

Review #2

Orbital CO₂ reconstruction using boron isotopes during the late Pleistocene, an assessment of accuracy – de la Vega et al

review by William Gray (william.gray@lsce.ipsl.fr)

Boron isotopes in planktic foraminifera are a widely used proxy for reconstructing ancient CO₂. As other studies have done previously, de la Vega et al use ice core CO₂ as a test of accuracy; to this end, they present new G-IG boron isotope data, including data from a new sediment core site in a location close to equilibrium with the atmosphere today. This is only the second record of its kind (i.e. a record from a site close to equilibrium and over multiple glacial cycles) and is a very welcome addition as it helps overcome the assumption that a single site (i.e. ODP 999) has remained in equilibrium with the atmosphere. They go on to assess the accuracy with which CO₂ can be reconstructed, and possible causes of the (albeit relatively minor) discrepancies (dissolution, second carb system parameter, calcite-borate calibration). Overall, I found the manuscript to be clear, well structured, and well-reasoned. We thank the reviewer for his constructive positive feedback and respond to each comment below.

I have the following suggestions:

I think a greater exploration/discussion of the thermal influence on the carbonate/borate system (via the dissociation constants) in paleo CO₂ reconstruction is warranted – basically, how sensitive is CO₂ to accurate SST reconstruction? More sensitivity tests could be implemented - including keeping temperature constant throughout. Overall, I think further exploration/discussion of the thermal effects are warranted as, my guess is, this could be an important source of bias/uncertainty and it's helpful to understand what we need to improve. The authors mention they also use our iterative pH correction, and briefly mention this in the text, but I think it warrants further discussion; there is very good evidence from culture studies that pH influences Mg/Ca, and this is an influence we know is going to covary with atmospheric CO₂.

The authors mainly rely on the calibration of Anand et al 2003 to derive Mg/Ca SSTs to use in their CO₂ reconstruction. Although widely applied I don't think the calibration of Anand et al accurately describes the relationship between Mg/Ca and temperature (see figures 5 and 6 in Gray et al 2018). This is very apparent if you compare the measured CTD temperatures at the sediment trap site used by Anand, and the temperature calculated using Anand's own Mg/Ca data and their calibration line (see figure below). Using the calibration of Anand et the Mg/Ca are almost always too warm, and the seasonal cycle is about half of what it should be as winter temperatures are 4 degrees too warm (there is no way to explain this by 'sampling issues' as the Mg/Ca SSTs are warmer than any individual CTD measurement ever taken at the site in winter, and this is BATS so it's about the most heavily sampled place in the ocean). Basically, if you use Anand et al on their own data, you get the wrong answer. If the authors want to use Anand et al, I think there needs to be more justification (and 'it gives a better fit to the CO₂ data/the core top SSTs' isn't a great reason).

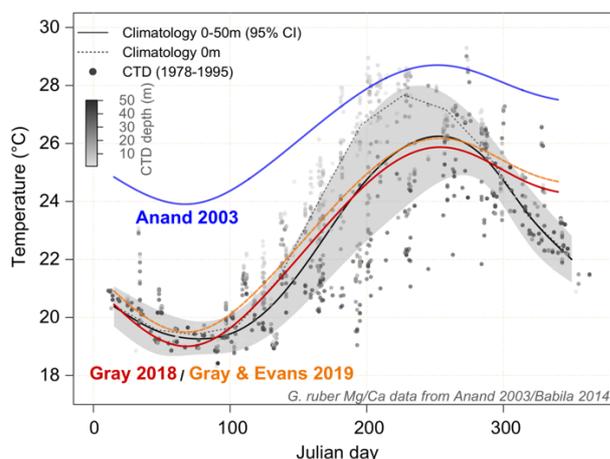


Figure above shows a comparison of Mg/Ca and CTD temperatures at the Sargasso Sea sediment trap site (the site used by Anand to derive their Mg/Ca calibration). I'm showing a LOESS fit to the Mg/Ca data (colored lines), rather than the individual data points to make it legible. The grey dots show all the individual CTD measurements ever taken at this site. The black line/grey shaded area is the expected temperature (and 95% CI) at the habitat depth of *G. ruber*.

