AC: We would like to thank both reviewers for their comments. Their suggestions have been extremely helpful in clarifying the scope of the work, strengthening the model evaluation, and improving the presentation of the manuscript. We have considered and implemented their recommendations where possible.

RC2: Overall Comment

The paper by Van der Zant presented an update to the previously published Radi early diagenetic model. This version of the model improves upon first version by including relevant processes related to the exchange of solutes across the sediment-water interface (SWI) as well as enhancing the model capabilities for simulating biogeochemical processes in highly permeable sediments. A key feature of this model and the overall paper is the showcasing of the adaptability of the model in computing benthic solute fluxes which can be used in coupled benthic pelagic application from a metamodel built from RadiV2. This parameterization of benthic solutes exchange from metamodel is important in the context of better representation of seafloor biogeochemical dynamics in global earth system model. Overall, the paper is well written and concise. The paper take care in explaining the relevant addition of the new development of Radi and showcasing its utilizing in estimating benthic fluxes across different marine environment. However, the paper will do well in at least providing supplementary text on the performance of this new version in capturing the variability of porewater profiles especially in coastal regions where it has previously not been tested. This will help inform users of the strength of the new model as well as the current limitation as highlighted in the latter part of the paper for their application. Overall, this is solid work, and I recommend for publication following some minor comments below.

AC: We thank the reviewer for this suggestion and agree that a porewater comparison in a coastal setting strengthens the evaluation of RADIv2.

In the supplementary data we have added a dedicated test case for the shallow Iberian Margin station 99-6 (Epping et al., 2002), which offers one of the most complete sets of constraints for organic matter kinetics and porewater chemistry for the chemical species included in the paper. This site was chosen because (i) it lies in a coastal margin environment, (ii) porewater profiles of O_2, NO_3^- and NH_4^+ are available, and (iii) the OM kinetics and flux are well constrained compared to many other coastal datasets. The other locations in this paper were either in the deep ocean, lacked porewater profiles, or combined information from multiple stations (e.g. the North Sea compilations).

We tuned only the organic matter parameters that are directly constrained by the observations: the FPOC flux and the total degradation rates (k_fast + k_slow, and made sure they match the reactivity supported from in-situ O2 microprofiles reported in Epping et al. (2002)). We did not perform any additional optimization to force the profiles to match the observed porewater data.

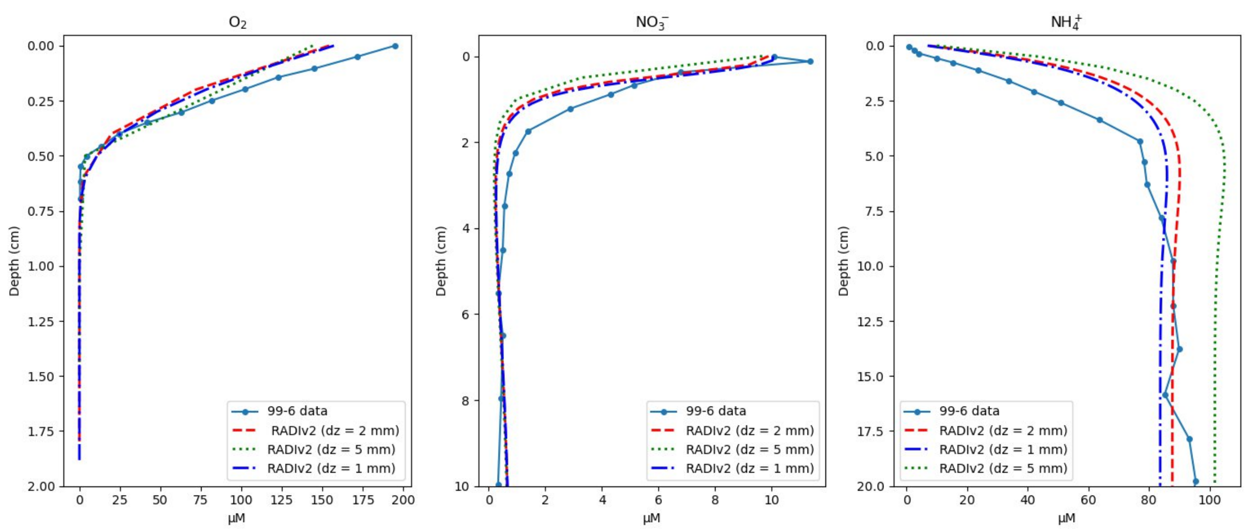
Physical settings such as bioturbation, irrigation and the global carbonate kinetics were kept at as universal parameters in RADIv2 and not adopted from Epping et al. (2002). We disabled dispersion as there was no permeability data available to constrain this, and dispersion led to too high O_2 pumping to the sediment, compared to observations.

