

Final Author Comments for ‘Intercomparison of carbonate chemistry measurements on a cruise in northwestern European shelf seas’ by M. Ribas-Ribas et al.

We thank both referees for their positive, thoughtful and constructive comments. Below we reply to the comments point by point. Pages and lines from the referee’s comments correspond to the BGD paper. Pages and lines from our replies (in brackets) correspond to the pdf of the revised manuscript.

Response to Anonymous Referee 1

1.1. P 2796 L 13: “data collection effort was not planned in advance as an intercomparison exercise”. This statement contradicts to some extent statement of previous sentence. Also, I have a hard time believing that no one envisaged this prior to the cruise, and it was only after the cruise that suddenly someone realized that an internal consistency and intercomparison exercise was possible.

We have rephrased the sentence to avoid misunderstanding. Now it reads: “secondly, the study was not designed at the outset as an intercomparison exercise, which normally involves placing all the instruments in one laboratory,” (P. 3, 25-27)

1.2. P 2797 L 20-21: provide values of the NOAA standards.

We provide the values for:

System 1: “National Oceanic and Atmospheric Administration (NOAA), which had values of 244.9 and 444.4 $\mu\text{mol CO}_2 \text{ mol}^{-1}$.” (P. 5, 16-17)

System 2: “NOAA, which had values of 251.6, 347.2 and 448.8 $\mu\text{mol CO}_2 \text{ mol}^{-1}$.” (P. 5, 32)

1.3. P 2799 L 9: provide manufacturer and model of the PT100 sensors

PT100 sensor for:

System 1: “Pico Technology, model PT100”. (P. 5, 4-5)

System 2: “a PT probe (Omega) with modified electronics“. (P. 6, 6)

1.4. P 2799 L 17: from a purely formal point of view, it's the precision of the Licor's $p\text{CO}_2$ measurements that's determined from the gas standards, not the actual seawater $p\text{CO}_2$ that will also depend on equilibrator performance, accuracy of temperature measurements, etc...

We agree and have changed the text accordingly: "The precision of both LI-COR's $p\text{CO}_2$ measurements was 1.0 μatm , established using standard gases". (P. 6, 16-17)

1.5. P2801 L 15: datasets were submitted to BODC but are not publicly available (or at least the link to the data are easily found). Please clarify.

BODC has now produced a doi-number for this dataset and the text has been amended accordingly: "The combined carbonate chemistry dataset is available via the British Oceanographic Data Centre at https://www.bodc.ac.uk/data/published_data_library/catalogue/10.5285/f56e35bc-635e-0ab5-e044-000b5de50f38/ with doi: 10.5285/f56e35bc-635e-0ab5-e044-000b5de50f38." (P. 8, 13-16)

1.6. P2802: Replace "refitted by (Mehrbach et al., 1973); Dickson and Millero (1987)" by "refitted by Dickson and Millero (1987)".

This has been done (P. 8, 29)

1.7. P2802: There are several sets of constants. Please justify the choice of Mehrbach et al. (1973) as refitted by Dickson and Millero (1987).

"We used the Mehrbach constants because they led to the smallest inconsistencies between different high-accuracy measurements in previous observational studies (Clayton et al., 1995; Lee et al., 1997; McElligott et al., 1998; Wanninkhof et al., 1999). However, we also compared against results calculated using constants from Roy et al. (1993), Lueker et al. (2000) and Millero et al. (2006) (Sect. 3.3)." (P. 8, 29- P. 9, 3)

1.8. P2802: "borate constants" as stated stand in fact for the ratio of total boron to salinity.

We agree and have changed that here and everywhere else in the manuscript. (P.9, 5 and Sect. 3.3)

1.9. P2802: Specify which dissociation constants of boric acid were used.

We assume the referee is only referring to one constant (not plural): “For the dissociation constant of boric acid we used Dickson (1990b),” (P. 9, 4)

1.10. P2802: specify which KHSO_4 was used.

This info was already in the previous ms: “for bisulphate ions Dickson (1990a)” (P. 9, 4-5)

1.11. P2802: The whole internal consistency exercise could have been also done using different sets of constants to determine which one would give the best fit.

Done. A new table and discussion have been added to Sect. 3.3, confirming that the best fit is Mehrbach.

1.12. P2804 L1-6: Please statistically test if the Y-intercept is significantly different from zero, and if the slope is significantly different from 1. Here and elsewhere.

