Geochronology Discuss.,
https://doi.org/10.5194/gchron-2019-19-AC1, 2020
© Author(s) 2020. This work is distributed under
the Creative Commons Attribution 4.0 License.

Interactive comment on "Miniature radiocarbon measurements (< 150 µg C) from sediments of Lake Åżabińskie, Poland: effect of precision and dating density on age-depth models" *by* Paul D. Zander et al.

Paul D. Zander et al.

paul.zander@giub.unibe.ch

Received and published: 4 March 2020

Response: Thank you for the review and comments on our manuscript. Our response to the comments are interjected in italic font below.

SUMMARY The manuscript describes the results of a case study in which radiocarbon ages obtained using gas-source technique are compared with radiocarbon ages of the conventional graphitized samples; both types of samples come from a number of selected depth intervals in a lake sediment core. Because this

core supposedly has a relatively well resolved varve-based chronology (albeit floating and not shown in this manuscript), the authors integrate varve counts and two types of radiocarbon ages into a simulated 'best age estimate' model. They further demonstrate a series of exercises in generating the synthetic age-depth distributions with a purpose of illustrating the effect of sampling density and sample size (mass carbon) on age model precision. According to the authors, the main idea of the work is an evaluation of how reliable gas source radiocarbon ages on miniature samples are for constructing age models. This is important for those lake records which lack enough datable material for the conventional radiocarbon analyses. The implications highlighted by the authors include (a) how to "improve sampling strategies" (the more age determinations the better, just as one may expect); and (b) what are the "expectations of age uncertainty". Among the benefits of skipping the graphitization step when using gas-source technique the authors cite "reduced cost", but there is no comparison provided for the respective costs for the two types of the techniques used.

NOTES The manuscript leans excessively toward theoretical evaluations of 'how things would be if: : :' and misses a discussion of several key points, which are named but not explored: depositional lags, outlier dates, examples of sample size effect on the radiocarbon date uncertainty as applied to a real core. This happens because the authors chose to (a) treat all their dates as equally good/likely; (b) use the 'best age estimate' for the sequence using everything at once, that is, they combined 3 varve count series + miniature+ regular + graphitized + gas source radiocarbon dates to make a single 'best age estimate'. No wonder there are no outliers if all these things are bundled together.

The reviewer is correct: our discussion of outlier ages and depositional lags is rather theoretical, but this is simply due to the lack of evidence for outliers in our dataset. From our point of view, it is notable that all these types of age information agree without any outliers that don't fit with the other age information. We find no statistically significant *evidence that any single age is an outlier. All ages from within a single level have overlapping 95% confidence intervals. We can envision several ways to test whether the varve count agrees with the ^{14}C ages. However because the varve count is floating, all of these methods rely on ^{14}C ages in some way, which does restrict our ability to detect outliers. If we tie the varve count to the combined ^{14}C ages of the uppermost dated level (732.5 cm), we find that the 95% confidence intervals of all ^{14}C ages overlap with the varve count age estimate when we consider the uncertainty of the varve count as well as the uncertainty of the age of the tie point. However, the selection of this tie point is arbitrary. Another approach is to use least squares minimization to minimize the offset between the varve count and all of the ^{14}C ages. We did this using the median calibrated age from combined ^{14}C ages within each dated level. We find that the 95% confidence intervals of every ^{14}C age overlap with the varve count (in this case they overlap even without considering the uncertainty of the varve-count-based age estimate). Please see the attached figure to see how these methods compare. Our OxCal V-sequence 'best-age estimate' yields the same result – all ^{14}C ages overlap with the median of this age-model. We will revise the manuscript to emphasize that the floating varve chronology is consistent with all of the radiocarbon ages.*

As a reader, going from the Introduction to Discussion I expected to see the Figures showing step by step how overlapping varve-based chronologies look like first and how their cumulative error changes with depth, then how a certain number of graphitized regular ages help improving these chronologies and errors, and then how adding gas-source ages on the regular-size samples improves this chronology further, and then how adding gas-source ages on the less reliable miniature samples may or may not improve it even further. Instead, I see a single red line as 'best age estimate' from the very start and then 9 software-generated arbitrary age-depth scenarios. One does not need a sediment core to generate these latter graphs.

