit in anymore.
We also explain the low saturation states in the paragraph on the Nordic Seas.

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?

It is correct that it should have been anthropogenic carbon. After the re-organisation of the introduction, this sentence has been removed.

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

We think that the table gives a visual aid for the reader to follow the rest of the manuscript, and would therefore like to keep it. We now write in what direction salinity affects pH.

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.

It has been deleted.

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 timer series stations that you mention in 4.1?

We have edited the figure after the suggestions of the reviewer. The only thing that we did not do was to remove the data-points outside of the Nordic Seas. We think that it is important to show how the choice of the region-boundaries impacts the amount of data.

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.

Nutrients were used for the CO2sys calculations, so we will keep them. We also added silicate and phosphate to the topic sentence. However, we did not use oxygen, so it has been removed.

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.

We think that our time series figures (Fig. 6 & 7), together with the map showing the regional distribution of samples (Fig. 1) well illustrate the time/space coverage of the data we use and provide enough information for the content of the paper. We have, however, added some figures showing the sampled seasons per year in the supplementary material (S8-S13), that will support our discussion on seasonal undersampling.

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 it the Norwegian ocean acidification monitoring program? Could you introduce it with one sentence?

We have divided the text into paragraphs and added some information on the Norwegian ocean acidification monitoring program.

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

We are afraid that we cannot be more precise than this because it has been varying in time. We think that our time series figures, together with the new supplementary figures S8-S13, provide the information on the time/depth coverage that is necessary for the manuscript.

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)?.

We have added the citations. With the choice of RCP 2.6, 4.5 and 8.5 we wanted to include one mitigation scenario, one stabilization scenario and one high-emission scenario (van Vuuren et al., 2011), which we now explain in the manuscript. RCP6.0 is another stabilization scenario, and is not included because it would make the paper too long.

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?

We removed this sentence because this paper does not discuss the performance of NorESM1-ME in the Nordic Seas.

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

We have added some text to provide a more quantitative assessment.

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 pCO 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.

We have added a timeseries figure with modelled pH from several CMIP5 models at three depth ranges in the supplementary material (fig. S5).

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?

We are grateful that the reviewer (and also reviewer 2) identified this problem. In the first version of the manuscript we added the modelled changes of pH and saturation states. In the revised version we have calculated future and past pH from 3D fields containing the GLODAP climatology plus the modelled changes of temperature, salinity, alkalinity, DIC and nutrients. This is described in Section 2.2.2.

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-dependece 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 the let the reader know here that this pre- industrial C T has these uncertainties.

We now provide an estimate of the uncertainty of the GLODAPv2 anthropogenic carbon.

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 measurmentes where not available?

We used nutrients for the CO2sys calculations. When measurements of phosphate and silicate were not available, their concentrations were set to 1 and 5 umol/kg, respectively, the errors incurred by this approximation are negligible. This is now clarified.

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.

We think it is important to refer to this work as it has shown that the ratio works well in the Nordic Seas, which validates our choice. We have revised the text to make it less confusing.

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.

We thank the reviewer for this suggestion, we think that this is a good idea, and are now estimating uncertainties with the CO2SYS uncertainty propagation (section 2.2.3).

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 regions, although this is a good start.

A description of the watermasses and the deep water convection is now in the introduction. However, one of the reasons behind the choice of these regions was to get a representation of the different water-masses. We will therefore keep this argument here. Another reason behind the choice of these regions was the data availability, which we now also clarify.

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.

The aliasing effect is one argument. As we wrote in the manuscript, we want to prevent including different water masses in the same region which could have large effects on the trends. Further, we wanted to constrain the regions to the areas with regular sampling, and avoid regions that are less sampled. We have elaborated the text on this.

We now clarify that it is the Eastern Fram Strait that we are working with. We have changed Fram Strait to Eastern Fram Strait throughout the manuscript.

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?

Summertime MLDs are normally less than 200m, but 200m were chosen anyhow as this is the absolutely lower limit for seasonal changes. We revised the text to make it clear.

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

We now explain how the significance was calculated in section 2.2.2.

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).

Also reviewer 2 commented on this, and that the results can be different if looking at [H+] instead of pH, because pH shows a relative change. However, in this study the pH-variations are relatively small, meaning that the results won't change much if looking at H+ instead. To verify this we made all plots showing pH change also for H+ (we included one representative plot in the Supplementary, Fig. S17). Also, we do want to stick to pH because we aim to make the paper accessible also for outside of science, such as people dealing with environmental policies, and pH is a more well known quantity.

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

We only use the abbreviations T and S in the equations, and therefore think that it is better to introduce them just before these.

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.

The figures of past changes show model output. We decided to stick to the model data to be consistent throughout the manuscript. The GLODAP preindustrial data is only shown in figure 2. We have now clarified this in the figure captions and in sections 2.1.3 and 2.2.2.

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?

No, we do not consider changes in nutrients as a direct driver of pH, due to their relatively low concentrations their impact is small. Nutrients are 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.

We provide a small introduction to pCO_2 *in the introduction of the revised manuscript (lines 37-40 and section 1.1).*

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?

We have revised the text as suggested by the reviewer. With good agreement, we mean that the mean pH lies within the spatial range (std) of pH of the two different sources.

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

It refers to the Nordic Seas surface pH. The sentence has been revised.

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 ?

The standard deviations are a measure of pH heterogeneity in the region, not uncertainty, we have revised the text to make this clear.. We do not compare the standard deviations, we are just showing that the mean of the two different data sources are within the standard deviations of each other, showing that they are overlapping.
The GLODAPv2 pre-industrial estimate is gridded.

