Author's response

I Point-by-point reply to the reviewers' comments

II Marked-up manuscript version

I Point-by-point reply to the reviewers' comments

Review No 1 (Dennis Mayk)

AC: We thank Dennis Mayk for his constructive and thorough review of our manuscript and addressed all his remarks and suggestions, which are listed below:

Raitzsch et al. present an interesting and timely manuscript about a comparative study of B isotopes in the benthic foraminifera Cibicidoides wuellerstorfi analysed using LA-MC-ICP-MS and SIMS. Despite the relevance of δ11B as a paleo-pH-proxy, very few studies have been published showing intra and inter foraminfera test (shell) δ11B variability as these analysis have proven to be challenging due to low B concentration and fragility of foraminifera tests. This study provides an interesting comparison between different heterogeneity levels within and between individual foraminfera which will be of widespread interest and should be published after revision of the issues listed below:

Main comments:
1. Data processing:
The manuscript lacks a general explanation of how the data were treated after collection. Fig. 2 shows a typical time-resolved laser ablation profile for a clean and a contaminated (clay filled) foraminifer.
In the caption, it is mentioned that some points have been removed from the ablation trend by a 2-sigma outlier test, however in the methods there is no explanation of the data processing involved. It would be important to mention the general data reduction routine that was employed.

AC: Right, the explanation of the data reduction routine was obviously too sparse, and is now complemented by more details (l. 127-132).

Furthermore, the ablation intensity profiles appear very bulgy and do not present apparent plateaus. Please report how the shell signal was extracted from the rest?

AC: The reason for the bulgy shape of the signal intensities in Fig. 2 is probably related to changing ablation efficiency for some samples, i.e. a more efficient ablation of material from a surface progressively getting rougher after a couple of "helix turns". As we always attempted to match the signal, i.e. to gain the same signal intensities between sample and standard, we often had to increase the ablation frequency to enhance ablation at the beginning of a measurement. Conversely, after some time, we often had to decrease the frequency, when ablation was liable to become too strong, resulting in a bulgy profile as shown in Fig. 2. For integration of the shell signal, we chose an interval, where the signal ratio clearly showed a smooth plateau (see Fig. 2), which is now better explained in the revised manuscript (l. 120-126).

2. Sample size estimation:

The estimation of the required sample size to resolve 0.1 pH unit is a very important part of the manuscript but the R function “combn()” used for that purpose lacks a detailed explanation in the manuscript – in addition it is unclear if the presumptions made in the manuscript are correct or lead to an underestimation of the required sample size.
In detail:
On line 237 it is reported that the sample size simulation is based on the assumption that the entire population (P) consists of the 18 shells analysed. Although this holds true for this particular study it is not a representation of the actual (true) population size which is what future studies would be interested in to estimate required sample sizes. In other words, the presumption of P =18 holds only true within this study but has no real world application. Instead it should be discussed what population sizes are realistic within similar pH-environments and simulations should be based on these.
Furthermore, it is not clear how the simulation are carried out using the 18 shells as they have not been measured in the same way according to the Supplemental Material. The “large crater” was analysed on 16 shells and the “umbilical knob” was also analysed on 16 shells suggesting that for the simulation using 18 shells two different measurement “types” were merged which further complicates its validity. It would be more informative to separate the two and report required sample sizes based on measurement type i.e., for measurements on the “umbilical knob” and for “large crater” measurements.

AC: This is a good point raised by Dennis, as only the 14 and 16 individuals were analyzed using the "umbilical knob" and "large crater", respectively. Hence, for the analysis of "sample size requirement" we have excluded the 2 individuals analyzed for inter-chamber variability, and only examine the separate variabilities based on "knob" and "large crater" analyses. This is explained in the main text (l. 246-249). Accordingly, we also modifed Fig. 7 showing the results for each "measurement type".

Considering the sample size of 16 or 18 it appears to be useful to consider the use of a conventional sample size estimation approach in comparison to a resampling approach as drawing from the same small population may result in errors. In the figure below, the estimated sample sizes required for e = 0.5 (2SD) and 1-α = 0.05 in relation to the population size is given as estimated by the R function “sample.size.mean()” (https://CRAN.R-project.org/package=samplingbook) for both measurement “types”. Given an overall population size of e.g., 500 specimens in the same pH-environment, it would require n = 87 specimens if the “umbilical knob” was measured and n = 40 if the “large crater” was measured (based on the variability observed in this study) to achieve the desired significance level. Even if the population size consisted of only 16 individuals, the estimated sample size would be n = 16 and n = 14, respectively and thus twice as large as reported in the manuscript.

AC: Also this is a very important point, which made us to rethink about this issue. Firstly, it is right that our approach is not the most appropriate resampling method, as the entire data population only consists of the 18 measured individuals. The R function "combn()" searches for all possible

combinations within this population and hence does not apply replacement of a sample, i.e. it does not resample one foram multiple times for generating one subsample. This ultimately results in an underestimation of the uncertainty as a function of measured individuals, which was also correctly pointed out by Dennis.
However, we think that the resampling approach proposed by Dennis (R function "sample.size.mean()") is not the most appropriate neither. That is because this function assumes that the value of a measurand from the entire population may be approached by measuring a random subsample, the size of which is dependent on the population size and the target uncertainty. In other words, this function allows for determining the required subsample size in order to gain the "true" average value of the entire population to within a quoted uncertainty, but it does not reflect whether the average value accurately records the influencing variable, in our case pH that influences δ11B. The output plot provided by Dennis implies that a few individuals are sufficient to gain an accurate value, if the population is small, but this is not true as each specimen has a large uncertainty in terms of the closeness of the agreement between the measured δ11B and the influencing pH. Consequently, the relationship between the "accuracy" and number of analyzed individuals must be independent of the population size.

Based on these thoughts, we modified our method, but which partly goes the same direction as the resampling approach suggested by Dennis. In the revised manuscript, we generated Monte Carlo simulations, where a large artificial population (n=10,000) is created by randomly generating δ11B values around the "true" δ11B value within the determined individual uncertainty of ±0.84 and ±1.38 ‰ (SD) for "large craters" and "umbilical knobs", respectively. On these populations, we applied the R function "combn()" for randomly resampled δ11B values to determine the 2SD uncertainty (= the potential error of δ11B) as a function of n (1 to 16) analyzed individuals. Interestingly, we come to the same required sample sizes as Dennis with his approach for the quoted uncertainty of ±0.5 permil, but is more realistic for fewer shells, as it is independent of the original population size. We have partly rewritten/complemented the text (l. 246-258) and redrawn Fig. 7, which is now without histograms.

Minor comments:
This is a non-comprehensive list of minor issues
Line 30: Consider removing the last clause of the abstract. "Vital effect" is a loaded term and since it is not further discussed (Line 185) of little value for this manuscript.

AC: Correct, removed (l. 29).

Line 35: Space missing between 27.2 and ±0.6 ‰

AC: Corrected (l. 34).

Line 57: Comma missing after "Also"

AC: Corrected (l. 56).

Line 69: Considering that this study looked at a total of 23 specimens the term “tens of specimens” seems excessive, better report the actual number of individuals.

AC: Right, changed (l. 68).

Line 179: Why was a non-parametric test used? Please specify what data the test was used on? Please report the Wilcoxon-Mann-Whitney test summary i.e., (W = XXX, p <0.001)

AC: We used a non-parametric test because it does not imply defined probability distributions a priori, but is open to the model structure. The WMW uses randomly selected values X and Y from two populations, and tests the null hypothesis whether the probability that X>Y is equal to the probability that Y>X. However, I am very happy that Dennis made this a subject of discussion, since I walked right into a trap when testing the null hypotheses. Because of the few datapoints for each chamber, I applied the statistical test on Monte-Carlo simulated d11B values (n=10) around quoted uncertainties, yielding p values smaller than 0.05, meaning that the differences in d11B between chambers f-1 and f-5 are statistically significant at a 95 % SL. This artificially increased population size, however, led to a biased uncertainty estimation, which was also subject to papers in mathematical journals (e.g. Lin et al. (2013), Too Big to Fail: Large Samples and the p-Value Problem, http://dx.doi.org/10.1287/isre.2013.0480). If just the original data are taken into account, both the WMW and Welch t-test yield p values ~0.07, and hence the d11B differences between the chambers are not statistically different at a 95 % SL, based on the small datasets of this study. I have to apologize for this incautios and naive application of statistical tests on our data. The text is corrected in the revised manuscript (l. 186-188).

Line 184: Space missing between “large-scale” and “suggesting”

AC: "large-scale" was a leftover from a former modfied sentence. Deleted (l. 191).

Line 186: “Somewhat” not useful, report how much δ11B was elevated in the umbilical knob

AC: The average elevation is ~0.5 ‰. Information added (l. 195).

Line 197: a total of 18 shells “were” used

AC: Corrected (l. 205).

Review No 2 (Lubos Polerecky)

AC: Lubos Polerecky's very helpful comments and suggestions, particularly from his analytical view, on our manuscript are very much appreciated and are all addressed in the revised manuscript (listed below).

Raitzsch et al. provide a detailed comparison between SIMS and LA-MC-ICPMS measurements of delta-11B in individual shells of benthic foraminifera. They show that intra-shell and inter-shell variability is significantly lower for the LA-based technique compared with SIMS, which they attribute to the larger volume sampled by the LA-based technique. Importantly, they show that both techniques yield very similar "average" values to those obtained by the traditional bulk measurements based on dissolved specimens. They conclude that the traditional bulk-based analysis is still the preferred approach for paleo applications, but demonstrate clearly the advantages and limits of the microanalytical techniques. The manuscript is well written and clearly organized. Also the figures are clear and of excellent quality. I only have a few minor comments and questions. I recommend the manuscript for publication after these minor issues have been resolved by the authors.

Technical comments/questions/suggestions:
l.69: Please formulate more clearly the *aim* of the study. 'What' do you want to achieve, and especially 'why'?

AC: That's right, the aim of the study was not clearly enough explained, but just vaguely outlined between l. 38-47. This is now more emphasized (l. 68-70).

l.154-155: Please clarify how this variability was calculated. Since 2*sigma is used, it may be confused with 2*sigma of the individual measurement's precision. And since permil units are used, it may be confused with the coefficient of variation (which is in percent). To avoid confusion, best would be to clarify in one sentence that 2*sigma here actually corresponds to 2*SD of n individual measurements (if I understand it correctly). Or is it 2*SE (standard error)?

AC: Yes, the reported 2*sigma variability is the 2-fold standard deviation, derived from the individual measurements. This is not to be confused with the measurement uncertainty (=precision), which is dependent on the ablation time and is also given as 2*sigma (=2*SD). So we use sigma as a statistical expression just to clarify that the SD is reported, and not the SE. We added a short notion in this sentence (l. 162).

l.157: unclear why such inistinguishability should affect variation in measured data. Please explain, or provide an alternative explanation.

AC: Right, this part is quite confusing and probably also not reflecting the truth. It is true that we observed the largest variability in the knob area that might be attributed to a signal mixture from multiple juvenile chambers, but it may also be due to the higher number of measurements compared to the other chambers. In addition, there were also similarly large variabilities found in chambers f-8 and f-9. So we have rephrased the according sentence (l. 164-166). Thanks for hinting at this inconsistency.

l.168-169: Please clarify how this was derived/deduced. Intuitively it is expected that variability in measurements is lower if larger volume is sampled. But it is unclear how you arrived to those values (e.g. ~0.3 permil).

AC: Thanks, we missed to describe how we calculated this. We simply applied the following function to estimate the variability reduction: u(V2) = u(V1)/sqr(V2/V1), where u(V1) is the variability for a quoted volume, and V1 and V2 represent the two different volumes that are compared. We have added this information as a short equation in parentheses (l. 177). Irrespective of this, in section 3.1, we encountered a few slightly wrong numbers related to variability, which were corrected (see track changes), but do not affect any conclusion.

l.176/fig.4: Please clarify representation of the data in polar plots in Fig. 4. I understand that the "phi" coordinate corresponds to the chamber, but it took me a while to figure out that the r-coordinate (scale -7 to 3) corresponds to Delta-delta11B. Also I am wondering whether it would be more beneficial/transparent to show each Delta-delta-11B datapoint rather than average Delta-delta-11B deviations derived from measurements of multiple specimens. Averages may be misleading, as we know.