We thank the reviewer for the suggestion to include a figure that shows the varve count

directly compared to the radiocarbon ages; we plan to include the attached figure in the revised manuscript. The suggestion to reorganize the study to start with varve counts, then use graphitized ages to improve the varve count, and then add gas-source ages takes the study in a different direction than we had intended. This would seem to help answer a specific question - how does adding 31 miniature gas-source radiocarbon ages improve an existing chronology based on varve counts and graphitized radiocarbon ages? However, most sediment cores will not be sampled and dated in the same way as our core, and we are using the varve counts to help develop the chronology rather than using the ^{14}C ages to check the varve counts. Our goal was to address more widely applicable questions about the tradeoffs between the number of radiocarbon ages and their precision. We do, however, show in Figure 2A how age models constructed using only gas-source ages or only graphitized ages compare to an age model with all radiocarbon ages, which is along the lines of the reviewer's suggestion. Finally, while it is true that one could simulate age-depth scenarios without a sediment core, our simulations are directly informed by empirical data from our core, providing a direct connection to real-world application, which we feel is valuable.

Depositional lags for organic fragments are discussed in a purely theoretical way. There appear to be three different varve chronologies, why not show each one of them and see which dates support which one (if any)? A test for potential age outliers would be more robust in this case.

The three varve counts are replications by 3 different people which we use to establish the master varve chronology and its uncertainty. The uncertainty is determined based on the agreement of all 3 counts for each individual layer using methods described by Bonk et al. 2015, and Åżarczyński et al. 2018. With the uncertainty range shown, it is redundant to plot or discuss each replicate count individually. Additionally, when viewing a plot of the 3 counts it can be difficult to assess the extent to which the counts agree at the scale of lamina, which is critical to the count uncertainty. It is possible for two counts to include the same number of total varves, but disagree about the location *of those varves. This type of disagreement is included in our uncertainty estimate, but is difficult to observe in a plot.*

Supposedly, as admitted by the authors, the younger the portion of the studied sequence, the more robust is the varve-based chronology. Why not take advantage of this and have a closer look at the potential depositional lags in the most reliable upper portion of the record?

We are uncertain of the exact meaning of this comment. The reviewer might be suggesting that we take a closer look at the potential depositional lags in the portion of the core published in Bonk et al. (2015) and Åżarczyński et al. (2018), which we briefly discuss this in lines 304-311. Alternatively, the reviewer might be suggesting that we obtain more ages from this section of the core, which would move the study toward the topic of depositional lags. However, the problem of depositional lags has been considered by previous studies, and is not the main focus of this manuscript, though it is somewhat relevant in that greater dating density (enabled by gas-source) may assist with detecting outliers.

What if the varve-only age models were used to compare with gas only and/or graphite only ages? The importance of mass for the reliability of the dates is stressed a number of times by the authors, but their Figures are not informative enough to illustrate this. For example, when discussing age offsets, why not show symbols of different size somehow proportionate to sample mass in Figure 3 and provide respective error bars for each of the dating points?

We thank the reviewer for the suggestion about Figure 3. This will be modified to show symbols of different size based on the mass of the sample. Including error bars would make the figure rather cluttered, and essentially equivalent information can be gleaned from Figure 2B.

If the sample mass is so important for the age date and bigger is definitely better (as shown in Fig.1), then is it really a good approach to consider all the dates

equal in constructing the 'best age estimate'?

It is true that bigger sample masses yield more precise dates. However, in our view, that alone is not a reason for removing an age from an age-depth model. We believe it is best to consider all dates as equally valid rather than removing dates without very strong reasons for doing so. Large analytical uncertainty does not indicate a date is invalid or unhelpful. The age modelling routines take into account the differences in precision through the use of probability density functions and thus give more weight to the more precise ages.

If the authors found room for nine simulated graphs in the manuscript, I think it would be beneficial to see two-three age-depth graphs using best dates, small-sample dates, and then all dates for comparison.

Age-depth models using gas-source ages (miniature samples), graphitized ages, and all ages are already included in Figure 2A.

