

RESPONSE TO REFEREE #1

Greenop et al. present a new $\delta^{11}\text{B}_{\text{sw}}$ -reconstruction over the past 23 million years, using the $\delta^{13}\text{C}$ difference between planktic and benthic foraminifers to predict the pH gradient between the surface and deep ocean, and then applying that pH gradient to infer $\delta^{11}\text{B}_{\text{sw}}$ from paired $\delta^{11}\text{B}$ analyses in planktic and benthic foraminifers from the same core sites and time intervals. The $\delta^{13}\text{C}$ vs pH gradient relationship in the modern is based on shipboard data, and the relative constancy of that relationship over the past 23 million years is estimated from carbon cycle models GENIE and CYCLOPS.

While this study is applaudable in the sense that it presents an elegant alternative approach to previous $\delta^{11}\text{B}_{\text{sw}}$ estimates, I have to admit that I find the practical results disappointing. Whereas the actual reconstruction results in a 10 permil $\delta^{11}\text{B}_{\text{sw}}$ spread over the past 23 million years, it is only reduced to something more feasible after applying a hefty smoothing factor. However, this does not diminish the effort presented in this study, reconstructing $\delta^{11}\text{B}_{\text{sw}}$ is difficult, and we do not have any direct archives for it, so any new evidence that allows us to home in on a consensus view is highly appreciated. In this regard it is encouraging that the reconstruction finds some significant synergy with previous efforts, but given the large uncertainties presented herein, and the fact that this is yet another indirect estimate, I recommend a somewhat more inclusive final figure of the current state of understanding of this parameter. I personally would feel extremely uncomfortable to apply this reconstruction as the only estimate.

To accomplish this goal, several significant revisions should be made to improve the presentation of these new data, but also the presentation of previously published data.

We thank the referee for the explicitly positive view on our work, and we particularly appreciate the thoughtful and constructive criticism voiced here. We agree that in our writing we were too much focused on shortcomings of previous approaches to reconstruct seawater boron isotopic composition, and in our discussion/conclusion we placed too much emphasis on the differences

between our results and previous reconstructions. We follow the referee's suggestion by significantly revising the manuscript in three specific ways: (1) in our description of the current state of the art we more clearly state that the description of shortcomings serves to motivate the more thorough approach that we describe rather than to invalidate previous results; (2) we include a wider set of reconstruction in our final figure that is central to our discussion; and (3) we now highlight the rather good agreement between our and the results of Raitzsch & Hönisch before discussing the differences. Overall, we have modified the tone of the manuscript with regard to prior work and we emphasize the progress towards a consensus on past seawater boron isotope change that is supported by different approaches, including our own.

To start from the beginning, the authors use the introduction to discredit all previous $\delta^{11}\text{B}_{\text{sw}}$ reconstructions, except their own (Foster et al. 2012). I encourage the authors to consider that this may be a useful strategy when proposing a study, but at this point their data are available, and given the uncertainty associated with their results, I recommend toning down the arguments, providing solid arguments if some data are to be dismissed, and being less judgmental in the presentation of previously published data. For instance, Paris et al. 2010 presented $\delta^{11}\text{B}$ from modern halites that reflect $\delta^{11}\text{B}_{\text{sw}} = -39.7$ permil, so they did present a reasonable and indeed promising proof of concept, and this should be acknowledged.

We will acknowledge this point in the revised manuscript and as noted above we have changed the tone of the manuscript.

I agree with the authors that the implications of the paleo- $\delta^{11}\text{B}_{\text{sw}}$ reconstructions based on halites are highly unlikely, and the data do not match other estimates. However, coeval halite data are highly consistent in their $\delta^{11}\text{B}_{\text{sw}}$, which suggests the data are likely due to a specific and ocean-

wide cause, and less likely to a variety of causes, as currently suggested by the authors.

In the study of Paris et al., (2010) coeval halites for the samples at 5.5 Ma show a spread of 3 per mil for a single location. While the authors of the paper suggest the lower values are attributable to a local fractionation following brine calcium enrichment or acidification, and discard the data, it is difficult to assess to what extent this process has occurred in each sample. However, the results do have utility as a minimum constraint. In the revised manuscript we will further stress this point.

