Reply to Referee #2. Our replies are indented.

Main referee comments:

Lougheed et al. address the influence of bioturbation in ocean sediments on the accuracy of sediment ages determined by 14C dating. Accurate ages are relevant for global correlation of sediment records and thus for a better understanding of the interactions of oceanography, climate, and the carbon cycle in the past and future. As such the subject of the paper is highly relevant. The modelling approach using the established SEAMUS model is clearly described and the many processes that can influence the 14C concentration of a sample of foraminifera picked from a discrete sediment layer are indicated.

> We thank Referee #2 for their kind words and the time they have taken to provide their insight on our manuscript.

It should be noted that the model results presented have been obtained under very idealized conditions, as clearly stated in the model assumptions. To demonstrate the value of SEAMUS in the real world it would be good to see results for common sized mono-specific foram samples (0.1 to 1.0 mg C, ∼30 to 300 shells) selected from a sediment section and modelled with sedimentation rates and species abundance as well as local surface reservoir age varying over time according to a realistic local scenario. [...] In conclusion, I agree with the authors that bioturbation needs to be considered when interpreting 14C ages obtained on planktic foraminifera samples but I find the present paper too far removed from reality to be published in Geochronology. In its present form it seems more appropriate for a modelling-oriented journal.

I would like to encourage the authors to invest a bit more time in this interesting work by running their model under (more) realistic conditions, as mentioned above, and comparing the results for the ^{14}C age that will be measured with the quantity sought in sediment dating, i.e. the time of deposition of the sediment section and the bulk of the foraminifera in it. Such a paper would be highly suitable for Geochronology and the 14C dating of deep-sea sediment archives.

> We respectfully disagree with the above comments. One of the major advantages of computer modelling analysis of complex systems is that a model can be used as an investigative tool to analyse the influence of an isolated parameter upon an entire system. Consider for a moment some of the classic modelling papers which have allowed us to better understand the Earth's climate system by keeping all climate parameters constant except for the one of interest (e.g. pCO_2). One could equally argue that such classic modelling papers are "removed from reality", but it is precisely for that reason that those papers are of great importance and why they provide valuable new insights into the functioning of a complex system.

> We also follow such a classic modelling approach. In our case we have taken advantage of computer modelling to construct an ideal experimental design whereby we can evaluate how the current ^{14}C state-of-the-art would work in the case of best-case conditions. Since these best-case conditions do not exist in reality, a computer modelling environment can uniquely be used to create such a best-case scenario, whereby the long-established and accepted understanding of bioturbation is incorporated into the model. This approach allows us to test the accuracy of the current ^{14}C dating state-of-the-art applied to deep-sea sediments on the most fundamental level. If it is demonstratively shown that the current state-of-the-art functions sub-optimally even in best-case conditions, then very important questions are raised, which are highly relevant for the readership of *Geochronology*.

> Of course we agree that in the case of "real-world" deep-sea sediment all input parameters do indeed exhibit constant temporal variation (which SEAMUS is capable of modelling), thus

increasing the degrees of freedom of practical experimental design for real-world studies in the field, which subsequently places strong interpretive limitations on such real-world studies. If we were to increase the degrees of freedom of the experimental design in our model study to mimic real-world conditions and their associated interpretive limitations, we would be unable to provide any new insights that have not already been provided by field-based studies. A huge advantage of a computer model is the ability to overcome such limitations by constraining multiple input parameters, allowing for new insights into the fundamental functioning of a complex system to be found, which we believe we have done. For example, we have shown that even in the case of best-case conditions, the current ^{14}C dating state-of-the-art will produce systematic age-depth artefacts which are a result a misrepresentation of the ^{14}C age distribution of the sample by the current state-of-the-art, combined with the mixing of single foraminifera from differing periods of the Earth's history.

Finally, we note that the referee did not actually find a significant fault with the modelling processes performed in our study.

Other referee comments addressed below:

A critical issue to be addressed for the use of single forams, that now may become possible as mentioned in Outlook, is the variability in the isotopic signal of individual coeval foraminifera. During the lifetime of a single foram the 14C concentration of the water surrounding the foram may vary due to varying ocean-atmosphere exchange, turbulent mixing with deeper layers, planktic bloom, and change in depth of the foram as it ages. The natural spread in 14C concentration, δ13C, and δ18O in a population of contemporaneous foraminifera needs to be determined and compared with the magnitude of the paleoclimatic signals expected to see what Information may be obtained. To a lesser degree this individual variability also needs to be considered in deciding to what extent a finite number of shells is representative for conditions at the ocean surface. 300 foraminifera may be representative, for a sample of 30 shells it may be an uncertainty factor.