AC: I agree, the so-called coxcomb chart seems to be difficult to read at the beginning, but once it's understood, it is a nice way of presenting such data. In the updated figure, it should now be clearer that red represents positive and blue negative Delta-delta values. I have tried a couple graph types, also plotting all individual datapoints, as Lubos suggested, but all resulted in quite confusing graphs due to the large number of measurements (at least for SIMS). For our aim to eventually observe trends, plotting averaged deviations seems to be the catchiest way.

Did you test whether the decreasing trend between f and f-5 is significant, or you can only state "the deviation tends to decrease"?

AC: Yes, we applied the Wilcoxon-Mann-Whitney approach to test the significance of d11B differences between chambers (l. 178 ff). Also, Dennis Mayk made this a subject of discussion, and I realized that I have walked right into a trap when testing the null hypotheses. Because of the few datapoints for each chamber, I applied the statistical test on Monte-Carlo simulated d11B values (n=10) around quoted uncertainties, yielding p values smaller than 0.05, meaning that the differences in d11B between chambers f-1 and f-5 are statistically significant at a 95 % SL. This

artificially increased population size, however, led to a biased uncertainty estimation, which was also subject to papers in mathematical journals (e.g. Lin et al. (2013), Too Big to Fail: Large Samples and the p-Value Problem, http://dx.doi.org/10.1287/isre.2013.0480). If just the original data are taken into account, both the WMW and Welch t-test yield p values ~0.07, and hence the d11B differences between the chambers are not statistically different at a 95 % SL, based on the small datasets of this study. I have to apologize for this incautios and naive application of statistical tests on our data. The text is corrected in the revised manuscript (l. 186-188).

l.200: I am wondering why the authors report median instead of, for example, the mean? If it leads to a different mean in comparison to the bulk-based analysis, it should be discussed why such a difference exists. In any case, I think it needs to be clarified why median was used. Similar on l.254.

AC: The differences between medians and means are small, e.g. the SIMS median is 16.08 ‰ and the mean is 16.19 ‰, while the LA median is 15.91 ‰ and the mean is 16.17 ‰. However, the median is less sensitive against outliers than the mean and also represents the average value of a non-uniform distribution. The median is also equivalent to the average shown by the violin/box plots in Fig. 5. We have inserted a sentence why we chose the median instead of the mean (l. 206-207).

l.211: yes, I agree, but it would be useful to expand this argument towards the *source* of this variability (e.g., shell-to-shell differences in the intra-shell heterogeneity?).

AC: We think this goes slightly beyond the scope of this study, and is difficult to answer, based on our database. Maybe the range of isotopic compositions of the trigonal BO_3 hosted in the calcite lattice is very large (not yet examined on the molecular scale), and hence better resolved the smaller the scale of the analytical technique is.

Figure 1: It is rather confusing to see signals for 10B and 11B centred on the same mass (10.25). Is it really so? And why? I am not familiar with the Daly detector principle.

AC: On a Nu Instruments multicollector ICP-MS, the Quad lenses are tuned in a way that the peaks of different isotopes (here 10B and 11B) coincide, i.e. the incoming ions hit the respective detectors simultaneously, where the one for high mass (11B) is on either and the low mass (10B) on the other side of the Center cup. Once the peaks coincide and the peak center is set, the information on the position of the peak center is read by the Center cup. Of course, 11B is measured on 11.009 amu and 10B on 10.013 amu, but the position of the peak center is in relation to the Center cup, and might slightly change on a daily basis, depending on the tuning parameters. We slightly modified the figure caption to clarify that the coincidence of 10B and 11B peaks appears at ~10.25 amu in the center cup.

Figure 3: Please verify the expression for 2*sigma in the graph (in red). First, the factor 1000 does not make sense if cps is in counts per second (perhaps it does if it is in kilo-counts per second). Second, if I substitute 300,000 and 300,000/4.9 for 11B and 10B, I get a factor 8.8, not 8.2. In my opinion, the formula should read as 2*sqrt(1/counts(11B)+1/counts(10B)), where counts(11B)=cps(11B)*1s*n and similarly for 10B. This is a formula for the Poisson error of 11B/10B based on counting statistics. In this formula the factor is then 0.00887 at B11=300,000. Please verify cps vs. kcps.

AC: The factor 1000 is because the boron isotopic composition is given in permil (see eq. 1 in the main text), so it's expressing the relative difference from the standard value. The measurement uncertainty (i.e. the internal precision) must hence be multiplied with 1000 to have the number in permil as well. We agree that the formula we provided is quite cumbersome, but it is mathematically correct. Based on the simpler formula provided by Lubos, we slightly modifed it to more easily enter the number of cycles (n), and multiply if with the factor 1000 to obtain the result in permil expression. The final formula is now 2*sqrt(1/cps(11B)+1/cps(10B)) / sqrt(n)*1000, which replaces the old one in the revised figure. Concerning the obtained factor at a countrate of 300,000 cps for 11B, the ratio between 11B and 10B is in nature in the order of 4, and not 4.9 (11B=80.1 %, 10B=19.9 %). Therefore, if 11B is measured at 300,000 cps, 10B is recorded at approximately 75,000 cps, thus giving a factor of 8.8 using the formula above. We added the expected counrate on 10B in the revised Fig. 3.

Editorial suggestions:
l.24: unclear why the word "presumably" is used in the abstract. It would help if the sentence is reworded to clarify what is certain and what is not (i.e., what is estimated).

AC: Right, "presumably" is a too careful term. Replaced with "estimated to be" (l. 23).

l.39: would be useful to cite few examples of such studies.

AC: These are the same references as in lines 32-33, but they are now listed here as well (l. 38-39).

l.104: perhaps it should read "45 cm3 *and* ablated"?

AC: Thanks, corrected (l. 105).

l.126: remove "and" before "that"

AC: Removed (l. 133).

In caption to Fig. 4, it should read "inset", not "inlet". Similar on l.195.

AC: Corrected, the same misspelling was in the figure captions (l. 198, 203, 219, 313, and 327).

l.184: remove "large-scale"

AC: Thanks, deleted (l. 191).

l.279: "French" - uppercase F.

AC: Corrected (l. 294).

Interactive comment (Kaoru Kubota)

AC: We appreciate the interactive comment of Kaoru Kubota and his careful reading of our manuscript, and will address his comments below:

Very nice work! It will be a great contribution to the community.

AC: Thank you.

Line 109: Should be 11B/10B?

AC: Well spotted! Corrected (l. 112).

Line: 184 Delete "large-scale"

AC: Thanks, deleted (l. 191).

Lines: 235-243: The readers may want to know more detail on the simulation.

Figure 7: It is interesting attempt, but I could not understand how it is simulated. If n = 4, count should be 52? (13*4) Why so much count is obtained in this simulation?

AC: Yes, it is an interesting approach, but it is not the most appropriate one, as also Dennis Mayk suggested, because it results in an underestimation of the uncertainties. We will thus apply the same R function, but on Monte Carlo simulated data sets in the revised manuscript, which gives similar results, but with correct uncertainty estimations (l. 246-258).

However, just for information on the combn() function. It uses an input population (e.g., A, B, C, D) and calculates the averages for all possible combinations among this population (k), for a given number of subsamples (n). For instance, for n=2 it calculates the averages AB, AC, AD, BC, BD, CD, so we get 6 possible combinations. To calculate the number of possible combinations (N) for any k and n, the binomial coefficient is used: $N=k!/(n!*(k-n)!)$. The possible combinations of 4 samples out of a total of 18, as in Kaoru's example, thus amount to 3060.

II Marked-up manuscript version

Technical Note: Single-shell $\delta^{11}B$ analysis of *Cibicidoides wuellerstorfi* using femtosecond laser ablation MC-ICPMS and secondary ion mass spectrometry

Markus Raitzsch[1,2,3], Claire Rollion-Bard[4], Ingo Horn[1], Grit Steinhoefel[2], Albert Benthien[2], Klaus-Uwe Richter[2], Matthieu Buisson[4], Pascale Louvat[4], Jelle Bijma[2]

[1]Institut für Mineralogie, Leibniz Universität Hannover, Callinstraße 3, 30167 Hannover, Germany
[2]Alfred-Wegener-Institut, Helmholtz-Zentrum für Polar- und Meeresforschung, Am Handelshafen 12, 27570 Bremerhaven, Germany
[3]MARUM - Zentrum für Marine Umweltwissenschaften, Universität Bremen, Leobener Straße 8, 28359 Bremen, Germany
[4]Université de Paris, Institut de physique du globe de Paris, CNRS, F-75005 Paris, France

Correspondence to: Markus Raitzsch (mraitzsch@marum.de)

Abstract. The boron isotopic composition ($\delta^{11}B$) of benthic foraminifera provides a valuable tool to reconstruct past deep-water pH. As the abundance of monospecific species might be limited in sediments, microanalytical techniques can help to overcome this problem, but such studies on benthic foraminiferal $\delta^{11}B$ are sparse. In addition, microanalytics provide information on the distribution of $\delta^{11}B$ at high spatial resolution to increase the knowledge of e.g. biomineralization processes. For this study, we investigated the intra- and inter-shell $\delta^{11}B$ variability of the epibenthic species *Cibicidoides wuellerstorfi*, which is widely used in paleoceanography, by secondary ion mass spectrometry (SIMS) and femtosecond laser ablation multicollector inductively coupled plasma mass spectrometry (LA-MC-ICPMS). While the average $\delta^{11}B$ values obtained from these different techniques agree remarkably well with bulk solution values to within ±0.1 ‰, a relatively large intra-shell variability was observed. Based on multiple measurements within single shells, the SIMS and LA data suggest median variations of 4.8 ‰ and 1.3 ‰ (2σ), respectively, where the larger spread for SIMS is attributed to the smaller volume of calcite being analyzed in each run. When analytical uncertainties and volume-dependent differences in $\delta^{11}B$ variations are taken into account for these methods, the intra-shell variability is estimated to be~~presumably~~ in the order of ~3 ‰ and ~0.4 ‰ (2σ) on a ~20 µm and 100 µm scale, respectively. In comparison, the $\delta^{11}B$ variability between shells exhibits a total range of ~3 ‰ for both techniques, suggesting that several shells need to be analyzed for accurate mean $\delta^{11}B$ values. Based on a simple resampling method, we conclude that ~12~~7~~ shells of *C. wuellerstorfi* must be analyzed using LA-MC-ICPMS to obtain an accurate average value within ±0.5 ‰ (2σ) to resolve pH variations of ~0.1. Based on our findings, we suggest to prefer the conventional bulk solution MC-ICPMS over the in-situ methods for e.g. paleo-pH studies. However, SIMS and LA provide powerful tools for high-resolution paleoreconstructions, or for investigating ontogenetic trends in $\delta^{11}B$~~, possibly due to "vital effects" during chamber formation~~.

1 Introduction

The boron isotopic composition ($\delta^{11}B$) of benthic foraminifera has been used to reconstruct deep-water pH (Hönisch et al., 2008; Rae et al., 2011; Raitzsch et al., 2020~~subm.~~; Yu et al., 2010) and to estimate the Cenozoic evolution of seawater $\delta^{11}B$ (Raitzsch and Hönisch, 2013). The underlying mechanism behind the boron isotope method lies in the constant equilibrium fractionation of 27.2 ± ~~±~~0.6 ‰ between the pH-dependent speciation of trigonal boric acid and the tetrahedral borate in seawater (Klochko et al., 2006), where only the borate ion is incorporated into the foraminifera test (Branson et al., 2015; Hemming and Hanson, 1992).