We thank the reviewer for this very important point drawing our attention to an alternative Mg/Ca calibration. This point is also mentioned by reviewer 1.

We have now conducted sensitivity tests with seven treatments of Mg/Ca or temperature: (1,2) Gray et al., (2018) with and without a pH correction, (3,4) Gray et al. (2018) using Mg/Ca corrected for depth-dependent dissolution, with and without a pH correction, (5,6) Anand et al. (2003) with and without a depth correction, and (7) constant temperature. The results are displayed in the Figure below (Figure R1)

The differences in SST and resulting CO₂ can be substantial: up to 6 degrees and ~50 ppm, respectively, between the Gray18 calibration uncorrected for pH and the Anand et al (2003) calibration corrected for dissolution. We have chosen the Mg/Ca treatments that accounts for pH effect on Mg/Ca and yields the closest agreement between coretop at both sites and modern temperature from Glodap v2 (Note coretop at 871 is not displayed and the most recent Mg/Ca from Dyez and Ravelo, 2013 was used). This treatments is with a pH correction and Mg/Ca corrected for depth/dissolution.

Choosing this approach is justified considering (1) the strong offset between Anand's multi-species Mg/Ca-T calibration and the more recent *G. ruber* compilation of Gray et al., (2018), as the reviewer notes; (2) the effect of pH correction as shown in Gray et al., (2018) and Gray and Evans (2019); (3) the suggested influence of dissolution on Mg/Ca (Dyez and Ravelo, 2013; Schmidt et al., 2006) and (4) the better agreement between coretop and modern SST at each site when using a pH and depth correction.

A full discussion on the various treatments of Mg/Ca is not the scope of this study but in the revised manuscript we will describe and present the different Mg/Ca calculation and resulting CO₂, and justify our chosen approach. We will note that the choice of Mg/Ca calculation is an added cause of potential CO₂ offsets and the way SST is calculated varies amongst boron isotope studies. However, the updated SST here considering the effect of pH and dissolution on Mg/Ca does not affect the main conclusions of the paper and the δ¹¹B downcore calibration still allows improvement to the fit to ice core CO₂.

Figures and data have been updated to account for the new Mg/Ca treatment and the estimated CO₂ are now slightly higher than previously calculated with an average updated combined offset to the ice core CO₂ of 13 +/- 46 ppm .

The comparison between fragmentation index and the updated CO₂ (pH and depth corrected) in Table R1 shows that periods of high fragments and high CO₂ offset coincide 65 and 75% of the time at core 999 and 871, respectively with the calibration of Henehan et al. (2013), and 5 and 33% of the time with the optimised calibration and pH-corrected Mg/Ca-derived SST. Note that almost all CO₂ offsets are within error (95 % confidence envelope) of the 0 residual line and that between 20 (ODP 999) and 55 % (ODP 871) of the CO₂ offsets are within 10 ppm (Table R1). CO₂ offsets below 10 ppm are effectively zero and not included in the criteria of positive or negative CO₂ offset. This evaluation should therefore be treated with caution.

CO2 points associated with high FRAGMENTS	999 (calib H13)		871 (calib H13)		999 (optimised calib)		871 (optimised calib)	
	Nb of points	% occurrence	Nb of points	% occurrence	Nb of points	% occurrence	Nb of points	% occurrence
negative CO2 offset	3	15	0	0	8	40	3	25
positive CO2 offset	13	65	9	75	1	5	4	33
within +/- 10 ppm	4	20	3	25	11	55	5	42
Total	20		12		20		12	

Table R1. Effect of various δ¹¹B calibration on the direction (positive or negative) of the CO₂ offset to the ice cores.

