General comments

This paper presents the inclusion of stable carbon isotopes, i.e. $^{13}C$, in the FAMOUS model. The authors have evaluated the effect of fractionation by air-sea gas exchange, biology and ocean circulation, and have tested different fractionation parameterisations for biology. This is a very useful development of the model that will be very valuable for paleo studies. This work is well presented and the reasoning is easy to follow. It has already demonstrated its usefulness by showing that the discrepancy between model results and data is likely to be due to biases in the simulated climate as well as the biogeochemical model.

*We would like to thank reviewer #2 for their feedback on our manuscript. We have considered all of the comments carefully and addressed them in turn below, with our responses in blue-italics.*

My main concern is that this model should not be used as it is for paleo studies but should be re-tuned, especially the biogeochemical module, as there are very large disagreements between simulated $\delta^{13}C$ and data. This is highlighted by the authors and re-tuning the model is clearly out of the scope of this study, but it might nonetheless be interesting to have a few additional sensitivity simulations to evaluate how much the results could be improved if the biogeochemical model was slightly modified, for example with a modified remineralisation, which could potentially help reduce the model-data disagreements.

*As noted by the reviewer, it is beyond the scope of the current study to retune the model, and as this project is no longer being funded, we are unable to conduct more simulations for inclusion in this manuscript. Nonetheless, we absolutely agree with the reviewer that tuning should be a priority, and will add their suggestion to conduct sensitivity studies with the biogeochemical model (in particular, modifying remineralisation) to section 3.4, where we discuss retuning the model.*

Specific comments

Abstract p.1 l. 7: do you mean “carbon isotopic ratios” instead of isotopic ratios?

Introduction p.2 l.2. The first sentence is almost the same as the first sentence from the abstract: maybe change it?

*In the abstract we were referring to isotopic ratios more widely. For example, $\delta^{18}O$ can be used as a tracer for density, temperature and salinity (Lynch-Stieglitz et al., 1999; Lynch-Stieglitz et al., 1999); $\varepsilon Nd$ can be used as a tracer for water mass provenance (Piotrowski et al., 2004; Rutberg et al., 2000); and $^{231}Pa/^{230}Th$ can be used as a tracer for the rate of overturning and scavenging (Marchal et al., 2000; Henry et al., 2016). However, for clarity, and to avoid repetition in the first sentence of the introduction, we will revise first sentence of the abstract to “Ocean circulation and the marine carbon cycle can be indirectly inferred from stable and radiogenic carbon isotope ratios ($\delta^{13}C$ and $\Delta^{14}C$, repsectively), measured directly in the water column, or recorded in geological archives such as sedimentary micro-fossils and corals.”*

p.2. l.2-10. This is not entirely true for $^{14}C$ as it also depends on radioactive decay: maybe you could say right at the beginning (after giving the percentages) that C14 is not studied here and only keep $^{12}C$ and $^{13}C$ in this part.

p2. l.9. You could give the complete definition of $\delta^{13}C$ mode explicitly as this is entirely on $\delta^{13}C$ inclusion it is worth reminding clearly the definition ($\delta^{13}C$ =. . .).

*In response to the above two comments, we will revise the manuscript to read “There are three naturally occurring carbon isotopes: the stable isotopes $^{12}C$ (98.9 %) and $^{13}C$ (1.1 %), and the radioactive isotope $^{14}C$ ($1.2\times10^{-10}$ %), which is also known as radiocarbon (Key, 2001). In this study,*

*we focus on the stable isotopes, with $^{14}C$ being discussed in detail elsewhere (Dentith et al., 2019, DOI: 10.5194/bg-2019-365), The relative proportions of $^{12}C$ and $^{13}C$ in a given oceanic pool (e.g. dissolved inorganic carbon, DIC, or particulate organic carbon, POC) are controlled by ocean circulation and mixing, and mass dependent fractionation during biogeochemical processes such as air-sea gas exchange (Lynch-Stieglitz et al., 1995; Zhang et al., 1995), photosynthesis (e.g. Sackett et al., 1965; Rau et al., 1989; Hollander and McKenzie, 1991; Keller and Morel, 1999), and calcium carbonate formation (Emrich et al., 1970; Turner, 1982; Ziveri et al., 2003). This is typically reported in delta (δ) notation, which is the heavy to light isotope ratio of a sample relative to a standard in per mil (‰) units:*

$$\delta^{13}C = \left( \frac{{}^{13}C/{}^{12}C_{sample}}{{}^{13}C/{}^{12}C_{standard}} - 1 \right) \times 1000. \quad (1)$$

