Substantial inter-model variation in OAE efficiency between the CESM2/MARBL and ECCO-Darwin ocean biogeochemistry models

Review comments and replies.

Preprint:
https://egusphere.copernicus.org/preprints/2025/egusphere-2025-3713/egusphere-2025-3713.pdf

Discussion:
https://egusphere.copernicus.org/preprints/2025/egusphere-2025-3713/#discussion

RC1:

We thank the reviewer for their constructive comments and suggestions !

This work is innovative and highly relevant. It addresses the complex challenge of comparing models when plume trajectories and CO_2 equilibration both differ and are interlinked. The proposed methodology is elegant and offers a valuable approach for disentangling model parameterizations. This study could inspire future multi-model comparisons of OAE. Overall, it represents a strong and meaningful contribution to understanding OAE model uncertainties. It requires minor clarifications, figure adjustments, and additional context for full clarity.

Specific comments:

Line 110: Sea ice should be mentioned here, given that it is studied in your work. You could also consider mentioning thermal effects (e.g., changes in $dDIC/dCO_2$ affecting equilibration rate constants) and salinity.

Yes, other factors can also play a role. We've added "the gas transfer velocities (which are a function *wind speed and sea-ice cover)*" and "Similarly temperature and salinity can affect the carbonate chemistry state." to line 110, to clarify this.

Table 1: How the optimization method of ECCO-Darwin may influence the results?

The optimization method of ECCO-Darwin uses the adjoint method for the physics, which adjusts initial conditions, surface-ocean boundary conditions, and 3-D time-invariant mixing coefficients; model biogeochemistry is optimized using a low-dimensional Green's Functions approach to adjust initial conditions and Darwin parameters. This results in a physically-consistent counterfactual solution with fully-closed property budgets (i.e., no nudging is used in the assimilation process). We note that ECCO-Darwin does not have a long spin-up period from pre-industrial conditions, as done in many forward-only ocean and

Earth System Models, but uses the ECCO data assimilation methodology to both reduce spin-up and drift in a data-constrained simulation that starts in 1992.

We've added this description to the manuscript. It is hard to anticipate what exactly the effect of these architectural differences will make. In general, the assumption is that constraining the model to data should yield a more realistic model. However, in this paper we're not trying to determine which model is the "best" compared to a suite of ocean observations but simply ask the question how large the variance is between two commonly-used models.

Line 151: Please include the corresponding figure in the supplementary materials.
We added a Figure showing this is indeed the case (Figure S3)

Line 188: Could you rephrase this sentence. It is unclear why this equilibration rate differs from Zeebe and Wolf-Gladrow (2001). Also, clarify that μ represents vertical dilution, not horizontal (if I understood correctly), and clarify z0.

The representation of the equilibration rate used here was developed in Zhou et al 2024. It was adapted to account for the fact that in the simulation gas exchange only occurs from the top simulation cell, whereas the equation in Zeebe and Wolf-Gladrow (2001) expresses the rate in terms of the mixed layer depth (using a simpler box model) which does not have a strict correspondence in the simulation, as the mixed layer typically spans many simulation layers and varies from place to place.

We've expanded this section in the manuscript to give more context on where this formulation came from and define more explicitly what the terms μ and z0 mean.

Lines 200–230: Equations (3) and (4) are referenced but not explicitly defined. Please add their definitions.

We're not sure what the reviewer is referring to - both equations (3) and (4) are explicitly defined on lines 171 and 175, respectively and show how [D] can be calculated. Perhaps this comment was referring to something else ?

To clarify further we've added explicit equations for how w_ij is calculated in a section further below.

Line 220: The weighting method appears to overestimate the plume extent since surface decay is not included. Will this plume attribution method remain valid in longer simulations with multiple OAE deployments?

This analysis is not intended to work with multiple OAE deployments - it is only designed to interrogate a single pulsed release.
We since reworked the analysis substantially and added a comparison that shows that the surface weight w_ij taken from alkalinity and from deficit are very similar and do not introduce a large change in the reconstructed rates. The surface deficit vs surface alkalinity, however, do diverge and for this reason we have reworked the analysis to be in terms of the

actual surface deficit, which should resolve this concern. We have added a plot (S4) that shows that the reconstructed rates match the actual rates closely.

