

## ***Interactive comment on “A record of Neogene seawater $\delta^{11}\text{B}$ reconstructed from paired $\delta^{11}\text{B}$ analyses on benthic and planktic foraminifera” by R. Greenop et al.***

**Anonymous Referee #1**

Received and published: 17 March 2016

Greenop et al. present a new  $d_{11}\text{Bsw}$ -reconstruction over the past 23 million years, using the  $d_{13}\text{C}$  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  $d_{11}\text{Bsw}$  from paired  $d_{11}\text{B}$  analyses in planktic and benthic foraminifers from the same core sites and time intervals. The  $d_{13}\text{C}$  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  $d_{11}\text{Bsw}$  estimates, I have to admit that I find the practical results disappointing. Whereas the actual reconstruction results in a 10 permil  $d_{11}\text{Bsw}$  spread

C1

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  $d_{11}\text{Bsw}$  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.

To start from the beginning, the authors use the introduction to discredit all previous  $d_{11}\text{Bsw}$  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  $d_{11}\text{B}$  from modern halites that reflect  $d_{11}\text{Bsw}=39.7$  permil, so they did present a reasonable and indeed promising proof of concept, and this should be acknowledged. I agree with the authors that the implications of the paleo- $d_{11}\text{Bsw}$  reconstructions based on halites are highly unlikely, and the data do not match other estimates. However, coeval halite data are highly consistent in their  $d_{11}\text{Bsw}$ , 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. I would recommend the authors look at the details of Liu et al. 2000 (their figure 6b) and consider the effect of decreasing  $[\text{Ca}]$  concentration throughout the Cenozoic. According to Liu's data this could explain a 3-4 permil difference in  $d_{11}\text{B}$  recorded by the halites, which is somewhat consistent with their much lower  $d_{11}\text{Bsw}$  estimates. More work clearly needs to be done to study the effect of  $[\text{Ca}]$  on halite precipitation, but

C2

there is value in the study of Paris et al., and it should not be discredited so lightly.

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. 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 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)).

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 ar-

C3

guments 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. 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.

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 Kakhana 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 Kakhana 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.

C4

2013). So even if the Kakhana 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.

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.

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.

C5

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  $p\text{CO}_2$  was applied once at 200 ppm and once at 1000ppm, or across the entire range (Line 368). Please list studied ranges for pH,  $\text{CaCO}_3$  saturation states and the biological pump in tabular form.

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.

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  $\sim 0.1$  permil per million years (line 398 in this manuscript).

C6

However, Lemarchand et al. also estimated up to 0.6‰ d11Bsw 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 whatever argument is brought up must be independent of an alternative d11Bsw estimate, or the argument becomes circular. Regardless, given that the smoothing function reduces the potential d11Bsw ranges estimated from this approach from 42-32 permil to ~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.

Similarly, Greenop et al. use the variability observed in their d13C 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 d11Bsw 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 d11Bsw trend, and discuss whether the synergies can help us to further improve our understanding of the boron isotope proxy.

C7

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 d11Bsw 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) d11Bsw estimates.

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 d11Bsw and compare it to the analytical uncertainty of d11B Line 66: please add a reference for the d11B 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 CO2 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 CO2 data used by Ridgwell et al. 2005 is derived using the boron isotope-pH proxy, leading to some circularity in the d11Bsw 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 Line 220: please specify that you assume there is no temperature effect on d11B. Ideally you should calculate d11B borate with and without the temperature effect, and estimate the associated uncertainty on the d11Bsw estimate Line 228/252: please specify whether trace element and d13C analyses were performed on the same foraminifer species and size classes as studied for d11B Line 231: please define "H" Line 249-251/281-286: please specify the influence of temperature on the d13C-DIC estimate Line 259: which benthic species was analyzed 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 Line 269: sp: outline Line 273: leads to a broad Lines 295/296: please rephrase, e.g. "CaCO3 dissolution increases CO3<sup>2-</sup> and pH". The sentence is difficult to read. Line 302: it would be nice to anticipate at this

C8

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  
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?  
Line 416/417: please rephrase to include that the benthic  $\delta^{11}\text{B}$  record is a function of both pH and  $\delta^{11}\text{B}_{\text{sw}}$   
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.  
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"  
Line 1026: sp: colours reflect  
Line 1033: sp: tests  
Line 1049: insert period after "details")  
Fig. 2: Please remove the two lower end Miocene  $\delta^{11}\text{B}_{\text{sw}}$  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.  
Fig. 3: It would be nice to insert a horizontal bar or arrow to indicate the smaller pH gradient in b.  
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*?  
Fig. 8: please indicate core site symbols in this figure. In particular, the large gradient at  $\sim 2$  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.  
Fig. 10: please add both curves from Raitzsch & Hönisch, including their uncertainty estimates  
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.

---

C9

Interactive comment on Clim. Past Discuss., doi:10.5194/cp-2015-177, 2016.

C10