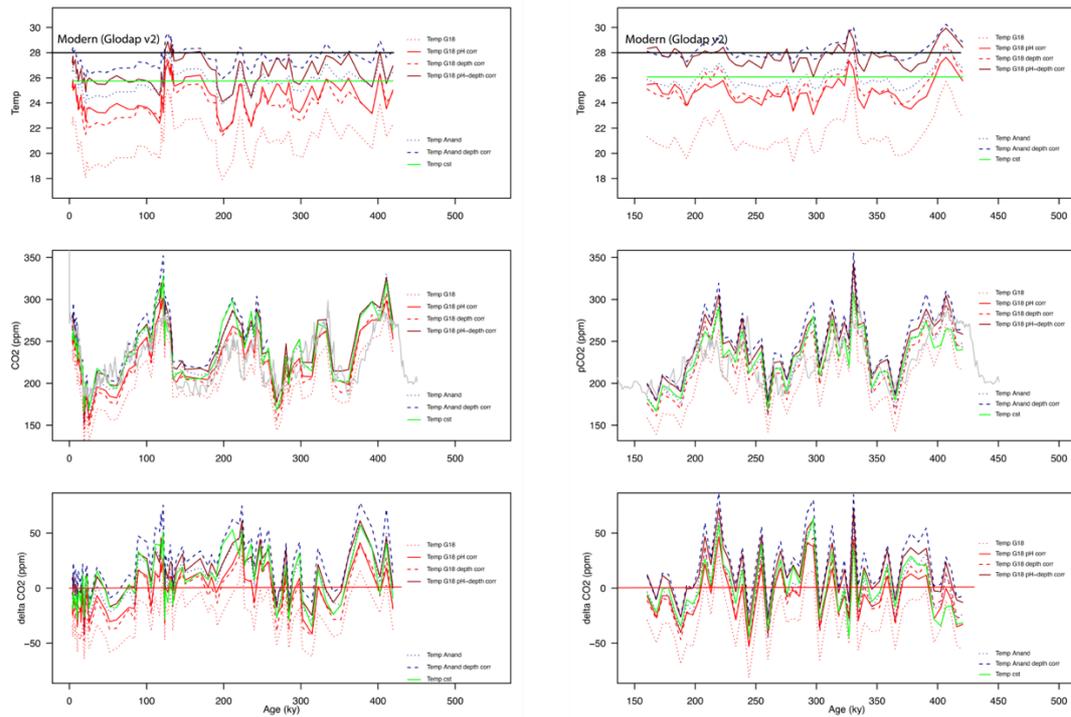


Figure R1. Effect of various temperature treatments (top), on $\delta^{11}\text{B}$ -derived CO_2 (middle) and CO_2 offset to the ice cores (bottom). Left: site 999, right: site 871. Temperature treatments are: G18 (Gray et al., 2018 no correction), G18 pH corr (Gray et al., 2018 with pH correction), G18 depth corr (Gray et al., 2018 with Mg/Ca corrected depth), G18 pH-depth corr (Gray et al., 2018 with depth and pH correction), Anand (Anand et al., 2003, no correction), Anand depth corr (Anand et al., 2003, Mg/Ca corrected for depth), Temp cst (constant temperature of 26°C).

core	Mg/Ca treatment	DpCO2	2sd	core	Mg/Ca treatment	DpCO2	2sd	Average DpCO2 for both records
999	T Gray18	-27.03	40	871	T Gray18	-28.21	52	-27.62
999	T Gray18 pH corr	-3.87	39	871	T Gray18 pH corr	-6.40	47	-5.14
999	T Gray18 depth corr	-6.77	43	871	T Gray18 depth corr	-3.30	54	-5.03
999	T Gray18 pH corr, depth corr	12.06	41	871	T Gray18 pH corr, depth corr	13.89	51	12.98
999	T Anand03	6.64	45	871	T Anand03	1.37	54	4.00
999	T Anand03 depth corr	23.28	48	871	T Anand03 depth corr	20.88	57	22.08
999	T constant	6.33	46	871	T constant	0.23	53	3.28

Table R2. Effect of various Mg/Ca-derived temperature calibrations on CO_2 offset (DpCO2). The line highlighted in green is the chosen updated calibration.