The section 4.4 "Recommendations: : :" is a disappointment as it states a number of trivial basic things about radiocarbon dating, which can be found anywhere and which are not supported by the data the authors present. For example: "we are convinced: : : that miniature samples: : : are better than bulk" – convinced based on what? There is no data presented to support this level of certainty. Indeed, it would have been a really nice test if they were to analyze at least couple bulk samples from the same horizons to see how they compare with those on sieved fragments. "Dating small amounts: : : is preferable to pooling : : :", "a rule of thumb is: : :" – again, there are no data in the paper supporting this conclusion. It appears that these didactic statements are pasted from elsewhere.

These are fair points: most of our recommendations are based on existing literature rather than data in this paper. We plan to revise this section to make the reasoning more clear. Even if some of these recommendations come from sources outside the *data in this paper, they are relevant to the topic of the paper – the usefulness of miniature sample masses for radiocarbon dating.*

"If ages do not agree well : : : youngest ae is most likely to be correct" - what about applying this principle to their own data set and showing how it works out in their studied portion of the lake record? It appears that in the paper the authors cite, Bonk et al. (2015) did just that and identified a number of outliers.

The key to the quoted recommendation is "If age results do not agree well...". In our case, the ages agree within the expected uncertainty. We cannot say that the older ages in our dataset are older due to a depositional lag, they are likely older due to the random variation expected with radiocarbon measurements (as defined by the analytical uncertainty). The difference between the Bonk et al 2015 study and ours is that their radiocarbon ages are generally more precise (larger sample mass), and their varve chronology is linked to the surface, which greatly reduces uncertainty and allows for easier detection of outlying ^{14}C ages.

Finally, the argument of 'cost reduction' for gas-source ages as compared to graphitized samples is used a number of times in the manuscript. Indeed, costs are lab-specific, however, it would be of interest to have at least some estimate in % since the authors repeatedly bring this issue up themselves.

We will revise section 4.4 to read as follows: Injecting CO2 into the AMS rather than generating graphite and packing a target substantially reduces the effort to analyze a sample following pre-treatment. Each sample also spends less time on the AMS when introduced as gas rather than graphite. These advantages are partly offset by the additional attention needed during gas source measurements. How these differences translate to per-sample costs depends on the pricing structures implemented in each lab. Cost estimates from two MICADAS labs at the University of Bern and Northern Arizona University range between around 15 and 33% lower costs for gas-source measurements compared to graphitized samples.

I suggest substantial revisions, not "major" but at the same time not "minor" or technical either.

Interactive comment on Geochronology Discuss., https://doi.org/10.5194/gchron-2019-19, 2019.

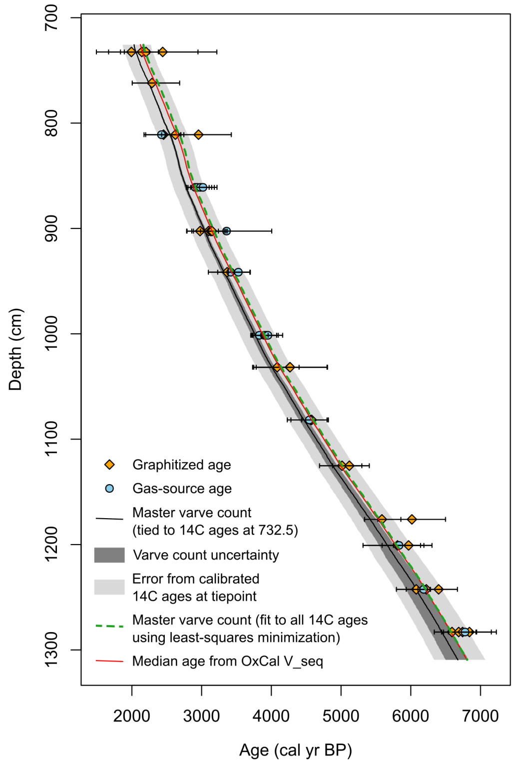

Fig. 1. All radiocarbon ages and their 95% calibrated uncertainties plotted versus the varve count. The gray bands show the varve count tied to the combined calibrated age of the uppermost 14C ages (at 732.5