“The 95 % confidence interval of the slope does include the value 1 and the 95 % confidence interval of the intercept does include the value 0.” (P. 11, 9-11)

1.13. P2804 L4: The Y-intercept of the regression is given at 0.01 μatm level which does not make sense given the uncertainty of the measurements.

We agree and have changed the equation accordingly: “The resulting equation of the regression is $p\text{CO}_2-1 = 0.9 (\pm 2.1) + 0.99 (\pm 0.01) \times p\text{CO}_2-2$ ”. (P. 11, 7-8)

1.14. P2804: the authors used “different sets of calibration gases” as stated but they were re-calibrated in the lab, before and after the cruise against NOAA standards (P 2797). This should have significantly reduced the differences in calibration gases, and should correspond to situation virtually identical to using the same calibration gases, if the Licors were working properly.

They were recalibrated in different labs with different sets of NOAA standards.

1.15. P 2804 L13: provide reference(s) to back statement “strong gradients in temperature, salinity and $p\text{CO}_2$ ”

The sentence now reads: “strong gradients in temperature, salinity and chemical variables (Rérolle et al., 2014)”. (P. 11, 18-19)

1.16. P2804 L 17: the residuals are given at 0.1 μatm level which does not make sense given the uncertainty of the measurements.

We have changed the ms accordingly (Tables and elsewhere in the ms).

1.17. P2804 L 19: provide reference(s) to back statement “well within the expected accuracy of $p\text{CO}_2$ calculated from C_T and A_T ”

We have added “(Millero, 2007).” (P. 12, 10)

1.18. P2805 L 1-13: Please also include an additional panel to Figs. 2 and 3, with the plots of in-situ temperature, equilibrator temperature of equilibrator 1, and equilibrator temperature of equilibrator 2.

Done (P. 35- 36)

1.19. P2806 L 6-7: this statement was already made by Körtzinger et al. (2000).

We added: “... in agreement with Körtzinger et al. (2000)” (P. 13, 22)

1.20. P2807: while I agree with the underlying idea, this needs to be worded differently because all the carbonate variables “are not fully independent”. A_T and C_T are not fully independent either, both depend strongly on $[\text{CO}_3^{2-}]$ and $[\text{HCO}_3^-]$. $p\text{CO}_2$ and pH do not make a good pair because $[\text{CO}_2]$ and $[\text{H}^+]$ are much smaller than the $[\text{CO}_3^{2-}]$ and $[\text{HCO}_3^-]$. Hence, relatively small errors on CO_2 and H^+ will propagate into relatively large errors on $[\text{CO}_3^{2-}]$ and $[\text{HCO}_3^-]$ if the system is computed from CO_2 and H^+ (e.g. $p\text{CO}_2$ and pH).

We have changed the wording as follows: “ $p\text{CO}_2$ and pH_T do not make a good pair for predicting other variables because CO_2 and hydrogen ion concentration are smaller than carbonate and bicarbonate concentration. Therefore, relatively small errors in CO_2 and/or hydrogen ion propagate into relatively large errors in carbonate and bicarbonate concentration when the system is computed from $p\text{CO}_2$ and pH_T .” (P. 14, 27-31)

1.21. P2808 L 25: Continental shelf break regions are also characterized by strong gradients due to enhanced vertical mixing (internal tides or upwelling).

This is a good point. We added: “for example in shelf sea regions with freshwater inputs, in continental shelf break regions with enhanced vertical mixing (internal tides or upwelling) and in regions with sea ice melt.” (P. 16, 27-29)

1.22. P2809: "complete shielding from light" will not prevent bacterial growth

We no longer mention microbial growth and now the sentence reads as follow: “c) prevention of phototrophic growth in the equilibrator by complete shielding from light.” (P. 17, 3)

1.23. P2809: Regarding measurements of pH indicator-dye techniques can the authors comment on the interference that could be expected from suspended matter and CDOM that are characteristically high in coastal waters?