It seems that plumes extent could be overestimated, with accumulation in the ocean and mixing occurring without any corresponding effect on the carbon flux.

Ideally the weighting function should average differences in equilibration parameters over the areas where equilibration is occurring. In that sense, yes, using the excess Alkalinity as the weighting function could potentially include areas where no more equilibration is occurring. However in practice we find that the horizontal extent of the plume is well tracked by ΔAlk and does not significantly exceed the area of ΔD. The resultant reconstructed rates match closely (Fig S4 dotted vs dashed)
By the time the equilibration is close to complete, however, the ΔD plume is very noisy and poorly defined, due to the vanishing denominator, whereas the ΔAlk-based plume definition remains numerically stable and well defined, and we use ΔAlk for the surface weighting.

It is not entirely clear to me weighting with ΔDIC is problematic, shouldn't the weights reflect both the trajectory and gas exchange history?

We have reworked the analysis to use ΔD, because it reproduces the actual gas exchange rates quite closely. However one disadvantage is that because ΔD depends on the gas exchange history, it somewhat convolutes model differences resulting from different mass transport modelling from differences in gas-exchange parameterization.

Please provide an explicit definition (equation) for the weights w_i.

An explicit equation was added.

Lines 295–305: This section suggests that all locations converge after seven years. Please rephrase for clarity.
The phrase "not observed in the North Atlantic" is confusing since Figure 1 indicates similar behavior is absent in the Western Sahara, Oman, and the U.S. East Coast.
Even Although you discuss these differences later in the text, consider rewording this paragraph to introduce the subsequent sections.

The locations where convergence is clearly absent are Norway, Iceland and U.S. East Coast. We've rephrased lines 295-305 for more clarity and included U.S. East coast in the exceptions to the general convergence trend.
We would argue that the differences are continuing to converge in Western Sahara and Oman, even though they are not converged by year 15. However the gradient at the end of the simulation of the dashed line (CESM) is steeper than that of the ECCO lines, suggesting the CESM prediction is slowly catching up to the ECCO one. We've added the proviso that we can't say for sure if these trajectories converge eventually since we have only simulated 15 years.

Lines 310–320: A short discussion on how interannual variability influences n_{ax}, and why this differs between injection sites and models, would be valuable.

A paragraph discussing the possible sources of differences was added.

Line 325: Can convergence can really be concluded from Figure 2? It is only evident in panel 2a.

Of course we don't know what happens after the extent of our simulations, in this case 5 years. But we can observe that the interannual and intermodel variance appears to decrease from year 1 to year 5 in cases a,c and d. Whether it will eventually converge entirely we cannot say for sure. We've updated this paragraph to express this in a more precise way.

Figure 3: Please move the legend away from the curves for readability. Consider adding a zoom on the first few months.

We updated the figure to zoom in on the first year for better clarity in virtually all the graphs of this paper and moved the legend such as not to obstruct the plot lines.

Line 340: The reported correlation is not immediately evident. It may help to show this relationship explicitly in an additional panel or inset. There appears to be sufficient blank space in Figure 3 to include both plots clearly.
Line 340: The phrasing "the effect is much stronger" suggests subduction causes the n(t) difference in Norway, but this interpretation seems premature. The correlation is not clear, Alaska shows similar n(t) curves between models despite larger surface fraction differences.

We shouldn't have used the word correlation here, that's perhaps too strong and of course subduction is not the only contributor to the difference as shown later.
What can be gleaned however by just qualitatively comparing the curves in Fig 1 and Fig 2, (old manuscript numbering) is that in general faster surface dilution corresponds to slower equilibration. We've rephrased this section to make it clearer that we're discussing a clear qualitative relationship between equilibration speed and surface dilution which holds in many locations but does not explain all the differences observed. In particular we've corrected the discussion of the Norway location in which subduction appears similar between the models but equilibration is nevertheless much slower in CESM2, owing to other model differences, which motivates the remainder of the analysis.