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?

Jiang et al., 2019 showed a strong correlation between the sea surface pH and the pCO_2, and suggested the low pCO_2 in the high latitude seas to be a result of the pCO_2 undersaturation. The pCO_2 variations can in turn be an effect of temperature. This is something we discuss in section 3.1

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.

We did check this. We think that it is enough to write it in text ,and that a figure illustrating this would be excessive.

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.

We have corrected this to 0.06. However, unlike the decreases in pH estimated from the observations, the modelled decrease does not have an uncertainty (unless we estimate a trend). This is because we have a full time-and space coverage of the model data, and we know the exact decrease. We go into regional differences later in the manuscript, and will not present it in this section. We have moved the comparison with the global ocean to its own paragraph.

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 valid here. Moreover, the trend in the observations seems to be much larger than the simulated trend. It would be helpful to sub-sample the model at the same places as the observations to make a proper comparison.

The standard deviations of the measurements show an estimate of the spread of the pH over the year and within the Nordic Seas, it is not an uncertainty. We have clarified this in the figure caption. As described before, we do not want to make an exact comparison, we just want to show that the two data-sources are overlapping. We want to include the whole Nordic Seas in the model data in this figure to give a global picture of the acidification of the Nordic Seas, which cannot be obtained from observations.

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.

Please note that after the revision, the trend estimates have changed a little as we 1) masked out the Baltic Sea and 2) found that the area went to far south in the script making Fig. 2, compared to the one defined in Fig. 1. The modelled trend and the trend from the observations are not significantly different anymore. Still, we do mention what can give rise to differences between the model and the observational data in the revised manuscript (lines 310-312).

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.

We have revised this part according to the suggestions of the reviewer, and now provide quantitative estimates.

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 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.

As suggested by the reviewer, we have moved the global-Nordic Sea comparison to its own paragraph (lines 323-332), and now provide quantitative assessment of the changes. It is correct that also the lower temperatures of the Nordic Seas play a role (giving them a lower buffer capacity). We have added this to the discussion.

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?

We agree that this is a hypothesis and that it is more of a discussion. To make this clear we are now writing " This is most likely...". We think that an analysis of the drivers of the pH change of the global ocean is out of the scope of this paper.

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 cabable of comparing 0.2 and 0.4.

We have restructured the text and provided standard deviations (spatial) next to the numbers.

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?

The reference is already in the manuscript, but it was misplaced. We have moved it so that it is after this statement.

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

We have put spatial standard deviation next to the numbers, and made sure that we use the same amount of significant digits throughout the manuscript.

Figure 2: I do not know if the global line is needed, it makes it more complicated.

We decided to keep the global line, we think it is important to show that the pH -dynamics of the Nordic Seas are different.

Please make the shading transparent so that the other scenarios can also be seen in the first half of the 21 st 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.

We have followed all the suggestions of the reviewer.

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.

We were also hesitating about making a Nordic Seas trend from the observations due to the reasons presented by the reviewer. However, we think it still can be there, if we mention the caveats of doing so, which we now do on lines 310-312.

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?

We found that a few grid-points of the Baltic Sea were included in the area (Fig. 1) when extracting the model data, which caused the larger standard deviation of the model data. After masking the Baltic Sea, the standard deviation of the model approaches that of the GLODAPv2 climatology.

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.

This would have been a good idea if the focus of the paper would have been different, i.e. if we would have used model and observational data together to understand a certain process in present climate. But for this, we would also have needed another model. ESMs are adapted for longer simulations spanning over several decades, and not for going into details on shorter time and space scales. The main reason being the relatively coarse spatial resolution, and that the climatic modes of the ESMs are out of phase with the observed ones (unless using data-assimilation).
The aim of our paper is to get an understanding of past, present, and future acidification using the best available information for the various time periods. We now clarify this in the introduction. The purpose of figure 2 is mainly to give an overview of the different time periods and data sources. For the application of our manuscript, we think that it is enough to show that the two data-sources are overlapping.

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?

We have now put a line in figure 3 showing the approximate boundaries between the main water masses.

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.

A definition of the water masses is provided (Atlantic Water has a salinity >34.5, polar waters have a salinity < 34.5). We have also made a thorough revision of this section and are now able to quantify the impact of the various drivers.

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.

This section has been almost completely revised, and we do no longer mention the relation between CT and salinity. Instead, we have written that the chemical properties co-vary as a result of the contrasting properties of the water masses in the Nordic Seas.

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 quantitively 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?

We have made a thorough revision of this section and are now able to quantify the impact of the various drivers.

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.

With the new analysis provided in the revised manuscript, these lines have been removed.

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.

We have moved it to Methods.

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 now what I would expect as the paragraph above does not really use the words polar waters and Atlantic waters.

We are now introducing polar and Atlantic waters earlier.

Your idea of presenting the results focusing on different watermasses is interesting. As we see it, there are two ways of structuring the paper:

1. *focusing at specific regions, as we do now*
2. *focusing on different water masses, as suggested by the reviewer. This could be done by constraining the analysis to measurements with specific temperature/salinity.*

However, if putting the focus on polar, Atlantic and Greenland water masses, we would mainly constrain ourselves to the surface. The depth analysis that we have done would not fit in anymore, unless we look into more water masses. The structure of the paper would change drastically.

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.

We think that these citations are appropriate. With the new analysis that we present in the revised manuscript, we are able to separate the effect of temperature on the solubility and show that this relationship holds.

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

The sentence has been deleted.

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?

We have changed this to be more quantitative.

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+]?

