We are grateful to the referee for their constructive feedback and the time they spent reviewing our manuscript. This helped us to improve the presentation, while results and conclusions remain unchanged. Below are our responses (in **bold**) to the referee comments (in *italics*).

Referee #2

1. Line 11: I feel that the term 'mystery interval' has become current without having a particularly clear meaning; it seems to be used to refer to a chronozone, for which there already is a name (Heinrich Stadial 1, etc…). Furthermore, the 'mysterious' part of the interval seems to be perceived differently by different people; is it the atmospheric radiocarbon decline, the proposed lack of marine radiocarbon activity increase, the entire 'mystery' of deglaciation? Not everyone shares the same notions regarding such 'mysteries', particularly regarding the marine radiocarbon inventory change, against which the term 'mystery interval' seems to have been directed. I would like to stick my neck out and suggest that this term has served its purpose in stimulating interest in a topic, and no longer serves a purpose for clear communication of a specific idea. I would therefore propose that the authors refer to other more clearly established chronozone designations, or even dates if these are trusted sufficiently.

We agree with the referee that using the term "mystery interval" to refer to the sharp drop in $\Delta^{14}C$ across Heinrich Stadial 1 ~17.5 to 14.5 kyr BP serves no purpose other than to stimulate interest. We will update the manuscript to be more precise, such that "mystery interval" is replaced by Heinrich Stadial 1.

2. Line 18: I think the word "more" can be dispensed with, here and elsewhere. One wonders: more than what?

Models allow us to investigate specific phenomena in more idealized settings compared to the "real world". However, we agree that, in this context, referring to such settings as "more idealized" rather than simply "idealized" is not very useful. The manuscript will be updated accordingly.

3. Line 40: here and throughout the manuscript I was not sure whether "millennial-scale" was a helpful designation, as it made me think of variability associated with Dansgaard-Oeschger events. Perhaps the term, or another such as "short term" etc…, can be defined clearly when first used?

The primary focus of this work is on the specific mechanisms responsible for variations in atmospheric $\Delta^{14}C$ on millennial time scales (i.e., time scale of thousands of years). We do not attempt to resolve more abrupt climate perturbations such as Dansgaard-Oeschger warming events, which is noted in lines 637-641 of the original manuscript.

To avoid confusion, we will add a note of caution in Sect. 2.4 when we introduce the carbon cycle scenarios considered in the model runs.

4. Line 60: same thoughts as above regarding the term "mystery interval"; if it is coincident with HS1, then we should use that term instead I think. At the time of the Broecker and Barker (2007) study there was proposed to be a lack of evidence for a radiocarbon depleted ocean interior at the LGM, and a subsequent increase in its radiocarbon activity; however, this is arguably no longer the case.

This comment has already been addressed in our response to comment #1.

5. Line 67: probably best to be more specific, e.g. "…used only high accumulation sites, and square barrel gravity cores with minimal sediment disturbance.."

We agree with the referee that it would be valuable for the reader if we elaborated on the coring and sampling methods that minimize the influence of drilling disturbance. This will be done in a revised manuscript.

6. Line 72, last sentence of the paragraph: I don't mean to suggest that ther is anyh I gotta incorrect about this sentence, but I found this to be an odd way of phrasing things. To me there is one question, "why was atmospheric radiocarbon activity so high during the last glacial (including well after the Laschamp excursion)", which entails a subsidiary question, "how much did production changes contribute to this elevated atmospheric radiocarbon activity".

We agree with the referee that it is unnecessary to make a distinction between the contribution of production changes to high glacial $\Delta^{14}C$ levels and their contribution to the deglacial $\Delta^{14}C$ decline. Our goal was to remind the reader that only if estimates of past changes in ^{14}C production are robust can one improve assessments of the relative importance of the two fundamental mechanisms responsible for glacial-interglacial $\Delta^{14}C$ changes (i.e., production and carbon cycle changes).

7. Line 95: I would say that the time required for ocean ventilation is not "up to", but rather "over" 1000yrs. Perhaps Primeau (2005) can be referenced for this. 8. Line 100: multi-millennial timescales?

While the ventilation time scale for the deep ocean is typically of order 1000 years, we note that the deep ocean ventilation time scale can exceed 1000 years, as demonstrated by the modelling study of Primeau (2005). This time scale depends on which Ocean General Circulation Model and tracer was used to predict the time scale of the penetration of water from the surface into the ocean interior.

9. Line 106: Andrey Ganopolski would disagree (see Ganopolski et al., CP, 2017). Perhaps this statement should be modified to say that it is currently not possible to do so without the use of any parameterisations of key processes, or something more specific?

While Ganopolski & Brovkin (2017) reproduce the overall trends and more general features of glacial–interglacial variability of climate, ice sheets, and atmospheric CO_2 concentration using only orbital forcing to drive the CLIMBER-2 model, the finer-scale temporal dynamics of the simulated CO_2 evolution do not match the reconstructions. In particular, the model fails to simulate the correct timing of the deglacial CO_2 rise. In addition, the model underestimates the magnitude of the deglacial decline in atmospheric $\Delta^{14}C$. Therefore, we think it is reasonable to conclude that models cannot yet reproduce climate and atmospheric CO_2 variations on the basis of orbital forcing alone.