RADIv2 reproduces the main structure of the observed NO_2^-, NH_4^+ O_2 profiles. We have added a new figure and a short description in the supplementary data ton this comparison and to clarify both the strengths and limitations of RADIv2 in this coastal setting.

In anticipation of the reviewer's follow-up question regarding vertical resolution, we also performed a sensitivity test in which we repeated the Iberian Margin 99-6 simulation with different uniform layer thicknesses in the upper sediment (dz = 1 mm, 2 mm, 5 mm).

The pore-water Figure and description are now added in the supplementary material (Sect. S1, Fig. S1). We refer to this in the text in **L323:** In addition, we illustrate the performance of RADIv2 for resolving depth profiles of porewater composition at a coastal margin station on the Iberian Margin (Sect. S1, Fig. S1).

And in the subsection of the Iberian Margin starting at **L362:** A dedicated comparison of simulated and observed O_2, NO_3^- and NH_4^+ porewater profiles at station 99–6 is provided in the Supplement (Sect. S1, Fig. S1).



General Comments

RC2: I agree with the first reviewer about the statistical validation of the model derived fluxes in comparison to observations. One other thing I might add as well is the lack of detail in how the five environmental drivers used in the empirical formulation of the metamodel was selected. Was there any stepwise selection screening to diagnose what

variables are sensitive to benthic fluxes of DIC, O2, TA? Were the selected variables based on prior literature findings (I can't find any cross-reference in the text)?

AC: We added a statistical test that quantifies whether the model shows a significant mean bias relative to the observations (via a Welch two-sample t-test) and whether the simulated fluxes fall within the bulk (25–75 %) and overall (5–95 %) range of observed values. Figures 2-5 have been updated and now show a boxplot of the site-wise fluxes and overlay points for each observational site, colour-coded by dataset, using the published site means. This representation shows where the bulk of the observational fluxes lie, how different datasets contribute to the overall spread, and how the RADIv2 distribution overlaps with (or deviates from) this cloud of observations.

We agree that the rationale for selecting the five environmental drivers in the metamodel should be described more clearly. In the revised manuscript, we now state explicitly that these drivers were chosen based on previous experimental, field, and modelling work that identifies bottom-water temperature, organic-matter and carbonate supply, saturation state, and near-bottom hydrodynamics as the dominant controls on benthic DIC, O2, and TA fluxes. **L420:** These predictors were chosen based on experimental, field, and modelling studies that identify them as the dominant physical and biogeochemical controls on benthic DIC, O2, and TA fluxes. Bottom-water temperature controls molecular diffusion coefficients and the rates of redox reactions (Yuan-Hui and Gregory, 1974), the POC flux drives organic-matter degradation and calcite dissolution (Emerson and Bender, 1981), while PIC concentration and the saturation state govern carbonate dissolution and precipitation (Boudreau, 2013). The current velocity modulates SWI exchange and carbonate kinetics via its effect on diffusive boundary-layer thickness (Boudreau, 2013; Sulpis et al., 2019; Glud et al., 2007), and dispersion (McGinnis et al., 2014). Within the RADIv2 ensemble, these five drivers explain most of the variance in the simulated benthic fluxes across the six calibration sites, while keeping the metamodel compact and compatible with the variables typically available as diagnostics in GOBMs.

RC2: Also, it will be beneficial to comment on the computational performance Radiv2 compared to version 1 as this version has exploited Julia's differential equations packages. I agree with the comment of improvement in computational efficiency of the newer version of Radi but details on how much and how well this improvement is will be useful to help guide potential users in their expectation as other competing framework are growing in this space (see SedTrace.jl – J Du, 2023).

AC: We agree that comparing the computational of RADIv2 to RADIv1 is a valuable addition to the manuscript. To keep the comparison as direct as possible, we re-ran the deep-sea Southern Pacific case (input file IC_SM7) from the original RADIv1 paper with both model versions, using comparable physics (no dispersion, constant DBL, identical forcing and boundary conditions). All timings below are for a single CPU core.