I would recommend the authors look at the details of Liu et al. 2000 (their figure 6b) and consider the effect of decreasing [Ca] concentration throughout the Cenozoic. According to Liu's data this could explain a 3-4 permil difference in $\delta^{11}\text{B}$ recorded by the halites, which is somewhat consistent with their much lower $\delta^{11}\text{B}_{\text{sw}}$ estimates. More work clearly needs to be done to study the effect of [Ca] on halite precipitation, but there is value in the study of Paris et al., and it should not be discredited so lightly.

In the revised manuscript we will expand on reasons for the difference in $\delta^{11}\text{B}_{\text{sw}}$ estimates between halites and other estimates. However, in the case of changing [Ca] concentration, estimates suggest that the concentration of [Ca] has increase by approximately 7 mmol (0.28g/L) (Horita et al., 2002) over the Neogene. This change is very small compared to range of [Ca] explored in the work of Liu et al., 2000.

Similarly, the data of Raitzsch & Hönisch 2013 are presented somewhat selectively in this study, and sometimes they are even misrepresented. Raitzsch & Hönisch 2013 used benthic foraminifers from all ocean basins for their study, to minimize the effect of local pH variations on their $\delta^{11}\text{B}_{\text{sw}}$ estimate. This fact alone deserves some acknowledgment and contrasts with the predominantly Atlantic focus presented by Greenop et al.

We will acknowledge this is the revised manuscript and where appropriate we highlight the good agreement between the two datasets.

Raitzsch & Hönisch's uncertainties (as displayed in Greenop's Fig. 5 but omitted in Greenop's Fig. 10) are averages based on 2-4 samples, where each individual sample is based on a monospecific benthic foraminifer sample from different ocean basins. The individual sample uncertainties are in fact comparable to the new individual data presented by Greenop et al., but what is presented in Greenop's Fig. 5b is the propagated uncertainty of all data normalized to the epibenthic *C. wuellerstorfi*. The statement on "large individual sample uncertainties" (Line 525) is therefore an inappropriate comparison.

In the revised manuscript we will make it clear where we are referring to the uncertainty in $\delta^{11}\text{B}$ measurements (which is comparable to this study) and the uncertainty in $\delta^{11}\text{B}$ borate (where the uncertainty is larger than for this study as multiple species are used).

In contrast, these propagated uncertainties should in fact be shown in Greenop's Fig. 10, because they do reflect the individual data uncertainty associated with the fit (in contrast to Greenop's claim that individual data uncertainty is not accounted for (Line 526/527)).

In the revised manuscript these propagated uncertainties will be added to the figure. It is important to note, however, the methodologies for propagating the uncertainty in the studies of Greenop et al., and Raitzsch & Hönisch are different and consequently not directly comparable.

But more importantly, Greenop et al. selected only one of the two solutions presented by Raitzsch & Hönisch, and specifically the one that shows greater discrepancy from their data. The reason why Raitzsch & Hönisch presented two estimates is because:

(a) the $\delta^{11}\text{B}$ vs pH sensitivity of all marine carbonates calibrated over a wide pH range (>0.5 units) shows a sensitivity that is less than that predicted by the aqueous fractionation factor of Klochko et al. 2006, and (b) because models and experimental data suggest that boron isotope fractionation is affected by temperature (Zeebe 2005, Rustad et al. 2010, Dissard et al. 2012, Kaczmarek et al. 2015, Liu et al. 2015). These two factors were placed on hold by Rae et al. 2011 because they could not evaluate them within their 0.3 pH unit coretop calibration, and Rae et al admitted that their “closeness to inorganic theory may be fortuitous”. We know now that the original arguments of Foster (2008) and Rae et al. (2011), that all marine carbonates follow a pH sensitivity similar to aqueous boron isotope fractionation when measured by MC-ICP-MS, is false (e.g. Henehan et al. 2013, Krief et al, 2010), and evidence for a temperature effect under otherwise constant environmental conditions is accumulating (see above). Raitzsch & Hönisch 2013 therefore applied two sensitivities, one similar to empirical carbonate calibrations established over a wide pH range, and the other using the aqueous fractionation after Klochko et al. 2006. Only the latter is shown by Greenop et al., but the first, which is more and more confirmed by newer experimental data as described above, has been omitted. Remarkably, the omitted record matches the estimates of Greenop et al. much better than the one they chose to present in this manuscript. I would urge the authors to include both, and provide appropriate arguments for each of them.