10. Line 113: this sentence seems to suggest that the main proposals for explaining glacial-interglacial CO2 involve exchanges with the solid earth, but this is not really true. Arguably, as has been sketched out many times before, including in a recent review (Galbraith 2020), the "ingredients" for glacial-interglacial CO2 change are well accounted for, it is their 'calibration' and organisation within an orbital pacing framework that remains elusive.

We agree with the referee. The text will be modified to highlight the role of ocean-based physical and biological mechanisms in explaining the glacial-interglacial variations in atmospheric CO_2, and to clarify that what is missing is a single framework in which these mechanisms are linked to each other in a predictable manner under the influence of orbital forcing.

11. Line 118: in idealised settings..

This comment has already been addressed in our response to comment #2.

12. Line 122: here and throughout the manuscript it would be best to suffix D14Catm, so that we know what reservoir is referred to.

We agree this notation would be useful for the reader and will apply it in a revised manuscript.

13. Line 125: is it not more accurate to state that the production rate is inferred from an atmospheric radiocarbon budget, combined with a range of hypothetical radiocarbon and carbon cycle scenarios?

We agree with the referee it would be more precise to state that our model-based 50,000-year reconstruction of the ^{14}C production rate is based on an atmospheric

radiocarbon budget that is put together by forcing the Bern3D carbon cycle model with reconstructed changes in atmospheric $\Delta^{14}C$ and CO_2 as well as carbon cycle scenarios.

14. Line 159: air-sea equilibration times are very different, which is potentially important...

The air-sea equilibration time scale for $\Delta^{14}C$ by gas exchange depends in part on the gas transfer velocity, which is investigated in the sensitivity experiments presented in Sect. 3.1.3. These simulations demonstrate a modest response of $\Delta^{14}C$ of approximately 4-8% to a 100% reduction of the gas transfer velocity at the north (> 60°N) and south (> 48°S) poles.

15. Line 169: perhaps Stuiver et al. 1978 should be referenced.

We cited Stuiver and Polach (1977) in lines 35-36 of the original manuscript, but we see no reason why we should not cite them again in Sect. 2.2 as suggested.

16. Line 170: would it be clearer to state that DI14C is simulated, separately from DIC?

We agree with the referee. We will modify Sect. 2.1 and 2.2 to clarify that CO_2, $^{14}CO_2$, DIC, and DI^4C are all carried by the model, and are used to diagnose atmospheric and oceanic $\Delta^{14}C$.

17. Line 189: I wonder if this is not a major part of the whole problem with simulating atmospheric radiocarbon in the past? If the modern (pre-industrial) state is in fact far from equilibrium then this would mean that production rates are all miscalibrated. Why not explore the possibility that production rates are higher than required for equilibrium, e.g. due to ongoing equilibration of sedimentation following the deglaciation and early Holocene? It seems to me that the very conclusions of this study require that this be explored as a possibility. More specifically, and perhaps I am not getting this right.. we might expect that, following the expansion of the terrestrial biosphere during the Holocene (and the removal of carbon from the atmosphere-ocean system, causing a slow reduction of 'young' carbonate sediment output from the ocean), the radiocarbon inventory of the ocean and atmosphere should be on a slow disequilibrium downward trend, so that a higher radiocarbon production would be needed to get today's radiocarbon activity as an equilibrium state. Is that correct? Or is it the opposite? In any event, one has a sneaking suspicion that this sort of thing might be important here.

This is a very interesting point, but our results suggest that such a disequilibrium effect is of relatively minor importance. Firstly, disequilibrium effects are fully accounted for in the model simulations where atmospheric CO_2 and $\Delta^{14}C$ are prescribed (see Sect. 2.5 and 3.4), given that the transient time evolution is modelled. Here, there is a major mismatch between the reconstructed production rates and those diagnosed from our simulations (see Fig. 10 and 11). Furthermore, as shown in Fig. 8a, the mismatch

between reconstructed and modelled atmospheric $\Delta^{14}C$ at the preindustrial is on the order of a few percent and scaling the production records accordingly would not remove the mismatch in atmospheric $\Delta^{14}C$ during the last glacial period. We refrain from such *a posteriori* scaling as the mismatch in atmospheric $\Delta^{14}C$ at the preindustrial is likely related to the mismatch between observed and modelled atmospheric CO_2 (see Fig. 8b). What we will say here is that an incorrect preindustrial ^{14}C production rate would introduce a potential bias, leading to systematic underestimates (or overestimates) of atmospheric $\Delta^{14}C$ values over time. However, increasing (or decreasing) the base level of our production rate would not fix the glacial $\Delta^{14}C$ problem, i.e., the persistent elevation of $\Delta^{14}C$ after ~33 kyr BP. This can also be understood by Fig. 9.

The uncertainty in the preindustrial production rate is on the order of 15% due to the uncertainties in the preindustrial ocean radiocarbon inventory (see Roth and Joos, 2013, Sect. 3.2). This potential systematic bias was not considered by our model simulations as it would not change the temporal evolution of atmospheric $\Delta^{14}C$; it would only lead to systematic deviations from the results presented in Fig. 7 and 8, moving the various mismatches with the reconstructions to other parts of the $\Delta^{14}C$ record.