For a 50-year spin-up, RADIv1 required 5 min 20 s, whereas RADIv2 (using Rosenbrock23 with a sparse KLU linear solver and abstol = 1×10^{-6}, reltol = 5×10^{-3}) completed in 40 s

For a 2000-year integration of the same set-up, RADIv1 required ≈ 4 h 30 min, while RADIv2 completed in ≈ 27 min

When we activate the new physics in RADIv2 (varying DBL and dispersion) for the same 2000-year run, the time remains of the same order (≈ 28 min), showing that the added physical realism does not make the system prohibitively expensive to integrate.

The main performance gain comes from the stiff ODE integrator with adaptive time-stepping and a sparse KLU solver. RADIv1 used a forward Euler scheme with a constant time-stepping strategy, which forces the model to use this timestep, even after the solution is close to steady state. RADIv2 takes many small steps initially, but then increases the time step once the system has largely equilibrated, which is particularly beneficial for long term simulations. Moreover, using a sparse LU factorization like KLU only uses the non-zero entries of the Jacobian, which reduce memory use and computation, so stiff simulations run faster.

With reltol = 5×10^{-3} and abstol = 10^{-6}, the solver keeps the local error per timestep below about 0.5% of the current concentration for typical state-variable magnitudes.

For most Julia ODE solvers (like Rosenbrock23), the local error per timestep is kept below

allowed error=max(abstol, reltol·|y|).

This means that the solver is allowed to be wrong by at most:

either 1e-6 (a hard minimum absolute tolerance),

or 0.5% of the current value (5×10^{-3} = 0.5%), whichever is larger.

For example, at 1 mol m^{-3} the allowed local error is 0.005 mol m^{-3}, and at 10^{-3} mol m^{-3} it is 5×10^{-6} mol m^{-3}. We consider this sub-percent accuracy sufficient for the equilibrium benthic fluxes, as the error will be larger in underlying assumptions than the solver. Tightening tolerances further mainly increases runtime without appreciably changing the fluxes.

We have added a new subsection "Performance of RADIv2 versus RADIv1" within the "Numerical methods and computational enhancements subection" where this comparison is described.

Specific comments

RC2: L71: How does RadiV2 steady state work. Julia differentialEquations.jl can deal with steady state but depending on the dynamic itself (eg oscillating equilibrium, unstable equilibrium), it can be difficult in finding solution space that is well-posed (Steady State

Solvers · DifferentialEquations.jl). The paper state nothing about the numerical scheme employed in this new version of Radi that makes it different from the previous version (which I believe was a first order Euler-based scheme).

AC: This is a great remark. Indeed, differentialEquations.jl can deal with steady state itself. You can change the line prob = ODEProblem(f, u0, tspan, (model_params=model_params,)) to SteadyStateProblem(f::ODEFunction, u0, p = NullParameters(); kwargs...)) in the model to make use of the build-in steady state solver. One disadvantage, as the reviewer points out, is that this can be a "black box" approach to steady state.

RADIv2 does not rely on the generic SteadyStateProblem interface of DifferentialEquations.jl and this is also not used for the runs in this paper. For the diffusive–reactive column with fixed boundary conditions considered here, we expect a unique stable steady state. Solutes equilibrate fast (years–decades), while particulate OM equilibrates much more slowly (hundreds–thousands of years) because it is buried and accumulates. So, when solid OM reaches a steady state, we assume that everything else is also at a steady state. For steady state, we require that the maximum change in the concentration of solid organic carbon (fast + slow) over the final 5 years of integration is below 10^{-3} mmol m^{-3} across the full depth of the simulated sediment column. We now describe this explicitly in the revised manuscript in **L324:** Ensemble simulations were integrated with the stiff Rosenbrock23 solver to steady state, using adaptive time stepping and a sparse linear solver. Dissolved species equilibrate on time scales of years to decades, whereas particulate organic matter adjusts more slowly as it is buried. We therefore defined steady state by requiring that the maximum change in depth-integrated solid organic carbon (fast + slow pools) over the final 5 years of integration was less than 10^{-3} mmol m^{-3} across the full depth of the sediment column.

We now explicitly state that both options for steady state are available in **L71:** Steady states can be obtained either by calling the steady-state solvers in the DifferentialEquations.jl framework or by time-integrating the transient system until a user-defined convergence criterion is met.

