Biogeosciences Discuss.,
doi:10.5194/bg-2016-66-RC2, 2016

Biogeosciences Discussions

Interactive comment on "Ocean acidification in the North Atlantic: controlling mechanisms" *by* Maribel I. García-Ibáñez et al.

Anonymous Referee #2

Received and published: 29 April 2016

Review of "Ocean acidification in the North Atlantic: controlling mechanisms" by Maribel I. García-Ibáñez, Patricia Zunino, Friederike Fröb, Lidia I. Carracedo, Aida F. Ríos, Herlé Mercier, Are Olsen, Fiz F. Pérez, as submitted to Biogeosciences Discussions in Feb. 2016.

This manuscript analyses a rich set of data of CO2-system measurements made in the subpolar North Atlantic Ocean over a period of 34 years, aiming to update the existing record of strong acidification of these waters and, newly, to apportion the total acidification to a suite of driving processes. The paper is excellently readable, modest and concise. I have no conceptual reservations with its general methodology. Minor criticism I voice is the rather strong likeliness to earlier work performed by some of the authors, which however is overridden by the new focus on attribution. Overall, I feel the manuscript warrants publication, with only minor revisions. Desired changes

mostly relate to lack of explicitness in the methods sections, and the confusing (or even confused?) use of statistical concepts, which is where this review focuses.

General comments:

– Consider adding "Subpolar" to North Atlantic in the title.

– 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.

– 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).

– 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.

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 alk reported a pH inaccuracy of 0.0055"? Is that a positive or negative bias? Systematic

for everyone or just for them? Do you compensate?

– line 113: less => minus

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

– 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).

– 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.

– 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 is 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).

– 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.

– 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.

– 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.

– 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.

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

– 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.

– 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 Cnat and Cant, but Eq2 does not.

– line 168: "real average value" => "observed linear trend" (???)

– 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?

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

– 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.

– Consider moving 3.1 to the introduction.

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

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

– line 212: can you qualify that "80%" in light of the mentioned ΔCdiseq? Is this what one would expect?

– 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?

– 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.

– 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.

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

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

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

– line 327: "salty" => "saline"

– line 333: consider "observe" => "infer". There's too many interpretative steps involved to call this "observe"

Interactive comment on Biogeosciences Discuss., doi:10.5194/bg-2016-66, 2016.