With the new figures in section 4.1 (Figs. 12 and 13), we now can better quantify the reasons behind the changes in pH.

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.

We are now discussing the patterns of change in section 4.1. With the new figures, we put less focus on changes in pCO_2 *undersaturation.*

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.

We realize that the use of the pharising “impact of acidification” is not appropriate here, as it can be understood as ecosystem impacts of OA. What is investigated here is only OA and not the impacts. We have reformulated the text.

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

We now present uncertainties related to this.

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.

We have gone through all the plots and adjusted the colorbars. For this specific plot, the choice was not related to the Baltic Sea, but it was to include the waters close to the Greenland coast. We have therefore kept the colorbar as it is.

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.

After the suggestion of reviewer 2 we have removed the maps of calcite.

Similarily, you might be able to better show changes (c,f,i) if you chose an adjusted range.
We have gone through all the plots and optimized the colorbars. The patterns of change look a bit different after redoing the calculations with the modelled changes of temp, sal, alk, DIC, and nutrients to the GLODAPv2 climatology.
Consider also to mask the Baltic Sea as ESMs are usually not made for these small basins.
We have masked the Baltic Sea.

Figure 4: Can you show the section on a map, maybe in figure 3?
We have put a line representing the section in figures 4,8 and 10..
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.
We have revised the colorbar ranges.
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?

We tried this out, but then we decided to put in lines showing the saturation horizons to reduce the number of plots.

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.

We have changed “Present trends” to “Present day changes (1981-2019)”
We have revised/restructured the section, and we think that it reads more easily.

Line 237: What do you mean by decreasing order?
We have reformulated the sentence.

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

Yes it is normal, we have added a sentence on this.

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

The uncertainties are standard errors, we explain this in the revised manuscript.

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.

The trend is calculated over the period of data. We do not extrapolate it. We want to keep the trend lines over the whole time period to make them more visible.
However, we have now removed the 95% prediction intervals showing the possible y-values when using the regression equation obtained in the regression analysis. We realized that it does not provide any useful information for this manuscript.

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

They do not provide any uncertainties of their trends. We realise that we cannot tell anything about statistical difference without this. We have therefore revised the sentence.

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 Skelvan 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.

We calculated the trend for the period 1981-2013, which ended up as -3.3+-0.48+-10^(-3), which overlaps with their estimate. We will not attempt to use exactly the same region as them as they were using circular regions.*
Although we cannot make an exact comparison, we still think that we should mention the difference, and give potential reasons to why they differ.

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

We have removed the "relatively strong".

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

It has been moved to the discussion.

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

We are now providing more quantitative estimates.

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?

We want to keep the same limits for all subplots to make them comparable. However, after the suggestion of reviewer 2, we have moved the trend estimates to tables.
The errorbars show standard deviations of the measurements in each year, which we explain in the figure caption. As explained earlier in this document, the 95% confidence intervals, showing the possible y-values that can be obtained with the equation obtained with the linear regression, have been removed.
The uncertainties in pH and saturation state obtained from the error propagation are discussed/showed separately in Table 2 and Figures S6-S7.

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.

The reviewer is correct that this has been presented before, we have therefore removed this sentence. The spatial standard deviations have been added.

Line 283: What exactly is a "Small region"

After the revision of the calculations of future pH and Omega, the future projections do not suggest that any region will be undersaturated in OmegaAr. This sentence has therefore been removed.

Line 284: Please show the changes in pCO 2 . You prominently highlight pCO 2 in the abstract, so I would like to see changes.

Acknowledging that changes in pCO_2 is not a driver of pH change, but only another indicator, we decided to put less focus on it in the discussion, and in the manuscript in general. To better understand the regional differences in projected pH changes, we instead made maps and cross sections of the different drivers. These are now shown in Section 4.1 (Figs. 12 and 13). Still, we think that it is still interesting to mention that the sea surface pCO_2 of the Nordic Seas does not perfectly track the atmospheric one. We therefore put a figure showing this in the supplementary material (Fig. S14).
Because of the large uncertainty in the pCO_2 trends for the present climate (1981-2019), we decided to remove the sentences on pCO_2 from the abstract.

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

Only the seafloor below 2000 was undersaturated.. we clarified this by writing "the entire seafloor ". It is correct that we should write the change in saturation state, and not the change in pH. This has been changed.

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

We have removed this sentence.

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.

We have put red-blue colorbars instead of a red-white-blue. We have revised the range of the colorbars to better comply with the range in the plots.

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?

We have merged the figures. We chose this section because it is in the middle of the domain and it nicely shows the impact of deepwater formation in the West, and the Atlantic surface layer in the East. Although the acidification rates may vary slightly, it captures the main physical features that are important for the Nordic Seas.

Lines 299-300: See comments above about the idea of changes in ocean pH expected from atmospheric CO 2 increase. It would be quite naïve to think that it the ocean pH depends only on the atmospheric CO 2 increase. Although it might be a useful comparison, I would not state it this prominently at the beginning of the Discussion.

We have removed the first lines 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.

We have also removed this part to avoid confusion. It was not necessary for the manuscript.

Line 307: Can you extend the buffer capacity explanation a little bit more?
We decided to remove this sentence. We realised that it can confuse the reader because it speaks about inter-basin variations, which we cannot assess with the observational data due to the temporal difference in data-coverage.

Line 312: In the Fram Strait and the Greenland Sea, the slope of atmospheric CO 2 and marine pCO 2 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?
After recalculating the expected change in pH after the suggestions of reviewer 2, and also checking the significance, the difference between the change in oceanic and atmospheric pCO_2 *is only significant in the Norwegian Basin and the Iceland Sea.*
As described in the manuscript, we will not attempt to compare the different regions, due to the difference in data coverage.