Due to the flexibility of DifferentialEquations.jl, different problem types can be paired with appropriate solvers depending on their stiffness. In the simulations presented in the paper, we found that the stiff Rosenbrock23 with adaptive time stepping and a sparse linear solver was efficient for these initial conditions. However, other different environments could perform better using different solvers. The setup used is described in more detail in our response on the numerical scheme and the RADIv1–RADIv2 performance comparison.

RC2: L100: Is the dz constant, if so, what interval is it? If it is constant dz over the depth domain, then the resolution of dz is important if RadiV2 is employed in coastal sediment as processes operating in the upper few cm of the sediment drive the computed benthic fluxes. Thus, not properly resolving the underlying diagenetic processes at finer spatial scale will be detrimental to the predicted benthic flux across SWI.

AC: In the standard configurations used in this study, we use a uniform vertical grid with a spacing of $dz = 5\times10^{-3}$ m, which is a user-defined input in the setup files. We consider this dz a compromise between accuracy and computational cost for the global set of applications presented here, matching the range of fluxes in the observations.

To assess the sensitivity to vertical resolution, we performed a dedicated test for the Iberian Margin coastal station 99–6,for which well-resolved porewater profiles and inputs for to constrain RADIv2 are available. In this test we repeated the simulation with dz = 1 mm, 2 mm and 5 mm, and compared both the porewater profiles. This Figure is added to the supplementary material (Sect. S1, Fig. S1). In these tests, O2 and NO_3^- profiles remain relatively similar across resolutions, while coarser grids mainly affect the deeper part of the NH_4^+ profile.

This is now mentioned explicitly in **L324:** In all configurations used in this study, we employ a uniform vertical grid with a spacing of $dz=5\times10^{-3}$ m. A sensitivity test for different dz values for the Iberian Margin coastal is presented in the supplement material (Sect. S1, Fig. S1).

We fully agree that future applications would benefit from, ideally, a non-uniform vertical grid, particularly in strongly stratified coastal sediments or in simulations that require a deeper sediment column while keeping computational cost low. RADIv2 is a "living" model, and in the revised manuscript, we now note that future RADIv2 developments focus on supporting non-uniform grids (for example, with finer dz near the SWI and coarser spacing at depth) to further improve the resolution of near-surface processes. **L466:** Fluxes across the sediment–water interface are controlled by processes that are strongly concentrated in the upper centimetres of the sediment. The vertical grid spacing must be sufficiently fine to resolve the associated reaction and concentration gradients, which can be computationally costly. Future applications of RADIv2 could benefit from a non-uniform vertical grid, with finer spacing near the sediment–water interface and coarser spacing at depth, particularly in coastal sediments or simulations that require a deeper sediment column. RADIv2 is a living model, and ongoing developments focus on supporting non-uniform grids (for example, with finer dz near the SWI and coarser spacing at depth) to further improve the efficiency of simulating of near-surface processes.

RC2: Table 1: My impression of RadiV2, is that it can simulate coastal early diagenetic processes. However, in these settings, the role of anaerobic pathway featuring metals (Iron in particular) coupled to the carbon cycle is strong. This will necessitate modelling secondary redox processes relating to metals (Iron and Manganese, sulfide and its minerals). In iron rich sediment, for example there can be significant routing of reduced sulfide to mineral phase (e.g pyrite) which can be either buried or undergo redox oscillation and ultimately alter alkalinity budget in marine sediment (Canfield, D. E., et al. (2005)). Is there a justification why this is omitted redox pathways are not included in the model?

AC: We agree with the reviewer that secondary redox processes involving Fe, Mn and sulfide minerals can play a significant role in coastal sediment biogeochemistry.

In the present RADIv2 configuration we do include dissolved Fe(II) and Mn(II) and their oxidised solid phases (Fe(III) oxyhydroxides and Mn oxides), and we include their role as electron acceptors in organic matter degradation. However, we do not yet resolve a full sulfide–mineral network. Sulfide produced by sulfate reduction is consumed via simple pathways (oxidation and precipitation). As a consequence, RADIv2 in its current form is better suited to capturing the first-order partitioning between aerobic and anaerobic remineralisation, than to closing Fe–S–TA budgets in iron-rich and bioturbated coastal sediments.