We have added to section 2.1.2: “In order to minimise absorbance interference by particulates an in-line filter (0.45 μm pore size, Millex HP syringe filter MilliporeExpress® (PES) membrane 33 mm diameter, Millipore) was placed at the entry of the sample tube. Chromophoric Dissolved Organic Matter (CDOM) only absorbs weakly in the visible where Thymol Blue absorbance is measured (<3 % at 435 nm and <1% at 596 nm) and is accounted for in the seawater blank. CDOM absorbance interference is thereby cancelled out. Additionally, measurements at the wavelength 750 nm (not affected by Thymol Blue indicator) were used to monitor for sample turbidity and instrument drift” (P. 7, 16-24)

1.24. Some of the coastal areas sampled are characterized by blooms of coccolithophores. Can the authors comment on how the CaCO_3 could have affected the C_T and A_T measurements?

We have added discussion of this: “During three days in the same coccolithophore bloom we collected additional samples of filtered C_T and A_T from the underway seawater supply. The filtering was carried out using an in-line filter (Sartorius Sartobran 300 Sterile capsule, 0.45 μm pore size). In this way we studied the effects of the presence of calcite mineral particles in the surface waters on C_T and A_T measurements. The average differences between unfiltered and filtered samples were 2.4 $\mu\text{mol kg}^{-1}$ and 3.7 $\mu\text{mol kg}^{-1}$ for C_T and A_T , respectively, with values in the unfiltered samples being higher. The differences between replicates of filtered samples were on average higher than those of unfiltered samples. The differences between measured pH_T and pH_T calculated from filtered and unfiltered C_T and A_T were -0.003 and 0.005, respectively. The difference between measured $\text{pCO}_2\text{-1}$ and $\text{pCO}_2\text{-2}$ and pCO_2 calculated from filtered and unfiltered C_T and A_T were -4 μatm in both cases. An

influence of dissolution of calcite particle on C_T and A_T measurements would be expected to lead to filtered C_T and A_T being lower than unfiltered with discrepancies twice as large for A_T as for C_T . This ratio was however not exactly observed in our study, and hence it is unclear from our observations whether dissolution of CaCO_3 particles affected C_T and A_T measurements. Filtration of samples for C_T potentially introduces a further error through CO_2 loss by turbulence and ebullition, which can affect the C_T measurement, although we took precautions to avoid bubbles in the filter” (P. 15, 15-32)

1.25. Most of the discussion on the two $p\text{CO}_2$ equilibrator systems relies on the Körtzinger et al. (2000) intercalibration. I’m surprised that the Körtzinger et al. (1996) is not discussed and cited since it provided intercalibration of equilibrators in the North Sea (possibly the very first intercomparison in coastal waters).

We added and discussed Körtzinger et al. (1996) in:

- 1) Sect. 1: “Körtzinger et al. (1996) carried out what may have been the first intercomparison study in coastal waters between two similarly designed underway $p\text{CO}_2$ systems. They found a remarkable agreement between the two simultaneously measured $p\text{CO}_2$ datasets even though the spatial variability in surface $p\text{CO}_2$ in the North Sea was high. The average difference was $0.2 \mu\text{atm}$ (standard deviation = $1.2 \mu\text{atm}$), indicating no systematic difference. The difference tended to be highest during the most pronounced $p\text{CO}_2$ gradients.” (P. 2, 17-23)**
- 2) Sect. 3.1: “This result is also comparable with a previous (the only other) coastal water intercomparison, described by Körtzinger et al. (1996). In this study, where there was highly variable spatial $p\text{CO}_2$ distribution in the southern North Sea, the average difference between observed values was $0.2 \mu\text{atm}$ (standard deviation = $1.2 \mu\text{atm}$)” (P. 11, 24-27)**

1.26. Also there have been numerous intercalibration exercises of $p\text{CO}_2$ systems in pools (NIES, etc...). It could be useful to mention these in the discussion.

We have added a brief mention of the Japanese intercomparison exercises: “There have been a number of intercalibration exercises of $p\text{CO}_2$ systems in an indoor seawater pool at the National Institute for Environment Studies, Japan (in 1993, 1998, 2003 and 2009)

(Katayama et al., 1999; IOCCP, 2004; Pierrot et al., 2009). Most of the instruments showed good agreement (within 2 μ atm).” (P. 11, 28-31)

1.27. Seminal paper of Wanninkhof et al. (1999) could be useful in the discussion since it provides an extensive “overdetermination study” in open ocean conditions that could be useful versus the coastal study reported here.

We thank the referee for pointing out this reference. We have now included it in the ms. (P. 9, 1) (P. 19, 6) (P. 19, 22) (P. 20, 10).