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.

We are now briefly mentioning the pCO_2 *undersaturation in the introduction. As explained earlier, we do not want to go into the same detail in the introduction as in the discussion because that would move the focus away from pH.*
We think that finding out the reasons behind the change in pCO_2 *undersaturation goes way beyond the scope of this paper. The reasons can be any, or a combination of, of the mechanisms discussed. To actually pin that down we would need extensive additional analyses and as such this requires a separate manuscript*

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.

We have evolved the discussion on this in the manuscript. The decomposition into a freshwater-driven component and a biogeochemical - driven component (Fig. S19) suggest that this strong increase in alkalinity is not a result of changing salinity (and thus a shift in water masses), but rather of biogeochemical processes or possibly uncertainties related to sampling. We discuss this in Section 4.1 at lines 495-507.

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 differen?

We tested this by plotting all available temperature data, and not only the ones taken at the same time as the DIC and alkalinity samples. When doing so we get a warming trend, and it is therefore clearly an issue of undersampling. We are now discussing this in section 4.1.

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.

The temperature increase is likely a result of a reduction in the deep-water formation, which has led to an increased presence of Arctic Ocean Deep water . We have added a number of references on this, including the one of Osterhus and Gammelsrod, in section 4.1 .

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

We mentioned it for the surface waters. Here we are discussing the layer below. We have revised the sentence to make this clear.

Lines 346-349: You point out regions where changes of C T and A T 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 T only temporarily slowing down acidification and once the A T stops, acidification will be even stronger than before? A lot of open questions that I would like to see discussed.

We have evolved the discussion around this, in particular for the Barents Sea Opening where the AT increase is strong and significant.

Lines 349-352: If the C T 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?

It would indeed be interesting to get a deeper understanding about the role of deep winter mixing in the transfer of ocean acidification to the deep Nordic Seas, and how this contributes to eventual regional differences. However, for this we would rather need a regional model (ESM's rarely reproduce the exact locations of deep water formation). Further, eventual observations of transient tracers would need to be combined with information about horizontal advection in order to understand how the deep water masses formed during convection spread at depth. We believe that such an extensive data analysis is out of the scope of this paper. Instead, we refer to studies that have already been done on this subject.

The literature on the Nordic Seas deep water formation and water masses do not indicate that the Lofoten Basin is a region of deep winter mixing. This takes place in the Greenland and Iceland Seas, and these water masses have then been shown to spread in the deeper layers of the Nordic Seas. The discussion about this was already there in the first manuscript, but we have revised the section in an attempt to make it clearer.

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 T 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).

Please note that also for present, CT is the main driver.

Indeed, in the old manuscript,figure 11 is based on observations, while S9-S11 are based on the model (Note that these figures have been replaced by figures 12 and 13 in the revised manuscript).

Because of the different time scales considered for the changes between past-present (145 years), 1981-2019 (39 years), and present-future (95 years), a comparison between the different time-periods should not be made.
At the timescales considered for past and future changes, the external forcing, in this case climate change, becomes an important driver. When it comes to CT being the most important driver, this is a result of the strong increase in atmospheric CO_2*, and thus the emission scenario. Here, eventual model deficiencies play a minor role. On the other hand, at the time-scale of our period of observations, climate variability can still play an important role. However, we do still see that CT is the main driver, and that the changes in CT are in good agreement with the change in atmospheric* CO_2*.*

As explained before, the aim of this manuscript is to assess the acidification in the different time periods, using the best available data, and we do not attempt to compare the different periods per se. An ESM is more adapted for long term studies of climate change stretching over a century, and not short-term studies. On the time scale of 40 years, climate variability can have an important impact on the observed changes, and the climate modes of an ESM are not in phase with the observed ones, which would complicate a comparison between observed and modelled changes.

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.

There is a recirculation pattern of polar waters in the Iceland Sea. With the new figure 12 in section 4.1 this becomes very clear.

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..

We have reformulated the paragraph and now also provide a figure in the supplementary showing the evolution of the pCO_2 difference between the ocean and the atmosphere (Figure S14).

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.

We are now providing a short calculation on this on lines 546-550.

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.

With the revision of this section, this phrase has been removed.

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.

We have quantified this.

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.

We have removed the text about the three deepest reefs.

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.

With the new uncertainty analysis the results for the RCP4.5 are not that clear anymore. We have revised the text to make the results more clear.

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).

We have removed the stars for waters below the surface.

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!

We now call this section Summary and Conclusions, and we have separated into two subsections. We think that it now reads better, and hope that the reviewer agree.

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

We have corrected them.

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

ok

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”?

ok

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

ok

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

we will revise this

Line 33: returns

after the revision of the introduction, this is not used anymore

Line 39: “Projected” instead of “expected” *ok*

Line 51: “are” instead of “is”

ok

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

ok

Line 86: Is “visited” the right word?

We tried to think of another word but could not find a good substitute.

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

They should be there.

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.

We have changed this to present day trends (1981-2019).

Line 184: An uncertainty of 0.00 looks weird.

This is what it is rounded to two digits.

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.

We have changed mpH to 1e -3 . We have also gone through the manuscript to make sure that we use the same amount of significant digits.

Line 187: I would start a new paragraph here.

ok

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

We have changed this.

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?

ok

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

ok

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

We have removed it.

