We much appreciate the reviewers’ detailed comments, constructive suggestions as well as the critical points. Below we list all the changes that will be made in the revised manuscript and our answers to reviewers’ questions, following the order of comments in the reviews.

In the following, reviewer comments start with a C: and are set in italics, while our responses start with a R: .

Review #1:

C: Ye et al. present in their paper the coupling of the process-based Model of Early Diagenesis in the Upper Sediment (MEDUSA v.2) to an ocean biogeochemistry model (FESOM v.2.1 with reduced model resolution + REcoM v.3 in reduced complexity) as well as first results from an pre-industrial simulation with prognostic atmospheric CO2.

The paper is clearly written and the coupling of a process-based sediment model to an Earth system model is outlined nicely and is an important step, especially when investigating the long-term carbon cycle.

Therefore, I support publication in GMD and hope the authors will find my few comments below helpful and consider their implementation.

R: We thank reviewer #1 for his/her support.

General comments:

C: While I understand that the coupling of model components and work on model code in general can be very time consuming and that the focus of this study is the documentation of this coupling, I still think it would be nice to see a little more results.

The authors state that they use a reduced complexity version of REcoM3, REcoM3p, that is targeted for paleo simulations. Also, ocean-sediment interactions become especially interesting when looking at long timescales, such as in paleo-simulations.

Therefore, I think it could enrich the manuscript to see some snapshots of the coupled model under, e.g., LGM conditions and update Tables 2 and 3 with the corresponding LGM values.

In such an exercise, carbon isotopes would be of interest as well...

R: The suggested simulations (LGM snapshots, carbon isotopes in sediments) are indeed also of interest to us and are in our research focus. However, both suggestions would require substantial additional model runs and tuning before being ready for publication. Furthermore, both suggested additional studies warrant to be analysed in more detail than can be done in a model description paper. We have chosen to submit the model description of the coupled model setup here to GMD in order to be able to have these other papers written up with less focus on the technical details and more on the scientific issues.

C: Further, during comparison with observational data, the coarse resolution of the PI mesh is mentioned as a limiting factor (for example l. 246-249, 280-289) and in the conclusions an outlook
is given to stay tuned for not only carbon isotopes but also higher spatial resolution. In my eyes, the manuscript could further benefit from including some results of this ongoing effort, if possible, and not save it all for future publications.

R: In response to the reviewer's comment, we also analyzed a simulation at a higher spatial resolution (126858 surface nodes, as in Gürses et al. (2023) and found that the POC sinking flux in shelf regions is a larger fraction (80%) of the global POC flux (close to the estimated 67-82% Muller-Karger et al. (2005); Burdige (2007)) than in the coarse-resolution (3140 surface nodes) runs presented in the manuscript, where this represented only 33% of the POC sinking flux. A paragraph about the comparison between runs in the coarse and fine resolution will be added in the model-data comparison and some results from the high-resolution run can already be found in Tab. [1]. As said in response to the previous comment, the results with carbon isotopes are in our view worth a separate paper, and could not be discussed here adequately.

Specific comments:

C: You could consider illustrating the carbonate chemistry by providing the governing equations for a better overview in the introduction.

R: We thank the reviewer for the suggestion and will improve the introduction by adding equations.

C: Figure 1: maybe add riverine + dust inputs to figure to close the loop?

R: These fluxes will be added.

C: Section 2.4.4 is not fully clear to me: Does the reported performance apply to REcoM3 or REcoM3p and to FESOM2.1 with reduced model resolution?

R: It applies to FESOM2.1-REcoM3p-MEDUSA with the reduced resolution which is presented in the manuscript. This will be stated more clearly in the revised version.

C: Section 3.1: you write that global vertical profiles in the model agree 'rather well' with observations from GLODAPv2. Could you give metrics? Are there differences between basins? Maybe add profiles for the different basins? Maybe also add section plots (model, obs, difference) for the main ocean basins.

R: This is a good point. Here we already show the zonal distribution of DIC, Alk, DIN and O2 in the Atlantic and Pacific ocean basins (Fig. [1]). More figures (including the basin-averaged profiles) and details of comparison with observations will be added in the revised version.

C: Mass conversation in the coupled model: It seems to me that this definitely needs to be addressed for longer (paleo-)simulations!

R: Yes, we totally agree with the reviewer that ensuring mass conservation is critical for longer paleo simulations. In the submitted version we mentioned that the increase of total Si in Rcoupled by 0.8% during the 1500 model years (i.e., 0.53% kyr⁻¹) is very likely caused by the temporal shift in fluxes during the asynchronous coupling procedure. We now further investigated the role of the coupling frequency for mass conservation by running a coupled simulation with a higher coupling frequency (every 10 years). Although only 300 years were finished before the submission of this reply, we observed an improvement in mass conservation: Si increases by 0.40% kyr⁻¹ during
Figure 1: Zonal averaged distribution of DIC, Alk, DIN and O$_2$ in the Atlantic and Pacific Ocean (mmol m$^{-3}$).
the simulation with the higher coupling frequency. Applying a high coupling frequency for long-term simulations, however, would require much higher computation time, making long experiments unfeasible. Therefore, we will calculate a correction factor of mass conservation based on the change of Si and do the correction to the bottom water concentrations. We will test this solution during revision and add a paragraph describing this solution and show results in the revised version.