However, while the number of studies on planktonic foraminiferal $\delta^{11}B$ to estimate surface-ocean pH has rapidly increased within the last decade, deep-sea pH reconstructions based on benthic foraminifera are relatively rare (Hönisch et al., 2008; Rae et al., 2011; Raitzsch et al., 2020; Yu et al., 2010). Possible reasons for this might be the lower abundance of benthic foraminifera, compared to planktonic species, and a limited selection of species that truly record bottom-water, rather than pore-water conditions (Rae et al., 2011). Fortunately, there are two suitable candidates, *Cibicidoides wuellerstorfi* and *Cibicidoides mundulus*, that cover a relatively large oceanographic and stratigraphic range, and which have a high boron content of ~12-27 ppm (Raitzsch et al., 2011; Yu and Elderfield, 2007). Although their high [B] may partly compensate for the low abundance in the sediments, in many cases the availability of enough specimens for $\delta^{11}B$ analysis remains limiting.

~~However, while the number of studies on planktonic foraminiferal $\delta^{11}B$ to estimate surface-ocean pH has rapidly increased within the last decade, deep-sea pH reconstructions based on benthic foraminifera are relatively rare. Possible reasons for this might be the lower abundance of benthic foraminifera, compared to planktonic species, and a limited selection of species that truly record bottom-water, rather than pore-water conditions (Rae et al., 2011). Fortunately, there are two suitable candidates, *Cibicidoides wuellerstorfi* and *Cibicidoides mundulus*, that cover a relatively large oceanographic and stratigraphic range, and which have a high boron content of ~12-27 ppm (Raitzsch et al., 2011; Yu and Elderfield, 2007). Although their high [B] may partly compensate for the low abundance in the sediments, in many cases the availability of enough specimens for $\delta^{11}B$ analysis remains limiting.~~

Here, microanalytical techniques such as laser ablation multicollector inductively coupled plasma mass spectrometry (LA-MC-ICPMS) and secondary ion mass spectrometry (SIMS) can help to overcome the problem of sample limitation. These techniques have already been successfully used for a variety of biogenic carbonates to gain information on biomineralization processes or seasonal pH variations (e.g., Blamart et al, 2007; Fietzke et al, 2015; Howes et al., 2017; Kaczmarek et al., 2015a; Mayk et al., 2020; Rollion-Bard and Erez, 2010; Sadekov et al., 2019). On the other hand, microanalytical analysis of $\delta^{11}B$ is usually afflicted with larger uncertainties in terms of repeatability and reproducibility, as well as of natural $\delta^{11}B$ heterogeneity within single shells and within a population. In addition, some recent studies using LA-MC-ICPMS suggest correction modes for measured $\delta^{11}B$ values because detected interferences on the ^{10}B peak, possibly due to scattered Ca ions from the carbonate sample, can result in large offsets from the expected value (Thil et al, 2016; Sadekov et al., 2019;

Standish et al., 2019), whereas in other studies this matrix-induced effect was not observed (Fietzke et al., 2010; Kaczmarek et al., 2015b; Mayk et al., 2020).

Also, the reported analytical reproducibility for $\delta^{11}B$ in biogenic carbonate using LA-MC-ICPMS differs considerably among different studies, ranging between ±0.22 and 1.60 ‰ (2σ =2 standard deviations), determined from repeated measurements of either a carbonate or glass standard (Fietzke et al., 2010; Kaczmarek et al., 2015b; Mayk et al., 2020; Sadekov et al., 2019; Standish et al., 2019; Thil et al., 2016). As there is no standardized protocol nor a commercially available homogenized $\delta^{11}B$ carbonate standard for determining the analytical uncertainty of LA-MC-ICPMS, this issue remains the most challenging task to compare the different labs and instruments. The most commonly used carbonate standards with well-constrained boron isotopic compositions are samples from a coral (JCp-1) and a giant clam (JCt-1), provided by the *Geological Survey of Japan* (e.g., Inoue et al., 2004; Okai et al., 2004). However, for microanalytical analysis the standard is usually powdered in a mortar and finally pressed to a pellet, which is produced individually in each laboratory, thus potentially resulting in different heterogeneities (e.g., through different grain sizes or applied pressures) in each pellet. This issue is also true for SIMS analyses, and the reported reproducibility is strongly linked to the in-house reference material used (e.g., Kaseman et al., 2009; Rollion-Bard and Blamart, 2014).

In this study, we investigate a ~~how $\delta^{11}B$ in *C. wuellerstorfi* varies within single shells and between shells of a~~ population of 23 ~~tens of~~ specimens of *C. wuellerstorfi*, which is a widely used benthic foraminifer species in paleoceanographic studies, to extend our knowledge of $\delta^{11}B$ variability within and between individuals. The aim of our study is to demonstrate the capabilities and limitations of $\delta^{11}B$ analyses in *C. wuellerstorfi* on a microscale. For this purpose, we used the femtosecond LA-MC-ICPMS and SIMS techniques and compared the results with bulk-solution MC-ICPMS. Finally, we examine the size of population required for targeted $\delta^{11}B$ uncertainty levels in paleoceanographic studies using LA-MC-ICPMS.

2 Material and Methods

2.1 Foraminifer samples

For this study, we used sediment samples from GeoB core 1032-3, taken in the Angola Basin on the Walvis Ridge at a water depth of 2505 m. From a Holocene interval (6-8 cm, 5.6 ka), 23 pristine (glassy) shells of the benthic foraminifer species *C. wuellerstorfi* from the size fraction >350 µm were picked and prepared for subsequent microanalytical analysis. Five large specimens (>400 µm) were embedded in epoxy and polished down to a planar surface for SIMS analyses, while the remaining 18 specimens were mounted on carbon tape for LA measurements. From these 18 individuals, two large tests were analyzed for detailed chamber-to-chamber variability, while the remaining 16 tests were used to measure quasi-bulk $\delta^{11}B$ by ablating large shell areas, preceded by measurements of the smaller umbilical knob area.

2.2 Secondary ion mass spectrometry

For the ion microprobe analyses, we used the same technique as described in Rollion-Bard et al. (2003) and Blamart et al. (2007). Boron isotopic compositions were measured with the Cameca ims 1270 ion microprobe at CRPG-CNRS, Nancy, France. A primary beam of $^{16}O^-$ ions generated using a Radio Frequency Plasma source (Malherbe et al, 2016) with an intensity of 50 nA was focused to a spot of about 20 µm. A mass resolution of 3000 was used for B isotope analyses, allowing the elimination of all isobaric interferences. Boron isotopes were analyzed in mono-collection mode using the central electron multiplier. The dead time of the electron multiplier was determined before the analytical session and set to 65 ns. A pre-sputtering of 120 s was applied before the analysis itself. The typical intensities of $^{11}B^+$ in foraminifer tests were between 2000 and 4500 counts per second (cps), depending on the boron concentration. The analysis consists of 60 cycles of 10 s for $^{10}B^+$ and 6 s for $^{11}B^+$, respectively. The reference material was a calcium carbonate with a B concentration of 22 ppm and a $\delta^{11}B$ of 16.76 ± 0.11 ‰, relative to the standard reference material (SRM) NIST 951 (WP22, value determined at IPGP using the method of Louvat et al, 2014). The reproducibility, as estimated by multiple measurements of the reference material, was 2.48 ‰ (2σ, n=8), and is very close to the predicted 2σ uncertainty derived from counting statistics.

2.3 Femtosecond laser ablation MC-ICPMS

Boron isotope measurements were performed using a customized UV-femtosecond laser ablation system coupled to a Plasma II MC-ICPMS (Nu Instruments) at the AWI, Bremerhaven. The laser ablation system is based on a Ti-sapphire regenerative amplifier system (Solstice, Spectra-Physics, USA) operating at the fundamental wavelength of 775 nm with a pulse width of 100 fs and pulse energy of 3.5 mJ/pulse. Consecutive frequency conversion results in an output beam with a wavelength in the UV spectra (193 nm) and a pulse energy of 0.08 mJ. The short femtosecond pulses were shown to have major advantages over nanosecond pulses for a wide range of element and isotope ratios with respect to laser-induced and particle-size-related fractionation, thus enabling non-matrix-matched calibrations (e.g., Horn and von Blanckenburg, 2007; Steinhoefel et al., 2009).

The sample and standard materials were mounted in an ablation chamber with an active volume of ca. 45 cm^3 and ablated in a helix-mode scan at a speed of 2 mm s^{-1} by using a laser spot size of ~40 µm. This technique allows producing ablation craters of almost any diameter, in this study ranging from ~80 µm for analysis of single-chamber to ~400 µm to cover whole shells. The aerosol was transported via a He gas flow (~0.5 L/min) and admixed with Ar gas (~0.5 L/min) before entering the MC-ICP-MS. Torch position, ion optics and gas flows were optimized to gain maximum signal intensity and stability on ^{10}B and ^{11}B peaks. The mass spectrometer was equipped with standard Ni sample and skimmer cones for dry plasma conditions. The radio frequency power was set to 1300 W. Boron isotopes were determined on Daly detectors, where high-mass D5 was used for ^{11}B and D0 for ^{10}B. Each measured sample $^{11}B/^{10}B$ was normalized to $^{11}B/^{10}B$ measurements of the glass standard NIST SRM 610 ($\delta^{11}B$=0 ‰ NBS 951), using the Standard-Sample-Bracketing technique. The analyses were performed at low mass resolution (M/ΔM ~2000, 5‰), which was sufficient to resolve all interferences.

We performed mass scans on the peaks of ^{10}B and ^{11}B for both gas blanks (laser off) and measurements on carbonate (laser on) (Fig. 1) to investigate possible effects by scattered ions of matrix elements as observed in some recent studies (Sadekov et al., 2019; Standish et al., 2019). For our set-up, we can exclude such matrix-induced effects, which is in line with Fietzke et al. (2010) and Mayk et al. (2020). Hence, there was no need to correct the raw LA data as done in the recent studies by Sadekov et al. (2019) and Standish et al. (2019). Before analysis, sample and standard materials were pre-ablated to remove potential surface contamination. Laser repetition rates ranged between 12 and 60 Hz to match the signal intensity between carbonate samples and standard material NIST SRM 610 (~300,000 cps on ^{11}B). As ablation efficiency and hence signal intensity may vary with progressively increasing surface roughness and crater depth, we adjusted the repetition rate, if required, to target intensity matching between sample and standard. Whereas this approach could result in bulgy time-resolved isotope signals, as shown in Fig. 2, clean calcium carbonate was identified from a plateau-like $^{11}B/^{10}B$ signal. Conversely, any contaminated phase from partial ablation of clay infillings, indicated by dropping $^{11}B/^{10}B$ ratios accompanied by rising [B], were excluded from further data treatment (Fig. 2).

~~Boron isotope measurements were performed using a customized UV-femtosecond laser ablation system coupled to a Plasma II MC-ICPMS (Nu Instruments) at the AWI, Bremerhaven. The laser ablation system is based on a Ti-sapphire regenerative amplifier system (Solstice, Spectra-Physics, USA) operating at the fundamental wavelength of 775 nm with a pulse width of 100 fs and pulse energy of 3.5 mJ/pulse. Consecutive frequency conversion results in an output beam with a wavelength in the UV spectra (193 nm) and a pulse energy of 0.08 mJ. The short femtosecond pulses were shown to have major advantages over nanosecond pulses for a wide range of element and isotope ratios with respect to laser-induced and particle-size-related fractionation, thus enabling non-matrix-matched calibrations (e.g., Horn and von Blanckenburg, 2007; Steinhoefel et al., 2009). The sample and standard materials were mounted in an ablation chamber with an active volume of ca. 45 cm^3 ablated in a helix-mode scan at a speed of 2 mm s^{-1} by using a laser spot size of ~40 µm. This technique allows producing ablation craters of almost any diameter, in this study ranging from ~80 µm for analysis of single-chamber to ~400 µm to cover whole shells. The aerosol was transported via a He gas flow (~0.5 L/min) and admixed with Ar gas (~0.5 L/min) before entering the MC-ICP-MS. The mass spectrometer was equipped with standard Ni sample and skimmer cones for dry plasma conditions. The radio frequency power was set to 1300 W. Boron isotopes were determined on Daly detectors, where high-mass D5 was used for ^{11}B and D0 for ^{10}B. Each measured sample $^{10}B/^{11}B$ was normalized to $^{10}B/^{11}B$ measurements of the glass standard NIST SRM 610 ($\delta^{11}B$=0 ‰ NBS 951), using the Standard-Sample-Bracketing technique. The analyses were performed at low mass resolution (M/ΔM ~ 2000, 5%), which was sufficient to resolve all interferences. We performed mass scans on the peaks of ^{10}B and ^{11}B for both gas blanks (laser off) and measurements on carbonate (laser on) (Fig. 1) to investigate possible effects by scattered ions of matrix elements as observed by some recent studies (Sadekov et al., 2019; Standish et al., 2019). For our set-up, we can exclude such matrix-induced effects, which is in line with Fietzke et al. (2010) and Mayk et al. (2020). Hence, there was no need to correct the raw LA data as done in the recent studies by Sadekov et al. (2019) and Standish et al. (2019). Before analysis, sample and standard materials were pre-ablated to remove potential surface contamination. Laser repetition rates ranged between 12 and 60 Hz to match the signal intensity between~~