Finally, the preindustrial ^{14}C production rate Q of 1.66 atoms cm^{-2} s^{-1} that is diagnosed at the end of the preindustrial spin-up agrees reasonably well with independent estimates from production rate models, e.g., Masarik and Beer (1999, 2009) (Q = 2.05 atoms cm^{-2} s^{-1} for a solar modulation potential of 550 MeV) and Kovaltsov et al. (2012) (Q = 1.88 atoms cm^{-2} s^{-1} for the period 1750 to 1900 AD), and from Roth and Joos (2013) using an earlier Bern3D-LPX model version (Q = 1.75 atoms cm^{-2} s^{-1} for the period 1750 to 1900 AD).

18. Line 220: "…levels, given available 14C production scenarios."

We agree with the referee it would be more precise to state that what we are interested in investigating is the extent to which changes in the ocean carbon cycle could explain high glacial $\Delta^{14}C$ levels, given available reconstructions of past changes in ^{14}C production.

19. Line 244: Why was benthic d18O chosen? It is a smooth, slow function that lags behind most of the climatic processes that were important for the carbon cycle. Although it might seem circular, I don't think it is any more ad hoc to scale these parameters to atmospheric CO2 instead.. having rapid jumps in HS1 and the YD, and a faster change than benthic d18O, might help with getting the deglacial CO2 change 'right' (for parameterised reasons).

We agree with the referee that a different scaling approach would be preferential when addressing the last glacial termination as benthic $\delta^{18}O$ lags the rise in atmospheric CO_2

and temperature as shown by Shackelton (2000). However, as our primary focus is on the last glacial period, a different scaling, e.g., by CO_2, would not change our conclusions.

20. Line 254: the cited study is based entirely on the 'plateau tuning' approach, which may be questioned. Perhaps best to also cite Skinner et al. (2017) who showed that the LGM ocean was 'older' pretty conclusively with a range of other data.

We agree with the referee that Skinner et al. (2017) would be a good study to cite here.

21. Line 283: It seems crucially important to me that the 10Be and 36Cl flux records from the ice cores are NOT consistent with the final age scale that they are all placed on. As far as I can tell from Adolphi et al. (2018), the ice core data were converted to fluxes based on each ice core's individual age scale, and then they were all placed on the GICC05 age-scale, whereas Channell et al. (2018) argued that this age scale implies very different fluxes. Surely the ice core cosmogenic nuclide data ALL need to be placed on the same age scale and THEN the fluxes should be calculated and 'stacked'. I think this is a really crucial thing, and I am really confused as to why the specialists working with these isotope records take a different approach that surely produces incorrect fluxes. A basic test I would propose is: are the individual ice core flux records consistent with the accumulation rates that are implied for each ice core by the GICC05 age scale? If not, they need to be corrected, surely. I suspect this will only make matters worse for reconciling everything, but it is still important to consider carefully.

We are afraid that there has been a misunderstanding. The referee is correct that all time scale revisions impact ice-core accumulation rates and hence fluxes. We want to point out, however, that, as described in Adolphi et al. (2018) (Sect. 3.1, first paragraph), all ice cores were first placed on the same time scale (GICC05) before fluxes were calculated. Channell et al. (2018), on the other hand, describe the differences that arise from using the old ss09sea time scale (where accumulation rates are based on an empirical relationship with $\delta^{18}O$) instead of GICC05 (where they are based on the annual layer count) – so this does not apply to the record by Adolphi et al. (2018). And yes, as demonstrated by our results, using the GICC05 accumulation rates does make it more difficult to reconcile ^{14}C and ^{10}Be as compared to the ss09sea accumulation rates. As mentioned in lines 70-72, ice-core accumulation rates remain the largest source of systematic uncertainty in the ^{10}Be-based production rate estimates. However, the largest systematic uncertainty in the calculation of accumulation rates comes from the correction of layer thinning through ice flow modelling, which is a slowly varying function of depth, and hence is relatively insensitive to minor corrections of the time scales themselves.

22. Line 355: note again that this conflicts with the premise that the modern state is at equilibrium!

This comment has already been addressed in our response to comment #17.

23. Line 448: my intuition tells me that air-sea has exchange may have a small effect, but depending on the circulation state. Is it not possible that changes in air-sea exchange might combine non-linearly with particular changes in the circulation geometry?

As noted in lines 448-449, air-sea gas exchange has only a small effect on atmospheric CO_2 as compared to ocean circulation, given that the time scale of deep ocean ventilation (of the order of several hundred years to 1000 years or more) is much longer than the time scale of air-sea equilibration for CO_2 by gas exchange (approximately one year). In other words, the rate limiting step that determines the kinetics of the oceanic uptake of CO_2 is ocean circulation, not air-sea gas exchange. We will clarify this point in the third and fourth paragraphs of Sect. 3.1.3.

24. Line 460: Although I see why the authors try to wiggle free from resolving the deglacial CO2 problem, I think it is it entirely possible to set it aside, and I also think it is basically not true that the study deals only with the glacial portion of the record. It is the glacial versus interglacial amplitude of atmospheric D14C that is of concern, and therefore the change across the deglaciation is entirely relevant! In fact, as suggested below, I would propose provocatively that this study shows that atmospheric radiocarbon can be explained reasonably well up until the deglaciation, and that it is the modern radiocarbon activity that defies explanation. I wonder what the authors think of this contention.