**Technical corrections:**

C: l. 50: More complex scheme → More complex schemes

C: l. 111: 2) selecting of processes → 2) the selection of processes

C: l. 112: 3) writing the resulting → 3) the writing of the resulting

C: l. 127: were than partitioned → were then partitioned

C: l. 267: are reproduced in the model → are reproduced to some extent in the model?

C: l. 278: The opal belt in the equatorial eastern Pacific is smaller and less pronounced in the model than observed → not visible to me

R: That is true. We apologize that the pattern was changed by changing the color bar and we did not pay attention to this detail. This figure has been redrawn with a color bar adapted to make the low-concentration areas visible. (Fig. 2).

![Figure 2: Opal content in sediment (weight %) in R_{medini} in the submitted version.](image)

C: Caption figure 3: Horizontal averages of → Global horizontal averages of

C: Figure 6 and 7: label 0 seems misaligned in colorbar

C: Caption table 2: Note that the units here are Tmol year⁻¹, not Pg year⁻¹ → Note the different units in the table

C: Table 2: add observational estimates where available

R: Thank you for the corrections - all technical corrections will be implemented.
Review #2:

Summary:

C: Ying Ye and colleagues report on the coupling of the early diagenetic model MEDUSA2 with the ocean biogeochemistry model FESOM2.1-REcoM3. In order to be able to spin-up the model for multiple millennia (i.e., until the sediment-water interface (SWI) in the deep ocean is in steady-state), the authors use a lower horizontal resolution and reduce the complexity of the marine ecosystem model (i.e., simply using one generic zooplankton and detritus class instead of two for each) compared to a very recent model development paper (Gürses et al., 2023). Currently, most global Earth system models poorly represent the coupling between ocean and sediment biogeochemistry. Because the presented setup explicitly addresses the coupling of these domains, it can potentially be a very useful tool – especially for paleo-applications and simulations studying climate and marine biogeochemical feedbacks over multiple thousands of years.

While the model coupling itself represents a substantial contribution to Earth system modelling, the manuscript, unfortunately, lacks a proper evaluation of the performance of the new model setup and has several other weaknesses, omissions, and confusing parts. Not much new model development has been done for the manuscript – as the authors report on the coupling of two existing models. This would be okay if extensive experiments of the new coupled model evaluate its performance properly and show the added value of the new setup. Unfortunately, neither is done here. The authors only perform and show two experiments: one with the previous one-box sediment representation and one where they couple MEDUSA2. Both experiments are run under pre-industrial pCO2 for 2500 years (the new coupled configuration is only run for 1500 years from the previous model setup). Then, the authors compare some features at the end of both runs (Fig. 6, 7). The lesson learned from the results – apart from that patterns and values are slightly different - is unclear to me (e.g., is it a crucial improvement?).

Therefore, I cannot support the publication of this manuscript in Geoscientific Model Development. I hope my general comments will help to improve the useful coupling exercise and the evaluation of the new configuration. I suggest reconsidering the manuscript after major revision.

R: We thank the reviewer for her/his comments, particularly those concerning critical points. We have considered these inputs to improve the manuscript and summarize below the changes that have been done or will be done for the revised version.

Changes that have already been done:

1. analysis of the impact of model resolution on the ratio between the sinking fluxes onto the shallow vs. the deep-ocean sediment;
2. corrected and more model-data comparisons of the sinking fluxes onto the top sediment and the burial fluxes (Tab. 1);
3. analysis of the spatial pattern and temporal development of the deep-ocean concentrations, sinking fluxes onto sediment top and diffusive fluxes out of sediment (Fig. 3);
4. analysis of the contribution of OM degradation by O2 and NO3⁻ (Fig. 4);
5. experiments that vary the rate of oxic degradation for the OM class with a lower C:N ratio (Fig. 7);
6. a new experiment with increased coupling frequency to investigate the role of the coupling frequency for mass conservation.

More details of the changes already done are given to each of the reviewer’s corresponding comments in Section ”General comments”.

Changes that will be done in the revised manuscript:

1. rerun the \( R_{\text{coupled}} \) with the tuned degradation rate of the low C:N OM;
2. list parameters of FESOM2.1-REcoM3p and MEDUSA2 in tables;
3. complete model-data comparisons of DIC, Alk and nutrients (basin-averaged profiles);
4. experiments with mass correction;
5. all the detailed technical corrections.

General comments:

C: The new coupled model is not thoroughly evaluated. I understand that the lower FESOM resolution and the reduced ecosystem complexity in RecoM represent an entirely new configuration. Therefore, a more in-depth evaluation of the model than comparing it simply to global depth profiles of DIC, DIN, DSI, and ALK (Fig. 3 – the Fig. caption states these are horizontal averages?) and showing a time-series of atmospheric pCO2 (Fig. 4) is necessary. It is not too surprising that the model gives a good match to the globally averaged GLODAP data, considering that it is initialized with it and the model is only run for 2500 years. It is necessary to show that the model – particularly seafloor conditions and the sediment-water interface (SWI) fluxes – is properly spun up. To evaluate if the model is in steady-state, I suggest including time-series plots of global properties such as global mean ocean \( O_2 \), nutrients, DIC, DSI, ALK; export production, settling and burial fluxes of POC, CaCO3, opal, SWI fluxes of dissolved \( O_2 \) and nutrients – potentially also concentrations of dissolved species at the seafloor. Maps of NPP and basin-averaged meridional-depth distributions of DIN, \( O_2 \), and ALK (compared to observations) would further increase the credibility of the model.