~~carbonate samples and standard material NIST SRM 610 (~300,000 cps). Torch position, ion optics and gas flows were optimized to gain maximum signal intensity and stability on ^{10}B and ^{11}B peaks. Each analysis consisted of 200 cycles with an integration time of 1 s and a prior on-peak gas blank measurement of 60 s, which was subtracted from the LA signal.~~

Each analysis was preceded by an on-peak gas blank measurement of 60 s on ^{10}B and ^{11}B, which was subtracted from the LA signal. The LA analysis itself was assessed by calculating the mean of the blank-corrected ^{11}B/^{10}B signal within an interval of up to 370 cycles (1 s each), where all data exceeding two standard deviations were removed as outliers. After analysis, B was washed out for 120 to 180 s until reaching background levels before a new measurement was started. A typical blank had ~7,000 cps on ^{11}B at the beginning of a session, but decreased to less than 3,000 cps during the course of a day. As signal intensity on ^{11}B was aimed at ~300,000 cps, the signal-to-noise ratio was in the order of ~100~~. Any contaminated phase from partial ablation of clay infillings, indicated by dropping ^{11}B/^{10}B ratios accompanied by rising [B], were excluded from further data treatment (Fig. 2)~~.

Accuracy of boron isotope measurements was frequently checked by ablating an in-house carbonate standard ~~and~~ that was also used for SIMS analysis (i.e. WP22, Rollion-Bard et al, 2003). The average δ^{11}B of 16.49 ± 1.26 ‰ (2σ, n=20) for WP22 is very close to the bulk solution values (δ^{11}B=16.60 ± 0.30 ‰ (2σ, n=6) measured at AWI, and δ^{11}B=16.76 ± 0.11 ‰ measured at IPGP). As the measurement uncertainty is mainly dependent on the ablation time, we report measurement uncertainties (as 2σ) for each δ^{11}B analysis as a function of analysis time, which was determined from multiple measurements of NIST glass standards and carbonate standards, and which is very close to the predicted uncertainty based on counting statistics (Fig. 3).

2.4 Bulk solution MC-ICPMS

After LA analyses, the 18 shells were carefully removed from the carbon tape and cleaned following the procedure outlined by Raitzsch et al. (2018). Briefly, foraminifer shells were gently crushed under a binocular between two glass slides and transferred to Eppendorf vials. After the clay removal and oxidative cleaning steps, the samples were leached in 0.001 N HNO_3, and finally dissolved in 60 µL of 1 N HNO_3.

Prior to boron isotope analysis, we used the micro-distillation technique to separate B from the calcium carbonate matrix (Gaillardet et al., 2001; Misra et al., 2014; Raitzsch et al., 2018; Wang et al., 2010). The distillate was diluted with 400 µL of 0.3 N HNO_3. The B concentration of a small aliquot was determined using a quick (20 s) on-peak measurement of ^{11}B on Faraday cup H9 using a Nu Plasma II MC-ICPMS (AWI, Bremerhaven). The remainder of the sample was then diluted to yield a solution with a [B] of 3 ppb and concentration-matched with the SRM NBS 951 to within ±3 %.

For isotope ratio measurements, boron was collected on Daly detectors, where high-mass D5 was used for ^{11}B and D0 for ^{10}B. Boron isotope data were measured in triplicate using the standard-sample-bracketing technique and reported in delta notation normalized to SRM NBS 951:

$$\delta^{11}B_{sample} = \left(\frac{^{11}B/^{10}B_{sample}}{^{11}B/^{10}B_{NIST\,951}} - 1 \right) * 1000 \qquad (1)$$

When 2σ of the mean derived from the triplicate was smaller than the long-term reproducibility (0.30 ‰), we report the latter as the measurement uncertainty. In addition, a small fragment of an in-house carbonate reference material WP22, used for our SIMS and LA-MC-ICPMS study, was cleaned and measured exactly the same way as the foraminifera sample to obtain a bulk $\delta^{11}B$ value for comparison (16.60 ± 0.30 ‰). This value is almost identical to that measured at IPGP using the bulk solution ICP-MS (16.76 ± 0.11 ‰).

3 Results and discussion

3.1 Intra-shell $\delta^{11}B$ variability

The results from SIMS measurements conducted on 5 large specimens reveal a high $\delta^{11}B$ variability ranging between 4.6 and 6.8 (mean 5.2) ‰ (2σ, 2 standard deviations of n individual measurements) within single shells, based on 8 to 19 single spot analyses on each shell. A similar variability of 4.4 ‰ (2σ) on average is observed for measurements within single chambers (Fig. 4). Since it is difficult to distinguish between the very small (i.e. the juvenile) chambers in the central part, we allocated these measurements to the umbilical "knob", which is also equivalent to the thick central part of the spiral side used for LA measurements. If $\delta^{11}B$ is averaged for each chamber (1 to 3 analyses per chamber), the mean variability between chambers is 4.2 ‰ (2σ) (Fig. 4). The two specimens measured chamber by chamber with LA also show variable $\delta^{11}B$, but with a much lower variation of ~1.1 ‰ (2σ), compared to the SIMS data (Fig. 4). The average $\delta^{11}B$ variability from all 16 shells measured multiple times is ~1.3 ‰ (2σ).

~~The results from SIMS measurements conducted on 5 large specimens reveal a high $\delta^{11}B$ variability ranging between 4.6 and 6.8 ‰ (2σ) within single shells, based on 8 to 19 single spot analyses on each shell. A similar variability of 4.4 ‰ (2σ) on average is observed for measurements within single chambers (Fig. 4). The largest variation is observed in the central part (i.e. the juvenile chambers) of the shell, which may be attributed to the difficulty in distinguishing between the very small chambers in this part. Hence, we allocated these measurements to the umbilical "knob", which is also equivalent to the thick central part of the spiral side used for LA measurements. If measurements are averaged for each chamber (1 to 3 analyses per chamber), the mean variability between chambers is 4.2 ‰ (2σ) (Fig. 4). The two specimens measured with LA also show variable $\delta^{11}B$, but with a much lower variation of ~1.1 ‰ (2σ), compared to the SIMS data (Fig. 4).~~

Here the question arises whether the difference in $\delta^{11}B$ variability between the two methods is due to differences in analytical uncertainty or different scales of natural heterogeneity. If we consider an average uncertainty of ±0.97~~7~~ ‰ for LA ~~predicted by Poisson counting statistics~~ (Fig. 3), intra-shell variability is reduced from 1.3~~1~~ ‰ to 0.4 ‰. As the 2σ measurement uncertainty for SIMS is roughly ±2.5 ‰, the remaining difference in variability between SIMS and LA methods of ~2.3~~3~~ ‰ is likely due to the different sampling volumes, and hence related to heterogeneous boron isotopic distribution in the test.

While the spot size for the SIMS method is ~20 µm and ~1 µm in depth, the laser-ablated volume ranges from 80 to 100 µm in diameter (Fig. 4) and approximately 10 µm in depth. Consequently, the ~200 times larger volume analyzed by LA would reduce the $\delta^{11}B$ variability detected by SIMS to ~0.2 (=2.3/√200) ~~3~~ ‰. Hence we argue that the "true" $\delta^{11}B$ heterogeneity is scale-dependent and assumedly in the order of ~3 and ~0.4 ‰ (2σ) on a ~20 and 100 µm grid, respectively.

To examine potential systematic trends in $\delta^{11}B$ among successive chambers, we calculated the residual boron isotopic composition $\Delta\delta^{11}B$ for each site within each shell by comparing the B isotopic composition of a single chamber $\delta^{11}B_{single}$ with the mean value of the shell $\delta^{11}B_{mean}$:

$$\Delta\delta^{11}B = \delta^{11}B_{single} - \delta^{11}B_{mean} \qquad (2)$$

The SIMS data suggest that $\Delta\delta^{11}B$ tends to decrease from the penultimate chamber (f-1) towards chamber f-5 by roughly 4 ‰ (Fig. 4), whereas no systematic change exists between chambers f-6 and the juvenile chambers. However, it is compelling that also the LA results suggest a decreasing trend in $\Delta\delta^{11}B$ from the final chambers towards chamber f-5 by more than 0.5 ‰, while in the earlier chambers no systematic change can be observed (Fig. 4). For both methods, Wilcoxon-Mann-Whitney tests and Welch's t-~~-~~tests suggest that the $\Delta\delta^{11}B$ change between the final chambers and f-5 is statistically insignificant at a 95 % significance level (p-values ≥ 0.07). ~~, but it should be kept in mind that we are at the limits in determining ontogenetic trends, due to the relatively high analytical uncertainties of the LA and SIMS techniques.~~ However, decreasing $\delta^{11}B$ from the final chamber towards earlier chambers would be in line with the LA study by Sadekov et al. (2019) showing a ~2 ‰ decrease along the last whorl of *C. wuellerstorfi*. A similar pattern was also observed for B/Ca, with the highest value in the final chamber (Raitzsch et al., 2011; Sadekov et al., 2019), ~~large-scale~~suggesting a strong biological influence or kinetic (i.e. growth rate) effect on boron incorporation. An in-depth discussion of biological/calcification processes is beyond the scope of this study, but the discovery of such high variability has implications for the use of $\delta^{11}B$-microanalytical techniques in paleoceanographic studies (e.g., Rollion-Bard and Erez, 2010).

Another notable feature derived from LA and SIMS is the elevated $\delta^{11}B$ (by ~0.5 ‰ on av.) of the umbilical knob, compared to the whole-shell $\delta^{11}B$. This is confirmed by supplementary ablation of the knob of individuals, which were used for whole-shell analysis in section 3.2. On average, umbilical knob $\delta^{11}B$ was ~0.4 ‰ higher than the value derived from the larger ablated area (see inset picture in Fig. 7), although this behavior is not systematic and was observed in only two thirds of the cases.

~~Another notable feature derived from LA and SIMS is the somewhat elevated $\delta^{11}B$ of the umbilical knob, compared to the whole-shell $\delta^{11}B$. This is confirmed by supplementary ablation of the knob of individuals, which were used for whole-shell analysis in section 3.2. On average, umbilical knob $\delta^{11}B$ was ~0.4 ‰ higher than the value derived from the larger ablated area (see inlet picture in Fig. 7), although this behavior is not systematic and was observed in only two thirds of the cases.~~

3.2 Inter-shell $\delta^{11}B$ variability

Apart from the seven specimens used for inspecting the chamber-to-chamber variability, 16 individuals of *C. wuellerstorfi* were laser-ablated using a large area of at least 300 µm in diameter to cover a major part of the spiral side, and in 14 specimens subsequently analyzed for the composition of the thicker umbilical knob using a smaller crater (inset picture in Fig. 7). This way, we approached quasi-bulk $\delta^{11}B$ values for single shells. ~~–~~Together with the $\delta^{11}B$ medians from the two specimens described in the previous section, a total of 18 shells were~~was~~ used for determining the inter-shell $\delta^{11}B$ variability using LA-MC-ICPMS (Fig. 5). It should be noted that we usually report the average as median, since it is less sensitive to outliers than the mean, and also reflects the average of a non-uniform distribution. For SIMS analyses, the medians of single-spot analyses were calculated for each of the 5 shells.