These are all very interesting details which we have often pondered ourselves, but go beyond this current study. Our study aims to evaluate the current state-of-the-art of ^{14}C dating of deep-sea sediments, by testing how the current state-of-the art performs in best-case conditions in deep-sea sediment. Including many more variable parameters in our simulation would increase the degrees of freedom of the experimental design and preclude us from making useful interpretations. Hypothetical questions about the intra-shell ^{14}C variability of single foraminifera as they transition through water bodies during their lifetime are well beyond the scope of our study. We refer the referee to many studies about changes in the ^{14}C activity of marine waters, i.e. studies about spatio-temporal changes in marine reservoir age.

In practical sediment dating, the aim is, generally, to establish the time of deposition of the particular sediment section by determining the 14C content of planktic foraminifera deposited coevally.

Respectfully, the oft-applied assumption that foraminifera retrieved from a given centimetre(s) thick interval of a deep-sea sediment archive were deposited at the sea floor coevally (*i.e.* at the same time) is one of the main assumptions we seek to re-evaluate in this manuscript.

Thus the demonstration that bioturbation may significantly affect the 14C content of planktic foraminifera in a sediment section (5.0 Conclusion) does not directly contribute to a better age determination of ocean sediments (6.0 Outlook).

Section 5.0 describes previously undescribed challenges to ^{14}C dating that we have highlighted in our study. Section 6.0 proposes future methods with which these challenges can be quantified when calibrating multi-specimen samples and/or overcome by using new analytical methods. We will make clearer that the Outlook section is describing potential future studies/remedies and not resolutions that we have produced in this study.

Of practical use would be modelling of the difference between the average planktic 14C content and the planktic 14C at time of deposition of a section.

We're not sure if we fully understand this comment. The ^{14}C activities (relative to 1950 CE) assigned to the simulated planktonic foraminifera do not change due to deposition.

Line 44: True difference in age is not the only possible cause of 14C age heterogeneity. Other causes as listed in lines 83-87 also come into play.

Indeed, as we mention in lines 83-87, as the referee points out. We will attempt to better foreshadow this in the text for the benefit of the reader.

Line 77: "14C history of the Earth" is too general. It is better to separate the atmospheric 14C history, which is largely global, from the oceanic 14C history, which isstrongly local.

Agreed, we will make this sentence more specific.

Line 90: The discussion does not differentiate between the probability distribution of the measured 14C concentration and that of the related 14C age although the latter follows nonlinearly, via e-log, from the first, which for old samples has significant consequences.

The ^{14}C calibration software embedded in SEAMUS is MatCal (Lougheed and Obrochta, 2016), which calibrates ^{14}C determinations in $F^{14}C$ space, so there is no trouble with calibrating older samples in the case of SEAMUS. We should indeed mention this somewhere for the benefit of the reader!

Line 118: 10^4 specimens represents ideal conditions compared with 30 to 300 foraminifera selected from the population of the 1-cm section.

We aim to investigate how well the current ^{14}C state-of-the-art performs under ideal conditions, and adding further degrees of freedom to our experimental design by including additional noise created by small and/or variable sample sizes would impede our aim.

Line 121: the primary parameter is F14C, an apparent 14C age follows from it. Although Marine20 is quite different from Marine13 beyond 14 ka, this is for the discussion of the technique, at present, not important.

Each single foraminifera within the simulation is first assigned a ^{14}C activity in conventional ^{14}C age, which is subsequently converted to $F^{14}C$. The reason for this approach is that we use the *Marine13* curve to assign ^{14}C activity, which is published by the IntCal group, who report activity as conventional ^{14}C age (i.e. in the downloadable "*Marine13.14c"* file). When it comes to assigning ^{14}C activity to single foraminifera, the two units (^{14}C age and $F^{14}C$) are readily convertible via a simple formula, so there are no consequences for our study.

Marine20 may indeed be different from *Marine13* for older parts due to the improved and extended Hulu Cave record, which we allude to in lines 225 to 230.