R: We agree with the reviewer that the model-data comparison could contain more details and be more extensive. Maps of NPP were shown in Fig. 6 in the submitted version and a comparison to observations will be added in the revised version, along with basin-averaged meridional-depth distributions of DIC, Alk, \( O_2 \) and nutrients (Fig. 1).

Fig. 3 in the submitted version shows a comparison of \( R_{\text{seabox}} \) with observations. In this model setup, the one-layer sediment was used, not the coupled sediment model MEDUSA2. In the one-layer sediment particles are fully remineralised and dissolved, and no accumulation or burial is considered. In such an ocean-sediment system, a steady state can be reached within a much shorter time than in a fully coupled ocean-sediment system. Based on our experience with the model behavior, the deep-ocean distribution of DIC, Alk and nutrients reach their steady state concentrations after 800 model years and processes mainly happening in the surface ocean, such as NPP, rather do adjust at a time scale of decades.

The reviewer is right about that 2500 years would be too short for reaching a steady state when talking about a fully coupled ocean-sediment system, since equilibration of the sediment takes place at a much longer time scale. To deal with this issue, we ran MEDUSA first for 100 kyears to fill
up the sediment ($R_{medinit}$) and reach a steady state at the ocean-sediment interface in equilibrium with the given sinking flux from $R_{sedbox}$. This state of sediments was used as the initial state for the coupled simulation $R_{coupled}$. We show the temporal changes of deep-ocean concentrations, sinking fluxes onto sediment top, diffusive fluxes out of sediments and burial during the coupled simulation in Fig. 3. One can clearly see that deep-ocean concentrations and SWI fluxes are not in a steady state but become slowly stabilised, except for the deep-ocean concentration of DIN which sinks further with time. The feature of nitrogen is probably related to the N loss through denitrification in sediments which will be explained below in our reply to a related comment. During the revision we will run $R_{coupled}$ (or rerun it with the tuned parameters) for a longer time to see how the temporal trend evolves.

Back to the reviewer’s comment to the model-data comparison, here we wanted to compare $R_{sedbox}$ to observations and show that the model is validated and produces reasonable sinking fluxes and bottom water properties which are needed for the initial MEDUSA run $R_{medinit}$. In the revised version we will underline that this comparison is based on results from $R_{sedbox}$ without a fully coupled sediment and state clearly that it needs much longer to reach a steady state in a fully coupled ocean-sediment system.

![Figure 3](image_url)

**Figure 3:** Temporal changes of deep ocean concentrations (DIC, DIN, DSI, Alk and O$_2$) (upper left), sinking fluxes of POM, calcite and opal onto sediment top (upper right), diffusive fluxes of solutes out of sediments (lower left) and burial fluxes (lower right) during the 1500-year coupled simulation $R_{coupled}$. Changes are in percentages relative to the values at the beginning of $R_{coupled}$.
C: The manuscript does not show and/or make use of the features that are added by coupling FESOM-REcoM to MEDUSA. MEDUSA2 simulates OM degradation with O\textsubscript{2} and NO\textsubscript{3}. It would be interesting to see a map of the fraction of aerobic OM degradation vs denitrification. This should show a clear difference between the deeper ocean and shallower sediments with more OM input.

R: This is a very helpful suggestion. Since our model setup only considers these two pathways of degradation, we calculated the fraction of oxic degradation in the total degradation, i.e., the sum of oxic degradation and denitrification. Fig. 4 (left) shows that oxic degradation contributes in large areas of the deep ocean sediments up to 100\%, whereas denitrification mainly takes place in shallower sediments, which results in very low concentrations of NO\textsubscript{3} in the shallower sediments at high latitudes (Fig. 4, right). A chapter will be added in the revised version to discuss these model results and compare global numbers of these two processes with observations.

Figure 4: Fraction (%) of vertically integrated organic matter degradation by O\textsubscript{2} in total degradation (left) and the ratio of NO\textsubscript{3} (R\textsuperscript{NO\textsubscript{3}}) concentration in pore water at the bottom of the bioturbated sedimentary mixed-layer (10 cm below the SWI) to its half saturation concentration (5 mmol m\textsuperscript{-3}).

C: As mentioned in my first comment, I would like to see maps of simulated SWI-fluxes (e.g., of O\textsubscript{2}, NO\textsubscript{3}, DIC, ALK). [2.1. As a side-note: What is done in MEDUSA2 when NO\textsubscript{3} is exhausted?]

R: Here we show the diffusive fluxes of DIC, Alk, DIN and O\textsubscript{2} in R\textsubscript{coupled} (Fig. 5) which will be added into the revised version.

With regard to the question ‘What is done in MEDUSA2 when NO\textsubscript{3} is exhausted?’: organic matter that reaches sediment depth devoid of NO\textsubscript{3} is simply preserved and buried. This happens in some high-latitude shallower sediments where both NO\textsubscript{3} concentration in pore water (Fig. 4, right) and denitrification rate are extremely low.