What we have tried to demonstrate with this work, especially by the analysis shown in Fig. 9, is that although models are able to reproduce successfully the high glacial $\Delta^{14}C$ levels associated with the Laschamp (~41 kyr BP) event, it is very difficult to explain the persistence of relatively high $\Delta^{14}C$ values after ~33 kyr BP, given available reconstructions of past changes in ^{14}C production and extreme changes in the ocean carbon cycle. We think that this may be crucial for explaining the deglacial $\Delta^{14}C$ decline, as the model representation of the mechanisms responsible for high glacial $\Delta^{14}C$ levels will determine the carbon inventories of the different reservoirs prior to the deglacial $\Delta^{14}C$ decline. And yes, the model fails to simulate the correct magnitude and timing of the deglacial $\Delta^{14}C$ decline. But given that we did not attempt to reproduce accurately the observed glacial-interglacial variations in atmospheric CO_2 and $\Delta^{14}C$, this work seeks to highlight the persistent elevation of $\Delta^{14}C$ after ~33 kyr BP as a major outstanding problem in our understanding of the atmospheric $\Delta^{14}C$ record.

In other words, we can reach the amplitude of the Laschamp-related $\Delta^{14}C$ change, but we cannot sustain the high levels during the last glacial nor can we get down low enough or fast enough during the last deglaciation.

25. Line 537, the discussion of simulated B-Atm values: why do the authors not refer at all to published data for comparison? The compilation of Skinner et al. (2017) estimated, with the available data, that the global average ageing of the ocean at the LGM was 'only' ~689 14C years. This is relevant here, and indeed it would suggest that all of the model scenarios produce rather extreme outcomes as compared to available data.

We agree comparison with measurement- and model-based estimates of radiocarbon reservoir age offsets from, e.g., Skinner et al. (2017) and Butzin et al. (2017), is a missed opportunity. It was a sacrifice made to reduce the length of an already very lengthy manuscript. Nonetheless, some intriguing points can be made by such a comparison, so we will incorporate it into Sect. 3.3 and Fig. 8.

Comparison of our LGM B-Atm age offset estimates from runs CIRC, VENT, and VENTx (range of 3682 to 3962 ^{14}C years) with the compiled LGM marine radiocarbon data of Skinner et al. (2017) demonstrate that the carbon cycle scenarios are extreme, although it should be noted that they consider a wider depth range (~500 to 5000 m) of the ocean than we do. Skinner et al. (2017) predict a global average LGM B-Atm value of ~2048 ^{14}C years, an increase of ~689 ^{14}C years relative to preindustrial. Turning our comparison to surface reservoir ages, we note that our global average LGM surface reservoir age of ~1132 ^{14}C years from runs VENT and VENTx is comparable to the ~1241 ^{14}C years obtained by Skinner et al. (2017) for the LGM. The model-based estimates of surface reservoir age from Butzin et al. (2017) indicate a much lower LGM value of ~780 ^{14}C years, and values ranging from 540 to 1250 ^{14}C years between 50 and 25 kyr BP. Note that these estimates are based on model-simulated values between 50°N and 50°S. If the polar regions are included in the calculation (see Fig. 8c), their surface reservoir age estimates become comparable to our glacial values (range of 911 to 1354 ^{14}C years), and between about 34 and 22 kyr BP can exceed them, including even those from model runs VENT and VENTx, unless atmospheric $\Delta^{14}C$ and CO_2 are prescribed (dashed colored lines in Fig. 8c). Interestingly, this is also roughly the time period where our deconvolutions of the IntCal13 and Hulu Cave $\Delta^{14}C$ records give production rate estimates that are about 17.5 percent higher than the reconstructions, which indicates at the very least this is an important piece of the puzzle of the glacial-interglacial $\Delta^{14}C$ problem, given that the effect of upper ocean stratification and/or sea ice on air-sea gas exchange is particularly important for surface reservoir ages.