1.28. Finally, there are several papers that have raised the issue of the impact of organic acids on computations of the CO₂ system in coastal waters. I’m surprised these aspects are not mentioned/discussed (Koeve & Oschlies 2012; Kim et al. 2006; Muller & Bleie 2008; Kim & Lee 2009; Hernández-Ayón et al. 2007).

We have added to the discussion: “Several papers have raised the issue of the impact of organic acids on computations of the CO₂ system in coastal waters (Kim et al., 2006; Hernández-Ayón et al., 2007; Kim and Lee, 2009). Dissolved organic matter produced by phytoplankton during photosynthesis can potentially make a significant contribution to seawater total alkalinity although we saw no evidence for this in our study (analysis not shown).” (P. 15, 10-14)

1.29. Please clarify how the recommendations P 2808-2809 converge or differ from those already given by Körtzinger et al. (2000) or by Dickson et al. (2007)

Done. (P. 16-17)

Response to Anonymous Referee 2

General comments:

The authors present an extensive carbonate chemistry dataset acquired during a cruise in the northwestern European shelf seas, during which the carbonate system was overdetermined by measurements of $p\text{CO}_2$, pH, A_T and C_T . Overall, measurements compare surprisingly well with each other and the authors are to be congratulated for the high quality of the dataset. While I greatly appreciate the efforts to unravel systematic discrepancies between different methods to characterise marine carbonate chemistry and the explanations provided for the remaining discrepancies, the manuscript currently still has some major deficits that should be fixed before final publication in BG.

2.1. My biggest criticism concerns the insufficient discussion (and citation) of previous work. Generally, rather few citations can be found in the discussion even though a lot of work has been done in this field. With respect to the intercomparison of different methods, several studies have provided similar levels of agreement between estimated carbonate chemistry parameters (e.g. McElligott et al. 1998, Luecker et al. 2000), so the conclusion that the “results show that it is possible to obtain good consistency between measurements” (P2810, L22-24) is neither surprising nor new. This impression can be avoided by referring to previous findings. Similarly, it remains unclear how much of the ‘best practise suggestions’ are novel. Please clarify which suggestions have been made before (and by whom).

We have added references to these papers to the discussion and we have also discussed and cited work by other researchers (Körtzinger et al., 1996; McElligott et al., 1998; Katayama et al., 1999; Lueker et al., 2000; IOCCP, 2004; Pierrot et al., 2009).

Our conclusion about good consistency is specifically for the case where data are collected under more normal conditions, and not as part of a preplanned intercomparison exercise, as was made clear in the original manuscript.

We now specify which conclusions are completely novel and which were made previously.

2.2. The authors seem to be undecided if they consider the dataset to prove high consistency between the different datasets (e.g. P2794, L12 and P2808, L2-6) or to contain systematic discrepancies (e.g. P2794, L15 or P2808 L9-11).

We observed high consistency between four datasets and a discrepancy with one $p\text{CO}_2$. We have clarified now: “We therefore conclude that four of the independent datasets of carbonate chemistry variables were of high quality” (P. 1, 27-28)

“Four out of the five independent datasets of carbonate chemistry variables are deemed to be of high-quality” (P. 16, 4-5)

2.3. Some of my confusion between these two conclusions seems to arise from the differences between “raw” and corrected $p\text{CO}_2$ values. Please distinguish between these two levels more clearly.

We think there is some misunderstanding here. The whole of the intercomparison exercise used the same dataset, the corrected one. We added a sentence to the methods section to try to clarify this point: “The intercomparison exercise was carried out on the datasets after they had been adjusted according to the procedures just described, including corrections to in-situ seawater temperature described above.” (P. 6, 28-30).

2.4. Also, isn't it common practise to correct measured $p\text{CO}_2$ levels for the differences between SST and temperature in the equilibrator? If so, why is the non-temperature corrected data discussed?

Yes, it is common practice and we have done that for both systems. We feel that the reviewer may have misread our manuscript. We do not use non-temperature corrected data nor discuss its use.

2.5. In the abstract (P2794, L12-14) as well as the discussion, the authors state that the present dataset is “suitable to be used as a basis for evaluations of the impact of OA on ocean biogeochemistry” (P2808, L5-6). It remains unclear how the authors come to this conclusion. How does the dataset provide information on Ocean Acidification and its effects on marine biogeochemistry? On which timescales do you expect in-situ OA to be measurable, also in view of measurement uncertainties as well as temporal and spatial variability of carbonate chemistry? Is this study thought to serve as a baseline for future investigations? Please clarify.