Beyond nitrate, MEDUSA2 has previously been used with coupled manganese and iron reduction and oxidation cycles in the sediment (Munhoven 2021). Sulfate reduction could be easily added as a further oxidative pathway for organic matter. This was, however, not considered for the coupled simulation in this study, as including these processes, which mainly occur in shallow shelf regions with enough organic matter input, requires a much finer vertical resolution in the sediment model: the application with Mn and Fe redox processes included that was presented in Munhoven 2021.
called upon a vertical grid of 340 nodes, compared to 21 here. Here our focus is mainly on how the coupling to a sediment model influences the simulated ocean carbon cycle. Therefore, the sedimentary input to the ocean is of interest. Since our model setup in the coarse resolution does not really resolve continental shelves shallower than 150 m where sulfate, manganese and iron reduction mainly take place (Jørgensen et al., 2019; Thullner et al., 2009), the global impact of these reactions on fluxes from sediments to the ocean would be negligibly small in our simulations. We agree that these reactions should be considered if interested in processes in sediments and using a much finer model resolution.

Figure 5: Diffusive flux of DIC, Alk, DIN and O$_2$ from sediment to ocean (mol m$^{-2}$ day$^{-1}$). Sources for the ocean are shown as positive values.

C: As described in section 2.1.2, MESUSA2 not only simulates a reactive surface sediment layer but also a core layer that records synthetic sediment cores which is a fantastic feature for paleo-applications. It would be very informative to show simulated sediment core layers for different ocean depths (e.g., a shelf vs deep ocean core) for instance during a carbon perturbation experiment.

R: It is also our interest to look at the simulated sediment core layers. As the reviewer states, this becomes interesting in non-steady-state simulations. Our main motivation to couple the ocean biogeochemistry model to a sediment model is to study transient changes in carbon reservoirs over long time scales, such as the last glacial termination. This is being investigated in the coupled Earth System Model (AWI-ESM) which uses the same ocean circulation (FESOM) and biogeochemistry (REcoM) model, as mentioned at the end of the manuscript. There we run the model with a higher
resolution to also better resolve processes in shelf regions. A drawback is that more computing time is required. To this end, however, we first need to show that the model is able to produce a reasonable pre-industrial steady state, which is what we do in this paper. In this steady situation, the simulated cores are simply not interesting enough to be included in this paper. Furthermore, results of these simulations will be too comprehensive to be included in this model description paper. Therefore, we here focus on describing the technical development of the models as a first step and then, in a second step, build the paleo-applications on this basis.

C: Related to the previous point: The authors want to make REcoM3 ‘fit for paleo-applications (see 2.2.2). Carbon isotopes are of particular interest for paleo-applications and, in my (and also the other reviewer’s) opinion, should be included in this manuscript and not in a (very short) additional publication of pretty much the same authors (Butzin et al., EGUsphere). The reduced complexity configuration of the model here is particularly useful as long spin-ups are necessary to reach an equilibrium for the isotope system. I understand that carbon isotopes are currently being developed in MEDUSA2 and this might not be straightforward in a vertically resolved diagenetic model. If this feature is not yet available, the sediment coupling of C-isotopes could for instance be simply realised by assuming no fractionation during OM remineralisation etc. in the sediment and calculating: $\text{DIC}_{\text{13C}} = \text{DIC}_{\text{flux OUT}} / \text{POC}_{\text{flux IN}} \times \text{POC}_{\text{13C IN}}$

R: We share the opinion of both reviewers that including coupled model setups and results for carbon isotopes is an important step to apply the model to paleo-situations. The aim of this model description paper is, however, to mainly focus on the coupling technically and the basic model performance (simulation of PI climate conditions). Therefore, we considered a coarse resolution which is commonly used for technical tests, allowing us to run a reasonable number of tuning experiments within a realistic time frame. The configuration described in the manuscript requires 2-3 weeks of computing time for 1000 model years. Including carbon isotopic traces would nearly double the computing time required of the ocean part alone. For the sediment part, including $^{13}$C alone for all the solid C-bearing components and considering a bulk DI$^{13}$C tracer only would require $\sim$1.8 times as much time as the current configuration. Furthermore, Butzin et al. (2023) do not just focus on $^{13}$C but also on $^{14}$C. For a merge of these two publications, $^{14}$C in sediments would have to be considered as well: adding $^{14}$C on top of $^{13}$C to would multiply its computing time by another factor of $\sim$1.6, notwithstanding the order(s) of magnitude longer equilibration times required for isotopes than for normal tracers. Rerunning the coupled model with carbon isotopes to equilibrium would therefore require additional work of several months at least. As already mentioned in the reply to Reviewer #1, paleo-applications of this model will be published separately based on results from the coupled Earth System Model AWI-ESM, with a stronger focus on scientific questions than technical developments and model validation.

C: The previous one-layer sediment model box, Rsedbox, is unclear to me. So Rsedbox is not really a reflective boundary but it is also not really a sediment model either – hence, there is no benthic preservation simulated with Rsedbox (lines 59-65). It would be good to clarify how the sediment box calculates the return fluxes differently compared to a reflective boundary condition. I just saw that this is described in the appendix of Gürses et al., (2023) but it would be good to also include it in this manuscript as it is necessary to understand Rsedbox.