-For the leaching experiment is *T. sacculifer* with or without sacc? This is really important as the lines of argument regarding the differential dissolution of the gametogenic calcite versus the rest of test only hold if it has a sacc (note there are similar arguments for a $\text{d}18\text{O}$ dissolution effect in this species). The authors mention they didn't assess weight loss (a shame), but are there really no TE data for these samples? Was an aliquot not analysed as part of the boron isotope analysis to check for cleaning etc? If TE data are available it would be really informative to show how Mg changes with the leaching, as having some metric to be able compare to real world samples would make these leaching experiments much more useful.

T. sacculifer was picked with sacc-like final chamber but as mention in our response to reviewer 1, we very regretfully could not measure elemental data for *T. sacculifer* due to machine down time for a significant period of time after samples were dissolved. However, weight and elemental data were obtained for the repeat leaching experiment on *G. ruber*. No change in weight and elemental composition were observed after 10 hours of leaching for *G. ruber*. Whilst this shows that indeed, no dissolution has occurred for this species, it shows some relative robustness to dissolution compared to *T. sacculifer* that showed a significant $\delta^{11}\text{B}$ fractionation after 6 hours of leaching. We however acknowledge weight and trace element data for *T. sacculifer* would be a very

welcomed addition. For this reason, we have decided to remove these results from the manuscript and we will aim to complement these data in the future.

-For the uncertainty and the discussion of changing $DpCO_2$ its worth noting that Earth System Models under glacial forcings typically simulate very small changes in $DpCO_2$ – this can be seen in the figures below which show the pH difference minus the mean ocean pH difference in the IPSL model between PI and LGM forcings (taken from Gray and Evans 2019) – basically most of the ocean just reflects the change in atmospheric CO_2 (95% range within ± 0.05 units, equivalent to about ± 40 uatm $DpCO_2$). There is almost no residual pH change at the two sites used in the present study. I think this is really encouraging for paleo CO_2 reconstruction from boron isotopes and I think we could exploit ESMs to understand/quantify the $DpCO_2$ aspect of the calculation. Doing this kind of exercise with a larger ensemble would be really good way to test this central assumption in surface ocean carbonate system based pCO_2 reconstructions in the future.

For the present study, I'd be happy to provide some version of the figures below (perhaps in $DpCO_2$ space, rather than pH) to include in the manuscript (say below Figure 1) if the authors thought it would be helpful and wished to redraw them and include them.

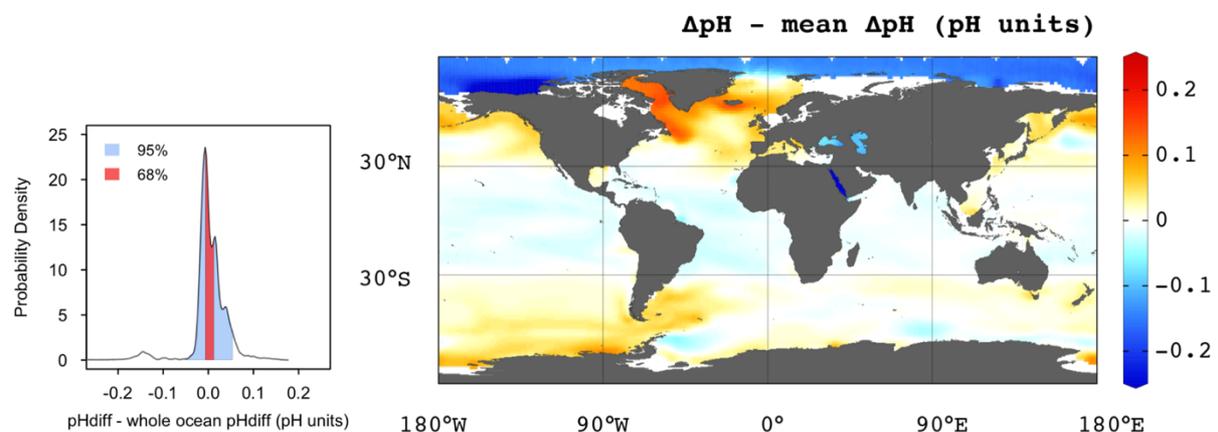


Figure above (from Gray and Evans 2019) pH difference minus the mean ocean pH difference in the IPSL model between PI and LGM forcings. Most of the surface ocean shows very little pH change beyond the impact of changing atmospheric CO_2 , which is reassuring for trying to reconstruct atmospheric CO_2 .