The SIMS data reveal a huge spread of single-spot $\delta^{11}B$ across the 5 specimens (section 3.1), but the $\delta^{11}B$ values averaged for each shell exhibit a narrower range between tests, with a median $\delta^{11}B$ of 16.08 ± 2.70 ‰ (2σ) (Fig. 5). In contrast, the single-site LA data across all 18 individuals show a smaller variation in $\delta^{11}B$ than the SIMS data, where the values averaged for each shell yield a median of 15.90 ± 1.62 ‰ (2σ). Both the average $\delta^{11}B$ measurement uncertainty for LA of ±0.9 ‰ (2σ) and the variation difference between foraminiferal shells and WP22 of ~0.4 ‰ suggest a residual inter-shell variability in the order of 0.4 to 0.7 ‰. Similarly, if an uncertainty of ±2.50 ‰ (2σ) for SIMS measurements is taken into account, the remaining inter-shell variability is only ~0.2 ‰. Therefore, we estimate the "true" variability between shells of a population to be ~0.4 ‰, which is the same as the variation estimated for the intra-shell variability (section 3.1).

For shells where both large areas and knobs were measured (n=14), it is interesting to note that if only the large LA craters are considered, the mean $\delta^{11}B$ is 15.87 ± 1.78 ‰ (2σ), while it is 16.27 ± 2.75 ‰ (2σ), if solely the small LA craters are taken into account (cf. Fig. 7, inset picture). As the volume of the large LA craters is ~3 times larger than the smaller ones, the resulting variability among means of 3 resampled small-crater values is 1.59 (=2.75/√3) ‰ (2σ), which is quite close to the 1.78 ‰ derived from large craters, and confirms our conclusion that the $\delta^{11}B$ variability is dependent on the scale at which it is measured.

~~It is interesting to note that if only the large LA craters are considered, the mean $\delta^{11}B$ is 15.87 ± 1.78 ‰ (2σ), while it is 16.27 ± 2.75 ‰ (2σ), if solely the small LA craters are taken into account (cp. Fig. 7, right SEM image). As the volume of the large LA craters is ~3 times larger than the smaller ones, the resulting variability among means of 3 resampled small-crater values is 1.59 ‰ (2σ), which is quite close to the 1.78 ‰ derived from large craters, and confirms our conclusion that the $\delta^{11}B$ variability is dependent on the scale at which it is measured.~~

3.3 Bulk solution $\delta^{11}B$

Both the SIMS and LA results reveal median values that match the bulk $\delta^{11}B$ of 15.99 ± 0.30 ‰ (2σ) measured in solution to within analytical uncertainties (Fig. 5). It should be noted again that the same specimens measured in solution had been measured before by LA, ensuring that we compare different techniques based on the same set of samples. Similarly, the

average $\delta^{11}B$ of 16.48 ± 1.26 ‰ (2σ) in the reference material WP22 determined with LA-MC-ICPMS is not distinguishable from the bulk solution value of 16.60 ± 0.30 ‰ (2σ), which confirms the robustness of the LA technique, and also the SIMS results, as the median foraminifera values are identical for LA and SIMS techniques.

The $\delta^{11}B$ values obtained from all three methods fit in with the calibration data set for *C. wuellerstorfi* from the study by Rae et al. (2011) (Fig. 6), and confirm that the boron isotopic composition in this species closely matches the one of borate of ambient seawater. Further, it proves that LA-MC-ICPMS and SIMS yield accurate results for $\delta^{11}B$, if the data set is large enough to overcome the issues of intra- and inter-shell variability (~0.4 ‰), and analytical uncertainty of micro-analytical techniques (~±0.9 and ±2.5 ‰ for LA and SIMS, respectively).

3.4 Implications for paleoreconstruction studies

The large intra- and inter-shell variations in $\delta^{11}B$ described in sections 3.1 and 3.2 raises the question whether microanalytical techniques such as SIMS or LA-MC-ICPMS can be used for analyzing $\delta^{11}B$ in *C. wuellerstorfi* to reconstruct past deep-water pH. The SIMS method requires careful embedding of foraminifer shells in epoxy and polishing down to a planar surface, which precludes further processing for e.g. bulk solution analyses. However, because the size of the beam spot is small (20 µm or less), it is still possible to measure some other elemental and isotopic ratios at the same location on the sample; e.g. the same foraminifera specimens were used to measure $\delta^{18}O$ (Rollion-Bard et al, 2008), $\delta^{11}B$ (Rollion-Bard and Erez, 2010), and $\delta^{7}Li$ (Vigier et al, 2015). SIMS technique is very useful for biomineralization studies (e.g. Rollion-Bard and Erez, 2010), but for paleoreconstruction of deep-sea pH, where high precision is necessary, it may not be the most appropriate technique for routine downcore $\delta^{11}B$ analysis. However, here we will inspect LA-MC-ICPMS as a potential tool for paleo-pH studies.

To attain information on the number of shells required for accurate LA analysis of $\delta^{11}B$ to within a target uncertainty, we applied a Monte Carlo approach to generate two data sets with 10,000 $\delta^{11}B$ data each, within a quoted uncertainty of ±1.68 ‰ and ±2.75 ‰ (2σ) for "large crater" and "knob" measurements, respectively, as given by the original data set (n=16 and n=14, resp.). The average $\delta^{11}B$ values are considered identical between large craters and umbilical knobs, as in the original data they agree to within analytical uncertainty. Then we applied the 'combn()' function of the R package 'utils v3.4.4' (R Core Team, 2018) on each of the simulated data sets. With this function, we can calculate the uncertainty by generating all possible combinations of n shells taken from the simulated populations. For instance, if we would randomly pick five shells from this sediment sample, the analyzed $\delta^{11}B$ would be accurate to within ±0.75 ‰ with a probability of 95 %, if large areas, and ±1.22 ‰, if only the knob areas were analyzed. If we targeted a standard uncertainty of ±0.50 ‰, which is equivalent to a pH uncertainty of roughly ±0.1, we would need to measure ~12 specimens with LA, if large areas, and ~14 specimens, if only the knob areas were analyzed (Fig. 7). The relationship between number of analyzed shells (n) and the estimated 2σ uncertainty is given by the quoted variability u_q, i.e. the measured $\delta^{11}B$ variation across a population (as 2σ), and n:

$$2\sigma = \frac{u_q}{\sqrt{n}} \quad (3)$$

To attain information on the number of shells required for accurate LA analysis of $\delta^{11}B$ to within a target standard uncertainty, we applied the 'combn()' function of the R package 'utils v3.4.4' (R Core Team, 2018). In this simulation, it is assumed that the entire population of *C. wuellerstorfi* consists of the 18 shells, for which we have measured $\delta^{11}B$ both using LA and bulk-solution MC-ICPMS (Figs. 5 and 6). If from this sample only n individuals were available, we can calculate the standard uncertainty by generating all possible combinations of n shells taken from the entire population. The histograms of resulting $\delta^{11}B$ values averaged from n shells are shown in Fig. 7. For instance, if we would randomly pick four shells from this sediment sample, the analyzed $\delta^{11}B$ would be accurate to within ±0.72 ‰ with a probability of 95 %. If we targeted a standard uncertainty of ±0.50 ‰, which is equivalent to a pH uncertainty of roughly ±0.1, we would need to measure 7 specimens with LA (Fig. 7).

Given that the analysis uncertainty of the same amount measured in solution is about ±0.3 ‰, bulk solution analysis appears to be the more convenient technique for reconstructing paleo-pH. On the other hand, the LA technique may be useful for generating high-resolution records, where sharp pH trends would partly compensate for the larger standard uncertainty or when only few foraminifera specimens are available. Further, LA, like SIMS, has the potential to gain insight into ontogenetic $\delta^{11}B$ variations, helping to better understand the biological uptake of boron during chamber formation.

5 Conclusions

Microanalytical methods such as SIMS or LA-MC-ICPMS are potentially powerful tools for studying biomineralization processes or possible alternatives to conventional bulk solution analysis of $\delta^{11}B$ in benthic foraminifera, if sample material is limited. For this study, we measured a population of 23 *C. wuellerstorfi* in total using SIMS and femtosecond LA-MC-ICPMS and compared the results with the bulk-solution $\delta^{11}B$, revealing consistent average values among the different techniques. While the medians agree to within ±0.1 ‰, a large intra-shell variability was observed, with up to 6.8 ‰ and 4.5 ‰ (2σ) derived from the SIMS and LA methods, respectively. We propose that the larger spread for SIMS, compared to LA, can be attributed to the much smaller volume (~200^{-1}) of calcite being analyzed in each run, and hence supposedly reflects a larger heterogeneity of $\delta^{11}B$ in the foraminiferal test on a smaller scale. When analytical uncertainties and scale-dependent differences in $\delta^{11}B$ variations are taken into account, the intra-shell variability is likely in the order of ±0.4 and 3 ‰ (2σ) on a 100 and 20 µm scale, respectively.

The $\delta^{11}B$ variability between shells exhibits total ranges of ~3 ‰ for both techniques, suggesting that a number of shells needs to be analyzed for accurate mean $\delta^{11}B$ values. We applied a simple resampling method and conclude that about 12~~7~~ shells of *C. wuellerstorfi* must be analyzed using LA-MC-ICPMS to obtain an accurate average value to within ±0.5 ‰ (2σ). Hence, we suggest that, based on this high number of required individuals, the bulk solution MC-ICPMS method remains the first choice for analysis of $\delta^{11}B$ in routine paleo-pH studies.

Data availability

The boron isotope data collected for this study are available from Table S1 in the supplement.

Author contribution

MR, CRB, IH, and JB conceived the study (conceptualization). MR, CRB, and PL carried out measurements, analyzed the data, and performed data statistics (data curation, formal analysis, investigation). AB, KUR, and GS maintained and provided access to analytical instruments at AWI (resources). JB, CRB, and IH raised funding for the French-German project 'B2SeaCarb' (funding acquisition). MR produced the figures for the manuscript (visualization). MR and CRB wrote the first draft of the manussucript (writing – original draft), and all authors interpreted, edited, and reviewed the manuscript (writing – review & editing).

Competing interests

The authors declare that they have no conflict of interest.

Acknowledgements

This research was carried out in the framework of the joint French/German project 'B2SeaCarb' and was supported by the Deutsche Forschungsgemeinschaft (DFG) grant number BI 432/10-1 to JB and DFG grant number HO 3257/5-1 to IH. On the Ffrench side, the project was supported by the French National Research Agency (ANR) grant number ANR-16-CE92-0010 to CRB. CRB thanks N. Bouden (CRPG) for his technical help, and the MARUM GeoB core repository is acknowledged for providing sediment samples.

References

Blamart, D., Rollion-Bard, C., Meibom, A., Cuif, J.-P., Juillet-Leclerc, A. and Dauphin, Y.: Correlation of boron isotopic composition with ultrastructure in the deep-sea coral *Lophelia pertusa*: Implications for biomineralization and paleo-pH, Geochem. Geophys. Geosyst., 8(12), Q12001, doi:10.1029/2007GC001686, 2007.

Branson, O., Kaczmarek, K., Redfern, S. A. T., Misra, S., Langer, G., Tyliszczak, T., Bijma, J. and Elderfield, H.: The coordination and distribution of B in foraminiferal calcite, Earth Planet. Sci. Lett., 416, 67–72, doi:10.1016/j.epsl.2015.02.006, 2015.

Fietzke, J., Heinemann, A., Taubner, I., Böhm, F., Erez, J. and Eisenhauer, A.: Boron isotope ratio determination in carbonates via LA-MC-ICP-MS using soda-lime glass standards as reference material, J. Anal. At. Spectrom., 25(12), 1953, doi:10.1039/c0ja00036a, 2010.

Fietzke, J., Ragazzola, F., Halfar, J., Dietze, H., Foster, L. C., Hansteen, T. H., Eisenhauer, A. and Steneck, R. S.: Century-scale trends and seasonality in pH and temperature for shallow zones of the Bering Sea, PNAS, 112(10), 2960–2965, doi:10.1073/pnas.1419216112, 2015.

Gaillardet, J., Lemarchand, D., Göpel, C. and Manhès, G.: Evaporation and Sublimation of Boric Acid: Application for Boron Purification from Organic Rich Solutions, Geostand. Newsl., 25(1), 67–75, doi:10.1111/j.1751-908X.2001.tb00788.x, 2001.