R: Rsedbox was run with a sediment layer with a thickness of 10 cm which is not further vertically resolved. Particles sinking out of the bottom water boxes enter this sediment layer and go through
remineralization (organic particles) and dissolution (calcite and opal). The fluxes of dissolved parts back to the bottom water boxes are regulated by remineralization and dissolution rates. However, O$_2$ is not considered in the sediment layer and thus not consumed by POM remineralisation and calcite dissolution is independent on the saturation state of the waters. Due to the remineralisation and dissolution rate, not all of the material instantaneously comes back into waters above and a small part of solids stays in the sediment layer before getting remineralized or dissolved. It has been shown in Kriest and Oschlies (2013) that this type of bottom boundary condition improves carbon and nutrient fields, compared to the slightly more simple reflective boundary condition. This sediment layer is part of FESOM2.1-REcoM3 and described briefly in Gürses et al. (2023). We will add more details about the differences between that sediment layer and MEDUSA in the revised version of the manuscript, including the different parameterisations of POM degradation and calcite dissolution.

C: Related to the previous comment – Section 3.1 and table 1 is confusing. The fluxes given in Table 1 are very confusing – it is not 100% clear if these are settling or burial fluxes. The title of Tabel 1 says sinking fluxes (so settling fluxes onto the seafloor?) but the text refers to “calcite burial fluxes (line 249, 252; but then I thought Rsedbox does not simulate any preservation?). So I suppose the model estimates are settling fluxes. But some of the observational estimates are clearly burial estimates (e.g., CaCO3 data estimates are burial fluxes, as stated in Cartapanis et al., 2018). Also POC data estimates stated (50 – 2600 PgC kyr-1) probably refer to burial rates and are confusing. First, where does the 50 actually come from? I know Cartapanis et al. state it but Burdige (2007) gives 160 PgC kyr-1 as the lowest value – and these are POC burial fluxes in these publications. Often cited POC settling rates are 2628 PgC kyr-1 (Burdige, 2007), 2290 PgC kyr-1 (Dunne et al., 2007), or 930 PgC kyr-1 (Muller-Karger et al., 2004). So if the model estimates are really settling fluxes, as suggested by the title of the table and the text (see, e.g., line 240), then these values are too low and not well distributed between shallow and deeper ocean. I find the argument that the coarse resolution is responsible not convincing. Previous models with similar or even coarser resolution are able to simulate POC (and calcite) settling fluxes and preservation on the shelves much better (e.g., Palastanga et al., 2011; Hülsen et al., 2018). Assuming that model estimates in Table 1 are settling fluxes – one cannot judge how well burial rates are simulated by the model. If the text is correct (in that these are calcite burial rates) then the CaCO3 burial rate in deep sea sediments (0.3 PgC yr-1) is 2 – 3 times larger than budget estimates (0.1 – 0.15 PgC yr-1). It would be helpful to know the mean surface sediment CaCO3 content (vs observed 34.8 wt%)

In summary: Please make sure the model estimates are compared to the correct observational estimates. Also, it would be very helpful to include export, settling and burial fluxes for POC, calcite and opal in the table. And also please distinguish between settling and burial fluxes for sediments shallower and deeper than 1000m, as done for the calcite data estimates in the table. This will hopefully help to understand what causes some of the mismatches in POC preservation.

R: We agree that Table 1 is misleading and greatly appreciate that the reviewer has taken time to provide numerous references to the literature with precision about the kind of fluxes for which they give estimates. For the revised version, instead of summarily reporting ranges from review papers, we have collected estimates directly from single studies and thoroughly compared the definitions of shallower and deep waters. For the comparison we adapted the definition of shallower sediments with 1000 m depth from literature and will use this differentiation for shallower and deep-ocean sediment consistently throughout the revised manuscript. The total POC sinking fluxes of
simulations presented in the submitted version are then lower than the estimated sinking fluxes and the modelled burial fluxes are within the range of estimates. For the too low sinking fluxes we retuned the model and Tab. 1 show numbers from the best tuned simulation in the low resolution and from the historical simulation in Gürses et al. (2023) which uses a much higher resolution with 126858 surface nodes (R_{high}). As mentioned in the reply to Reviewer #1, a clearly larger shallow to deep sediments ratio is found in the high-resolution run since the shallow-water regions are much better resolved and the ratio agrees well with estimates by Muller-Karger et al. (2005) and Burdige (2007). The total sinking flux of POC in the low-resolution run (≈700 PgC kyr^{-1}) is still lower than the lower end of data-based estimates (930 PgC kyr^{-1}), while in the high-resolution run (≈840 PgC kyr^{-1}) it is closer to the estimated range. This underestimation of sinking fluxes is not surprising since FESOM2.1-REcoM3 and our simplified setup both have lower global primary production compared to observations (Gürses et al. 2023). In the revised manuscript we will show new model-data comparisons of both settling and burial fluxes for the retuned R_{sedbox} and the new R_{coupled} which will be rerun with tuned oxidation rates for the two classes of organic matter in MEDUSA2 (see reply below and Fig. 7).
Table 1: Sinking fluxes onto the top of sediments and burial fluxes of POC (PgC kyr\(^{-1}\)), calcite (PgC kyr\(^{-1}\)) and opal (Pmol Si kyr\(^{-1}\)) in simulations and measurement-based estimates, reported for the global ocean and ocean regions deeper than 1 km. Numbers for \(R_{\text{coupled}}\) will be renewed in the revised version since experiments are still running. References for observations are summarized here and will be sorted in the revised version: [Burdige, 2007; Cartapanis et al., 2018; Hayes et al., 2021; Tréguer et al., 2021; Tréguer et al., 1995; Tréguer and De La Rocha, 2013; Jahnke, 1996; Muller-Karger et al., 2005; Seiter et al., 2005; Dunne et al., 2007; Cartapanis et al., 2016; Sarmiento et al., 2002; Nelson et al., 1995; Hilton and West, 2020].