We thank the reviewer for this input. This is an encouraging result that adds confidence to the fundamental tenet that a given site has remained in equilibrium with the atmosphere in the past. We will refer to this published figure when introducing the modern ocean-atmosphere pCO_2 map (Figure 1)

For the second carbonate system parameter the authors use ALK, taking a flat $175 \mu\text{mol/kg}$ range about the modern value – how is this range distributed around modern value? Is it weighted more heavily to higher values to account for likely higher glacial ALK à la Martinez- Boti 2015? They suggest the ALK variations needed would be too large to be sole cause of discrepancy, but we I guess cannot rule out more minor/systematic ALK changes over G-IG cycle cannot explain some part of discrepancy.

The uncertainty of $175 \mu\text{mol/kg}$ around alkalinity is equal on either side of the central value. The figure R2 bellow shows the CO_2 output for TA uncertainty distributed uniformly (black) or weighted more heavily to the higher value (red). The results for both treatment are close and overall show a slightly higher CO_2 offset to the ice core in the case of asymmetrical uncertainty (+ 17 ppm offset on average for both cores, vs. +13 ppm for a uniform uncertainty). This is due to the higher number of simulation in the Monte Carlo that will lie above the central alkalinity value in case of a -25/+75 envelope, and this has the effect of increasing CO_2 .

Whilst the reasoning on a higher uncertainty bond during glacials is sensible (based on model outputs, Toggweiler, 1999; Hain et al., 2010), we think it is more conservative to keep a symmetrical large uncertainty to allow for the uncertain factors in contributing to alkalinity change over G-IG changes (carbonate burial, carbonate pump, soft-tissue pump).