Line 284: Delete “interestingly”.

ok

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

ok

Responses to James Orr (Reviewer 2)

We thank James Orr for his suggestions and all the work he has done. The manuscript will greatly improve from this, both scientifically and structurally. Please find below our responses (in blue italics) to the comments (in black). References to line numbers, figures and tables refer to the revised manuscript, unless stated otherwise.

This manuscript uses observations and a model to assess the regional details of acidification of the Nordic Seas during the industrial era through to the end of this century. The authors find that during 1981-2019, the change in surface ocean pH is larger than would be expected from the corresponding change in atmospheric CO_2 . They ascribe the cause to an evolution of surface ocean pCO_2 , which while remaining undersaturated with respect to the atmosphere, increases at a rate faster than atmospheric pCO_2 . They suggest that the main driver of the change in pH is the DIC increase associated with ocean uptake of anthropogenic CO_2 . They also find that observed pH changes may be detected down to 2000 m in some parts of the Norwegian seas. The authors further focus on corresponding changes in the saturation state of waters with respect to aragonite and corresponding changes in the aragonite saturation horizon and what those changes may mean for cold water corals. In their model, most cold water corals would not be exposed to waters that are undersaturated with respect to aragonite under the low-end RCP2.6 and mid-range RCP4.5 emissions scenarios. But under the high-end RCP8.5 emissions scenario, most of those corals would be exposed to such conditions, which are are unfavorable for their long-term survival.

Overall the authors have addressed an important topic, the details of acidification of the Norwegian Seas, a regional focus that has not been addressed previously. They appear to have used all the best data available for this assessment, thanks to the many coauthors with observational expertise in the Norwegian Seas. Also included are coauthors who are experts in using the chosen model routinely to assess ocean acidification and related aspects of ocean biogeochemistry. The Abstract and Introduction (sections 1-3) generally establish the need for this study, the Methods section appears to provide sufficient detail except for the final subsection, and the Results and Discussion sections reveal much effort being devoted to the analysis.
Yet despite these positive aspects, there is also much room for improvement.

CONCERNS in order of importance:
(1) Unfortunately, there seems to be a complete lack of understanding of what a pH change actually means. Although pH offers a convenient way to represent the hydrogen ion concentration, its log scale means a pH change actually represents a relative change in [H^+], not an absolute change (Kwiatkowski and Orr, 2018). That relative change is unlike the change in any other CO_2 system variable, all being absolute. Focusing only on pH and not [H^+] can give a completely wrong impression, e.g., as in this manuscript when it is used to compare changes at different depths and at different locations (Fassbender et al., 2020). Looking only at pH change, as in the manuscript, we cannot know what part of the change is due to a change in [H^+] and what part is actually due to differences in the reference [H^+], the starting point. The manuscript neglects this key point entirely, not even mentioning hydrogen ion concentration. Remedying this problem will require major revisions.

Thank you for pointing this out, it is a very important remark. Please note that the pH variations in this study are relatively small, and our results do therefore not look significantly different if analyzing H^+ instead (we remade all our plots showing pH change to verify this). Because the manuscript is intended to address people outside of science, we prefer staying with pH in the plots in the main manuscript. We believe that this quantity is more well known than H^+ concentration. However, we have carefully revised the text in the manuscript to avoid any misinterpretations related to this, and to discuss this issue. The revisions include:

i) link between pH and H^+ in the introduction (lines 21-23, and section 1.1)
ii) A short paragraph on the potential caveats of looking at pH change, with the references, that you mentioned (lines 204-207)
ii) one representative figure in supplementary material, with accompanying table, (Fig. S17 and Table 7), showing the trends in H^+ calculated from the observations.

(2) Projections with only one model are unreliable. Model projections are hard to publish nowadays without using multiple models and for good reason. One model can give very different results from others. A range of models provides an estimate of model uncertainty, and the model mean typically performs better than any given model. Because the ocean component of the NorESM1-ME model relies on a dynamic isopycnic vertical coordinate, we might expect it to have very different results in simulated deep-ocean anthropogenic carbon concentrations relative to most other CMIP models. Modeling centers such as the one where some of the authors of this manuscript are associated seem to now have access to and experience working with the CMIP5 or CMIP6 models. All analyses in the current manuscript need NOT be repeated with all models. But it will be needed to show at least where the NorESM1-ME model is situated relative to other Earth system models, in terms of the depth distribution of anthropogenic carbon concentrations (and perhaps also [H^+] and Ω_{Ar}) in the different regions of the Norwegian Seas.

Thank you for this comment, we agree with this. In order to show where the NorESM1-ME model is situated, we additionally made a timeseries plot showing the simulated pH at three different depths of the Nordic Seas for an ensemble of CMIP5 models that we put in the *supplementary material and that we mention in the discussion (Fig. S5). Furthermore, we decided to analyze the output of these CMIP5 models for the section on future saturation horizon and cold-water corals (Section 4.2).*

Unfortunately, the CMIP5 database only contains output for the RCP8.5 when considering emission driven runs (which we used for the manuscript), which is why we could not make a sensitivity analysis for the other scenarios. Considering that NorESM1-ME1 simulates among the higher acidification rates in deep waters in our model-ensemble, it is likely that the estimates of the future saturation horizon from NorESM is in the upper bound.

(3) The description of the decomposition of the drivers (namely the equations in section 4.4) is weak and incomplete.

a) Eq. (1) comes from Takahashi et al. (1993) and is fine except that the authors will need to replace the Greek delta δ with the correct partial sign ∂ in all the partial derivatives. This is not a major problem, just the convention of multivariate calculus. The δ is used for something else (inexact differential). Please don't confuse them.'