<table>
<thead>
<tr>
<th>Run</th>
<th>Sinking fluxes</th>
<th>Burial fluxes</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>POC</td>
<td>calcite</td>
</tr>
<tr>
<td></td>
<td>all</td>
<td>&gt;1 km</td>
</tr>
<tr>
<td>(R_{\text{sedbox}})</td>
<td>696</td>
<td>464</td>
</tr>
<tr>
<td>(R_{\text{coupled}})</td>
<td>501</td>
<td>326</td>
</tr>
<tr>
<td>(R_{\text{high}})</td>
<td>841</td>
<td>159</td>
</tr>
<tr>
<td>Observations</td>
<td>930–5739</td>
<td>310–1029</td>
</tr>
</tbody>
</table>

**C:** The POC wt\% and the spatial distribution look not very convincing! Large areas show POC wt\% > 5 (what are the maximum, mean values in these areas?) where observations show much lower values – mainly at high latitudes. In contrast, other areas were the data shows higher POC wt\%, e.g., the major eastern boundary upwelling zones and the Arabian Sea, \(R_{\text{medinit}}\) does not simulate any OC preservation (Fig. 5).

**R:** Our model results indeed show a strong contrast of POC content and too high preservation in shallower sediments at high latitudes. We followed the suggestion by the reviewer and have carried out additional tuning experiments with varying the degradation rates for the two OM classes in MEDUSA2. The results are described below.

**C:** [Also why are the observations compared to \(R_{\text{medinit}}\) and not to the final results of the coupled model? I know \(R_{\text{medinit}}\) is compared to \(R_{\text{coupled}}\) in Fig. 8 but I don’t find this comparison very informative.]

**R:** The main difference between the two runs is that \(R_{\text{medinit}}\) was run for 100 kyr to equilibrium with given boundary conditions from \(R_{\text{sedbox}}\), while in \(R_{\text{coupled}}\), fluxes between ocean and sediment change during the 1500 coupled years, although the changes are small due to low sedimentation rates. Therefore, with the short simulation time of 1500 years, \(R_{\text{coupled}}\) does not differ much from \(R_{\text{medinit}}\). This can also be seen in the diffusive fluxes out of sediments and burial fluxes ((Fig. 3), lower panel). Fig. 8 and the comparison in the submitted version should demonstrate the similarity of the two sediment simulations. Further, coupling to MEDUSA2 causes some changes in nutrient supply (Fig. 7 in the submitted version) but no substantial changes in productivity (Fig. 6 in the
submitted version). That also explains the small differences between sediment results of the two runs. The similarity and differences will be made clearer in the revised version and R_{medinit} in the comparison of sediment content with observations will be replaced by R_{coupled}. Additionally, we will start a 5000-year MEDUSA run with constant sinking fluxes and boundary conditions from the average of the last 50 years of R_{coupled} and check how sediment contents further evolve with time and discuss the results in the revised version.

C: I suspect that the simplification to only use one class of detritus (line 78) in the water column might be partially responsible for the poor representation of POC wt% in the sediments (but it is impossible to be sure since no maps of POC settling fluxes are shown). The main reason however might be organic carbon degradation as simulated in MEDUSA2. MEDUSA2 simulates two classes of organic matter to approximate the different C:N stoichiometry of POC, right? What are the degradation rate constants for these classes? Is the more C-heavy class remineralized more slowly? I would argue that more tuning of parameters (degradation / dissolution rate constants and/or other boundary conditions) are necessary to improve the model-data fit.

Figure 6: POC sinking flux (in logarithm for easier comparison with Dunne et al. (2007)) and calcite sinking flux on sediment top in R_{sedbox}.

R: We agree with the reviewer that this point requires further analysis. As a starting point (we will add more details about this in the revised manuscript) we have produced maps of POC and calcite sinking fluxes, shown on Fig. 6. Sinking flux of POC compares well with Dunne et al. (2007) and that of calcite also shows similar range and pattern as in Hayes et al. (2021).

We furthermore checked the C:N ratio in sinking flux of POM (Fig. 7, upper left) and found out that C:N is generally low in shallow sediments, particularly at high latitudes, and higher in many regions in the open ocean.

The elemental composition (C:N ratio) of organic matter certainly influences its degradation time scale (Amon and Benner 1994, Martin et al. 1987) and in the water column (FESOM2.1-REcoM3) we also considered a faster remineralisation of nitrogen compared to carbon. In MEDUSA2 we indeed applied the same oxidation rates for the two organic matter classes in the submitted manuscript. We have now started several tuning experiments where we adopted a faster degrada-
tion rate of the low-C:N organic matter class in MEDUSA2 than for the higher-C:N class. The preliminary results already show some improvements of the sedimentary POC content (see the lower panel in Fig. 7): in regions adjacent to the continents at high latitudes, POC contents still exceed 5% but the affected area is strongly reduced, while POC content in some open ocean regions decreases with the increasing degradation rate. The contrast between shallower waters and open ocean can be then mitigated. Calcite and opal content are only slightly affected by changing the degradation rate of organic matter. The upper right plot in Fig. 7 shows the original POC content in Rmedinit for comparison. This figure looks somewhat different to that in the submitted version since the color bar was adapted here to make the low-concentration areas visible. We will continue with some more tuning experiments and replace in the revised version Rmedinit and Rcoupled with the best tuned simulations and change the model-data comparison correspondingly.