Crucially we have significantly expanded the analysis to include the relationship of early equilibration with the mixed layer depth (MLD) which is found to be substantially different in the two models, with a clear trend showing ECCO-Darwin having much shallower MLDs which accounts for much of the observed equilibration differences. Figures added are Figure 6, Figure 7 and Figure S1 and S2 (new manuscript numbering).

Line 370: Did Xie et al. use coastal or offshore injection? Can this change their interpreatation?

Xie et al. also used coastal injections for the most part (most of which did not show much of a resolution dependence), except for their South China location (which, likewise, showed almost no resolution dependence) and the two southern ocean deep water formation areas (which did show some resolution dependence). There's not a clearly discernable correlation between coastal distance and resolution dependence in their data.

Line 402: The sentence “This is consistent with a general understanding that ocean carbonate chemistry is well understood and therefore not a major contributor to the final efficacy of OAE” seems to be missing a word—perhaps “uncertainty” or “variability”?

Changed to “This is consistent with a general understanding that the ocean carbonate chemistry is well understood and therefore not a major contributor to the inter-model-variance of OAE efficiency”

Lines 410–420: Could you comment on the effect of wind on MLD?
Wind stress can increase vertical mixing in the upper ocean, which is parameterized using the GGL90 scheme. We added a sentence to this section. Wind speeds do differ slightly in the two models, however the differences in MLD are greater than what that difference would cause.

Line 433: The discussion on biological influence is insightful. However, the cause–effect relationship in this sentence is unclear: “However, the shape of the OAE efficiency curves is, in principle, independent of the quantity added, so in that sense the efficiency curves appear to be indeed relatively independent of biological activity’. Consider rephrasing for clarity.

We’ve removed this unnecessary and confusing sentence. We added more discussion on how the biological model may interact with alkalinity and DIC in general and how this may be reflected in the analysis of equilibration rates.

Line 451: It would help readers if you reiterated which parameters are being referred to here.

We replaced “(k and β)” with “(gas-exchange velocity k and carbonate sensitivity β)” to make this more clear.

Also, it is not clear how parameters vary in Figures 9–12 or the magnitude of these variations. This information could be added to the Methods section.

We have completely reworked this analysis and changed the way individual parameters are compared. Only a single plot is presented which would be easier to follow.

Line 462: Please specify the magnitude of the change in β.
This entire section was reworked, this doesn’t apply anymore.

Line 464: Please add a reference to a figure supporting this result.
We referenced the subfigures in the text where appropriate.

Figures 9–11: Add x-axis labels and include the black line in the text legend.
Figures were replaced with a different analysis.

Figure 9: Were ECCO-Darwin parameters also changed to CESM2 ones?
No, we only did the comparison one way.
In the new analysis the parameters were changed in the ECCO-Darwin → CESM2 direction

Does the 3-hourly versus 6-hourly wind forcing matters?

We don't have any good way to determine this one way or another, though our suspicion is that this is not important relative to the differences in the wind speeds themselves.

Line 490: One might argue that the remaining challenge lies in accurately determining the location of the plume.

Indeed. We discuss this in the following section and in the overall discussion. Accurately predicting the plume trajectory given a particular deployment will require models to be improved.

Line 502: Providing specific locations would make the statement clearer to readers.

Added a specific reference to Oman instead of the vague "some places".

Figure 13: You might consider moving this figure earlier in the manuscript.

Good idea! We moved this into the relevant methods section and added a paragraph to motivate the rate constant framework based on distinguishing plume and parameter-based differences between models:

"This is illustrated in Figure \ref{plume_trajectories} for an alkalinity release near Alaska. The extent of the alkalinity plume from three different runs is overlaid on the local gas-exchange parameter k. One can see how the intersection of the plume with k (as well as the other gas-exchange parameters) will determine the overall equilibration rate. Different models will not only have different gas-exchange parameters, they will also predict different plume trajectories - both contributing to the overall observed variance between models. The goal of the following section is to develop a framework to be able to attribute the differences to the various contributing aspects."