In our modified figure 10 we now show both scenarios from Raitzsch and Hönisch (2013). However, we disagree with a number of points raised above. Specifically, our reading of the literature suggests that whilst the aqueous fractionation factor for boron is well-known (Klochko et al. 2006, Nir et al., 2015) there is little consensus regarding the universality or not of the $\delta^{11}\text{B}$ response of marine carbonates to changing pH. Furthermore, there is no experimental evidence supporting a *significant* effect of temperature on the aqueous fractionation factor (Klochko et al., 2006). Indeed, a recently published inorganic calcite precipitation study has shown no measureable effect (at the ± 0.3 ‰ level) of temperature on the fractionation factor when

growth rate is controlled (within the temperature range investigated 12 °C to 32 °C; Kaczmarek et al. 2016). This finding is entirely in agreement with the conclusions of Rae et al. (2011), Henehan et al. (2013), Martinez-Boti et al. (2015), Henehan et al. (in review) based on field calibrations. As noted by the reviewer, Dissard et al. (2012) do find that the $\delta^{11}\text{B}$ in *Acropora* coral cultured at 22 to 28 °C is temperature dependent. However, the observed dependency (~0.15 permil per °C) is within uncertainty of the expected relationship given the known dependence of pKB on temperature (~0.13 permil per °C) and hence is not evidence of a significant temperature influence on the isotopic fractionation factor. It is of course likely that the fractionation factor is temperature dependent, current evidence however suggests it is smaller than analytical uncertainty. In a comprehensive study of the theoretical determination of the boron isotopic fractionation factor Zeebe (2005) states that:

“Given the range of outcome for $\alpha_{\text{B3-B4}}$ [fractionation factor] at 300 K calculated in the current paper, no recommendation will be made regarding α 's temperature dependence, which equally depends on the frequencies/methods chosen.”

So we feel that currently the safest option, and one that is supported by the field calibration datasets (Rae et al., 2011; Henehan et al., 2013; Martinez-Boti et al., 2015; Henehan et al. in review) is to not apply a temperature correction. Nonetheless, for completeness we will show both curves from R&H (2013) whilst also reiterating these points in the revised manuscript.

There are certainly shortcomings in the approach applied by Raitzsch & Hönisch, as there are with all indirect $\delta^{11}\text{B}_{\text{sw}}$ estimates, but looking for synchrony between different approaches, and learning from differences would be powerful way to improve the paleoceanographic community's confidence in this valuable proxy; insisting only on differences will do the opposite.

The agreement between the two approaches is indeed encouraging. In the revised manuscript we endeavour to find the common ground between the studies whilst also highlighting the uncertainties (including those in our own study).

Finally, the estimates of Pearson & Palmer (1999, 2000) are dismissed based on their N-TIMS analytical procedure and the use of the boron isotope fractionation factor after Kakihana and Kotaka (1977). I disagree with both of these arguments: Foster et al. (2013) have shown that N-TIMS and MC-ICP-MS report different numbers in absolute terms, but the relative differences between samples are similar, such that they will yield the the same amplitude of $\delta^{11}\text{B}$ and pH change. The comparison of benthic $\delta^{11}\text{B}$ data measured by N-TIMS by Raitzsch & Hönisch (2013) and Greenop in Figure 5 confirms this notion and argues against the analytical procedure creating a systematic bias. Foster and his colleagues have discredited N-TIMS analyses long enough with this unfounded argument, and it should finally be put to rest. Similarly, the $\delta^{11}\text{B}$ vs. pH sensitivity implied by the Kakihana factor actually matches the sensitivity of empirical carbonates very well, both using N-TIMS (e.g. Sanyal et al. 1996,2001, Hönisch et al. 2004) and MC-ICP-MS (e.g., Krief et al. 2010, Henehan et al. 2013). So even if the Kakihana fractionation factor incorrectly describes boron isotope fractionation in seawater, it accidentally describes marine carbonates, and in particular planktic foraminifers very well. There may be analytical issues with the specific N-TIMS method applied by Palmer, but we will not be able to evaluate this because that particular method is no longer applied in any labs reconstructing pH from marine carbonates. The only reasonable and significant argument that can and must be applied for the data of Pearson and Palmer (1999, 2000) is that they did not know about vital effect offsets between different foraminifer species at that time, and consequently applied the same calibration curve for a variety of them. Given what we know about vital effects on $\delta^{11}\text{B}$ today, that is a fundamental

shortcoming of those studies, and I agree that they should be considered with caution on this basis.