Figure 7: Upper panel: left: C:N ratio in POM sinking flux on sediment top in Rsedbox; right: POC in sediment (weight %) in the submitted version. Lower panel: left: POC content in the tuning runs with a 5-time enhanced degradation rate for the low C:N organic matter and right: a 10-time enhanced degradation rate.

C: E.g., what about sedimentation rate: The terrestrial clay input of 2.5 E-8 mol cm-2 y-1 is spatially uniform. But should this not, as a first estimate, decrease with distance from the continents?
This might help with the unrealistic distribution of carbon burial between the shallow and deep ocean (as stated in table 1 and the related text).

**Figure 8:** Clay flux on the top of sediments.

**R:** In MEDUSA2, ‘Clay’ is a generic denomination for inert materials to fill the sediment. In our model setup, dust input (prescribed following Albani et al. (2014)) is primarily considered as input of lithogenic material into sediments. In addition to dust, we introduced a small uniform background flux similarly to Heinze et al. (1999) in order to represent input of lithogenic material redistributed by ocean internal processes, such as mixing and resuspension. This background flux is ten times greater than in Heinze et al. (1999) and is a result of our tuning experiments. The clay input in Fig. 8 thus derives from the sum of dust and this constant background flux. With that we partly considered a gradient decreasing with the distance from continents. We will add this figure and some explanatory sentences into the revised manuscript to describe the tuning experiments.

Rivers deliver several petagrams of suspended sediments to the ocean each year (Peucker-Ehrenbrink, 2009; Milliman and Meade, 1983). Most of these particles are deposited close to the continental margins and estimates of their strength and distribution are not well constrained. Therefore, this source of lithogenic particles to marine sediments is not yet considered in our model setup. We do, however, agree that an additional riverine input of lithogenic material could reduce the too high POC fractions in shallower sediments in our simulations, and will test some assumptions in future experiments.

**C:** The manuscript does not include any parameter values. A comprehensive table stating parameter names, values, units, and references is necessary to understand how the model is set-up. The same applies to the riverine (i.e., weathering) inputs stated in Table. Please indicate where the values come from and how they compare to observational estimates.

**R:** A table of parameters will be added in the revised version.

**C:** I suggest that the loss of N (i.e., 0.8% over 1500 years of simulation) will very likely be a problem during longer model runs if not compensated for via N2 fixation and/or weathering input – especially during paleo-applications with larger contributions of denitrification. I suggest to fix the N-leak.
**R:** N loss of 2% kyr$^{-1}$ is mentioned in the submitted version. And we agree with the reviewer that it needs to be compensated by nitrogen fixation or weathering input. Our long-term plan is to complete the N cycle in REcoM by adding N$_2$ fixation and denitrification in the water column. For simulations in the near future, we thought of two options: the pathway via weathering would add N into the surface ocean by rivers which might have a strong effect on biological productivity, while adding N into the deep ocean (similar to the mass correction for Si mentioned above) might not significantly change the nutrient availability. We will try both options during the revision process and present results of mass correction for both Si and N in the revised version.

**C:** Sedimentary source of iron: The text says (line 82ff.) “The sedimentary source of iron can be calculated in two ways: 1) in a fixed ratio to degradation of particulate organic nitrogen (PON) in the benthic layer as described in Gürses et al. (2023, Eq. A77 in Appendix A) or 2) in a fixed ratio to the diffusive flux of dissolved inorganic nitrogen (DIN) calculated by MEDUSA2 in coupled simulations.” Can you please provide a justification for these representations and how well they approximate realistic SWI fluxes of iron. Also how well do they represent Fe fluxes under anoxic conditions – where Fe-cycling behaves very differently. This might be important depending on the paleo-applications the authors have in mind. Also, A77 does not exist in the Appendix of Gürses et al. (2023).

**R:** Iron contained in organic particles is released into pore water during degradation of organic matter. This is the same process happening in the water column as well. Elrod et al. (2004) demonstrated a clear correlation between the iron flux out of sediments and the oxidation of organic matter on shelves, with a Fe:N ratio that is much higher than typical Fe:N ratios in sinking organic matter, implying that a large fraction of the Fe flux out of the sediment is from lithogenic material, and is mobilized by redox reactions in the sediment. Under anoxic conditions the flux of iron is increased owing to the higher solubility of ferrous iron. To represent this effect, we applied a higher Fe:N ratio (3 µmolFe : 20 mmolN) for the flux of iron from the sediment to the water column than the Fe:N ratio that we used for remineralisation in the water column (1 µmolFe : 30 mmolN). We will add this information into the revised version. The equation for nitrogen degradation should be A67 and will be corrected in the revised version.

**Minor comments:**

**C:** A better motivation & explanation could be included why this model is appropriate for paleo-applications; also giving potential applications. Ideally this would be compared to existing Earth system modeling approaches, highlighting the benefits of this new model configuration.

**R:** This is a good point. We will revised the introduction by highlighting 1) the flexible stoichiometry in REcoM3 and thus a more realistic presentation of growth limitation and marine biological carbon pump which is important for determining the sensitivity of the carbon cycle to changing climate conditions; and 2) the sediment module MEDUSA which enables simulation of the key archive of marine proxies.