This limitation is identified as a target for future model development and stated in the Discussion starting at **L467:** Both RADIv1 and RADIv2 are optimised for simulating carbon and oxygen cycling. In the present configuration, RADIv2 includes dissolved Fe(II) and Mn(II) and their oxidised solid phases, and represents their role as electron acceptors in organic-matter degradation. However, it does not resolve a full Fe–Mn–S mineral network. Sulfide produced by sulfate reduction is treated with simplified oxidation and precipitation terms (see Table 1), and processes such as explicit FeS/pyrite formation, burial and re-oxidation are not included. These missing pathways can be important for the alkalinity budget in iron-rich, bioturbated coastal sediments, because they control whether sulfide is buried as pyrite or repeatedly re-oxidised and recycled back to sulfate through Fe–S redox cycling (these alternative pathways have different net TA yields (Canfield et al. 2005)). Future extensions of RADIv2 should focus on implementing more comprehensive iron and manganese cycles with explicit sulfide–mineral pathways, following formulations such as (Fossing et al., 2004) and (Canfield et al. 2005), to improve applications in such settings.

RC2: Section 2.1.2, I think the organization of the paper section can be improved. Generally, the paper can provide more information on the computational cost associated with the model improvement relative to the previous version.

AC: We added a RADIv1-RADIv2 comparison in the revised manuscript, see previous response. We moved section 2.1.2 to the end of the method section (now section 2.2.3), as the tests refer to the DBL thickness and the dispersion, which were not introduced in the paper yet.

RC2: L180 Eqn 6, what does b0 mean? No mention of the term in the text and it appears again in eqn 12 (L215). My assumption is it related to surface bioturbation rate, but I might be wrong.

AC: Yes, that is correct. It is the bioturbation constant coefficient at the surface from Archer et al 2002, calculated as a function of the POC rain as: 0.0232810e-4*(100*Fpoc)^0.85. This is now specified in the revised manuscript starting in **L181.**

RC2: Same for eqn 8 (L200). The terms in the equation are not clearly defined. What does C2 mean? Concentration in the second layer (or what).

AC: Thank you for noticing this is a mistake in the equation. C2 should be Cw, the concentration of the bottom water. This is corrected and both terms Cw and Cz are explained in the text in **L200.**

RC2: L258 I think the cross-reference equation numbers are wrong. Eqn 10 is related to bioirrigation and has little to do with bottom water viscosity. ***v*** **should be substituted in eqn 14. Am I right in this assertion? If so, please fix the numbering.**

AC: Yes that is correct, thank you. *v* (eq 15) should be substituted in Equation 14. Also, Equation 16 is said "to be expressed in Equation 12 as", while this should be Equation 14. Both cross-references have been corrected in the revised manuscript.

RC2: In the practical application, is U in Eqn 14 derived from the terms (*k, v, u) and calculated implicitly or is it defined by user. Browsing through the code posed in zenodo, it seems U is imposed by the user or application. I suppose both options are valid.**

AC: The current velocity (U) is defined by the user, as it is a more easy to constrain input than the friction velocity directly. From U and T the friction velocity u* is calculated following Eq. 16, which is then used calculate the DBL thickness and the dispersion.

RC2: Table 2: It seems that the depth considered here only validates the model for continental shelf/and adjacent coastal as well as deep Ocean. For delta and shallow coastal area, how will the coverage of the ensample of tested region perform.

AC: This is a good point raised by reviewer 2 and aligns with the comment from reviewer 1. The locations listed in Table 2 were selected because they (i) provide sufficient data to constrain and validate RADIv2 and (ii) jointly span the range of conditions used to train the metamodel. At this stage, we did not include deltaic or very shallow coastal sites, because with the current limited number of calibration sites a single such location would have had strong leverage on the regression coefficients, unless it was artificially given less weight in the regression. For deltaic and very shallow coastal environments, specialised site- or region-specific metamodels may be more appropriate than a single universal formulation. However, and as result, RADIv2 has not yet been systematically tested in deltaic or very shallow coastal sites, and the metamodel likewise lacks explicit representation of these settings (as is also the case for carbonate-rich tropical shelves noted by Reviewer 1).