Line 576: It may be worth reiterating that vertical DIC and alkalinity gradients need to be appropriately resolved, as this is highly dependent on the biological representation.

Yes, it is only during the pulse trajectory that modelling biology makes little difference to the uptake curve. We added a sentence here clarifying this fact. "However, the biological model is critical to setting up the correct starting state of the ocean, in particular the vertical Alk and DIC gradients."

Limitations paragraph: Could you discuss limitations specific to ocean-only models? For instance, are internal variability and feedback processes a concern?

We've added a short paragraph discussing these limitations.
"The inherent limitations of using ocean-only models, vs full earth-system models have also not been explored sufficiently yet, where reservoir feedbacks or long term changes to calcification rates at large deployment scales could potentially play a role. However many of

these effects occur on longer timescales and may not directly influence the equilibration speed of individual OAE deployments."

Ensure consistent labeling of subplots : use i), ii), iii) in figure, referred to a) b) c) in the text legend.

Thank you for catching this discrepancy. However, the figures were replaced entirely in the latest version of the manuscript and only use single numbering a) b) c).

Fig 11 & 12 : Should it be "ECCO minus CESM2" in figures inset legend?

Figures were replaced with a different analysis.

RC2:

We thank the reviewer for the thorough review and the great suggestions !

Review of "substantial inter-model variation in OAE efficiency between the CESM2/MARBL and ECCO-Darwin ocean biogeochemistry model"

This study presents a model intercomparison examining how ocean alkalinity enhancement efficiencies in virtual alkalinity deployments differ due to differences between ocean biogeochemistry models. The upshot is that uncertainties (or biases) in vertical transport in the upper ocean are associated with pretty substantial uncertainties (or biases) in OAE efficiency in popular global ocean biogeochemical models. This is a valuable contribution to the literature, building on recent work to map OAE efficiency and understand its sensitivity to deployment location and ocean circulation regime. It should be published after revision.

Comments/suggested revisions line-by-line:

Table 1. If you're going to include this Biological Ablation Experiment, I think you should include some reference to and discussion of the actual equations in Darwin and specifically the coupling between Alk, Dic and biology, and whether it is realistic. In general, I think you should have more explicit caveats about potential missing processes, e.g. small-scale plume processes, microscale processes on particles, etc.

We added a more extensive discussion of the features of the Darwin biology model, its dependence on the carbonate state and processes not modelled here.
In the Darwin equations, the productivity rates are not explicitly dependent on DIC or Alk. Therefore our null hypothesis was indeed that the effect of removing biology should be very small if any. However, it was useful to numerically confirm that it is.

Fig 8: I don’t think this Biological Ablation experiment is worth publishing. See my comments above regarding the model formulation. The lack of sensitivity is unsurprising to me, and I’m not sure the Darwin model adequately captures the sensitivities of biology to DIC and Alk (to the extent these sensitivities are significant at all). (I'm more surprised by the difference near Oman)

The importance of modelling biological processes when simulation OAE has so far not been explored sufficiently in the literature and we have had many conversations with colleagues who have argued either way. Therefore, we believe this negative result is worth publishing and of interest to the community.
Indeed when we showed some early ablation results in presentations, many listeners were very surprised by these results. It is reasonable to think that because the biological pump sets up the vertical DIC and ALK gradients in the first few 100 meters, simulating this feature during the pulse trajectory is critical. However we show that as long as these resulting gradients are present in the starting conditions, it is not necessary to simulate the processes that maintain them for short-medium length simulations during which the OAE pulses equilibrate - not necessarily an intuitive result for everybody.

Furthermore, as we discussed in lines 520-525 the fact that for this sort of efficiency calculation you can ignore the biological tracers, and thus a very substantial part of the calculation, is a useful fact for practitioners to know, as it can save a lot of compute time (advection of the 31 Darwin tracers consumes almost half the compute time).

