

Reply to Referee #2

We thank referee #2 for the helpful comments. We have addressed the referee's concerns as explained below.

General comments:

- Consider adding “Subpolar” to North Atlantic in the title.

Added.

- I believe your results are occasionally strongly affected by the TTO data (particularly in the Irminger basin), conceivably worsened by your time-interpolation performed to ‘provide weight to old cruises’. I recommend publication of your results with exclusion of TTO data, or least mention in the text of how such exclusion would affects results.

Thank you for your suggestion. We recalculated all the trends excluding TTO data and the difference between pH trends with and without TTO were significant. Hence, we decided to exclude the TTO data from our study, with all the changes that it entails.

- The sections plots of Fig1 and Fig2 are unrealistically noisy. The captions suggests the “mean distributions” are plotted, but these are not averages, but rather all data of all cruises thrown into a single section, with inappropriately short influence radii for the contouring (or whatever the equivalent terminology is for DIVA gridding). They thus represent not natural spatial heterogeneity, but temporal aliasing. This leads to disturbingly jittery artifacts (particularly evident in Fig2d as blue/purple/pink patchwork). Consider either contouring true averages, or simply increasing the influence radii (i.e., smooth it more).

Thank you for your suggestions. You are right, what we did is plotting the data of all cruises in a single section, rather than averaging the data. Since the purpose of describing Figure 2 is giving a general view of the properties along this section, we think that describing a single cruise is enough. For this reason, we decided to describe the general distribution of the main variables along the section using the 2004 cruise as reference. This cruise represents the mean year of the studied period. In this way we avoid interpolating the data from all cruises to a single grid and then averaging them to build Figure 2, with all the errors that this would entail.

- Consider adding a visually catchy and informative summarizing section plot (one for pHobs, or perhaps one per pH-driver), showing per water mass the rate of pH change. In each, surface layers would most red, as would DSOW, with intermediate layers slightly lower, and Iceland on average lower than Irminger.

Thank you for your suggestion. We added the suggested pHobs figure to the supplementary information (Figure S1).

Specific comments:

- Consider capturing some more cruise details in your Table 1. For instance, please tabulate the type of measurements performed on each cruise (which had pH directly, which calculated it – i.e., you lines ~75–100). What is the consequence of the rather seriously sounding, but nonchalantly made remark in line 91 “However, Carter et al reported a pH inaccuracy of 0.0055”? Is that a positive or negative bias? Systematic for everyone or just for them? Do you compensate?

We added an extra column that specifies the measurements performed on each cruise. What we wanted to highlight with the sentence in line 91 is the fact that although it is possible to achieve high reproducibility in pH measurements, all the measurements will have an inherent uncertainty of 0.0055 due to the uncertainty in the determination of the constants of the tris-buffer. Therefore, the 0.0055 quantity is an uncertainty that affects all pH measurements, which we cannot compensate. We have clarified this point by changing the highlighted sentence to: "However, Carter et al. (2013) reported an inherent uncertainty of spectrophotometric pH determinations of 0.0055 pH units, associated to the tris-buffer used for calibration". (See also answer to comment about line 131).

– line 113: less => minus

Corrected

– line 120: “advantages” relative to what? ΔC^* ?

Yes. We added the following statement at the end of the sentence: “relative to the previous method proposed by Gruber et al. (1996)”.

– line 121: I can't follow. The suggestion is that no Cant-free reference waters are required, but it's not clear why that is. Consider explaining more clearly or not at all and only keeping the reference to VR2012).

We changed the statement to make it clearer: “And second, the parameterizations of A_T^0 and ΔC_{diseq} are determined using the subsurface layer as reference (Vázquez-Rodríguez et al., 2012a), where the age of the water parcel and, therefore, its C_{ant} concentration is estimated using CFC measurements (Waugh et al., 2006)”.

– line 131: explain why you consider 0.0055 the “accuracy” of the pH measurements. Again, if Carter thinks this is a /systematic/ error of the method, that would not affect detectability of trends.

You are right; this uncertainty will affect all pH measurements and, therefore, will not affect the detectability of trends. What we wanted to highlight with this statement is that the data we are using have high reproducibility (higher than the accuracy of the measurements) and thus are suitable for determining trends. We added the following statement to the manuscript: “The high reproducibility, an order of magnitude better than the uncertainty (0.0055 pH units, Carter et al. (2013)), is suggestive of high quality data”. (See also answer to comment about line 91).