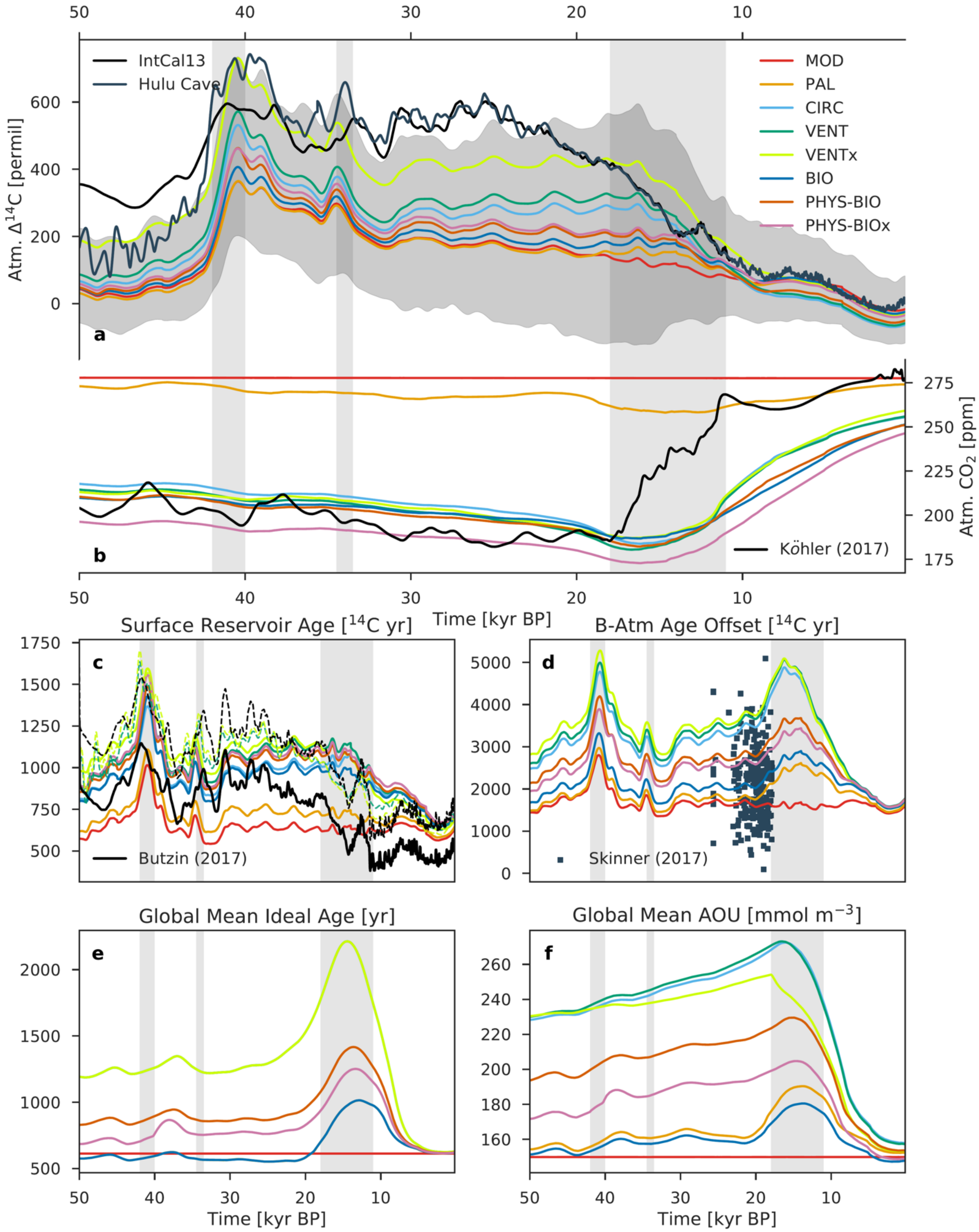

Fig. 8. Modelled records of atmospheric (a) $\Delta^{14}C$ and (b) CO_2, compared with their reconstructed histories (black and dark blue lines). Also shown are modelled records of the global average (c) surface reservoir age and (d) B-Atm ^{14}C age offset, compared with a recent compilation of LGM marine radiocarbon data (dark blue squares) by

Skinner et al. (2017) and model-based surface reservoir age estimates between 50°N and 50°S (solid black line) and across all latitudes (dashed black line) from Butzin et al. (2017), as well as (e) ideal age and (f) apparent oxygen utilization (AOU). Colored lines show the results of model runs using the mean paleointensity-based ^{14}C production rate and the eight different carbon cycle scenarios described in Sect. 2.4 and Table 1. The gray envelope in (a) shows the uncertainty (2σ) from all production rate reconstructions and carbon cycle scenarios, providing a bounded estimate of $\Delta^{14}C$ change. The dashed colored lines in (c) show the surface reservoir age results from VENT and VENTx where atmospheric $\Delta^{14}C$ and CO_2 are prescribed. Radiocarbon ventilation ages are expressed here as radiocarbon reservoir age offsets following Soulet et al. (2016) which are used extensively by the radiocarbon dating community.

26. Line 545: I think it is worth specifying in what ways these indirect methods are also potentially inaccurate, due to different processes affecting e.g. oxygen and radiocarbon.

A comparison of modelled apparent oxygen utilization (AOU) with the model ocean's deep-water reservoir age (B-Atm age offset) is not meant to be taken as a direct comparison. The goal of showing the parallel occurrence of depleted ocean interior oxygen levels (i.e., increased AOU) was to provide the reader with additional (indirect) evidence that deep water ageing is occurring in the model runs that consider reductions in ocean circulation and air-sea gas exchange (e.g., scenarios CIRC, VENT, and VENTx). A significant reduction in deep ocean ventilation permits the enhanced accumulation of remineralized carbon in the ocean interior, and therefore the progressive consumption of dissolved oxygen, as well as an increase in the radiocarbon disequilibrium between the deep ocean and the atmosphere, due to a decrease in the rate of transport and mixing of younger (higher $\Delta^{14}C$) waters. These observations (increased AOU and increased B-Atm age offset) taken together suggest that deep water ageing is occurring. We will clarify this point further in the third and fourth paragraphs of Sect. 3.3.

27. Line 567: ..is a dedicated 'control knob', in the model.

We agree with the referee it would be prudent to clarify that air-sea gas exchange is a principal "control knob" governing atmospheric $\Delta^{14}C$ in a model framework.

28. Line 605: viewed as tentative, perhaps. The viewing is not tentative; the results are.

Agreed.

29. Line 676: is it worth stating by how much this polar bias would have to be in order to reconcile everything? Is that magnitude reasonable?