Hemming, N. G. and Hanson, G. N.: Boron isotopic composition and concentration in modern marine carbonates, Geochim. Cosmochim. Acta, 56(1), 537–543, doi:10.1016/0016-7037(92)90151-8, 1992.

Hönisch, B., Bickert, T. and Hemming, N. G.: Modern and Pleistocene boron isotope composition of the benthic foraminifer *Cibicidoides wuellerstorfi*, Earth Planet. Sci. Lett., 272(1–2), 309–318, doi:10.1016/j.epsl.2008.04.047, 2008.

Horn, I. and von Blanckenburg, F.: Investigation on elemental and isotopic fractionation during 196 nm femtosecond laser ablation multiple collector inductively coupled plasma mass spectrometry, Spectrochim. Acta Part B At. Spectrosc., 62(4), 410–422, doi:10.1016/j.sab.2007.03.034, 2007.

Howes, E. L., Kaczmarek, K., Raitzsch, M., Mewes, A., Bijma, N., Horn, I., Misra, S., Gattuso, J.-P. and Bijma, J.: Decoupled carbonate chemistry controls on the incorporation of boron into *Orbulina universa*, Biogeosciences, 14(2), 415–430, doi:https://doi.org/10.5194/bg-14-415-2017, 2017.

Inoue, M., Nohara, M., Okai, T., Suzuki, A. and Kawahata, H.: Concentrations of Trace Elements in Carbonate Reference Materials Coral JCp-1 and Giant Clam JCt-1 by Inductively Plasma-Mass Spectrometry, Geostand. Geoanalytical Res., 28(3), 411–416, 2004.

Kaczmarek, K., Horn, I., Nehrke, G. and Bijma, J.: Simultaneous determination of $\delta^{11}B$ and B/Ca ratio in marine biogenic carbonates at nanogram level, Chem. Geol., 392, 32–42, doi:10.1016/j.chemgeo.2014.11.011, 2015a.

Kaczmarek, K., Langer, G., Nehrke, G., Horn, I., Misra, S., Janse, M. and Bijma, J.: Boron incorporation in the foraminifer *Amphistegina lessonii* under a decoupled carbonate chemistry, Biogeosciences, 12(6), 1753–1763, doi:10.5194/bg-12-1753-2015, 2015b.

Kasemann, S. A., Schmidt, D. N., Bijma, J. and Foster, G. L.: In situ boron isotope analysis in marine carbonates and its application for foraminifera and palaeo-pH, Chem. Geol., 260(1–2), 138–147, doi:10.1016/j.chemgeo.2008.12.015, 2009.

Klochko, K., Kaufman, A. J., Yao, W., Byrne, R. H. and Tossell, J. A.: Experimental measurement of boron isotope fractionation in seawater, Earth Planet. Sci. Lett., 248, 276–285, doi:10.1016/j.epsl.2006.05.034, 2006.

Louvat, P., Moureau, J., Paris, G., Bouchez, J., Noireaux, J. and Gaillardet, J.: A fully automated direct injection nebulizer (d-DIHEN) for MC-ICP-MS isotope analysis: application to boron isotope ratio measurements, J. Anal. At. Spectrom., 29(9), 1698–1707, doi:10.1039/C4JA00098F, 2014.

Malherbe, J., Penen, F., Isaure, M.-P., Frank, J., Hause, G., Dobritzsch, D., Gontier, E., Horréard, F., Hillion, F. and Schaumlöffel, D.: A New Radio Frequency Plasma Oxygen Primary Ion Source on Nano Secondary Ion Mass Spectrometry for Improved Lateral Resolution and Detection of Electropositive Elements at Single Cell Level, Anal. Chem., 88(14), 7130–7136, doi:10.1021/acs.analchem.6b01153, 2016.

Mayk, D., Fietzke, J., Anagnostou, E. and Paytan, A.: LA-MC-ICP-MS study of boron isotopes in individual planktonic foraminifera: A novel approach to obtain seasonal variability patterns, Chem. Geol., 531, 119351, doi:10.1016/j.chemgeo.2019.119351, 2020.

Misra, S., Owen, R., Kerr, J., Greaves, M. and Elderfield, H.: Determination of $\delta^{11}B$ by HR-ICP-MS from mass limited samples: Application to natural carbonates and water samples, Geochim. Cosmochim. Acta, 140, 531–552, doi:10.1016/j.gca.2014.05.047, 2014.

Okai, T., Suzuki, A., Terashima, S., Inoue, M., Nohara, M., Kawahata, H. and Imai, N.: Collaborative Analysis of GSJ/AIST Geochemical Reference Materials JCp-1 (Coral) and JCt-1 (Giant Clam), Chikyu Kagaku, 38(4), 281–286, 2004.

R Core Team: A Language and Environment for Statistical Computing, R Foundation for Statistical Computing, Vienna. [online] Available from: https://www.R-project.org, 2018.

Rae, J. W. B., Foster, G. L., Schmidt, D. N. and Elliott, T.: Boron isotopes and B/Ca in benthic foraminifera: Proxies for the deep ocean carbonate system, Earth Planet. Sci. Lett., 302(3–4), 403–413, doi:10.1016/j.epsl.2010.12.034, 2011.

Raitzsch, M. and Hönisch, B.: Cenozoic boron isotope variations in benthic foraminifers, Geology, 41(5), 591–594, doi:https://doi.org/10.1130/G34031.1, 2013.

Raitzsch, M., Bijma, J., Bickert, T., Schulz, M., Holbourn, A. and Kučera, M.: Eccentricity-paced atmospheric carbon-dioxide variations across the middle Miocene climate transition, Clim. Past Discuss., https://doi.org/10.5194/cp-2020-96, 2020. ~~submitted.~~

Raitzsch, M., Hathorne, E. C., Kuhnert, H., Groeneveld, J. and Bickert, T.: Modern and late Pleistocene B/Ca ratios of the benthic foraminifer *Planulina wuellerstorfi* determined with laser ablation ICP-MS, Geology, 39(11), 1039–1042, doi:https://doi.org/10.1130/G32009.1, 2011.

Raitzsch, M., Bijma, J., Benthien, A., Richter, K.-U., Steinhoefel, G. and Kučera, M.: Boron isotope-based seasonal paleo-pH reconstruction for the Southeast Atlantic – A multispecies approach using habitat preference of planktonic foraminifera, Earth Planet. Sci. Lett., 487, 138–150, doi:https://doi.org/10.1016/j.epsl.2018.02.002, 2018.

Rollion-Bard, C. and Blamart, D.: in Biomineralization Sourcebook: Characterization of Biominerals and Biomimetic Materials, pp. 249–261, CRC Press, Taylor & Francis Group, Boca Raton, FL., 2014.

Rollion-Bard, C. and Erez, J.: Intra-shell boron isotope ratios in the symbiont-bearing benthic foraminiferan *Amphistegina lobifera*: Implications for $\delta^{11}B$ vital effects and paleo-pH reconstructions, Geochim. Cosmochim. Acta, 74(5), 1530–1536, doi:10.1016/j.gca.2009.11.017, 2010.

Rollion-Bard, C., Chaussidon, M. and France-Lanord, C.: pH control on oxygen isotopic composition of symbiotic corals, Earth Planet. Sci. Lett., 215(1), 275–288, doi:10.1016/S0012-821X(03)00391-1, 2003.

Sadekov, A., Lloyd, N. S., Misra, S., Trotter, J., D'Olivo, J. and McCulloch, M.: Accurate and precise microscale measurements of boron isotope ratios in calcium carbonates using laser ablation multicollector-ICPMS, J. Anal. At. Spectrom., 34(3), 550–560, doi:10.1039/C8JA00444G, 2019.

Standish, C. D., Chalk, T. B., Babila, T. L., Milton, J. A., Palmer, M. R. and Foster, G. L.: The effect of matrix interferences on in situ boron isotope analysis by laser ablation multi-collector inductively coupled plasma mass spectrometry, Rapid Commun. Mass Spectrom., 33(10), 959–968, doi:10.1002/rcm.8432, 2019.

Steinhoefel, G., Horn, I. and von Blanckenburg, F.: Matrix-independent Fe isotope ratio determination in silicates using UV femtosecond laser ablation, Chem. Geol., 268(1), 67–73, doi:10.1016/j.chemgeo.2009.07.010, 2009.

Thil, F., Blamart, D., Assailly, C., Lazareth, C. E., Leblanc, T., Butsher, J. and Douville, E.: Development of laser ablation multi-collector inductively coupled plasma mass spectrometry for boron isotopic measurement in marine biocarbonates: new improvements and application to a modern *Porites* coral, Rapid Commun. Mass Spectrom., 30(3), 359–371, doi:10.1002/rcm.7448, 2016.

Vigier, N., Rollion-Bard, C., Levenson, Y. and Erez, J.: Lithium isotopes in foraminifera shells as a novel proxy for the ocean dissolved inorganic carbon (DIC), Comptes Rendus Geoscience, 347(1), 43–51, doi:10.1016/j.crte.2014.12.001, 2015.

Wang, B.-S., You, C.-F., Huang, K.-F., Wu, S.-F., Aggarwal, S. K., Chung, C.-H. and Lin, P.-Y.: Direct separation of boron from Na- and Ca-rich matrices by sublimation for stable isotope measurement by MC-ICP-MS, Talanta, 82(4), 1378–1384, doi:10.1016/j.talanta.2010.07.010, 2010.

Yu, J. and Elderfield, H.: Benthic foraminiferal B/Ca ratios reflect deep water carbonate saturation state, Earth Planet. Sci. Lett., 258(1), 73–86, doi:10.1016/j.epsl.2007.03.025, 2007.

Yu, J., Foster, G. L., Elderfield, H., Broecker, W. S. and Clark, E.: An evaluation of benthic foraminiferal B/Ca and $\delta^{11}B$ for deep ocean carbonate ion and pH reconstructions, Earth Planet. Sci. Lett., 293(1–2), 114–120, doi:10.1016/j.epsl.2010.02.029, 2010.

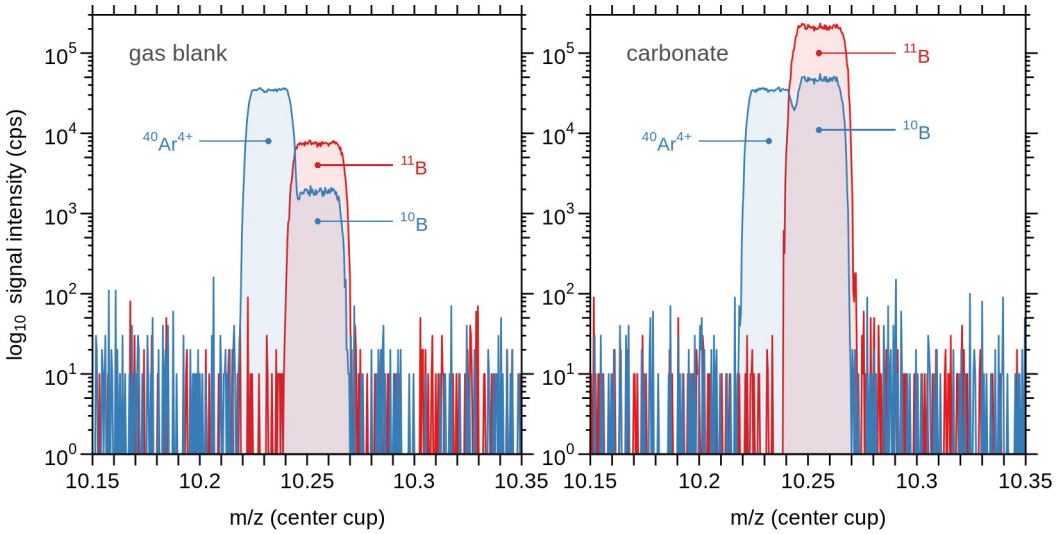


Figure 1: Mass scans over atomic masses 10 (blue) and 11 (red) u, centered at ~10.26 amu using Daly detectors, where peak center coincidence appears at ~10.25 amu in the center cup. Left: Gas blank (laser off), showing the typical double peak of $^{40}Ar^{4+}$ and ^{10}B, and the ^{11}B peak. Right: Signal of ablated calcium carbonate (laser on). The baseline exhibits only electronic noise from the Daly detectors, but no sign of unresolved interferences on ^{10}B as matrix-induced scattered Ca ion. Note that the signal intensity is on a logarithmic scale.