**C:** Table 2: Why does seafloor deposition (POC + CaCO3) not equal diffusive C flux out of sediment for Rsedbox?

**R:** If we understand the reviewer correctly, the difference between 75.4 and 76.4 Tmol year$^{-1}$ was asked. This is explained by the net outgassing of 0.9 Tmol year$^{-1}$ (in the same table) which is about 0.01 PgC year$^{-1}$. That means the whole system is very close to but not completely in a
steady state.

C: Section 3.3.3: Unclear where the 402 PgC come from an how it compares to the 21 PgC in table 3. Unclear what should be learned from the last bit of the section, i.e., the discussion of the carbon, alkalinity and silicon inventories not being in steady-state in the coupled run.

R: We thank the reviewer for pointing this out. The discussion in the submitted version is very short and might be unclear. The 21 PgC is carbon stored as POC in the reactive layer of sediments and the 402 PgC (i.e., 381 PgC + 21 PgC) is the sum of calcite and organic carbon. This 402 PgC has two sources: 1) sediments are filled during the 100 kyears of Rmedinit and 2) due to the accumulation and burial in sediments, less carbon is released back to the ocean which can be seen by comparing the diffusive and burial fluxes in Rsedbox and Rcoupled (Tab. 2 in the submitted version). Therefore, the carbon storage is more shifted from the atmosphere to the deep ocean and sediments, resulting a lower atmospheric CO2 when coupled with MEDUSA2. A more detailed discussion will be added in the revised version.

The reason that we discussed the mass conservation of silicon in this manuscript is that it could be violated by asynchronous coupling where flux exchange between models is temporally shifted. We want to use this as a measure of the effect of asynchronous coupling on the inventory change for other tracers (e.g., carbon, nitrogen). Since carbon and alkalinity have external sources (weathering fluxes) and are not in a closed system, their inventories also change during Rcoupled. Thus, it is tricky to determine how much of changes in carbon storage in different reservoirs during Rcoupled is caused by sedimentation. In the revised version, we will analyse results with mass corrections and a longer simulation time, so the problem with changing inventory will be minimized.

C: Some of the methodology is unclear – I thought “FESOM2.1 was run for 1000 years to spin-up ocean circulation.” (line 201) why does the 2500 year run in Fig. 4 show again ocean circulation stabilisation?

R: Our description here is indeed not entirely clear. During the first 1000 years, only FESOM2.1 (the ocean model without biogeochemistry) was spun up. The results of this run are not shown in the manuscript. After that, the model was started with the physical fields produced in the first 1000-year run and REcoM3p with its original one-layer sediment, and run for another 1000 years to spin up the biogeochemistry (Rsedbox). After these 1000 years, one simulation was continued with the same setup as Rsedbox and another one was run with the coupled MEDUSA2. Both were run for 1500 years, so that Rsedbox has 2500 model years in total. We will explain this better in our revised version. Fig. 4 shows the stabilisation of atmospheric CO2 concentration, not ocean circulation. The air-sea gas exchange is affected by the DIC concentration in the surface ocean which is regulated by the marine carbon pumps (described in the introduction). After the spin-up, the whole system is approaching a steady state and the atmospheric CO2 (290.5 ppm) comes also close to its steady-state concentration (293 ppm). From 1000 to 2500 years, Rsedbox further reaches its steady state and Rcoupled undergoes a perturbation through storing carbon in sediments and lowering sedimentary input of carbon to the ocean. Therefore, more CO2 is taken up by the ocean and the atmospheric CO2 decreases, until the gas exchange is balanced after several hundreds to about 1000 years. During this period, ocean circulation remains stabilised and only the distribution of carbon is changed from one to another state.

C: It is also confusing that two experiments are named Rsedbox – one described in 2.4.1 and then
one described in 2.4.3 (and shown in Fig. 4?) The description of \( R_{coupled} \) is also not 100% clear. The text states: “(2) a coupled simulation \( R_{coupled} \) was conducted for 1500 model years first using the output from \( R_{medinit} \) as sedimentary input of DIC, Alk and nutrients.” (lines 231-214) I suppose this means, you start with the SWI-exchanges calculated in \( R_{medinit} \) (i.e., for the first 50 years), after that MEDUSA is called every 50 years.

**R:** Yes, it is correct. As explained in our reply to the previous comment, \( R_{sedbox} \) has actually run for 2500 years in total, where the first 1000 years come from the biogeochemistry spin-up stage. That seafloor deposition fluxes and the bottom water concentrations obtained at the end of the biogeochemistry spin-up were used for the first MEDUSA2 run \( R_{medinit} \) (whose purpose was to pre-fill MEDUSA2’s sediment in an approximate steady-state way); \( R_{sedbox} \) was continued further 1500 years to produce results that are comparable (i.e., that cover the same simulation length) to those obtained with MEDUSA2 coupled to FESOM2.1-REcoM3 (\( R_{coupled} \)). We will rephrase the description of experiments and clarify this in the revised version.

**C:** Lines 268 ff. What does this refer to? How is this different to Fig. 5?

**R:** The model-data comparison of sinking fluxes was not shown in the submitted version. They are added here (Fig. 6) and will be shown in the revised version as well.

**References**