Thank you for noticing this, we have changed the delta to the correct partial sign.

b) Eq. (2) is added by the authors but is unnecessary. That equation comes from Metzl et al. (2010), who expanded each partial derivative in Eq. (1) to get at so-called "known quantities". Such complexity is no longer necessary because all of the partial derivatives in Eq. (1) are now easy available as precise quantities in "derivnum", an add-on package to CO2SYS-MATLAB (Orr et al., 2018). See https://github.com/jamesorr/CO2SYS-MATLAB
The simpler choice, just deleting Eq. (2), is preferred and avoids unnecessary complexity that can lead to mistakes in implementation. For instance, the authors four definitions that immediately follow Eq. (2) are ambiguous because they are missing key parentheses. Hopefully their actual code is less ambiguous. There is no longer any need to introduce all these extra terms.

We have followed the recommendations of the reviewer and removed equation 2, we agree that this is not needed, and we instead added some short text on how the derivatives were calculated. The derivnum-package would have been very useful for us, and is something we will implement in future studies. However, for this study we did not use it as the scripts and analyses were written/starting before its release.

c) Eq. (3) should be recast in the same pattern as Eq. (1), i.e., replacing fCO_2 with $[H^+]$. It should not be cast in terms of pH (as in the current manuscript) for reasons mentioned in (1) above. The partial derivatives of $[H^+]$ are also available in derivnum. That routine is called with the same arguments as CO2SYS, with one argument added in the beginning to specify what the user wants to take partial derivatives with respect to. This further move towards simplicity will avoid the old-fashioned complexity that is now in the manuscript. Furthermore, this change will help avoid misinterpretation of what changes in pH mean.

We have rewritten equation 3 in terms of d[H$^+$]/dt to prevent any confusions of what pH changes mean. Because we did not use the derivnum as explained above, we will keep the same pattern as it currently has.

d) An equation is missing in Section 4.4 concerning the freshwater Taylor-series decomposition, results of which are presented in Fig. S8. With that equation, the appropriate citations need to be given, starting with Lovenduski et al. (2007). For the associated salinity normalization, the authors must also specify their choices of the regional salinity references and if those remain constant or change with time. Furthermore, the authors will need to mention why they generally seem to prefer to use the older, less complicated decomposition from Takahashi et al. (1993).

You are correct that we forgot to add this, we have added a description with the necessary equation and references in section 1 in the Supplementary material. We prefer to stay with the older version because the net impact of freshwater is generally negligible due to the opposing effects on DIC and alkalinity (we explain this at lines 241-243).

e) Another equation is missing in Section 4.4 concerning what the authors call “pHperf”. Currently that term is mentioned in the short final paragraph of section 4.4, where the authors attempt to describe how they compute “the pH change in seawater that perfectly tracks atmospheric CO_2 ”. Unfortunately, the current description does not tell us exactly what the authors have done. For instance, in the calculation of pHperf, do the authors use i) the actual atmospheric pCO_2 as the reference value along with the atmospheric pCO_2 change or ii) the oceanic pCO_2 as the reference value, to which they add the change in atmospheric pCO_2 ? The importance of this question is illustrated with a simple example. Suppose atmospheric pCO_2 is at 400 µatm and oceanic pCO_2 is at 300 µatm. Although a 1
µatm change in pCO_2 starting at 300 µatm produces only a 0.7% greater change in [H$^+$] when compared to starting at 400 µatm, the corresponding change in pH is 30% greater in the former relative to the latter. The reason is that a change in pH represents a relative change in [H$^+$], i.e., relative to the [H$^+$] of the starting point. These numbers slightly depend on the other reference conditions, which I have arbitrarily set to T=2 ◦ C, S=35, ALK=2300 µmol/kg, nutrients=0. If the authors have used approach (i), the results will be wrong. The authors should be able to resolve this issue by using approach (ii) and by adding an equation and improving the text to avoid ambiguities.

Thank you for the illustrative example showing the importance of this. We used approach (i). We fully agree with the reviewer that approach ii) should be used. We therefore remade the calculations and the plots and we have added some text at lines 244-249 to make the description clearer. Because the calculation is relatively simple, we think that an equation is not necessary.

A related minor question: Do the authors actually use atmospheric xCO_2 (ppm) or do they first convert that to atmospheric pCO2 (µatm), making corrections for water vapor pressure and atmospheric pressure?

We did not make any corrections for water vapour and atmospheric pressure because the rates of change for xCO_2 and pCO_2 are the same, this is now clarified.

(4) The section on cold-water corals is too cursory. The authors' analysis of the change in the aragonite saturation state to which cold-water corals are exposed has potential, but the authors devote only one rather short paragraph to describing the results, which are presented in one figure. They authors also neglect to clearly attribute previous studies that have attempted the same type of exercise using model projections and cold-water coral positions. Additionally, the data set used in the manuscript for coral positions is not cited adequately, and the authors do not give enough information about their procedure for extracting the saturation state from the model. For instance, is the model sampled at the depth of the coral (as provided in the data base) or is the depth taken to be that of the model's bottom depth at a coral's latitude and longitude? More discussion of results and the addition of uncertainties from a multi-model analysis would seem critical.

We agree that this section was very short. We have revised it by:

i) working on the text and including more details about our results
ii) adding results from several ESMs
iii) adding a few references using a similar approach

We additionally provide more details about the dataset and calculations in Section 2.1.4.