Inorganic carbon system variables are the basis for any ocean acidification study (whether experimental or observational). In the context of the UK Ocean Acidification Sea Surface consortium, our data is used by other researchers in the consortium to evaluate ocean acidification impacts. For instance, in order to compare the CaCO_3

shells of organisms living in more acidic locations to those living in more alkaline locations, one must first know the pH along the transects ((Clark et al., 2014; Hopkins and Archer, 2014; MacGilchrist et al., 2014; Poulton et al., 2014; Richier et al., 2014; Young et al., 2014), this issue). We have deleted the phrase in the abstract and modified the phrase in the discussion for clarity: “...suitable to be used as a basis for evaluations of the impacts of ocean acidification by other scientists on the some cruise, see for instance the papers by Poulton et al. (2014) and Young et al. (2014)” (P. 16, 6-8)

Specific comments

2.6. P2794- P2796: The introduction reads as a list of various related studies rather than an introduction to or a rationale for the presented manuscript. Please restructure/rewrite and put the different studies (more) in context with each other.

We have rewritten the introduction in a more appropriate way. (Sect. 1)

2.7. P2794, L11-18: If diurnal cycles with differences as high as 41 μatm were observed, can you really state that the datasets “were all of really high quality”? What would be your definition of “really high quality”?

We now attach a quantitative criterion to the statement: “All comparison resulted in MR less than or equal to accuracy (with the exception of predictions based on $p\text{CO}_2$ and pH and A_T from C_T and $p\text{CO}_2$ -1). In terms of RMSE, all comparisons resulted in RMSE less than twice the uncertainty, except those involving measured $p\text{CO}_2$ -2. On this basis we conclude that there is generally good agreement between measured and calculated variables, except those comparison involving measured $p\text{CO}_2$ -2 or calculations from pH_T and $p\text{CO}_2$.” (P. 10, 18-23).

2.8. P2794, L12-14 & P2796, L10-11: How can this study be used to the evaluation of OA impacts?

See response to 2.5 above.

2.9. P2797, L17-19: Sentence sound a bit clumsy, maybe change to “The $p\text{CO}_2$ -1 and $p\text{CO}_2$ -2 systems undertook 6187 and 26671 measurements of surface water $p\text{CO}_2$ during the cruise, respectively.”

The sentence has been improved. (P. 4, 24-26)

2.10. P2798 L29-P2799, L3: This is a very long sentence, please consider rewording.

We have reworded the sentence: “The seawater flow to the container laboratory was highly variable throughout the cruise. This was due to the location of the container downstream of an intermittently large water demand for an experiment. The water flow was regulated to a maximum of 1.8 L min⁻¹, to avoid flooding of the equilibrator and CO₂ analyser during sudden spikes in supply.” (P. 5, 32- P. 6, 3)

2.11. P2801, L13-14: Please state the number of (CRM?) measurements used to gain these values.

We have added: “(159 CRMs analysed in duplicate).” (P. 8, 13)

2.12. P2801, L24: Please state that this is the Matlab version.

Done. (P. 8, 26) (P. 10, 3)

2.13. P2801, L24-29: Why were different borate constants, but not carbonate equilibria constants compared?

Referee 1 also raised this point (1.11). We have now also compared the results obtained using different carbonate equilibria constants (Table 3 and Sect. 3.3).

2.14. P2803, L6-9: Please reword this sentence in order to improve grammar.

We have reworded this sentence as follow: “In this study, uncertainties in calculated values were determined by a Monte Carlo approach as follows: 1) The original carbonate chemistry variable values in the dataset were input into the CO2SYS program (MATLAB version) (Van Heuven et al., 2011); 2) Artificial random errors (normally distributed according to the central limit theorem, with a mean of zero and standard deviation equal to the accuracy of measurement) were calculated using a random number generator; 3) New carbonate chemistry variable values (the originalones plus the randomly generated errors) were input into CO2SYS.” (P. 10, 1-7)

2.15. P2803, L24 – P2806, L7: It is not clearly stated which paragraph of this section refers to which level of correction of raw $p\text{CO}_2$ measurements. Maybe it would help to change the order of the paragraphs, starting with raw data followed by corrected data.