This is a limitation of the present work, and in the revised Discussion we now clarify environmental coverage explicitly, and we highlight training on a wider, yet physically consistent, ensemble of simulations as a key avenue for future improvement, particularly for more localized applications in deltas and shallow coastal systems in **L502:** However, the present training and validation set primarily samples temperate shelf, margin, and deep-sea settings, and does not explicitly include environments such as deltaic systems, very shallow nearshore regions, or carbonate-rich tropical shelves. Expanding the training ensemble to a wider, yet physically consistent, range of bottom-water conditions and particulate fluxes, that can also be validated with in-situ observations, represents an important future improvement, especially for more localized applications, rather than global estimates.

RC2: Are the values for pFast/slow as well as kinetic constant literature based? from previous modelling/experimental work done in these sediments or are they arbitrarily chosen?

AC: The reactivity of the OM pools can be set manually, for example based on literature values or measurements, or it can be calculated from the empirical flux relationship used in RADIv1 (tuned for deep-sea settings). We have now explicitly stated this in the revised manuscript starting from **L109:** The rate constants for the fast and slow organic carbon pools can be specified directly by the user, for example based on literature values or in-situ measurements, or they can be diagnosed from the empirical flux relationship used in RADIv1 (tuned for deep-sea settings (Sulpis et al., 2022)).

For the values in Table 2, wherever site-specific constraints are available (e.g. the Iberian Margin runs), we base pfast/pslow and the kinetic constants on existing modelling and experimental studies for those sediments. In other cases, the parameters are fitted so that RADIv2 reproduces the observed ranges of benthic fluxes, but this tuning is performed within the envelope of degradation rate constants compiled by Arndt et al. (2013) for 2-G/3-G diagenetic models.

Discussion

RC2: In the new version of Radi, it was stated that it can perform transient simulation, however in some of these regions there can be significant variability in physical (tide, resuspension/erosion) and biogeochemical forcing (changing boundary condition) that either need to impose or modelled. Did you test some configuration of this new version of the model in simulating these transient dynamics? Perhaps for on some stations. How were the results. Perhaps including some of that in the supplementary text will be useful.

AC: We agree that testing RADIv2 under explicitly time-varying physical and biogeochemical forcing is important, especially in coastal settings where tides and seasonal productivity strongly modulate benthic fluxes.

For the sites listed in Table 2 we focused on steady-state simulations. This choice was driven by the primary goal of building and the steady-state metamodel. To demonstrate the transient capabilities of RADIv2 and how it responds to time-dependent forcings, we have added a dedicated transient test case in the Supplement material. In this experiment we impose both a physical and a biogeochemical forcing, with two different frequencies. (i) An M2 tidal current signal that modulates the diffusive boundary layer thickness and dispersion, and (ii) two prescribed phytoplankton blooms that modulate the reactive organic matter supply.