– I find the use of statistical terminology confusing. The terms “standard deviation”, “confidence interval” and seem to be used loosely or even interchangeably while they each have a clearly defined use case. (Line 99-100 seems to suggest that you equate “two standard deviations” to “confidence interval”). If the use of these terms is nonetheless correct then certainly the employed confidence level should be mentioned to make sense of the stated confidence intervals. I particularly object to referring to the standard deviation of depths in a defined depth (or density) range as the ‘confidence interval’ of depths (first column of T1).

Thank you for your comment; we noticed that we were using the term ‘confidence interval’ incorrectly. In lines 99-100 the term ‘confidence interval’ was not correctly used and, therefore, it was deleted. In the revised manuscript we

changed the term ‘confidence interval’ with ‘standard deviation’. We, therefore, change the numbers presented in the tables accordingly. We also eliminated the standard deviation of the pressure data in Table 2 as you suggested. We only used the confidence intervals in figures 3-6, where we defined what we consider as confidence interval $(2 \times \text{standard deviation})/\sqrt{N}$, where N is the number of samples), which corresponds to a 95% confidence interval since our samples are independent.

– I hold the whole of line 125-140 to constitute a slight misuse of statistical numbers. The reasoning here seems to be “the spread between the means of cruises is smaller than the spread within each cruise, and thus we believe we can detect trends between cruises”. Although the closeness in cruise means is certainly comforting, that alone does not make for detectability of trends. It would at best provide a lower bound for the detectability of trends (i.e., trends within the ranges given in T1 would go unnoticed but might nonetheless exist). Consider adding a small statement that indicates these results are suggestive of high quality, and try to avoid suggesting to provide evidence thereof.

Thank you for your suggestions. We agree with your comment. We changed lines 139-140 as follows: “The high reproducibility, an order of magnitude better than the uncertainty (0.0055 pH units, Carter et al. (2013)), is suggestive of high quality data. Using these standard deviations for the seven cruises, and taking into account the 25 years considered in this study, the threshold of detectability of pH trends at 95% of confidence is 0.00012 pH units·yr⁻¹, which renders confidence to the estimated trends”.

– line 150: this ‘replacement’ process is a little rash. I can imagine ignoring these shallow data, but simply overwriting them with data that has less sensitivity to seasonality without providing a compelling case for doing so is not warranted. I do not believe that ignoring the 100m surface layer would yield a vastly different result to what you now got. If that is indeed so, I recommend using that ignore-approach, to avoid the suggesting that you’re fudging.

Thank you for your suggestion. We followed it by removing the photic layer (pressure < 75 dbar) from our study, so as to avoid the seasonality effect. This depth was determined by the depth of the seasonal nutrients drawdown. Thus, we replaced sentence in line 150 by the following sentence: “To reduce the influence of seasonal differences in sampling on the inter-annual trends, only samples with pressure ≥ 75 dbar were considered. The 75 dbar level was determined by the depth of the seasonal nutrients drawdown along the section”.

– line 154: please be specific in how you “take into account thickness and separation”. I presume that the average of tall profiles with large spacing to east and west get higher weight in average-of-averages? Do you have a specific reason for not using the alternative approach of averaging per layer the ‘grid boxes’ of Fig2? Presumably you did not grid the data for analysis but only for figure making.

You are right when highlighting the different weighting given to the profiles in function of their spacing, but, intuitively, horizontally gridding with a linear interpolation would give the same result as the method used in our study.

We did not grid the data horizontally but vertically, and this vertical gridding was taken into account for both figure building and calculations. Averages were then performed in an area-weighted basis, so that accounting for the irregular spacing between stations. This was performed by using a trapezoidal integration, which

produces the same results as the horizontally gridding, which consumes more computational time.

– I'm not clear on what you've done here. I agree that pressure-adjustment here is necessary, although heaving of even 500 m would not even produce a pH shift larger than 0.02 units. However, you'd do this also simply to reduce the range of pH values within each (non-horizontal) water mass. I would, however, believe the proper procedure to be (i) "recalc pH for each sample in water mass to pH at the single mean depth of that watermass (for the cruise or the whole dataset, that shouldn't matter much) and then (ii) calculate average of these recalculated pH values. This might be what you did, but the way I read it, you first calculated the average pH, and then shifted that average pH to the 'correct' depth. If so, consider redoing more appropriately, or explain for daft readers like myself why the used approach **is** appropriate.