The reason for the small difference near Oman continues to be unclear. We investigated this a little more since publishing the preprint, suspecting perhaps an effect of the Oman location being the only location at the mouth of a marginal sea (Arabian gulf). However, we tried several marginal sea locations since, and found virtually no difference between the regular and the ablated model. Conversely we tried the offshore Oman location as well as the red sea and here too there appears to be a small dependence on the biological model being present.

L121: State the definition of the brackets [..]
Added *(where the square brackets mean "concentration of")*

Eq1-2 / Supplemental Material. I didn’t check all the calculations completely. That said, they feel out of scope for this paper. Can you justify why this is needed or appropriate in this context? Why can’t you just use available tools? Is there some reason why you can’t make these calculations using existing infrastructure in pyCO2SYS, e.g. finite differences or automatic differentiation?

We don’t disagree with the reviewer that some of the exploration of convenient approximations are auxiliary and supplemental. However, we believe that these approximations are valuable to the community and should exist in the publication record. They have, to our knowledge, not been published before and the various approximations from the literature have never been compared systematically. We’ve heard from colleagues that although they would like to use them, they can’t unless they can give a citation.

Inclusion here in the supplement solves this issue for the community. It was included in this manuscript because they were developed and used as part of the work presented here. They are considerably faster and sometimes more convenient to use compared to running PyCO2Sys. Finally we see little downside in including these derivations here - they do not obstruct the main paper in any way but may be of value to those who are interested.

Is this critical to making your computations efficient, i.e. speeding up pyCO2SYS for this application?

Yes we used the approximations when calculating the deficit for every voxel and timepoint in the simulations simply due to the speed advantage of obtaining ∂DIC/∂Alk.

Approximations that don't require solving the carbonate system and are based on Alk and DIC alone (fundamental tracers that are readily available from most biogeochemical simulators) are very convenient and useful because there's no need to use pyCO2Sys.
We have seen the previously published approximations for (dALK/dDIC ~~ 3-2DIC/Alk) used by other practitioners for this reason. The fact that eta_max is almost independent of anything but DIC/Alk makes this possible, an intriguing fact that has, to our knowledge, not been explicitly published. However the previous approximation retains some inaccuracy especially when borate is taken into account with a deviation of up to a few percent (Fig S2, b)) for typical ocean DIC/Alk values. Given that our new softplus-based equation is just as easy to calculate we felt that we should share this insight with the community.

Are sensitivities like d DIC/ dCO2 or d ln DIC/d ln CO2 not already available from existing outputs of pyCO2SYS?

It is true that dDIC/dCO2 can also be calculated from gammaDIC and CO2 from pyCO2 like so: dDIC_dCO2 = csys["gamma_dic"]/(csys["CO2"]). We didn't use this in our approach, but used the full explicit formula given here - the two equations give the same result.
The advantage of our formula is that it only requires HCO3 and CO2. If these are already saved from the simulation itself, \beta can be calculated immediately, conveniently and expediently, without needing to invoke the slow PyCO2Sys again.

Fig S2. If you're going to keep this material, it would help to discuss how frequently/where DIC/Alk > 1 occurs in the real ocean and thus whether this parameterization improvement beyond Tyka et al. 2022 has practical relevance. The approximation from Tyka et al. 2022 is nice to have here, but how important are these adjustments.

We've changed the plot of the approximations (preprint Fig S8) to span only realistic values of Alk/DIC, taken from OceanSODA, rather than taking a full grid of Alk and DIC values. This gives a more practically relevant comparison. The new approximation is shown to be quite a bit better than the old one for high values of DIC/Alk, which do occur (0.95-1.01).
Older approximations do more poorly across bigger ranges of eta_max

Also, I'm wondering if OAE could push the system into low DIC/Alk< 0.5, where the approximations start to fail? (I don't see how it could push it to high DIC/Alk)

Large scale applications of OAE could push the system into very low DIC/Alk but only very transiently and locally to the injection site. Since eta_max is primarily concerned with the ultimate, long-term equilibration point, inaccuracies in transient low DIC/Alk are not particularly important. We've added some discussion of this to the supplement.