Interesting point. However, we would rather not discuss the polar bias further as we do not think that it can really reconcile everything. Firstly, the geomagnetic field reconstructions do not suffer from a polar bias and yet, cannot explain atmospheric $\Delta^{14}C$ either. Secondly, as shown in Fig. 7c, the difference between reconstructed $\Delta^{14}C$ and modelled ^{10}Be (or RPI)-based $\Delta^{14}C$ is changing over time and the largest changes of this difference occur between ~35 and 30 kyr BP and then during the last deglaciation, not during the Laschamp event as one might expect if these mismatches were due to a polar bias. Instead, production rates (as inferred from ^{10}Be and RPI) were relatively stable across these two periods. Hence, it seems difficult to explain the mismatch by the presence of a polar bias alone.

30. Line 703: in this paragraph the realism of the implied sea ice changes is discussed, but again no mention is made of what existing marine radiocarbon data imply. These are really important constraints to mention, surely.

This comment has already been addressed in our response to comment #25.

31. Line 726: I couldn't help but feel that the conclusion of the study might be more hard hitting if we had a more specific 'shopping list' of things that could help to resolve this puzzle. For example, constraining the global marine radiocarbon inventory change across the deglaciation, estimating any gradient in cosmogenic nuclide production across latitudes (i.e. polar bias, perhaps from tropical ice cores?), estimates of global carbonate/POC export rates (which already exist incidentally; Cartapanis et al., 2016; 2018), etc…

What may help to resolve the glacial radiocarbon problem is progress in several different areas. Additional records of glacial atmospheric $\Delta^{14}C$ would help to further refine the IntCal $\Delta^{14}C$ record. Cosmogenic isotope production records may be improved, e.g., by refining estimates of ice accumulation, by developing a better understanding of ^{10}Be transport and deposition during the glacial, by recovering additional long and continuous records from Antarctic ice cores and including marine ^{10}Be records, and by obtaining additional geomagnetic data. An expanded spatiotemporal coverage of $\Delta^{14}C$ of DIC in the surface and deep ocean would allow one to narrow the time scales of surface-to-deep transport and air-sea equilibration of $\Delta^{14}C$, carbon and nutrients, and thereby guide model-based analyses. Models should be improved to better represent the glacial cycles of carbon and radiocarbon, by taking into account exchange with sediments and the lithosphere, by better representing coastal processes, and by representing a wide variety of paleo proxies such as $\delta^{13}C$, Nd isotopes, carbonate ion concentration, lysocline evolution, and biological productivity proxies in a 3-D dynamic context. What is also missing are methods to quantify how the global ocean carbon inventory, which co-determines the $^{14}C/C$ ratio and thus $\Delta^{14}C$ value in the ocean, has changed over the last 50,000 years.

32. Table 1: it would be helpful to specify here which simulations have active sediments included. Incidentally, why was the rain ratio changed in one simulartion?

As mentioned in lines 209-210 of the original manuscript, all transient simulations are performed with Bern3D model configuration ALL, which is the atmosphere-ocean-land-sediment model configuration. Hence, transient simulations include the 10-layer sediment model of Heinze et al. (1999) and Gehlen et al. (2006). We will clarify this point in Table 1 caption.

As discussed in lines 267-276 and summarized in Table 1, the $CaCO_3$-to-POC export ratio was changed over time in model scenarios BIO, PHYS-BIO, and PHYS-BIOx in order to investigate the impact of biological carbon pump changes on atmospheric $\Delta^{14}C$. While changes in the $CaCO_3$-to-POC export ratio are important for achieving glacial atmospheric CO_2 drawdown, our model results demonstrate that biogeochemical changes alone (scenario BIO) do not lead to an improved simulation of high glacial $\Delta^{14}C$ levels as compared to model runs invoking only physical changes (i.e., changes in ocean circulation and/or air-sea gas exchange). This is clearly illustrated by Fig. 8 and 9.

33. Fig 3, caption: I think it is more mathematically correct to state <100m and >1500m, no?

Yes, this is a typo that will be corrected in a revised manuscript.

34. Fig 7, caption, line 1203: I think it would be helpful to state ". . .using the mean reconstructed palaeointensity.."

We agree it would be more precise to state that RPI-based $\Delta\Delta^{14}C$ is the difference between reconstructed $\Delta^{14}C$ and model-simulated $\Delta^{14}C$ based on the mean RPI-based ^{14}C production rate.

35. Fig 8: shouldn't all the simulated D14Catm traces start at the same value and end at different values? Although this might look nasty, it suggests a different outlook in my view. Incidentally, the outputs in plots c and d are obvious candidates for comparison with existing data (e.g. Skinner et al., 2019, 20176), perhaps for a future study if not this one.

Since different carbon cycle scenarios (and therefore processes) were used to force the model into a glacial state over a 50,000-year integration, during which the glacial drawdown of atmospheric CO_2 was achieved, the model runs start from different global $^{14}C/C$ distributions, and therefore different values of atmospheric $\Delta^{14}C$, at 50 kyr BP. The analysis presented in Fig. 9 effectively normalizes the various $\Delta^{14}C$ records so that they are comparable, using two different "corrections".

36. Fig 9: this is a fascinating figure, though I find it slightly problematic. First, what is the rationale for normalizing to the average D14Catm value 0-50ka? I think that plots a and b should be replaced with normalization to the final 'modern' value, and that plots c and d should be extended up to the present. The latter is surely important, as it shows how we (well, you!) can do a pretty good job at simulating the amplitude of D14Catm change in the glacial when tweaking all the model's knobs, but that we can't subsequently get the deglacial change to the modern value, just as we can't quite get the deglacial change in CO2. I feel this must be significant... I wonder what the authors think.