While the interlaboratory comparison of Foster et al., (2013) did suggest relative variation in $\delta^{11}\text{B}$ are recorded by different analytical methods, the study also found some indication that for measurements by NTMIS the deviation from the interlaboratory mean increases with decreasing B/Ca ratio. This is important for the study of Pearson & Palmer (1999, 2000) where the whole foraminiferal assemblage is used, with a range of B/Ca, as the gradient between different species may in part be influenced by these analytical issues. While the lack of offset between the benthics from Raitzsch and Honisch (2013) and this study further confirms that the offset between MC-ICPMS and NTMIS is smaller in foraminifera with higher B/Ca, this good agreement is in contrast to that observed by Rae et al. (2011) who documented a ~ 1.2 permil difference between *C. wuellerstorfi* measured by MC-ICPMS and NTIMS (by Honisch et al. 2008). We see no offsets between our dataset and Rae et al. (2011) and so the NTIMS vs. MC-ICPMS offsets in absolute sense are clearly not that well understood. Similarly, at ODP 668 in the Equatorial Atlantic, we see a 2.6 permil difference between similar sized Holocene *G. sacculifer* (comparing data in Foster, 2008 and Honisch and Hemming, 2005), but the Atlantic benthic data in Raitzsch and Honisch (2013) are in good agreement between methods. It is for this reason that we are cautious about over interpreting the agreement between our estimates of $\delta^{11}\text{B}_{\text{sw}}$ and those of Pearson and Palmer (1999, 2000). In addition, there are also uncertainties associated with the difficulty of assigning a depth habitat to individual foraminiferal species and species-dependent isotope effects (as mentioned above) Pagani et al., (2005). However, given the comments of this reviewer, in the revised manuscript we do include the Pearson and Palmer (1999,2000) estimates in Figure 10 and use them to support the emerging view of the evolution of $\delta^{11}\text{B}_{\text{sw}}$ (albeit caveated with the points above).

Site locations: A map should be provided with the locations of the studied core sites, including a justification why these sites lend themselves to this particular study. Their water depths should be mentioned, and the preindustrial $\delta^{13}\text{C}$ and pH gradient for each core site should be added to Figure 4c, so the reader can evaluate how well these sites fit the fundamental premise of this study, i.e. that these sites can be used to reconstruct the $\delta^{13}\text{C}$ and pH gradients over the past 23 million years. It is particularly noteworthy here that sites 758 and 761 are from the Indian Ocean and therefore differ in their hydrography from Atlantic sites 926 and 999. Importantly, this difference is somewhat dismissed in the later part of the study, where the slope specific to Atlantic sites is selected and applied to estimate the pH gradients for all sites. A map will be added to the revised manuscript. These sites were initially chosen as they have been shown to faithfully reconstruct surface water pH in previous studies (Foster et al., 2008; Foster et al., 2012; Martinez-Boti et al., 2015). ODP Site 758 was used in Pleistocene to extend coverage outside of the Atlantic Ocean. In the supplement we present the data using both the “Atlantic only” and the “low latitude” gradient. In the revised manuscript we will plot the data from both these scenarios in the main part of the manuscript in order to more fully explore the data.

Boron isotope reconstructions: I have been looking for the original $\delta^{11}\text{B}$ data measured for each sample, but cannot find them in tabular form and Fig. 5 only displays them in the form of $\delta^{11}\text{B}$ borate, i.e. already one step removed from the original data. It is important to list and display the original data and their uncertainties. Have different foraminifer species been measured on the same samples, to cross-calibrate ancient species versus modern *G. ruber*? This is an important aspect for evaluating the appropriateness of the applied $\delta^{11}\text{B}$ vs $\delta^{11}\text{b}$ borate calibrations.