Sorry for the misunderstanding of our methodology. What we did was calculate an average pH for each layer and year from the average values of DIC and A_T, and referred to the mean pressure of the layer over the studied time period. We rewrote this explanation in the manuscript: "The exception comes with pH_{Tis}, which is pressure sensitive, and for which we needed to define a unique reference pressure to remove pressure effects due to varying sampling strategies. pH_{Tis} was calculated using the layer average values of DIC and A_T for the considered year but using the time-averaged pressure of the layer over the studied time period as reference pressure". We have also corroborated that the methodology you propose and the methodology that we applied give the same results. This is because the variability of the physicochemical properties within each layer is relatively low, and therefore the system is linear. Following this line of reasoning we added the following sentence in the first paragraph of section 2.2: "The advantage of working in layers is the relatively low variability of the physical and chemical properties within the layers, allowing us to assume linearity in the ocean CO₂ system". We, then, decided to keep our methodology since it is simpler and working with averaged DIC and A_T (conservative) is desirable instead of working with recalculated pH (not conservative) values for each sample and then averaging them. Besides, the recalculated pH for each sample that you propose will not reproduce the pH shown in Figure 2c.

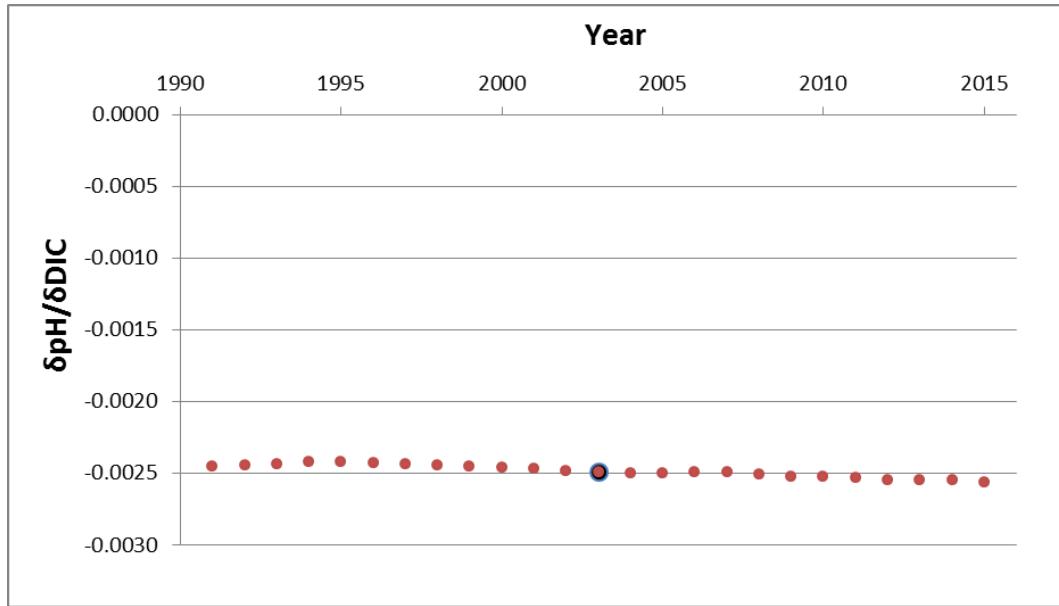
– line 157: remove last four words "over the pH trends"

Done.

– I can't fully comprehend what the approach is that was followed in section 2.3. The idea is clear "keep all but one parameter constant and see how pH changes. The sensitivity of pH to an increase in DIC would be sharper in 2015 than in 1981. Is that accounted for in the method? Specify the calculation routine you used.

We are aware that the 'buffer factors' change with time and that the system is not linear. However, since we split the studied region in different layers, the range of variation of the parameters within each layer during the studied time period is small (the ± of Table S1 gives insights about this small variability), and then, we can assume linearity. Also because of the small range of variation of the 'buffer factors' is also small, and we can neglect their change when calculating the change in pH due to each of the proposed controlling mechanisms. I am enclosing an example about the change of 'δpH/δDIC' in the SPMW layer of the Irminger basin. The red

points show the ‘ $\delta pH/\delta DIC$ ’ values for each year, and the black point is the ‘ $\delta pH/\delta DIC$ ’ used in the study. As you can see, the change of ‘ $\delta pH/\delta DIC$ ’ over time is negligible, and the value used in the study is the mean value for the period 1991-2015.



We added the following statement to clarify this question in the manuscript: “Given that the variability of the physicochemical properties within each layer is relatively low (see standard deviations of the averaged values in Table S1), we can assume that these derivatives are constant over the studied time period and use a constant derivative value for each layer”.