The reason for subtracting the mean value from the $\Delta^{14}C$ records shown in Fig. 9a,b was to remove the offset/trend and emphasize the fluctuations in the $\Delta^{14}C$ data about the overall trend. This is effectively an offset correction normalization. Here, we can see that all model runs do a good job of reproducing the magnitude of the Laschamp-related $\Delta^{14}C$ change, but none are able to sustain the high $\Delta^{14}C$ levels after the Mono Lake excursion or capture the sharp decline in $\Delta^{14}C$ during the last deglaciation. We agree with the referee that the $\Delta^{14}C$ records shown in Fig. 9c,d should be extended up to 0 kyr BP.

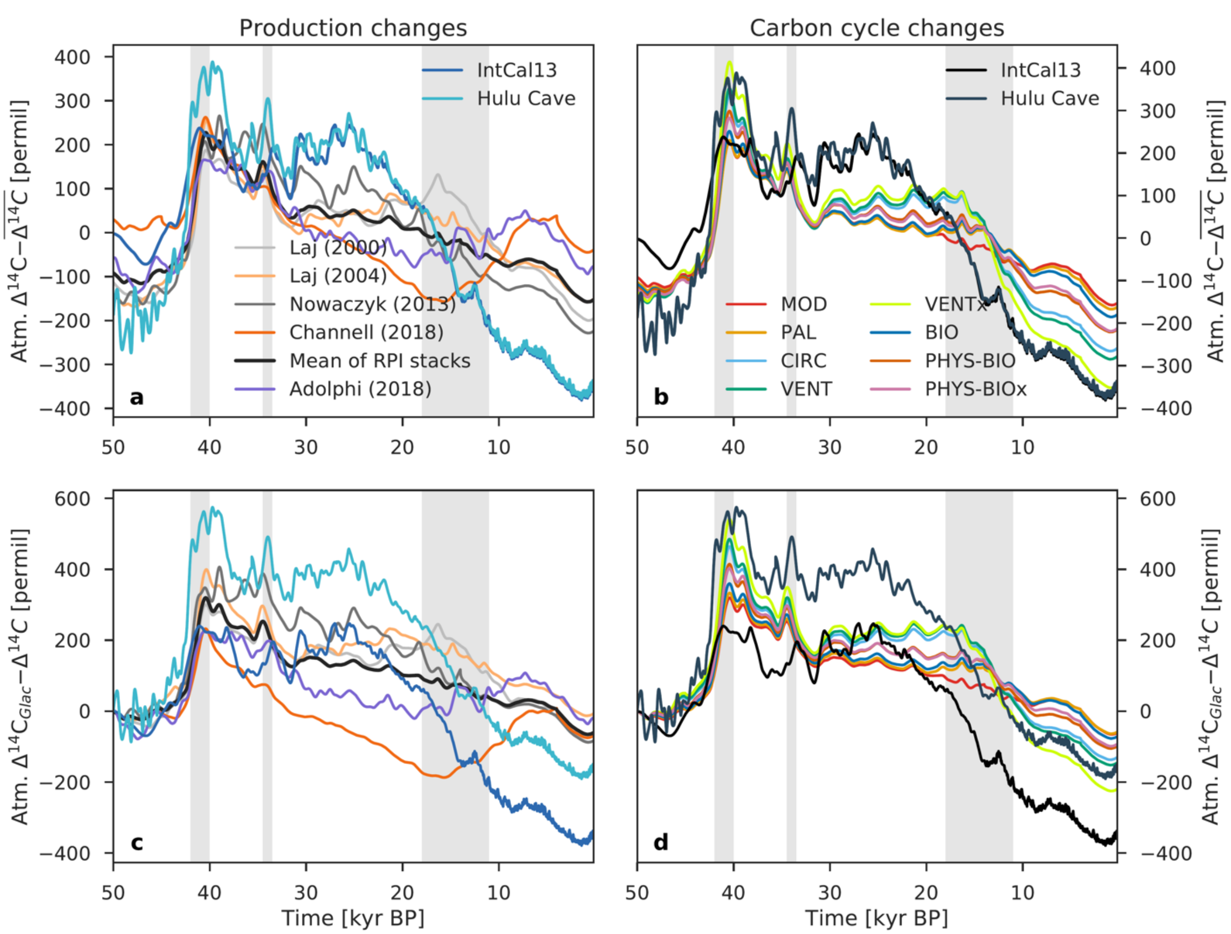

Fig. 9. Comparison of atmospheric $\Delta^{14}C$ variability caused by changes in the ocean carbon cycle (b, d) with production-driven changes in atmospheric $\Delta^{14}C$ using scenario MOD (a, c). For the analysis of carbon cycle changes, only the results of model runs using the mean paleointensity-based ^{14}C production rate are shown. The $\Delta^{14}C$ records in the upper panel (a, b) have been detrended by removing the mean, whereas the lower panel (c, d) shows $\Delta^{14}C$ anomalies expressed as differences relative to the $\Delta^{14}C$ value at 50 kyr BP. Three vertical light gray bars indicate the Laschamp (~41 kyr BP) and Mono Lake (~34 kyr BP) geomagnetic excursions, and the last glacial termination (~18 to 11 kyr BP).