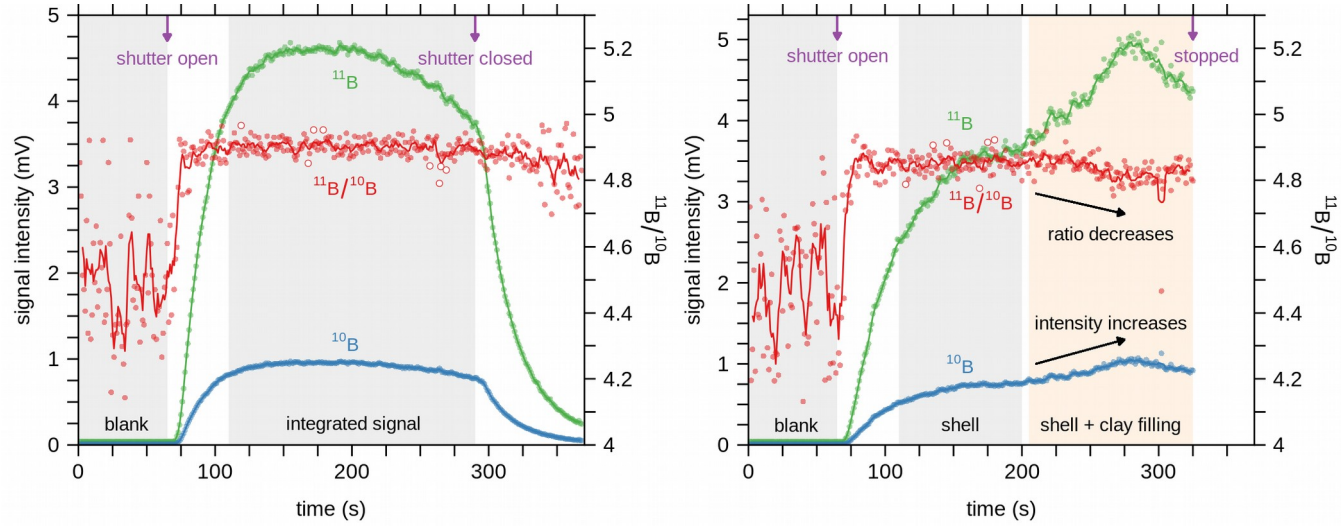
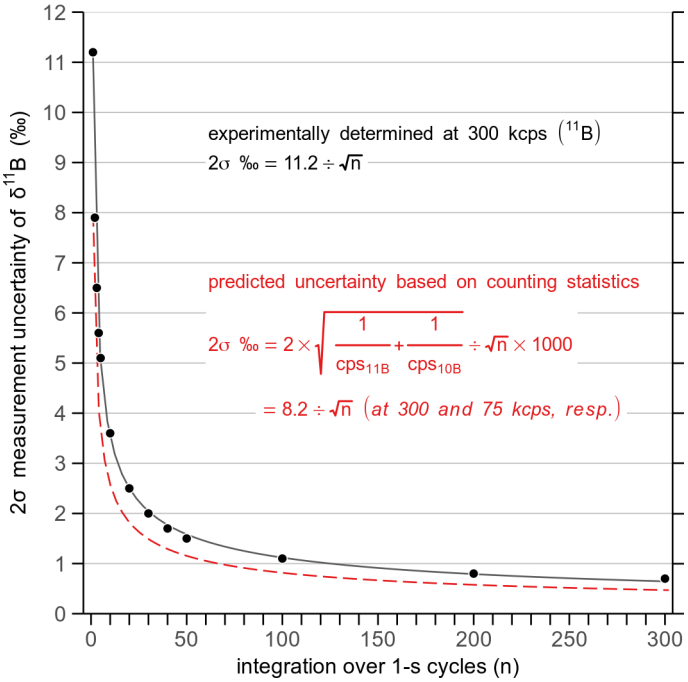
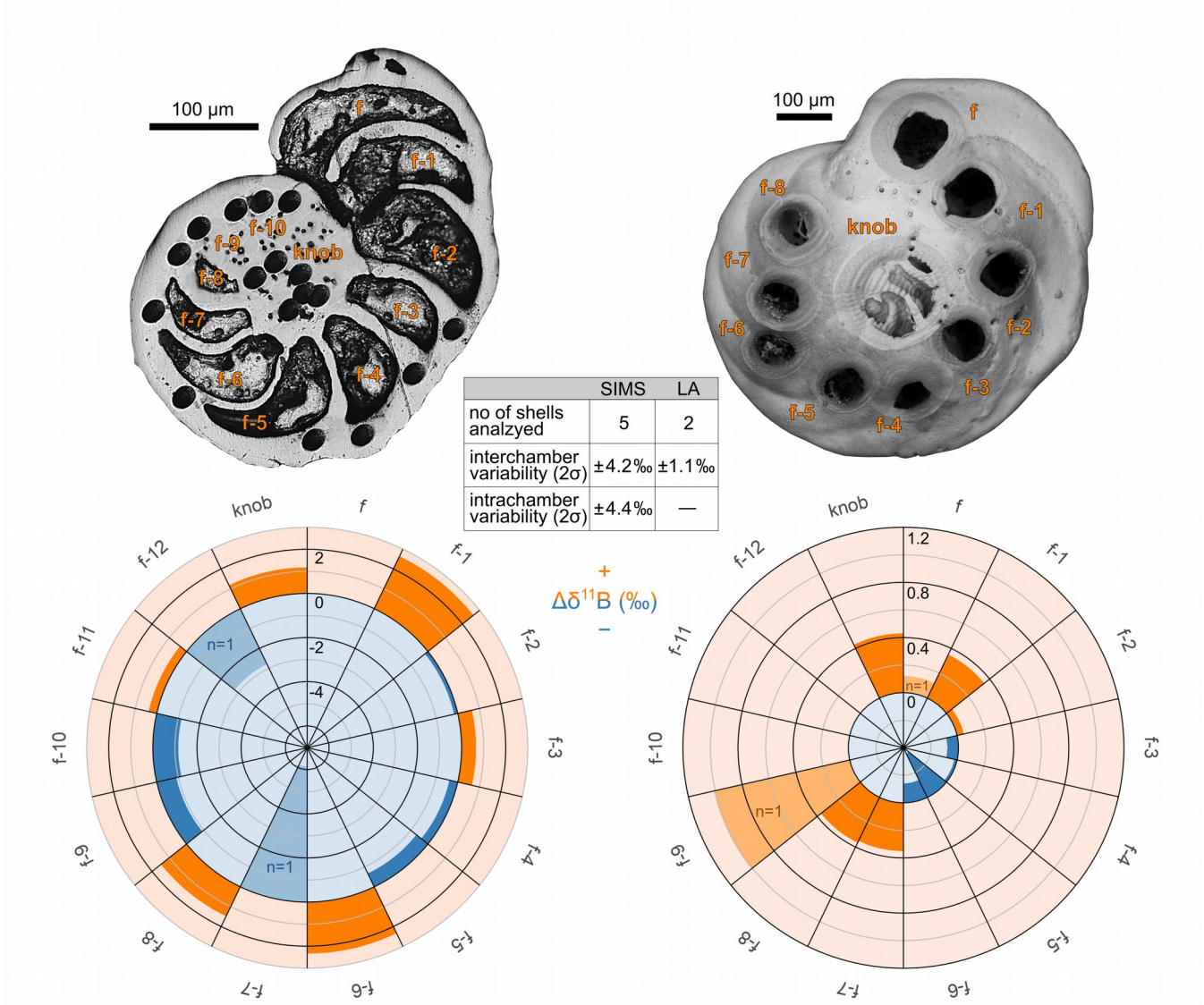


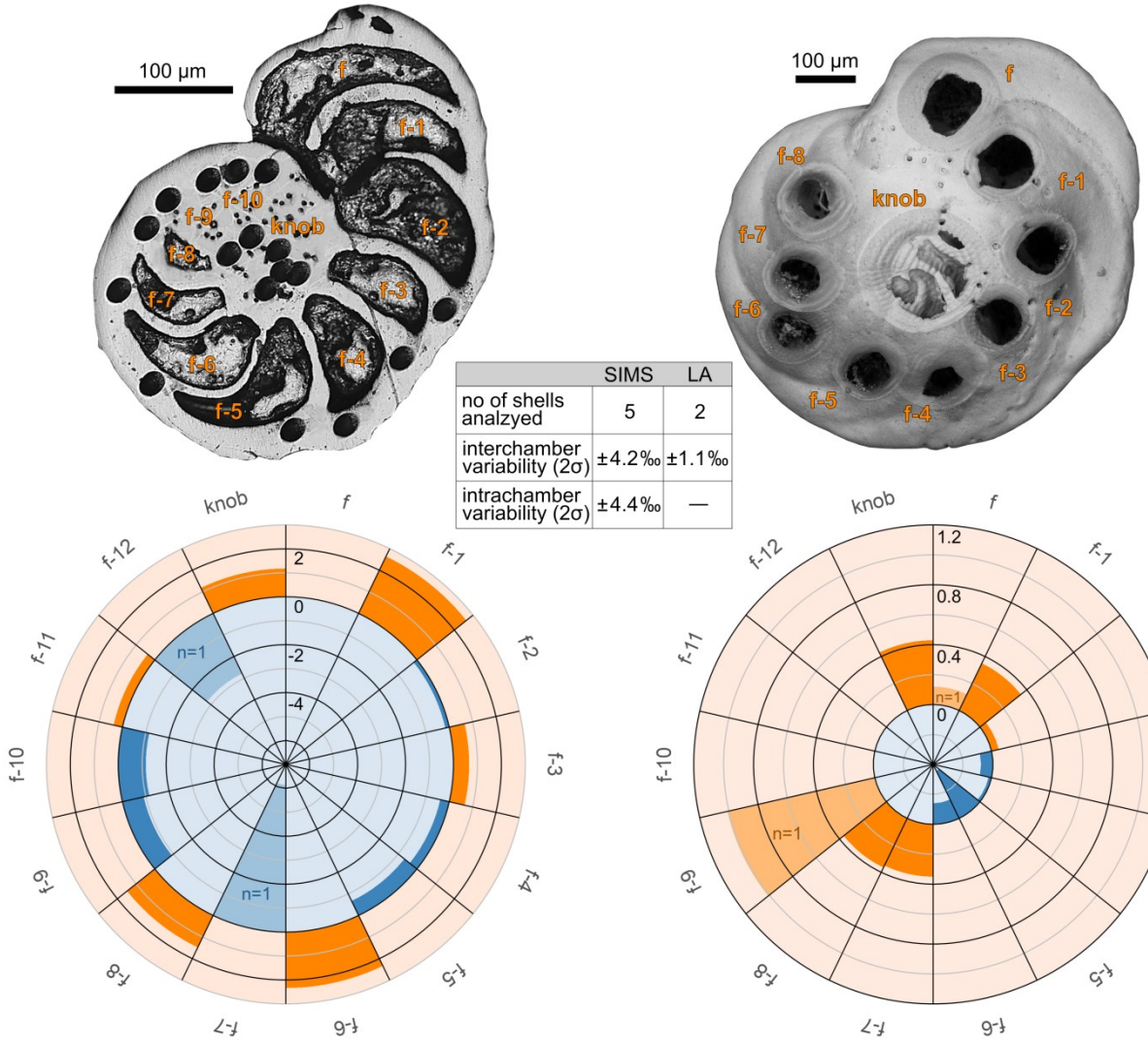
Figure 2: Left: Example of~~Typical~~ time-resolved laser ablation analysis for ^{10}B and ^{11}B of a *C. wuellerstorfi* shell using Daly detectors, preceded by a ~60 s blank measurement. Dots represent 1-s cycles, and lines 5-pt running averages. Open symbols are data that are excluded by the 2σ outlier test. Right: Example of a shell that was penetrated by the laser beam, resulting in the ablation of clay infillings.