– Consider restructuring 2.3 into a distinct paragraph for the determination of time trends and one for inferring strength of individual drivers. Your TABLE3 mentions the “sum of drivers” or “model”, which terminology is nowhere used in the text, please harmonize. Also, T3 separates influence of C_{nat} and C_{ant} , but Eq2 does not.

We included a table in the Supplementary Information that shows the strength of individual drivers (Table S2). Since time trends are represented in figures 4-6, rebuilding section 2.3 would involve citing figures 4-6 before figure 2, which we think this is not a logical order for figures in the manuscript. However, we added the following statement to the results section to clarify where the reader can find the derivative values: “The values of each term of $\frac{\partial pH_{Tis}}{\partial var}$ and $\frac{dvar}{dt}$ (where var refers to each of the drivers) described in Sect. 2.2 can be found in the Supplementary Table S2 and in Figs. 4-6, respectively”.

We changed dpH/dt_model by dpH/dt_total , and introduced this new term to Eq. (2) and Table 3. We also added some explanatory comments throughout the manuscript to distinguish more clearly between dpH/dt_model and dpH/dt_total .

We slightly changed Eq. (2) to clearly indicate the separation between C_{ant} and C_{nat} :
$$\left(\frac{dpH_{Tis}}{dt}\right)_{total} = \frac{\partial pH_{Tis}}{\partial T_{is}} \frac{dT_{is}}{dt} + \frac{\partial pH_{Tis}}{\partial S} \frac{dS}{dt} + \frac{\partial pH_{Tis}}{\partial A_T} \frac{dA_T}{dt} + \frac{\partial pH_{Tis}}{\partial DIC} \left(\frac{dC_{ant}}{dt} + \frac{dC_{nat}}{dt}\right)$$

- line 168: “real average value” => “observed linear trend” (???)

What we meant is the average value calculated for each layer and cruise. We changed the sentence to the following to make this clearer: “To estimate $\frac{\partial \text{pH}_{Tis}}{\partial \text{var}}$ (where var refers to each of the drivers: T_{is} , S , A_T and DIC) we calculated a pH_{Tis} for each layer and year using the layer average value of var for each year but keeping the values of the other drivers constant and equal to the time-average value for the layer over the studied time period”.

- TABLE1: I believe “confidence interval” here is “standard deviation”, or is it truly CI? Then state the confidence level. I’m not sure the CI of the average of averages, or however one would call the last row, has any statistical meaning – why not simply provide SD in that row? You use pH25 in table 1, while stating in the text that pHisT and pH25 are not easily compared – why the sudden use of pH25 here?

We believe you are referring to Table 2. We changed ‘confidence interval’ by ‘standard deviation’ and we changed the results accordingly. The purpose of this table is to give an estimate of the reproducibility of the analysis and calculation methods, which we take as an indicative of the goodness of the data for trend analysis. Regarding the scale of pH, since trends are performed using pH at in situ conditions, we followed your suggestion and we changed the pH reported in Table 2 by pH at in situ conditions.

- FIGURE1: some of the contour intervals have at sig1 or sig2 label, while caption and text suggest cutoffs were based on sig0.

Sorry for the confusion, $\text{sig}\theta$ would refer to all the sigma levels, either referenced to 0, 1000 or 2000 dbar. To make this clearer we changed the figure caption by (“referenced to 0 dbar, σ_0 ; 1000 dbar, σ_1 ; and 2000 dbar, σ_2 ; all in $\text{kg}\cdot\text{m}^{-3}$ ”)

- I generally very much like your other (time-tested) figures. Perhaps increase coverage of fig 1a to provide a view of distance to land on eastern extent of section.

We updated Figure 1a following your suggestions. We increased the eastward extension of the map.

- Consider moving 3.1 to the introduction.

We appreciate your suggestion. However, and despite the qualitative nature of this section, we think it fits better in the results section rather than in the introduction, since it is a description based on our data. Besides, it provides useful information for understanding the sections coming next. In view of this, we decided to keep this section as it is.

- line 209: “almost homogenous” – sections plots suggest otherwise, see earlier comment on influence radii

After changing Figure 2 as suggested in the previous comment regarding this figure, now the statement is true.

- line 211: “because they are correlated” – that relationship is not causal, please rephrase.

You are right. We eliminated this statement, so that, in our opinion, the explanation of property distributions is improved.