We think there is some misunderstanding here. All of the intercomparison exercise has been done with the same, corrected dataset.

2.16. P2804, L-12-16: In addition to the average difference, the RMSE should be discussed here, as the average difference alone does not provide enough information on the performance in relation to “strong gradients of temperature, salinity and $p\text{CO}_2$ ”

We have added the RMSE value: “we conclude that an average difference of 2 μatm and a RMSE of 10 μatm were good outcomes” (P. 11, 21-22)

2.17. P2804, L-15-16: Why do you judge this result to be only “reasonable” (I would say it is pretty good!)? If the average difference in your study is comparable to the one described by Körtzinger et al. (2000), wouldn't that mean that the “less ideal” settings of your intercomparison do not seem to negatively affect overall consistency?

We have adopted the reviewer's suggestion: “we conclude that an average difference of 2 μatm and a RMSE of 10 μatm were good outcomes. The non-ideal conditions of our intercomparison did not negatively affect the overall consistency (we obtained comparable results to the study by Körtzinger et al. (2000)).” (P. 11, 21-24)

2.18. P2804, L22-25: Please state average and maximum differences.

This has been added: “The average difference ($p\text{CO}_2\text{-1}$ minus $p\text{CO}_2\text{-2}$) was -2 μatm and the maximum difference was 41 μatm .” (P. 12, 14-15)

2.19. P2805, L14ff: Is this temperature corrected data or not? Did you check for autocorrelations between PAR and temperature differences (I would expect that temperature offsets could be influenced by exposure of the equilibrator to bright sunlight)?

The data is temperature corrected. “A correlation coefficient of 0.47 between the predictor variables (ΔTeq and PAR) indicated that they are not strongly correlated. Tolerance (or the inverse of the variance inflation factor) is 0.78 indicating no collinearity problems as this value is well above the tolerance threshold of 0.1 (Quinn and Keough, 2002).” (P. 13, 6-10)

2.20. P2805, L27 – P2806, L1: But $p\text{CO}_2\text{-2}$ was the one being closer to the values calculated from A_T and C_T (cf. P2804, L18). Please comment on this in the manuscript.

This has been clarified in the manuscript: “because $p\text{CO}_2\text{-2}$ compared to calculated from C_T and A_T had a smaller MR but higher RMSE and lower r than $p\text{CO}_2\text{-1}$ (Table 1). Lower MR does not necessarily mean that there is a better agreement, because

positive values may compensate negative values, as is the case here. The ranges of the residuals were (-7 to 14) μatm and (-29 to 39) μatm , for $p\text{CO}_2\text{-1}$ and $p\text{CO}_2\text{-2}$ respectively.” (P. 12, 5-10)

2.21. P2806, L1-L7: Are these novel suggestions? If not, please reference the statements appropriately.

Referee 1 also makes this point (1.19 and 1.29). Some of the suggestions are novel and some are not. We have made it clear which ones are and which ones not (Sect. 3.2).

2.22. P2807, L8-15: There are much more differences between both studies (e.g. measurement quality/overall uncertainty, number of samples, sample volume, scientific background of conductors, etc.) which could be discussed here. From my perspective, the presented dataset should compare well with other field-based intercomparisons (e.g. Luecker et al. 2000) rather than the OA-related lab-based datasets presented by Hoppe et al. (2012).

We have deleted the reference to Hoppe et al. (2012) and discussed instead Luecker et al. (2000).

2.23. P2808, L6: How can this study be used to the evaluation of OA impacts?

See response above to point 2.5.

2.24. P2808, L9 – P2809, L27: Are these novel suggestions? Otherwise please refer to other publications.

See 2.21.

2.25. P2809, L21: Does “We estimated...” refer to Rérolle et al. 2013? Then please state “They estimated...”

Yes, it refers to Rérolle et al. (2013) and we have changed the text accordingly. (P. 17, 27)

2.26. P2811, L6: This will prevent phototrophic growth but not microbial growth in general.

Referee 1 also makes this point (1.22). We have changed that here and elsewhere: “prevent phototrophic growth in $p\text{CO}_2$ equilibrators by completely shielding them from exposure to light.” (P. 20, 21-22)

2.27. P2811, L8-10: Please provide references for the statement.

Done: “As found in other studies (Millero, 1995; Cullison Gray et al., 2011), the variables pH_T and $p\text{CO}_2$ are far from an ideal pair for calculation of C_T or A_T ,” (P. 20, 26-28).