(5) The writing needs improvement. Getting through this manuscript was not easy. Although there are few if any errors in English, and individual sentences generally work well, the manuscript would benefit if the authors could redouble their efforts to improve flow between sentences. That is, connections between sentences are often rough, causing the reader to slow down and sometimes stop. Also lacking is coherence in many individual paragraphs. My recommendation would be for the authors to consult the book by J. M. Williams (Style The Basics of Clarity and Grace), and in particular the short chapter on Cohesion and Coherence. Then they could go through the manuscript trying to improve both aspects. If one cannot borrow this book from a library or colleague, older editions only cost about 10 euros. It offers the potential to dramatically improve one's writing by applying a few basic principles.

We are grateful for the book suggestion. We have worked on the flow of the manuscript , and think that it now reads better and we hope that you agree.

(6) The figures need improvements. Some figures appear to have too many panels, some figures should be combined, and some figures should be deleted. There are also other issues.

The figures have been revised:

- *Fig. 1: Addition of colorbar and position for time -series stations*
- *Fig. 2: The spatial standard deviation of future scenarios are in transparent*
- *Maps and sections: color ranges have been revised, plots for calcite have been removed*

- *Figs: 6 & 7: the parabolic lines showing the uncertainty range of the regression model has been removed, and the trend-numbers have been moved to tables*
- *The sections showing future changes have been combined to one figure (Fig. 9 in the revised manuscript)*
- *figure 11: the pH expected from the atmospheric change in pCO2 has been removed from deeper layers*
- *figure 14: white space has been removed, and we have added results from several CMIP5 models*

a) In Fig. S6, it seems that only 3 out of the 6 regions seem to show a trend in surface ocean pCO_2 that is significantly greater (statistically speaking) than that of atmospheric pCO_2 . Thus I am unconvinced by the statement that it is only the Barents Sea Opening does not follow this pattern. More care is needed when handling this subject in the revised manuscript.

The reviewer is correct, we did not pay enough attention to the significance of the trends and their differences when we were discussing this. After checking the significance, we found that the trend is greater than the atmospheric one in the Norwegian Basin and in the Iceland basin. Based on the comments from reviewer 1 we also made an attempt to estimate the impact of seasonal undersampling on the trends by calculating the trends for the productive season only (April-September, and June-August, respectively). We found these results to be sensitive to which months that are used for the calculation, showing that there are large uncertainties related to the seasonal distribution of the sampling. We therefore decided to put less focus on the pCO_2 change (although we still discuss it), and we also removed this from the abstract.

b) In Figs. 3, 7, and 9, the third row of maps for Ω_{Ca} should be deleted because it exhibits the same patterns as for Ω_{Ar} in the second row, only differing by a constant. Their constant relationship could be briefly mentioned once in the text rather than wasting space in each of those three figures. Likewise, Fig. S5 for Ω_{Ca} should be deleted because it shows exactly the same patterns as Ω_{Ar} in Fig. 6.

Thank you for this remark, we have removed the figures showing Ω_{Ca}.

c) In Figs. 5 and 6, the numbers given in each panel for the slope and uncertainty should be moved to a table, where it will be easier to compare numbers between regions and depth layers. The same goes for the corresponding supplementary figures (Figs S1-S4). In addition, there are often too many significant figures in the slope and uncertainty, and the number of digits is not always consistent. Furthermore, in those same supplementary figures, the slopes have the wrong units. In regards to these C6and other figures, when statistical significance is mentioned in the text, that should be backed up with a statement of how it was determined. Such is not the currently the case in the manuscript, but it is critical, e.g., when discussing if oceanic pCO_2 is increasing more rapidly than atmospheric pCO_2 (Fig. S6).

We have adopted the suggestions of the reviewer:
i) the numbers for the slope and uncertainty have been moved to tables, both in the manuscript and in the supplementary material

ii) we have made sure that the number of significant digits is consistent
iii) we have put in the right units

A description of how the statistical significance is calculated is now found at lines 229-231.

d) Figs. 8 and 10 should be combined.

Done

e) Fig. 11 includes some details that might need to be deleted, and corresponding supplementary figures should also be refined. What is the rationale for including the dashed line and black stars in subsurface layers? Those layers have been isolated from the atmosphere for some time and we would not expect them to track atmospheric CO_2 . Showing these details in subsurface layers will confuse the reader. Moreover, would it not be better to devote a separate figure just to the subject of ocean pCO_2 tracking atmospheric CO_2 rather than trying to squeeze that information here into a very small space? Fig. S6 fills this need well. That could be brought up into the main paper. Only the top level (0-200 m) of Fig. S6 would need to be shown as we do not expect subsurface levels to track current levels of atmospheric CO_2 . I also worry about how representative the 0-200 m layer is of surface ocean pCO_2 . Some discussion on that and perhaps a modified figure seems necessary.

We agree that the dashed line and stars can lead to confusion, and we therefore removed them from the deep layers. After redoing the calculations and calculating the significance, we found that the trend is only significantly greater than the atmospheric one in two of the basins (NB and IS). We additionally tried to do the same for a thinner surface layer (0-50m), then the trend is only significantly greater in the NB. Initially, we thought that it was a good idea to bring the pCO_2 *figure into the manuscript, but since we cannot state with any certainty that the trend is larger than that of the atmosphere, we do not think that it would fit in anymore, and would bring the reader away from the focus on pH.*

In corresponding supplementary figures for the model (Figs. S9-S11), the authors miss the opportunity to compare the model results over the same 1981-2019 period as used for the model.
These figures have been removed in the revised version of the manuscript. Instead we made contour plots (maps and sections) showing the drivers associated with the modelled change, which we put in the main manuscript (Figs. 12 & 13). We think that these are more illustrative.