We are changing figure to show planktic $\delta^{11}\text{B}$ foraminifera in the revised manuscript. No $\delta^{11}\text{B}$ -pH calibration is applied to the benthic $\delta^{11}\text{B}$ so this is the original data in agreement with Rae et al. (2011). For the majority of the

samples we use either the species-specific *G. ruber* or *G. sacculifer* calibration (from 0 to 22 Ma). In the case of samples at the Oligo-Miocene boundary the absence of either of these species means we are unable to do this. Here we use *G. praebulloides* making inferences about its life habit from $\delta^{18}\text{O}$ and $\delta^{13}\text{C}$ gradient with size fraction (Pearson and Wade, 2009). We currently do not have a sample where *G. praebulloides* and *G. sacculifer* co-exist and this is an area that needs to be further researched in the future. We hope to display this information clearer in the revised manuscript.

Modeling the $\delta^{13}\text{C}$ vs pH relationships: It would be nice to provide the exact parameters and their ranges applied for the model estimates in a table, so any reader could repeat these experiments without having to contact the authors. Also, references should be provided for all estimates, and please reconsider some of the wording. For instance, it is unclear whether atmospheric pCO_2 was applied once at 200 ppm and once at 1000ppm, or across the entire range (Line 368). Please list studied ranges for pH, CaCO_3 saturation states and the biological pump in tabular form.

We add this table to the revised manuscript.

The slopes of the $\delta^{13}\text{C}$ - pH relationships (Fig. 7) need to be better explained. I take from the text that the global estimate includes all data, then data are systematically restricted first to the low latitudes, and then to the low latitude Atlantic only. That should make the data density sparser from step to step, but while this seems somewhat correct for the dense cloud between 1.7-3 permil $\delta^{13}\text{C}$, the low-latitude and Atlantic-only subsets (b and c) display data that are not included in Fig. 7a. Please explain where these data are coming from, and why they are not included in the complete data set (7a)? To allow the reader to better evaluate the consequence of this restriction, the data in Fig. 4c should be colored to highlight global low latitude and low latitude Atlantic data, and the respective regressions should be shown. Furthermore, it needs to be

evaluated how the choice of slope affects the pH gradient estimates for the Indian Ocean cores. Specifically, the two 758 samples at around 2 Ma appear to have a large effect on the Pleistocene $\delta^{11}\text{B}_{\text{sw}}$ estimates, in particular after the smoothing function is applied. It seems like the Indian Ocean sites should be estimated using either the global low latitude slope (Fig. 7b) or a slope specific to the Indian Ocean, but not the one for the low latitude Atlantic.

In Figure 7 there is one estimate of the slope per individual ensemble member simulation (i.e. the regression for that simulation). In the “global” case we regress all 18 model boxes, in the “low latitude” case we regress only the low latitude surface/deep boxes (4x surface and 4x deep; 8 boxes) and in the “Atlantic-only” calculate the slope from only the low latitude Atlantic surface and deep box (2 boxes). That is, each of the panels has the same number of estimates for the slope of the relationship, but they are evaluated from the model ensemble in different ways. All experiments are consistent between the three panels of Figure 7. To prevent any confusion we have modified the relevant text and expanded the caption of the figure.

Smoothing the $\delta^{11}\text{B}_{\text{sw}}$ estimates: Fig. 9 demonstrates that irrespective of the pH gradient chosen, the data uncertainty encompasses the entire record, and only application of the smoothing function allows to discern a $\delta^{11}\text{B}_{\text{sw}}$ trend. Greenop et al. argue that the smoothing function is justified because Lemarchand et al. 2002 calculated the $\delta^{11}\text{B}_{\text{sw}}$ rate of change is ~ 0.1 permil per million years (line 398 in this manuscript). However, Lemarchand et al. also estimated up to 0.6% $\delta^{11}\text{B}_{\text{sw}}$ change per million years for the Cretaceous, suggesting that the 0.1 permil/Myr change is not the limit and depends on the model parameterization. Remarkably, Greenop et al. eventually argue that their record shows more variability than the model estimates of Lemarchand et al., and suggest that the model relevant boron inputs and outputs are not fully understood (Line 532/533). This means that Greenop et al want it both ways – they use the small rate of change to justify their hefty smoothing, but then cast doubt on Lemarchand’s estimates when they show less variation than these new estimates. The smoothing obviously