37. Figure 10 and 11: I would suggest including a narrow plot at the base of each of these showing the offsets between simulated and observed values over time.

This is a difficult comparison to make as there is no one true (correct) target value. Nonetheless, we agree that such a comparison would allow the reader to more easily visualize the time periods where disagreement between the model- and measurement-based production rate estimates is highest, i.e., between 32 and 22 kyr BP.

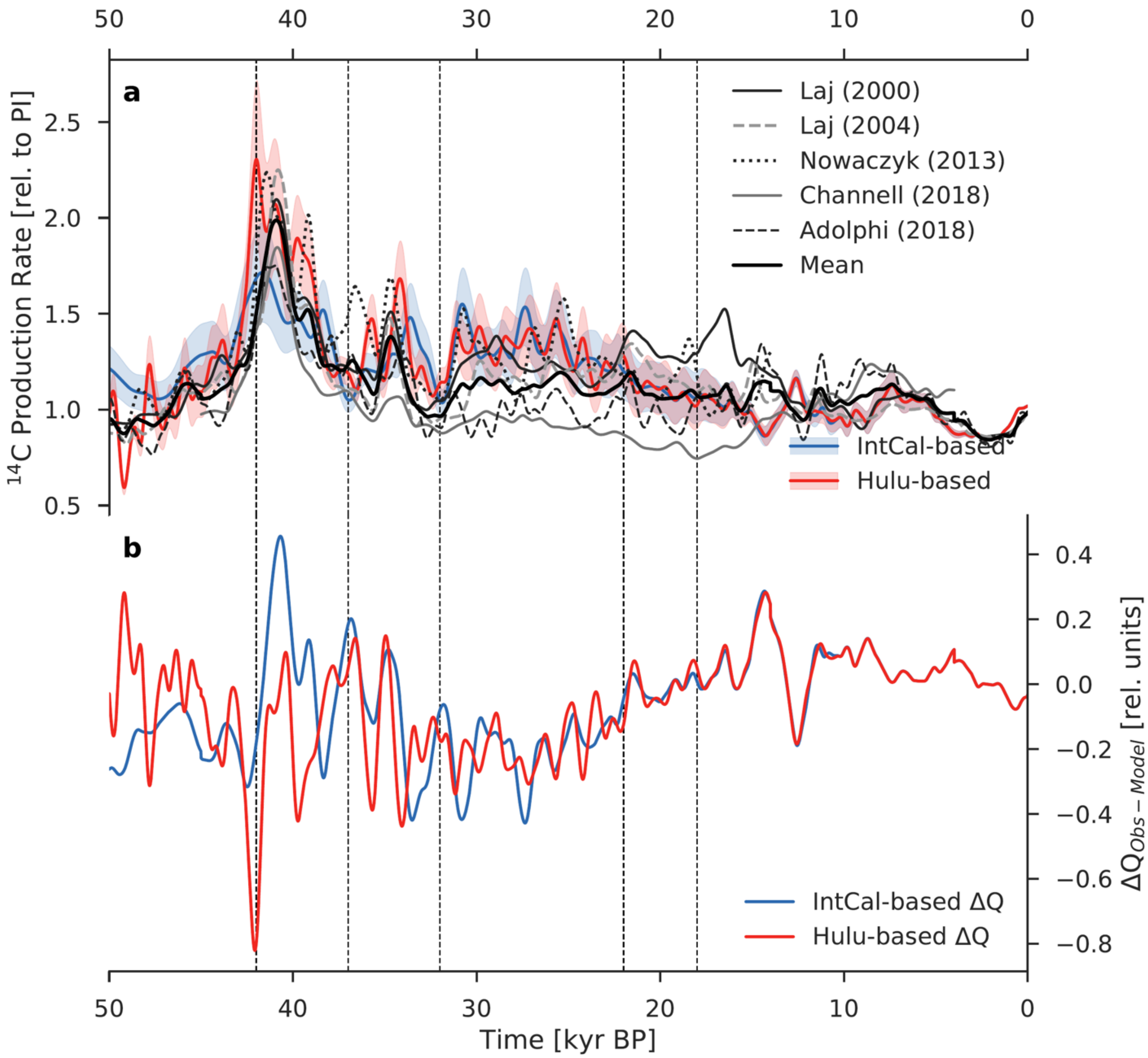

Fig. 10. Comparison of model-based estimates of ^{14}C production with estimates inferred from paleointensity data and from ice-core ^{10}Be fluxes. (a) Model-based estimates are determined from an atmospheric radiocarbon budget, which is put together by forcing the Bern3D carbon cycle model with reconstructed variations in atmospheric Δ^{14}C and CO_2 as well as seven carbon cycle scenarios. Results of model runs using the IntCal13 calibration curve are shown in the light blue envelope (2σ), whereas the light red envelope (2σ) shows the results obtained using the composite Hulu Cave (10.6 to 50 kyr BP) and IntCal13 (0 to 10.6 kyr BP) Δ^{14}C record. The heavy black line is the mean of five available production rate reconstructions: Laj et al. (2000), Laj et al. (2004), Nowaczyk et al. (2013), Channell et al. (2018), and Adolphi et al. (2018). (b) Difference between the mean reconstructed production rate and estimates inferred from IntCal13 (IntCal-based ΔQ; blue) and Hulu Cave (Hulu-based ΔQ; red) Δ^{14}C data.