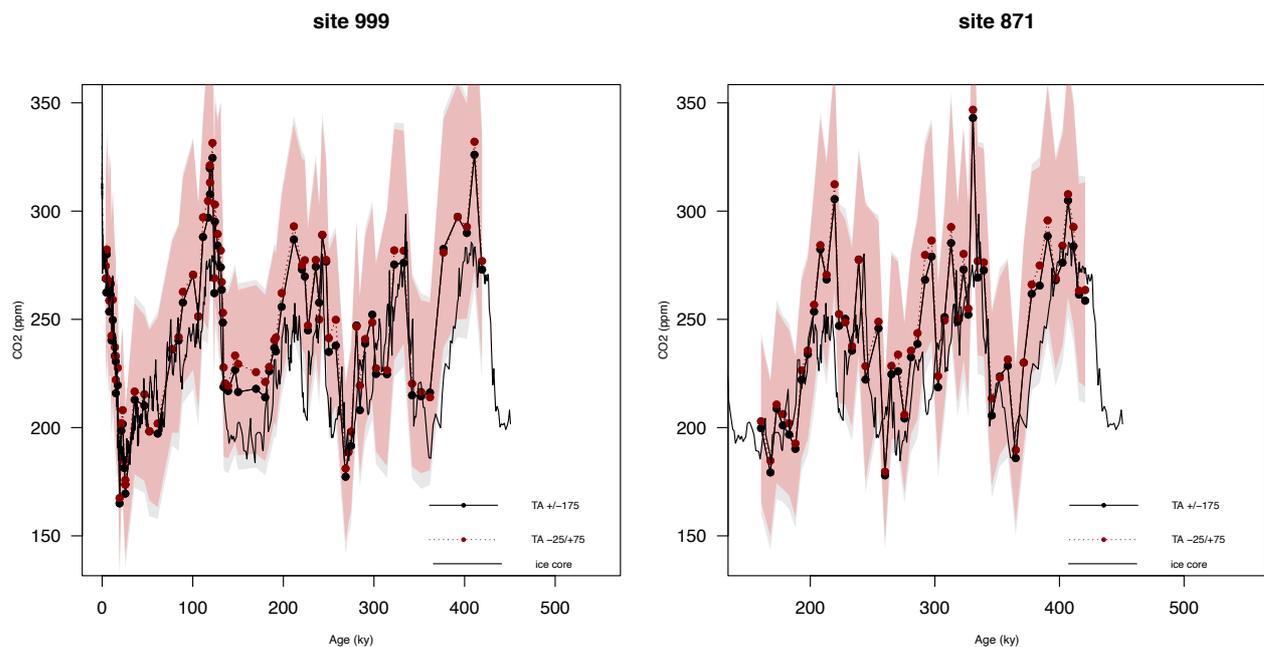


Figure R2. Effect of total alkalinity (TA) uncertainty on reconstructed CO₂, symmetrical +/- 175 μmol/kg (black line and grey envelope) and asymmetrical -25/+75 μmol/kg (dotted red line and pink envelope).

Detailed comments

line 58 – dissolved inorganic carbon (DIC). “Total” has been removed.

line 138 – I think you could sure up this assumption or try to quantify the likely uncertainty in DpCO₂ using ESM model output (see point above). | line 483 – I really think ESMs can be useful here. | line 273 – why not use the ESM output to try to quantify the likely DpCO₂ uncertainty? We will refer to this model output that adds confidence to the assumption that a given site has remained in equilibrium with the atmosphere in the past.

line 153 – Rebotim et al 2017 (biogeosciences) is a good reference for *G. ruber* habitat depth. We thank the reviewer, this reference has been added.

line 198 – is 0.25% the relative concentration of the TPB? It could still have a big effect if it has a funky d11B, so better to report the absolute per mil values of correction. Yes, 0.25% is the relative concentration of the TPB given a typical TPB contribution of 50 pg and a sample size of 20 ng. We however use a long term median of TPBs measured over multiple years (-7 permille) rather than the δ¹¹B of TPB measured on the day of column since a single δ¹¹B measurement has poor confidence given the small amount of boron in a TPB samples (40-100 pg) and the analytical methods used here. We also update a mistake in the manuscript that omitted to mention TPB correction conducted on the new samples from ODP site 999, for which TPB values ranged 70 to 100 pg. This resulted in a TPB fraction of 0.3 to 2.3% of the samples and a δ¹¹B correction of 0.1 to 0.7 permille. This will be added to the text in the revised manuscript.

line 237 – I think using Anand needs some real justification (see discussion/figure above) line 241 – an aside as you are not accounting for dissolution downcore, but it should really be parameterized as a function of Omega rather than depth. | line 246 – I think more details are need of the method you used for the pH corrected approach, and greater discussion of the results are warranted later. | Line 461 – why? We have now included a more complete approach by presenting various treatments of Mg/Ca and resulting CO₂ calculations. See detailed response above.

line 254 – how is the ALK range distributed around modern value. The uncertainty of 175 μmol/kg is distributed equally on either side of the central alkalinity value. This will be specified.

line 201 – with sacc or without sacc? | line 390 – sacc or no sacc? | line 524 – w/ or w/o sacc?

T. sacculifer was measured with sacc, but these leaching test results will be removed (see reply above and response to reviewer 1)

line 545 – really no TE data? Unfortunately not for *T. sacculifer* but available for *G. ruber* showing no changes in elements/Ca for *G. ruber* for different duration of leaching.

line 431 – I'm not sure this is the reason Pacific CaCO₃ preservation increases during glacial, as there is a lot of evidence for reduced ventilation of the deep Pacific in glacial (e.g. Anderson et al 2019). An ALK increase due to the reduction of CaCO₃ burial in the deep Atlantic seems more likely (e.g. Cartapanis et al 2018), but a lot of this discussion seems somewhat superfluous. It's enough to say that the fragmentation should give an indication of relative changes in dissolution and describe the patterns seen. We thank the reviewer for this added information and have rephrased the point about improved preservation during glacial. Even if this is not the main scope of the study we think it is important contextual information and have summarised the text as follow.