Regarding the suggestion about a comparison between the drivers in the model and observations, this was also suggested by reviewer 1:

The aim of this manuscript is to assess the acidification in the different time periods, using the best available data, and we do not attempt to compare the different periods per se. Such a comparison would thus not fill any function in the manuscript, with its current storyline. Further, an ESM is more adapted for long term studies of climate change stretching over a century, and not short-term studies. On the time scale of 40 years, climate variability can have an important impact on the observed changes, and the climate modes of an ESM are *not in phase with the observed ones, which would complicate a comparison between observed and modelled changes.*

By the way, why is this time span often referred to in the text as lasting 40 years; actually, it lasts only 39 years.
correct, we have corrected it in the manuscript
My impression is that relative to the observations, the model is dominated even more by the change in DIC, based on the analogous plots for the previous and subsequent time periods. These supplementary figures concern the model, but readers will be confused because ’OBS’ is used to designate the model result, both in the caption and in the figure itself. Please change ’OBS’ to ’MOD’.
These plots have been replaced by Figures 12 and 13 in the new manuscript.

f) Fig. 12 has too much white space.

We edited the figure to adjust for this.

g) The supplementary figures are mentioned out of order.
This has been corrected for.

Responses to reviewer 3

We are grateful to reviewer 3 for all the work done. The comments are identifying weak points of the manuscript that will help us to improve it. Please find below our responses (in blue italics) to the comments (in black). References to line numbers, figures and tables refer to the revised manuscript, unless stated otherwise.

This is an interesting and ambitious manuscript with important goals. I commend the authors on their substantial efforts in synthesising all the varied data streams (real and modelled) to put together the trend analyses, projections, and regional maps. While these synthesis figures don't deliver revolutionary new insight in a purely academic sense, they are extremely important and highly sought after in more policy-oriented applications. These results are certainly worthy of eventual publication.

I write this having also read the two existing peer reviews of this manuscript. I agree with the concerns of the other reviewers that many results are presented with either no or insufficient quantification, and/or too vague or incomplete conceptual explanation. This is my main concern about the manuscript as it is.

This review is so brief because there are not many points left to make without simply repeating the thorough work of the other reviewers. Other than the main concern noted above, I have only a few minor additions:

Abstract: is "window to the deep ocean" the right metaphor here? A place through which the deep ocean can be observed – is that the intended meaning?

With this phrase we wanted to refer to the strong connection between surface and deep waters through deep water formation. The introductory sentence of the abstract has been revised.

Abstract: sensitivity to OA in the Nordic Seas is not directly due to high latitude, but rather due to low water temperature?

You are correct, the two introductory sentences of the abstract have been replaced with the following:"With prevailing low temperatures, deep winter mixing, and cold-water coral reefs, *the Nordic Seas is vulnerable to ocean acidification." An explanation on the impact of low temperatures is given later in section 1.1.*

Sections 1 and 2, and probably also 3, are very much introductory material and I would also suggest to consider combining them, as mentioned by another reviewer.

It is a good idea to do some merging of these sections. After some thinking, we ended up dividing the introduction into two sections; one general introduction on ocean acidification and the Nordic Seas, and one on "theoretical background".

In Section 2 and Table 1, an important aspect of discussion is absent, that is about the timescale of the effects shown in Table 1. Are you showing instantaneous effects of T/S/DIC/TA increases, or effects after CO_2 re-equilibration with a constant atmosphere? Looks like it's the former – is that really appropriate, given the context?

We agree that this is an important aspect that should be discussed. Table 1 shows the instantaneous effects without any gas exchange, we now make this clear in the label and when we refer to it. We think that it is important to show this rather than the effect after CO_2-re-equilibrium because the Nordic Seas surface waters are not equilibrated, and therefore somewhere in between the instantaneous and secondary effect. We now discuss this also in Section 3.1.

Methods: given the relatively low temperature of your observations, why not use the Sulpis et al. (2020) carbonic acid constant parameterisation for your CO 2 system calculations?

The Sulpis et al (2020) paper was published in late July 2020, when all our analyses had been done, which explains why those parameterizations were not used. This could indeed be important for our estimates of the aragonite saturation horizon. But, looking at their figure 8, we see that the differences in the aragonite saturation state they get with the Sulphis parameterization and the Lueker parameterization, is about the same size as our uncertainty estimates (Table 2). These uncertainty estimates become even larger for past and future when taking into account the mapping error. This indicates that the results will not change substantially if switching between the two parameterizations. The Sulphis parameterization is however something that we will consider in our future work.

Throughout: pH is dimensionless; pH values do not need the word "units" after them, and $\times 10^{-3}$ can be used in place of "mpH".

We have removed the "units" and replaced mpH with $\times 10^{-3}$ *.*

Line 205: "DIC also relates to salinity" is very vague, please explain the mechanism– including timescale considerations noted above for Section 2 / Table 1. See e.g. Wu et al. (2019).

This section has been strongly revised in the new version of the manuscript, and we now also discuss the importance of timescales. The phrase "DIC also relates to salinity" is no longer in use.

Line 424 “both” implies two options when there are three (past, present and future). I am not sure that the chain of causality is properly represented in this and the subsequent sentences (i.e. which are drivers and which are responses in terms of air-sea CO 2 disequilibrium, pH change and hydrographic conditions), please be careful with the exact phrasing here.

In the revised manuscript we have put less focus on changes in pCO_2, and we have removed this paragraph. It is, indeed, not a driver of pH change, but just another indicator of a changing carbonate-system.

The request for more research at the very end of the manuscript is very unspecific and is unexpected given that the rest of the paragraph implies that all the observed phenomena have indeed been explained here.

We have removed this sentence.