2.28. Table 1: How can the calculated uncertainty be identical for the two $p\text{CO}_2$ -measurements if you know that there are systematic differences between the two systems (light exposure, length of inlet system etc.)?

The way that the uncertainty of a calculated variable such as calculated $p\text{CO}_2$ is determined is not affected by $p\text{CO}_2$ measurement uncertainties but rather by C_T , A_T and pH_T measurements uncertainties. We now clarify this point in the text (P. 10, 7-9) and the Table 1 caption: “Calculated $p\text{CO}_2$ -1 and calculated $p\text{CO}_2$ -2 have the same uncertainty because they depend only on the accuracies of the variables from which they are calculated. They are therefore identical for both $p\text{CO}_2$ systems”

2.29. Figure 2: Caption should read "... here defined as the residual of $p\text{CO}_2$ -1 (a) or $p\text{CO}_2$ -2 (b) and another measured or calculated...”

We have reworded figure 2 captions.

2.30. Figure 3: Is this corrected data? If so, how do you end up with a RMSE of only $10\mu\text{atm}$? Caption should read "Comparison between the $p\text{CO}_2$ between instruments 1 and 2 (μatm ; white circles) and the photosynthetically active radiance (PAR (W m^{-2}), in black circles) measured over five days“.

We double checked the figure and the RMSE calculation and all are correct. We did the zoom in the worst days (see fig. 2a). We have changed the caption.

2.31. Figure 5: Given that you identify more problems with equilibrator 2, you should also show these plots for $p\text{CO}_2$ -2. Please add panel identifications (a-d). Caption should read “Box-and-whisker plots of the residuals between measured $p\text{CO}_2$ (a), pH (b), C_T (c), A_T (d) and the respective estimates calculated from different pairs of measured variables (denoted on the x axis) for the two sets of borate constants.

Done.

References:

- Clark, D. R., Brown, I. J., Rees, A. P., Somerfield, P. J., and Miller, P. I.: The influence of ocean acidification on nitrogen regeneration and nitrous oxide production in the North-West European shelf sea, *Biogeosciences Discuss.*, 11, 3113-3165, 10.5194/bgd-11-3113-2014, 2014.
- Cullison Gray, S. E., DeGrandpre, M. D., Moore, T. S., Martz, T. R., Friederich, G. E., and Johnson, K. S.: Applications of *in situ* pH measurements for inorganic carbon calculations, *Mar. Chem.*, 125, 82-90, 2011.
- Hopkins, F. E., and Archer, S. D.: Consistent increase in dimethyl sulphide (DMS) in response to high CO₂ in five shipboard bioassays from contrasting NW European waters, *Biogeosciences Discuss.*, 11, 2267-2303, 10.5194/bgd-11-2267-2014, 2014.
- Hoppe, C. J. M., Langer, G., Rokitta, S. D., Wolf-Gladrow, D. A., and Rost, B.: Implications of observed inconsistencies in carbonate chemistry measurements for ocean acidification studies, *Biogeosciences*, 9, 2401-2405, 2012.
- IOCCP: Ocean surface pCO₂, data integration and database development, 14-17 January 2004, IOCCP report 2, IOC Workshop Report 211, available at: http://www.ioccp.org/images/01Underway/WR211_eo.pdf (last access 28 April 2014), UNESCO Tsukuba, Japan, 2004.
- Katayama, H., Karasudani, T., Ishii, K., Marubayashi, K., and Ueda, H.: Development of a gas-liquid equilibrator for estimating CO₂ flux at the ocean surface, *Journal of Atmospheric and Oceanic Technology*, 16, 1450-1455, 1999.
- Körtzinger, A., Thomas, H., Schneider, B., Gronau, N., Mintrop, L., and Duinker, J. C.: At-sea intercomparison of two newly designed underway pCO₂ systems—encouraging results, *Mar. Chem.*, 52, 133-145, 1996.
- Körtzinger, A., Mintrop, L., Wallace, D. W. R., Johnson, K. M., Neill, C., Tilbrook, B., Towler, P., Inoue, H. Y., Ishii, M., Shaffer, G., Saavedra, R. F. T., Ohtaki, E., Yamashita, E., Poisson, A., Brunet, C., Schauer, B., Goyet, C., and Eiseid, G.: The international at-sea intercomparison of fCO₂ systems during the R/V *Meteor* Cruise 36/1 in the North Atlantic Ocean, *Mar. Chem.*, 72, 171-192, 2000.
- Lueker, T. J., Dickson, A. G., and Keeling, C. D.: Ocean pCO₂ calculated from dissolved inorganic carbon, alkalinity, and equations for K₁ and K₂: Validation based on laboratory measurements of CO₂ in gas and seawater at equilibrium, *Mar. Chem.*, 70, 105-119, 2000.
- MacGilchrist, G. A., Shi, T., Tyrrell, T., Richier, S., Moore, C. M., Dumousseaud, C., and Achterberg, E. P.: Effect of enhanced pCO₂ levels on the production of DOC and TEP in short-term bioassay experiments, *Biogeosciences Discuss.*, 11, 3701-3730, 10.5194/bgd-11-3701-2014, 2014.
- McElligott, S., Byrne, R., Lee, K., Wanninkhof, R., Millero, F., and Feely, R.: Discrete water column measurements of CO₂ fugacity and pH_T in seawater: A comparison of direct measurements and thermodynamic calculations, *Mar. Chem.*, 60, 63-73, 1998.
- Millero, F. J.: Thermodynamics of the carbon dioxide system in the oceans, *Geochim. Cosmochim. Acta*, 59, 661-677, [http://dx.doi.org/10.1016/0016-7037\(94\)00354-O](http://dx.doi.org/10.1016/0016-7037(94)00354-O), 1995.
- Millero, F. J.: The marine inorganic carbon cycle, *Chem. Rev.*, 107, 308-341, doi: 10.1021/cr0503557, 2007.
- Pierrot, D., Neill, C., Sullivan, K., Castle, R., Wanninkhof, R., Lüger, H., Johannessen, T., Olsen, A., Feely, R. A., and Cosca, C. E.: Recommendations for autonomous underway pCO₂ measuring systems and data-reduction routines, *Deep-Sea Res Pt. II*, 56, 512-522, 2009.
- Poulton, A. J., Stinchcombe, M. C., Achterberg, E. P., Bakker, D. C. E., Dumousseaud, C., Lawson, H. E., Lee, G. A., Richier, S., Suggett, D. J., and Young, J. R.: Coccolithophores on the north-west European shelf: calcification rates and environmental controls, *Biogeosciences Discuss.*, 11, 2685-2733, 10.5194/bgd-11-2685-2014, 2014.
- Rérolle, V., Floquet, C. F., Harris, A. J., Mowlem, M. C., Bellerby, R. R., and Achterberg, E. P.: Development of a colorimetric microfluidic pH sensor for autonomous seawater measurements, *Anal. Chim. Acta*, 786, 124-131, 2013.

Rérolle, V. C. M., Ribas-Ribas, M., Kitidis, V., Brown, I., Bakker, D. C. E., Lee, G. A., Shi, T., Mowlem, M. C., and Achterberg, E. P.: Controls on pH in surface waters of northwestern European shelf seas, *Biogeosciences Disc.*, **11**, 943-974, doi: 10.5194/bgd-11-943-2014, 2014.

Richier, S., Achterberg, E. P., Dumousseaud, C., Poulton, A. J., Suggett, D. J., Tyrrell, T., Zubkov, M. V., and Moore, C. M.: Carbon cycling and phytoplankton responses within highly-replicated shipboard carbonate chemistry manipulation experiments conducted around Northwest European Shelf Seas, *Biogeosciences Discuss.*, **11**, 3489-3534, 10.5194/bgd-11-3489-2014, 2014.

Van Heuven, S., Pierrot, D., Rae, J. W. B., Lewis, E., and Wallace, D. W. R.: CO2SYS v 1.1, MATLAB program developed for CO₂ system calculations, ORNL/CDIAC-105b. Carbon Dioxide Information Analysis Center, Oak Ridge National Laboratory, U.S. DoE, Oak Ridge, TN., 2011.

Young, J. R., Poulton, A. J., and Tyrrell, T.: Morphology of *Emiliana huxleyi* coccoliths on the North West European shelf – is there an influence of carbonate chemistry?, *Biogeosciences Discuss.*, **11**, 4531-4561, 10.5194/bgd-11-4531-2014, 2014.