“The preservation cycle in the Pacific (Farrell and Prell, 1989), shows improved (poorer) preservation during glacial (interglacial) and this pattern seems to have developed after the mid Pleistocene transition (MPT) (Sexton and Barker, 2012). The origin of these cycles could be a combination of enhanced ventilation during glacial in the Pacific (Sexton and Barker, 2012), or increased burial due to enhanced global alkalinity following a decrease in burial in the Atlantic (Cartapanis et al., 2018). However glacial periods seem to have also been accompanied by a diminution in oxygenation in the deep Pacific (Anderson et al., 2019) that may have impacted preservation.”

line 467 – in Gray et al 2018 we found no systematic Mg/Ca offset between morphotypes across the Indian and Atlantic. This reference has been added.

line 595 – worth noting optimized calibration is still very similar to other planktic calibrations i.e. it doesn't require something radically different to what we might expect. We agree with this statement and will specify this in the text. This highlights that a small change in the slope is enough to improve the fit to the ice core and diminishes the apparent correlation between high fragmentation and CO₂ offset.

line 623 – the disequilibrium pump has garnered a lot of attention recently and is worth noting (e.g. Eggleston and Galbraith, 2018). Thank you for this input, this reference will be added.

figure 1 – better to use red-white-blue colour scheme for positive and negative anomalies. why not also add a map of modelled LGM-PI DpCO₂ differences below with the core sites indicated? The colour scheme will be updated to colourblind-friendly. A reference to the delta pH map of Gray and Evans (2019) will be added.

Figure s14 – why not plot on the other planktic lines to enable comparison with optimized line

This is a good point and the figure has been updated as follow (Figure R3).

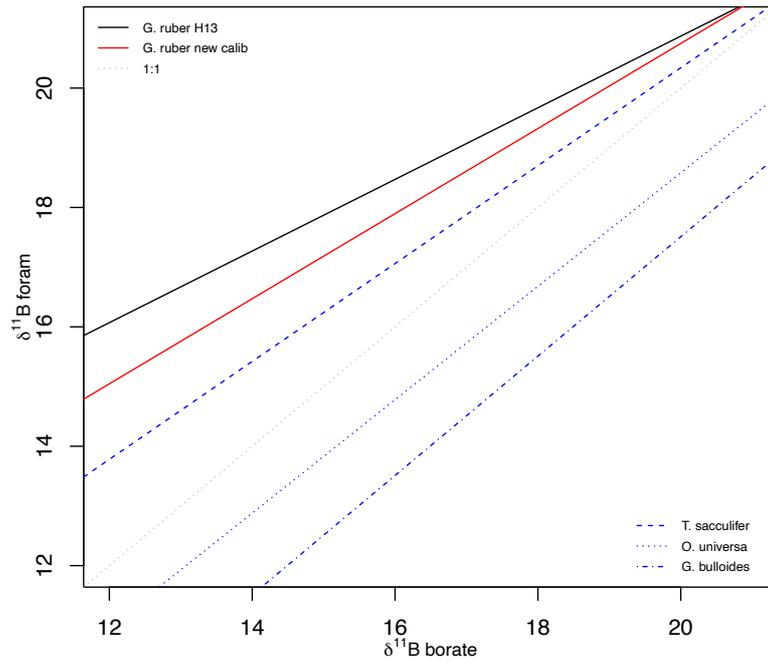


Figure R3. Comparison of the optimised $\delta^{11}\text{B}$ foram/borate calibration with the published calibration slope of Henehan et al. (2013) along with other calibration of spinose foraminifera *T. sacculifer* (Martínez-Botí et al., 2015), *O. universa* (Henehan et al., 2016) and *G. bulloides* (Martínez-Botí et al., 2015)