cannot be justified with Lemarchand's estimates, and what-ever argument is brought up must be independent of an alternative $\delta^{11}\text{Bsw}$ estimate, or the argument becomes circular. Regardless, given that the smoothing function reduces the potential $\delta^{11}\text{Bsw}$ ranges estimated from this approach from 42-32 permil to \sim 40-37 permil implies that application of the smoothing function largely dismisses the actual data, and therefore the heart of the reconstruction. This data treatment may be appropriate for a dataset with hundreds of data, but with just 20 data from not even a handful of sediment cores I am having a hard time finding the smoothed estimate convincing. In order to justify our use of smoothing we refer to the study of Lemarchand et al., (2000). Here we determine an approximate rate of change for $\delta^{11}\text{Bsw}$ from the inputs and outputs of boron into and out of the ocean relative to the boron concentration. This is therefore a fundamental constraint on the behaviour of the boron isotope system in the oceans given what we know about the modern inputs and outputs. Where we have concerns about the record of Lemarchand et al., (2000), is in regards to the extent to which changes in the magnitude of these inputs/outputs through time are understood. Many of the processes that determine the $\delta^{11}\text{Bsw}$ are poorly constrained through time (e.g. crustal permeability, lifetime of water-rock interactions, and expansion rate of the oceanic ridge; Simon et al., 2006). We make minor changes to the text to clarify this distinction.

This reviewer also suggests that by smoothing is somehow dismisses the actual data. In the revised manuscript we will detail why smoothing is both needed (since the uncertainties in each $\delta^{11}\text{Bsw}$ reconstruction is large due to propagation of uncertainties in all parameters involved) and justified. The latter is simply because our $\delta^{11}\text{Bsw}$ data exhibit a larger variability than is possible given the amount of boron in the oceans and the likely inputs and outputs. As noted by reviewer 2, given the long residence time of boron in the oceans it should have a smooth evolution and one that, given what we know about the boron cycle currently, should not result in large changes (i.e. < 0.1 permil per Ma). In order to further support our decision to smooth the

estimates we explore a number of other scenarios (e.g. binning the data, using an algorithm to smooth, and assume a spline fit) and then focus the subsequent discussion on what aspects of the evolution of $\delta^{11}\text{Bsw}$ are consistent across these scenarios and hence robust to the nature of our chosen smoothing (e.g. that $\delta^{11}\text{Bsw}$ is around xx permil lighter than modern in the middle Miocene, and much of the change in ratio occurred during the interval xx to xx Ma).

Similarly, Greenop et al. use the variability observed in their d^{13}C and pH gradient estimates to argue the variable pH gradient estimates give a better result than the constant pH gradient approach. With error bars that essentially overlap for all samples, and two estimates randomly discarded, I am rather uncomfortable believing any arguments for preferring one estimate over the other. Given the large uncertainty of this approach, I would recommend Greenop et al. plot all their estimates, and do not bury half of them in the supplement. This entire range of possible d^{11}Bsw estimates should be shown in Fig. 10, and not a subjective subset. Ideally, Figure 10 would present all available estimates not only from this study, but also from others (Raitzsch & Hönisch 2013, Lemarchand et al. 2000); and then focus the discussion on synergies between different data sets, find a consensus d^{11}Bsw trend, and discuss whether the synergies can help us to further improve our understanding of the boron isotope proxy.

Again, we want to thank the referee for the thoughtful comments along these lines, which we agree with and used to guide manuscript revision throughout. In the revised manuscript we will include a figure of all our $\delta^{11}\text{Bsw}$ records using the full range of different assumptions compared with published data and now put emphasis on how a collective view of the available $\delta^{11}\text{Bsw}$ records can aid our understanding.