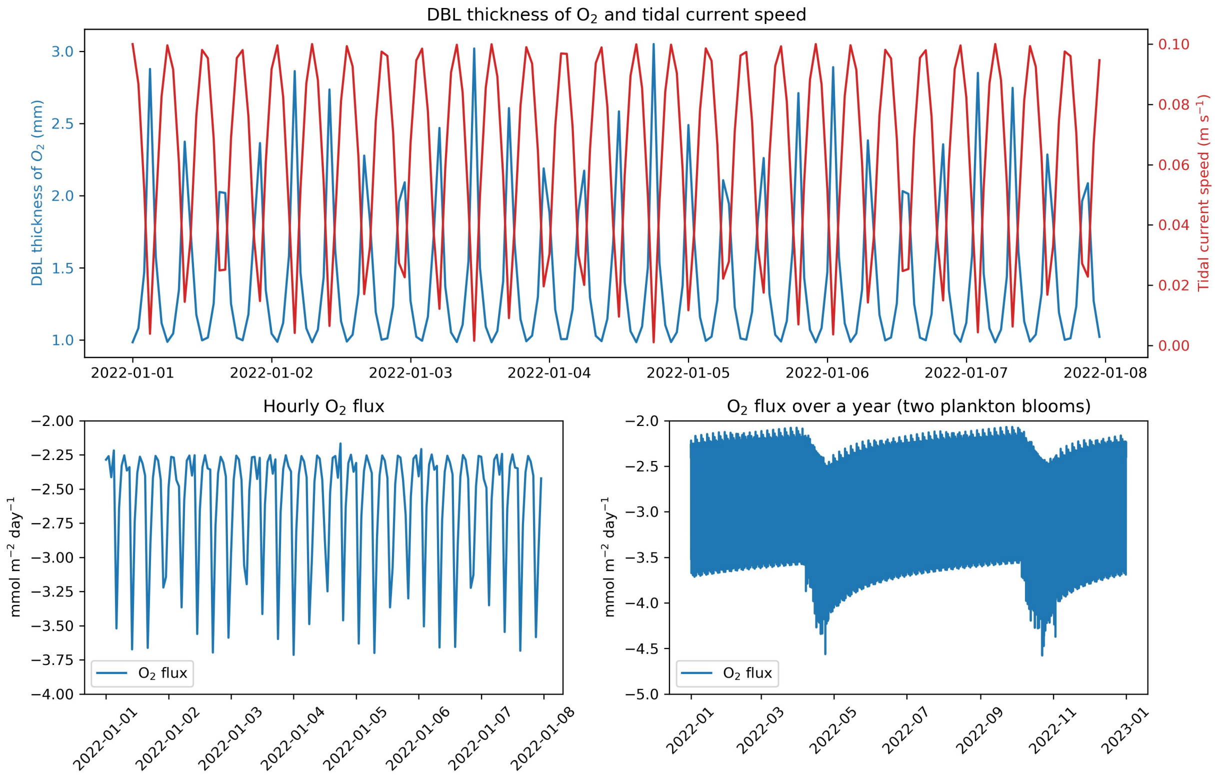
In this set-up (see Figure below), hydrodynamics exerts a clear control on the O_2 flux, which varies by roughly 50-60% over the tidal cycle. This magnitude of change agrees with Glud et al., (2007), finding that the O_2 uptake in Aarhus Bay could vary by 30% or more because of changes in the DBL. The non-linearity in O_2 uptake arises primarily from the dispersion term, which scales non-linearly with current speed (Equation 19 the manuscript). However, his range is much smaller than the ∼25-fold variability over tidal cycles reported by McGinnis et al. (2014), (This variability is hydrodynamically driven by dispersion and the DBL effect is not treated explicitly). Their results rely on more sophisticated turbulence and bottom-drag parameterisations than those currently implemented in RADIv2. In designing RADIv2, we deliberately limited the number hydrodynamic "dials" (e.g. roughness length, TKE dissipation, eddy size as they are hard to constrain inputs).

In addition, this experiment highlights an interesting interplay between current speed and benthic fluxes in RADIv2. Higher current velocities enhance dispersion and pump oxygen into the sediment, weakening the bottom-water–porewater gradient, while at the same time thinning the DBL. In permeable, high-energy environments, however, solute exchange across the SWI is not purely molecular diffusion but spans over different regimes (e.g. Voermans et al., 2018). In RADIv2, these regimes, and the conditions under which each dominates tracer exchange, are not yet fully included. Extending the flux formulation towards such a framework would further improve SWI exchange in permeable, high-energy settings.

During the bloom periods, the O_2 uptake increases by about ~20%, relative to baseline conditions. The magnitude and timing of this increase depend on the reactivity of the OM pulse: highly reactive carbon produces a stronger, near-instantaneous enhancement of O_2 uptake, whereas more refractory carbon is retained longer in the sediments and remineralizes gradually, leading to a more sustained increase in O_2 uptake over longer timescales.

We now describe this configuration and its main results in the Supplementary Text and refer to it in the manuscript starting at **L325**: Although all validation experiments in this study are run to steady state, RADIv2 also supports transient simulations with time-dependent boundary conditions and forcings. This transient configuration with tidal current variability and seasonal plankton-bloom forcing is presented in the Supplement (Sect. S2, Fig. S2).

And in **L447**: A transient test case with tidal currents (Supplement material, Sect.S2, Fig.S2) illustrates how the present formulation in RADIv2 translates such hydrodynamic forcing into time-varying benthic O_2 fluxes.



RC2: Summary of RadiV2 performance: Here, a short sentence on the computational performance of RadiV2 as compared to the previous version would be useful.

AC: In the “Summary of RADIv2 performance” section we focus specifically on how RADIv2 compares to in-situ observations and to previous published model–data estimates, which are not simulated in RADIv1. We have added a separate subsection where we present an as much as possible "apples-to-apples" computational performance comparison between RADIv2 and RADIv1 for identical test cases. There we report the relative speed-up and numerical behaviour of RADIv2 with the new stiff ODE framework and solvers, see response to previous comment.

We have also renamed this section to "Summary of RADIv2 performance against in-situ observations and to previous published model–data estimates" to avoid confusion.