*Oceanic $\delta^{13}C$ is primarily used to track individual water masses (Curry and Oppo, 2005), study past changes in the carbon cycle (e.g. de la Fuente et al., 2017), and investigate changes in ocean circulation on glacial-interglacial timescales (e.g. Spero and Lea, 2002; Campos et al., 2017). It has also been used to constrain air-sea gas exchange rates (Gruber and Keeling, 2001) and to estimate the uptake of anthropogenic carbon by the global oceans (Quay et al., 1992, 2003).”*

p.3 l. 5-8. You should also include the LOVECLIM model. What about Genie?
*Our intention was to provide illustrative examples of $^{13}C$-enabled models across a range of complexities as opposed to a complete list of all $^{13}C$-enabled models, but we are happy to add these additional examples in the revised manuscript.*

p.3 l.19. There is an arrow that should be deleted between “studying” and ”complex” .
*This will be deleted from the revised manuscript.*

p5. Line 16. Sea ice does not change salinity: this is out of the scope of the study but probably needs to be modified in the model. . .
*Acknowledged. We note that iceberg meltwater is included, however, and will amend the text as follows to incorporate the reviewer’s point: “…do not affect salinity distributions (an area for future model improvement), although an iceberg meltwater flux is represented in the model (Smith et al., 2008).”*

P5. Line 27-28. This seems at odds with what is said later. From what I understand from this paper both the physical model AND the biogeochemical model are responsible for carbon isotopes mismatch between simulation results and data and disentangling between the two is not done here.
*We can see the possible confusion/contradiction, and will revise the manuscript as follows:*

- *Section 2.1 (page 5, lines 27 to 28): “Previous studies have found that errors in biogeochemical simulations are largely driven by biases in the physical ocean circulation (i.e. inaccuracies in the climate or ocean model to which the ecosystem model has been coupled; Doney, 1999; Doney et al., 2004; Najjar et al., 2007). Thus, simulating carbon isotopes in a more complex ecosystem model would not necessarily yield substantially better results.”*
- *Section 3.4 (first paragraph): “In this contrast with earlier studies, we have demonstrated that the new carbon isotope scheme in FAMOUS is sensitive to biases in both physical and*

*biogeochemical processes. The simulated $\delta^{13}C_{DIC}$ distributions reflect known physical inaccuracies (such as over-deep NADW and weak convection in the sub-polar North Pacific Ocean) and have allowed us to identify previously undisclosed biogeochemical biases (e.g. in the representation of remineralisation). The new tracer therefore offers excellent potential as a holistic tuning target for recalibrating FAMOUS in the future."*

Results and discussion

p. 10 l. 15 / Figure 4. I would start with standard results before looking at the sensitivity experiments to be able to compare these sensitivity experiments with the standard one. So, on Figure 4 I would add the standard simulated $\delta^{13}C$ first as (a) and then the other 3 sensitivity simulations as b-d, which would also be more coherent with having the 4 simulations on Figure 5.

*We have chosen to present the results of the sensitivity experiments before the standard (std) results because this is, first and foremost, a model development study and we want to highlight that our new $^{13}C$ tracer is responding as expected to the physical processes (which have known biases) and the biogeochemical processes (where we have identified new biases). We think that beginning with the std results immediately raises questions regarding how they compare to observations, and consequently, why the simulated values are higher than observed. This would detract from the validation of the new isotope scheme. Therefore, we prefer not to restructure the results or figures as suggested.*

p.10 line 19. Is this a simulation that you actually did to verify this or just discussion? Please specify.

*We conducted this simulation to verify that our code is correct (i.e. that no carbon is being created or destroyed), but we did not include the results in our manuscript because they matched what we expected to see ($\delta^{13}C_{DIC}$ equilibrating at -6.5 ‰) and so are not very interesting. For clarity, we will revise this sentence to "If there is no fractionation during either air-sea gas exchange or photosynthesis, the ocean equilibrates at a uniform value of -6.5 ‰, in line with the atmosphere (simulation not shown)."*

p.11 line 8. Could you quickly remind the reader what this simulation is (to avoid looking for it earlier in the text)?

*We will revise this sentence to "When only biological fractionation effects are included (no-asgx-fract), $\delta^{13}C_{DIC}$ values in the surface ocean range between -7.65 ‰ in the eastern equatorial Pacific and -3.89 ‰ in the eastern equatorial Atlantic (Figure 4c), representing a shift of -1.15 ‰ to +2.61 ‰ relative to no isotopic fractionation."*

p.14 l.18. Could you test your hypothesis for the cause of the model-data mismatch due to the export ratio and remineralisation rate vs biases in ocean circulation by running additional sensitivity experiments? Testing the ocean circulation is probably more difficult, but modifying the export ratio and/or remineralisation to evaluate if this could have a large contribution to the mismatch is probably easier.

*There is a lot of scope for future tuning and exploration with this model set up, and we are excited to see this work carried forward. However, this effort is beyond the scope of this illustrative study. We will, however, add these suggestions to section 3.4.*