The map in Fig S1 would be better in the main text.
Figs 1/4: Move the map in Supplemental to the Main Manuscript.

Moved figure to main text as suggested.

Eq (2): I'm a little wary about this. Be careful to discuss units on these. The Revelle factor d ln CO2/d ln DIC has the advantage of making this type of sensitivity unitless.

The partial derivative ∂[DIC]/∂[CO2] in equation 2 is also unitless, since it's concentration divided by concentration. As this is the term explicitly appearing in the gas exchange equation, rather than the Revelle factor (∂ln pCO2/∂ln DIC), we chose this to be the sensitivity that's being investigated and compared.
We've added a sentence in the methods when these equations are introduced to remind the reader that they are unitless.

Section 2.6: What does "surface dilution" versus "surface distribution" mean? This is not very intuitive. Does it mean vertical vs lateral distribution of the plume? Be more precise and maybe show and discuss examples.

Yes, they are referring to vertical and lateral distribution. These concepts are made more rigorous later in the methods section. We've substantially reworked the analysis in section 2.6 and explained these terms much more explicitly. We've added equations defining them exactly.

L169/Eq 3: To me, "fully re-equilibrate with the atmosphere (not accounting for reservoir feedbacks (Tyka 2025))" is not quite the right wording. It is not clear what you mean by "equilibrate" or "re-equilibrate". The equation is also a bit imprecise here, because d DIC/dAlk varies spatially and temporally, as the plume moves. So it is not really a partial derivative. I think of this more precisely as representing a theoretical scenario in which co2 flux adjustment to surface alkalinity perturbations are instantaneous and leave pCO2sw unchanged, in which the partial derivatives make more sense. However, I appreciate the desire to frame it as the limit of a long-term adjustment when the ocean is fully mixed, which is related because the partial derivatives do not vary that much in space and time… Try to frame better.

We would argue that this framing isn't really new and goes back to Renforth and Henderson 2017, and was used in He et al 2023 and further developed in Zhou et al 2024 and several other papers. The equation is a precise description of the changes in the carbonate system in small parcel of water for small changes in ΔAlk, over which ∂DIC/∂Alk is a constant. Re-equilibration in this context means that if we start with a water parcel that was previously at some pCO2, and we add a small amount of alkalinity, then equation 3 gives the amount of DIC that needs to be added that offsets that amount of alkalinity such that the pCO2 returns to its previous value ("re-equilibrates"). While it is true that ∂DIC/∂Alk can change over time

for a given parcel, it still gives an estimate of the final equilibration point relative to the alkalinity perturbation at any given moment in time, for a given water parcel. The consequence is simply that the final equilibration point can change somewhat over time, we don't see this as a problem - it remains a useful tool: Transforming the alkalinity perturbation into a DIC deficit is a useful framework that conceptually unites Ocean Alkalinity Enhancement and mCDR methods that remove DIC from the ocean.

We have, however, extensively reworked Section 2.6 and developed the deficit concept more carefully, including graphs in the supplement showing that the approximations made allow a reasonable reconstruction of the overall equilibration process offline from its parts.

Related to the above/Eq 7 I think the relationship between pCO2sw and DIC is non-linear, so expressing beta as a partial derivative can introduce errors for large D. By contrast, the partial derivative in eta_max is reasonably linear and thus not so sensitive to large perturbations.

In equation 7 we state the differential equation relating the rate of change of the deficit with the remaining deficit itself, at any given moment in time.
If the terms k, beta, z0 and mu are independent of D itself, we have a simple first order differential equation. This assumption is reasonably true for k and mu for any value of D (and z0 is a constant). But as the reviewer points out, for beta this assumption is only true when D is small, i.e. where the additional alkalinity is small. While this is true in principle, here even the largest deviations in alkalinity caused by the pulsed alkalinity injections are in the low micromolar range over which beta is pretty linear, and quickly drop to below µM as the plume spreads.