Comparison to other isotope systems: This is a nice comparison and it looks convincing after the true data variability has been discarded, but it would be nice to expand this analysis and include all possible $\delta^{11}\text{B}_{\text{sw}}$ scenarios in it, including the actual data before smoothing. Suggesting weathering and catchment patterns from correlating the smoothed record to actual data seems premature given the significant uncertainty of the original (unsmoothed) $\delta^{11}\text{B}_{\text{sw}}$ estimates.

Clearly the database for this analysis is not fully satisfying at this point, both for boron isotopes and the other isotope systems. We will make it clearer in the revised manuscript that these interpretations are tentative and subject to further work, as the seawater records of all the isotope systems are improved. We take the position that the apparent correspondence among the isotopic change of the different elements is interesting and thought provoking, in particular with respect to future modelling work to better understand the evolution of $\delta^{11}\text{B}_{\text{sw}}$, even if it is still tentative.

Line 28: sp: consistent

Line 46: sp: Martínez-Botí

Line 54: please add Lemarchand et al. 2000. They were the first to note the change in $\delta^{11}\text{B}_{\text{sw}}$ and compare it to the analytical uncertainty of $\delta^{11}\text{B}$

Line 66: please add a reference for the $\delta^{11}\text{B}$ of B inputs Line 68: please specify whether mineral precipitation or rainfall is meant by “precipitation”

Line 76: please replace Zeebe & Wolf-Gladrow by Hemming & Hanson 1992, the CO_2 book only reports the earlier suggestion Line 83: please add

Lemarchand et al. 2000

Line 90/91: please rephrase “currently...at present”

Line 115/116 please rephrase: a small subset of the CO_2 data used by Ridgwell et al. 2005 is derived using the boron isotope-pH proxy, leading to some circularity in the $\delta^{11}\text{B}_{\text{sw}}$ method. However, the pH estimates applied by

Raitzsch & Hönisch (2013) from Tyrrell & Zeebe (2004) are based on GEOCARB, and the circularity problem does not apply.

Line 216: Why is it important to specify “at 0.7 Ma” here? Please explain or rephrase

All of the above typos and revised sentencings will be carried out in the revised manuscript. We would argue, however, that the suggested rephrasing of Line 115/116 is slightly misleading. Despite the GEOCARB pH estimates not containing any estimates derived using the boron isotope pH proxy, the $\delta^{11}\text{B}_{\text{sw}}$ record is going to be heavily dependent on the surface water pH input. Consequently, if this $\delta^{11}\text{B}_{\text{sw}}$ record is going to be subsequently used in pH calculations, the output pH will strongly reflect the GEOCARB input.

Line 220: please specify that you assume there is no temperature effect on $\delta^{11}\text{B}$. Ideally you should calculate $\delta^{11}\text{B}$ borate with and without the temperature effect, and estimate the associated uncertainty on the $\delta^{11}\text{B}_{\text{sw}}$ estimate

For the reasons described above we have elected not to apply a temperature correction to the calculated $\delta^{11}\text{B}$.

Line 228/252: please specify whether trace element and $\delta^{13}\text{C}$ analyses were performed on the same foraminifer species and size classes as studied for $\delta^{11}\text{B}$

All trace element data was analysed on an aliquot of the $\delta^{11}\text{B}_{\text{sw}}$ samples. In almost all cases $\delta^{13}\text{C}$ was also analysed on the same foraminifera. Where this was not possible another surface dweller/benthic foraminifera was used. All $\delta^{13}\text{C}$ will be made available on Pangaea on successful publication of the manuscript.

Line 231: please define “H”

This will be added to the revised manuscript.

Line 249-251/281-286: please specify the influence of temperature on the d13C-DIC estimate

We have not corrected the d13C-DIC estimates for temperature and will acknowledge this in the revised manuscript. Culture work by Bemis et al., 1999 suggests that in the symbiotic foraminifera *O. universa* the d13C is insensitive to temperature under low light levels and the d13C of high light shells decreased slightly with temperature (0.05 per mil/°C). Consequently, we expect changes in temperature to have a minimal effect on the d13C-DIC estimates.

Line 259: which benthic species was analyzed for d13C?