370 | **Figure 3: Measurement uncertainty of $^{11}B/^{10}B$ (2σ) at count rates of 300,000 cps (^{11}B) as a function of the laser ablation time. The uncertainty of each boron isotope measurement is calculated based on this relationship (black solid line). A major portion (~70 %) of the measurement uncertainty is related to Poisson-distributed counts (red dashed line).**



375 **Figure 4: Intra-shell variability of $\delta^{11}B$ using SIMS (left panel) and LA-MC-ICPMS (right panel) on selected large individuals of *C. wuellerstorfi*. The residual $\Delta\delta^{11}B$ (difference between single spot and mean $\delta^{11}B$, eq. 2) averaged from all analyzed specimens is shown for each chamber (f is the final chamber, f-1 the penultimate one, and so on). Orange color stands for higher-than-mean and blue for lower-than-mean values. Lighter colors indicate data that are based on only one measurement. The inset table summarizes the measured intra-shell variability derived from the two techniques.**



380 **Figure 5: Intra-shell variability of $\delta^{11}B$ using SIMS (left panel) and LA-MC-ICPMS (right panel) on selected large individuals of *C. wuellerstorfi*. The residual $\Delta\delta^{11}B$ (difference between single spot and mean $\delta^{11}B$, eq. 2) averaged from all analyzed specimens is shown for each chamber (f is the final chamber, f-1 the penultimate one, and so on). Orange color stands for higher-than-mean and blue for lower-than-mean values. Lighter colors indicate data that are based on only one measurement. The inlet table summarizes the measured intra-shell variability derived from the two techniques.**

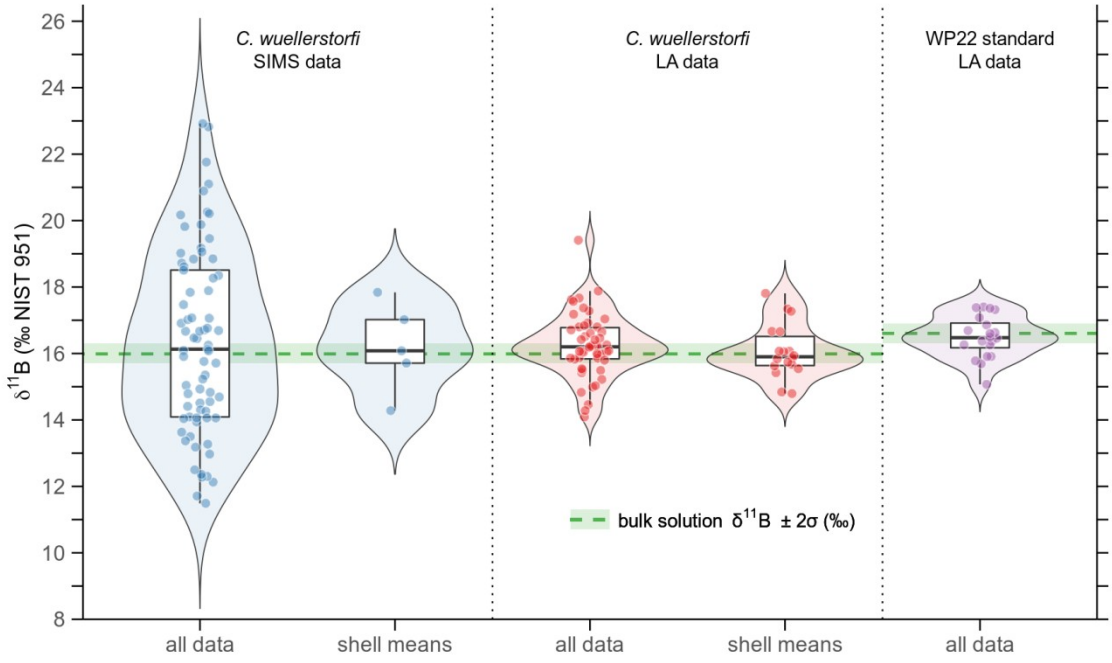


Figure 6: Violin, box and jitter plots showing the distribution of all single-site $\delta^{11}B$ values and single-shell means, both for the SIMS and laser ablation techniques. For comparison, the distribution of $\delta^{11}B$ values measured on the in-house reference material WP22 is displayed as well. The green dashed lines and bars represent the bulk solution $\delta^{11}B \pm 2\sigma$ values.

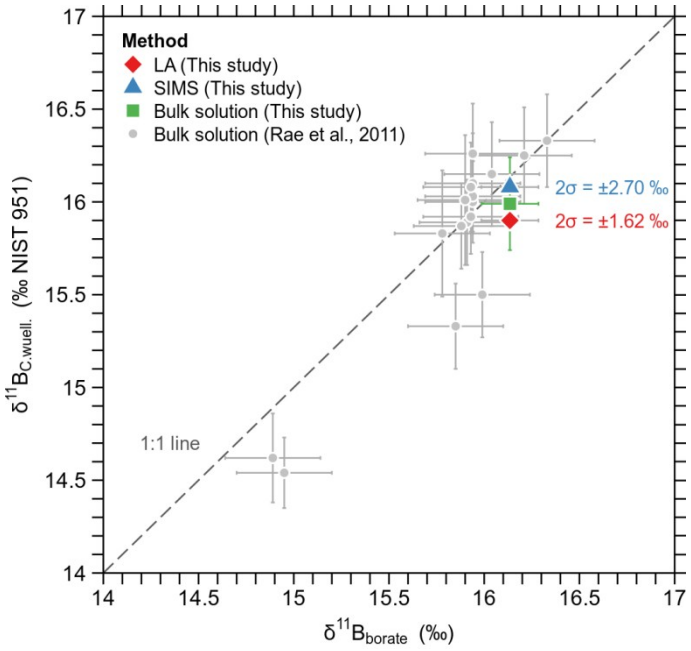
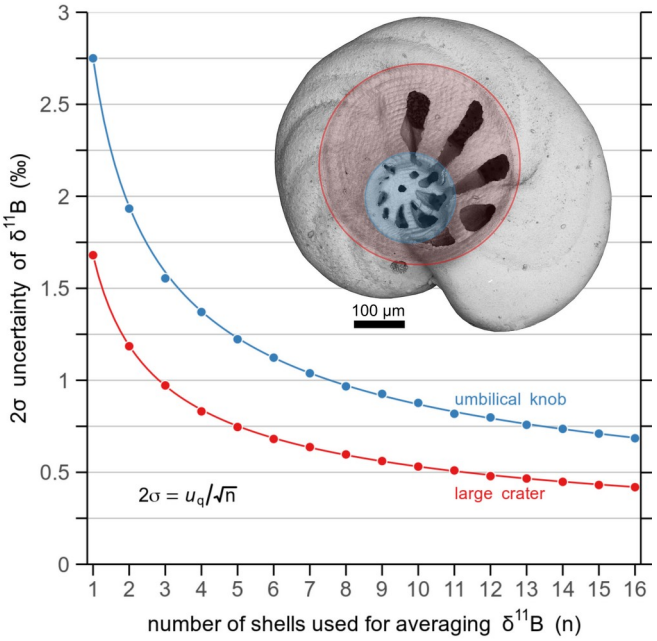
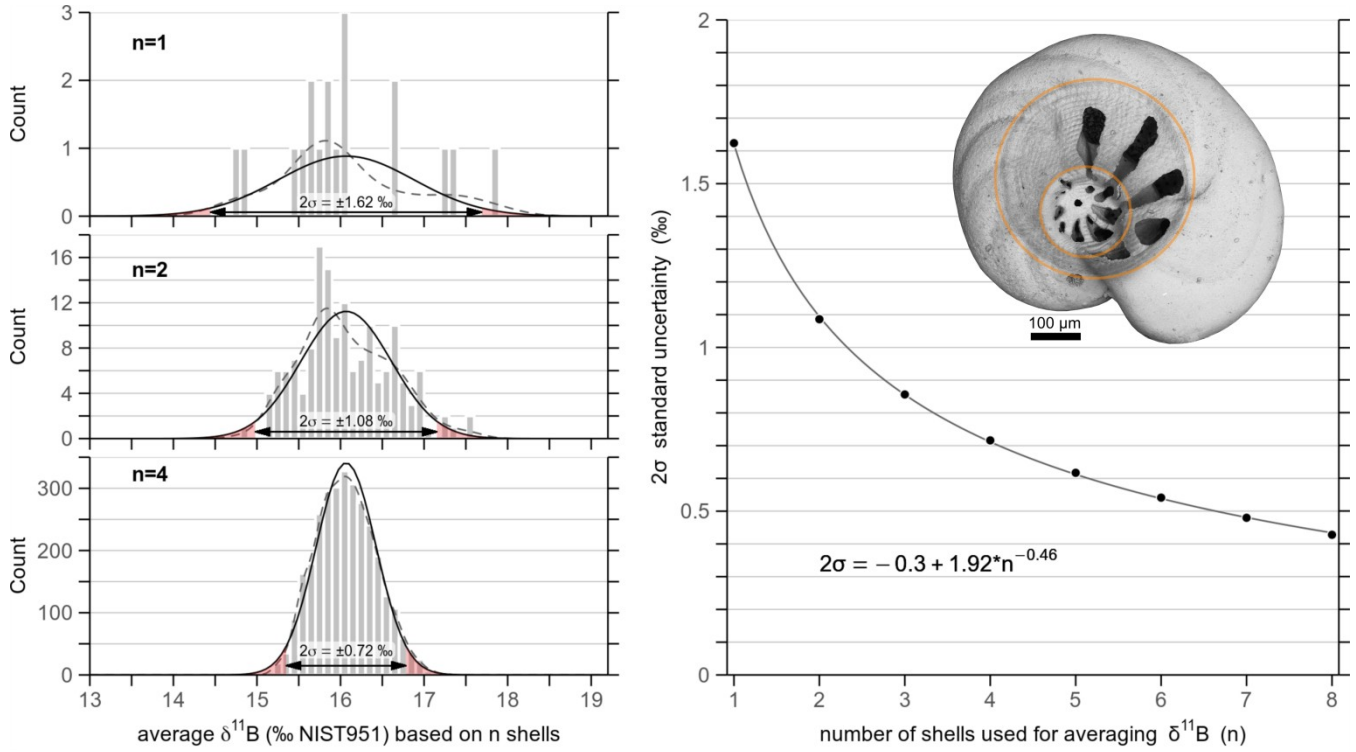


Figure 7: Median $\delta^{11}B$ of Holocene (5.6 ka) *C. wuellerstorfi* from GeoB core 1032 (Walvis Ridge, South Atlantic) measured with different techniques, shown along with the core-top calibration from (Rae et al., 2011). Note that the bulk solution analysis of this study was carried out on the same population measured before with laser ablation. Pooled $\delta^{11}B$ uncertainties for SIMS (n=5 shells) and LA-MC-ICPMS (n=18 shells) are shown as numbers, as error bars exceed the y-axis scale.



395 **Figure 8: Results from Monte Carlo simulations of 2σ uncertainty for $\delta^{11}B$ using LA-MC-ICPMS in relation to the number of analyzed *C. wuellerstorfi* shells (n). In red is the estimated uncertainty based on "large crater", and in blue on "umbilical knob" measurements (cf. inset SEM picture for different areas). The estimated 2σ uncertainty can be described by a function of the quoted uncertainty (u_q) and n (eq. 3).**



400 **~~Figure 9: Simulation of 2σ uncertainty for $\delta^{11}B$ using LA-MC-ICPMS in relation to the number of analyzed *C. wuellerstorfi* shells. The grey bars represent the numbers of mean $\delta^{11}B$ values in 0.1 ‰ bins from all possible combinations of n shells out of the original population (n=1). The dashed lines are density curves, and the solid lines normal distribution curves. The more shells are used for analysis, the smaller gets the 2σ uncertainty and the higher the probability to attain the "true" mean value. As an example, the $\delta^{11}B$ 2σ uncertainty is ±0.6 ‰ for the average out of 5 shells from the population. The inlet SEM picture shows a specimen measured for "whole-shell" $\delta^{11}B$, typically implemented by a large crater covering a major shell part, and a smaller one on the thicker umbilical knob.~~**