Line 130: There seems to be confusion on the meaning of blank value. 14C convention is that only definitive 14C values (measured minus background) that exceed twice their uncertainty should be given. This does not mean that foraminifera older than this limiting age/concentration have all the same 14C concentration. They don't; their 14C content keeps decreasing but we can no longer reliably measure it.

We thank the referee for making us aware that we need to describe this part of the method in a more clear way. We are of course aware that ^{14}C activity continues to decrease with time according to the principles of radioactive decay. The reason we assign, within the simulation, all foraminifera older than the blank value the same ^{14}C activity (in $F^{14}C$) as the blank value itself, is so that we can simulate the practical ^{14}C dating (AMS analysis) upon said foraminifera, i.e. the aim of our study. In our simulation the analytical blank value is set to 46806 ^{14}C yr BP (assigned as 0.0029477 $F^{14}C$). This represents the signal that older foraminifera will contribute to the measurement process in practice, irrespective of their age. So, for example, for a given virtual sample containing 1000 foraminifera of between 85,000 and 90,000 years old, we would calculate the virtual AMS date by taking the mean of all their individually assigned $F^{14}C$ values (the 'blank' value in this case), which would result in an AMS determination with a mean value of 0.0029477 $F^{14}C$, analogous to a real-world AMS determination of such material.

Line 155: Near the blank value the age uncertainty will be asymmetric and generally significantly larger than 200 years because not only the uncertainty in the measured14C but also that of the blank to be subtracted has to be considered.

As mentioned in a reply to a previous comment (Line 90), ^{14}C dates within the simulation are calibrated in $F^{14}C$ space using MatCal (Lougheed and Obrochta, 2016), so are not affected by such asymmetries.

As we stated to Referee #1, we are running a 'best-case' model evaluation of ^{14}C dating, including best-case AMS analysis with a very optimistic uncertainty of 200 ^{14}C yr for very old dates. We can re-run the simulations with an uncertainty of ~700 ^{14}C yr for very old dates, although it would have little bearing on our main conclusions.

Line 209: Note that one is usually seeking the time of deposition of the section and thus the 14C age of the foraminifera raining down at time of deposition. The bias of measured age relative to this will be towards older.

We don't fully understand this comment. In practical terms, ^{14}C is used in palaeoceanography to determine the average (calibrated) age of the foraminifera contained within an interval of retrieved sediment core, so that an accurate age can be assigned to whatever climate/oceanography proxy was also retrieved from foraminifera contained within the same interval. Is the referee commenting on the transit time of a dead foraminifera from the ocean water surface to the seafloor? In that case the transit time is mere weeks, see following study: doi:10.1038/ncomms7521

Line 245: Is the second decimal in 95.45 % relevant? Usually only one is given.

There is actually no reason why we give a second number after the decimal point. We are unaware of any decimal convention for reporting HPD intervals of calibrated ages. We can change to one decimal point if necessary, and considering that the calibrated ages are rounded to one year, two decimal places is probably indeed a bit optimistic anyway!

Line 287: Are the artefactually young 14C ages the result of assigning a constant "blank" 14C concentration to older foraminifera (see line 130 above)? Modelling could be changed.

We refer to our response to the comment for Line 130. Note that in the text we refer artefactually young "***measured*** ^{14}C age" and that the x-axis of Fig. 4 is labelled as "Discrete-depth ***AMS*** ^{14}C age". Consider a sample containing 70% foraminifera that have a ^{14}C activity greater than the blank value, and 30% foraminifera with a ^{14}C activity less than the blank value. In our simulation, the latter 30% will simply be assigned the blank value as their ^{14}C activity, thus biasing the mean sample activity of the sample calculated within the simulation towards a too high activity (= younger ^{14}C age). This is analogous to what would happen in a real-world laboratory determination of such a sample (what we seek to simulate in our study), because the laboratory cannot measure below the blank value. We will attempt to make this point clearer in the final version, because at the moment it does appear confusing to the reader.

Line 297: 1% contribution of 14C free carbon is equivalent with 1 % decay, meaning 80 years too old. In the age range mentioned here, this is well below the measurement uncertainty (i.e. fortunately negligible).

For other % contributions we refer the referee to Fig. 4.

Line 355: The statement that considering bioturbation could improve dating accuracy certainly is true. More realistic modelling is, however, needed to demonstrate the potential of SEAMUS to produce significant improvements.

We refer to the rest of our reply.

Thanks again to the referee for taking the time to review the manuscript and providing their insight.