C. wuellerstorfi or *C. mundulus* was analysed for d13C

Lines 261-265: symbiotic planktic foraminifers record heavier d13C than surrounding seawater, so the vital effect adjustment for estimating d13C-DIC should be negative, not positive

This section will be rephrased as “We use a carbon isotope vital effect for *G. ruber* (+0.94 ‰; Spero et al., 2003), *T. sacculifer*/*G. praebulloides* (+0.46 ‰; Spero et al., 2003; Al-Rousan et al., 2004;), *C. mundulus* (+0.47 ‰; McCorkle et al., 1997) and *C. wuellerstorfi* (+0.1 ‰; McCorkle et al., 1997) to calculate dissolved inorganic carbon (DIC).

Line 269: sp: outline

Line 273: leads to a broad

Lines 295/296: please rephrase, e.g. “CaCO₃ dissolution increases CO₂- and pH”. The sentence is difficult to read. Line 302: it would be nice to anticipate at this point which of the described effects will be considered in the remainder of the paper

Line 304: please define the Suess effect and provide a reference

Line 306: please add the uncertainty of the slope

All of the above typos and revised sentencings will be carried out in the revised manuscript.

Line 399: please analyze and explain possible reasons for the small $\delta^{11}\text{B}$ differences between the planktic and benthic forams excluded here. Is there any indication/reason why these data are anomalous?

It is possible that preservation is not so good at these intervals in the core and the planktic foraminifera are affected by partial dissolution (Seki et al., 2010). Alternatively, localised circulation changes at the core site may be affecting the benthic $\delta^{11}\text{B}$ signal.

Line 416/417: please rephrase to include that the benthic $\delta^{11}\text{B}$ record is a function of both pH and $\delta^{11}\text{B}_{\text{sw}}$

Will update in the revised manuscript.

Fig. 4 caption: Please explain the difference between open and closed symbols, and different symbol shapes in 4c. It would be advisable to replot this figures according to recommendations made above, instead of just copying it from Foster et al. 2012.

The figure will be replotted and the caption made clearer.

Line 1007: sp: Suess. To the best of my knowledge Dr. Seuss never wrote about carbon isotopes.

Line 1015: please explain what is meant by “paired measurements”

Mg/Ca measurements are conducted on an aliquot of the sample measured for $\delta^{11}\text{B}$.

Line 1026: sp: colours reflect Line 1033: sp: tests Line 1049: insert period after “details)”

This typo will be removed.

Fig. 2: Please remove the two lower end Miocene $\delta^{11}\text{Bsw}$ estimates of Paris et al., which were identified as outliers by the original paper. Including them makes the record look worse than it is, which seems unfair.

Following Paris et al., (2010) we will put these data points in brackets and explain their exclusion in the figure caption.

Fig. 3: It would be nice to insert a horizontal bar or arrow to indicate the smaller pH gradient in b.

This can be added to the figure.

Fig. 5: Please add a second time axis at the top of the figure for better comparison throughout the figure. Horizontal lines in each panel might also help to estimate values at each time interval. Please also indicate the planktic foram species measured for each data point, e.g. by different color symbols. Are there any data pairs that allow cross-calibration of *praebulloides* and *trilobus*?

The 2nd time axis and horizontal line will be added. Colour coding of species can also be applied. Unfortunately, there are no data pairs that allow cross-calibration of *G. praebulloides* and *G. trilobus* but this is a good avenue for study in the future.

Fig. 8: please indicate core site symbols in this figure. In particular, the large gradient at ~ 2 Ma is due to two samples from one site in the Indian Ocean. It is unfortunate to select two samples that are close in time from one site and then no more for the rest of the record. Those samples have the potential to bias the record at that time.

Core site samples can be added and, as noted above, we will fully explore the influence of these two samples on our smoothed record;

Fig. 10: please add both curves from Raitzsch & Hönisch, including their uncertainty estimates

We will add both curves from Raitzsch & Honisch, while explaining the caveats associated with including the record calculated with an empirical fractionation factor.

Fig. 12: at a minimum, I would suggest to remove the Indian Ocean cores from this estimate. I suspect the data will match much better without those estimates.

Here we elect to keep in the data from the Indian Ocean, however, we will put a greater emphasis in the manuscript on the record calculated using the global low latitude $\delta^{13}\text{C}$ -pH relationship and explore this further.