However even if beta is a function of D, the differential equation can still be valid for any given point in time, the only difference is that its integration (i.e. the time evolution of D from its initial value back to zero) won't be a simple exponential as would be expected for a first order ODE but a slightly deviating curve.
In this case the rate constant isn't really a constant in time. Indeed, all of the variables here (k, mu and beta) are dependent on time (and location). For simplicity of notation we did not make every variable here an explicit function of time (k(t), mu(t), etc..) however they all are.

We're using this approach to "reconstruct" the equilibration rate from its components, offline, allowing us then to replace one term of the reconstruction in turn to probe its influence on the overall equilibration rate. In this context, a slight dependence of beta on D is not fundamentally different than the dependence of beta, k or mu on time and location.

I'd appreciate seeing some initial plots of 1/D dD/dt and r, which should not be the same due to the various approximations, maybe in the appendix. Showing errors and giving some intuition. It feels like you jump to the model comparisons without illustrating this initial methodology.

We added a figure in the supplement plotting a "reconstructed" rate based on Equation 8 (now equation 12). I.e. For each surface cell we calculated its deficit D times the gas exchange parameters, which gives the rate of gas exchange in that cell. Summing over all surface cells yields an estimate of the total gas exchange rate in mol/s.

The two curves are in close agreement (Fig S4). This suggests that our conceptual model of using the virtual quantity D is a reasonable model of the gas exchange during the simulation and can therefore be used to compare models. We have substantially reworked Section 2.6 to explain this approach.

Fig 3: It would help if you could show zoomed in the first 1-2 years by making the x axis non-uniform. Legend in (a) could be moved to the top right of the panel.

We've implemented these great suggestions throughout the paper.

Fig 3-5: We need a more precise definition and maybe plot showing example of how "surface fraction" and "surface-ocean excess alkalinity fraction" are defined. See my comment above about this.

We've added an equation defining the "surface-ocean excess alkalinity fraction" in the methods section. It is simply the fraction of the total excess alkalinity (the alkalinity added in the pulse) that is currently residing in the top simulation layer.

Fig 6f: perhaps use eta_max on the x axis rather than the partial derivative? also what are the units in panels e-f?

The partial derivatives ∂DIC/∂CO2 and ∂DIC/∂Alk= eta_max are both unitless quantities. We've indicated them as such in the figure.
We changed the y-axis label to eta_max in 6f as suggested.

Section 2.7.1/Figures 9-10: I'm confused. I don't follow what is interchanged and what the 4 lines on the plot represent. Shouldn't there be a matrix of 9=3 x 3 scenarios for each parameter and virtual deployment? 3 different physical plumes (reflecting same changes in both numerator and denominator in each scenario) and 3 different ratios reflecting different values in numerator and denominator (ECCO 99/CESM99; ECCO92/CESM99; ECCO99/ECCO92) or more if oceanSODA is included as an option for k and beta? Please clarify what is happening here and in the comparisons in Fig 9-10.

Figs 9-10: What is the black line? (See my related comment in Section 2).

I get a bit overwhelmed by the examples here. I suggest: a) try to reduce the content in Figs 9-12 to a table if possible and/or b) synthesize the narrative and reduce the number of plots. Questions to think about: What is the message you are trying to convey here? Could most of this be in supplemental? Why 6 examples or 12 examples or 2 examples? Why just time series?

Based on the reviewer's comments (and those of the other reviewer) we decided to considerably rework this section, changing the approach to the analysis somewhat.
The result is Figure 13 and section 3.8 which are, we believe, much easier to follow.
We focus attention on comparing just ECCO-1999 with CESM2-1999 and compare the effect of replacing single terms in the rate reconstructions for wind, carbonate system, ice and horizontal plume extent. The presentation is now a single figure.

Why no plots of depth and time if vertical distribution is so important?

Based on the reviewer's comment we've added a completely new analysis of vertical mixing processes: the mixed layer depth differences, the vertical distributions of alkalinity, the surface dilution of alkalinity and deficit and the relationships between them. They have been added in Section 3.3 ("Subduction") , Figure 6, Figure 7 and Figures S1 and S2.

Fig 10/12: Use of the term "tropical" rather than "near-equatorial" is more appropriate for these regions.

Changed accordingly