- line 212: can you qualify that “80%” in light of the mentioned ΔC_{diseq} ? Is this what one would expect?

The existence of a saturation lower than 100% in the surface layer is expectable for an oceanic uptake of C_{ant} (see Matsumoto and Gruber, 2006). Therefore, a disequilibrium in the C_{ant} content with respect to the atmospheric C_{ant} is expected.

Matsumoto, K. and Gruber, N.: How accurate is the estimation of anthropogenic carbon in the ocean? An evaluation of the ΔC^ method, Global Biogeochem. Cycles, 19, doi:10.1029/2004GB002397, 2005.*

- line 239: can you speculate on the possible causes for the supposedly spuriously high rate of pH decrease observed by Bates et al at IrmSTS?

We added the explanation given by Bates et al. (2014): “Bates et al. (2014) linked the high acidification rate found at the Irminger Sea time-series to the high rate of increase in DIC ($1.62 \pm 0.35 \text{ } \mu\text{mol}\cdot\text{kg}^{-1}\cdot\text{yr}^{-1}$) observed at this site, which is almost twice our rate of increase in DIC ($0.64 \pm 0.07 \text{ } \mu\text{mol}\cdot\text{kg}^{-1}\cdot\text{yr}^{-1}$, Fig. 5c). This is based on data from only one site, further north than our section, and indicates that spatial variations are substantial in this region”. We do not have enough information to elucidate other possible explanations.

- line 244: you mention a tropical Pacific time series station, and contrast it with your work and a subpolar Pacific TSS, latter two match nicely. Add a brief sentence attributing that contrast.

Now that we are not including the TTO cruise data in our study, our trends are in agreement with the trends found in the tropical Pacific and are slightly higher than the trends found in the subpolar Pacific. We changed the manuscript accordingly to these changes, and we added the following explanation to the low trends found in the subpolar Pacific: “Wakita et al. (2013) attributed the lower than expected pH trends to an increasing A_T trend”.

- line 246-254: your statement “renders direct comparison difficult” does not stand up to scrutiny. Recalculating pH to different temperatures does not changes the slope of a pH trend. That is, slopes can be compared (i.e., VR12’s Fig3ab vs your Fig3ab), even if absolute values cannot. I recommend more work is made of this comparison, particularly if results between studies differ.

The fact that precludes comparing trends is not the difference in the temperature at which pH is referenced, but the fact that Vazquez-Rodriguez et al. (2012b) normalized pH values to potential temperature, salinity, silicate and AOU in WOA05. This normalization eliminates part of the influence of these parameters (potential temperature, salinity, silicate and AOU) on the pH trends. This is why direct comparison between their pH trends and our pH trends is difficult. However, we added some text comparing the trends reported by both works: “This normalization, combined with the different temporal coverage (1981–2008), causes the rates reported by Vazquez-Rodriguez et al. (2012b) differ from those obtained in the present work. The pH_N trends reported for the SPMW and uLSW layers of the Irminger basin and for the ISOW layer of the Iceland basin are very similar to our pH_{Tis} trends for these layers. However, the pH_N trends reported by Vazquez-Rodriguez et al. (2012b) for the cLSW layer in both basins and for the ISOW layer in the Irminger basin are significantly different from our pH_{Tis} trends for these layers, but are very similar to pH changes derived from C_{ant} changes

$(\frac{\partial pH_{Tis}}{\partial DIC} \frac{dC_{nat}}{dt}$ in Table 3). In the case of the DSOW layer, the pH_N trend is also in agreement with $\frac{\partial pH_{Tis}}{\partial DIC} \frac{dC_{nat}}{dt}$ trends. This suggests that the normalization carried out by Vazquez-Rodriguez et al. (2012b) could remove some of the impact of the natural component (represented here by C_{nat}) over pH changes, essentially due to the use of AOU in the normalization”.

- line 322: if anything, these are 4 decades (80s 90s 00s and 10s). Consider “34-year period” or similar

We changed it by “25-year period”, since now we are not considering the TTO cruise.

- line 323 (and likely elsewhere): “separate and increase into its drivers” is slightly sloppy English. Consider rephrasing

We changed the sentence into: “From the study of the main drivers of the observed pH changes...”.

- line 325: “However” => “thus”. Reduced rate of decrease of pH is what one expects with increasing alk.

Changed.

- line 327: “salty” => “saline”

Changed.

- line 333: consider “observe” => “infer”. There’s too many interpretative steps involved to call this “observe”

Changed.