Responses to Reviewers' Comments for Manuscript egusphere-2025-6078

Solving calibration and reanalysis challenges of ocean biogeochemical dynamics with neural schemes: a 1D vertical model case-study.

Addressed Comments for Publication to

Biogeosciences

by

Jean Littaye, Laurent Memery and Ronan Fablet

Dear Dr. Banerjee,

Please find enclosed our responses to your comments on our previous submission entitled "Solving calibration and reanalysis challenges of ocean biogeochemical dynamics with neural schemes: a 1D vertical model case-study.". In this document, we have carefully addressed all of your questions. A further revised resubmission will incorporate the changes you suggested to enhance clarity.

Sincerely,

Jean Littaye, Laurent Memery and Ronan Fablet

Authors' Response to the Editor

General Comments. This work paper addresses calibration and reanalysis challenges in ocean biogeochemical dynamics using neural networks in a 1D vertical model framework. The preprint is generally well-motivated and the topic is timely, especially at the interface of data assimilation, ocean modelling, and biogeochemistry. I find the manuscript promising overall. However, I have a few scientific concerns, mainly regarding the consistency of the DA formulation, the interpretation of the "weak-constrained 4DVar" setup, clarification of the closed-system vs authors' nitrate nudging argument and whether the 2-year spin-up is sufficient to ensure a repeatable annual cycle and negligible year-to-year drift in the nitrogen inventory.

Response: Thank you for your remark.

Your concerns pertain to the lack of clarity in the descriptions of the experiments and the oversight regarding the assumptions. The ensuing discussion methodically explores and clarifies each of these aspects.

Comment 1

Section 3.1, page 9 line 218-225: "the considered forcings are U* since exact forcings U are not available" the problem is posed due to noisy forcings and the target operator is meant to coestimate parameters, states, and corrected forcings as well. But in section 3.2 (page 11, lines 265-275), the author states DA uses forcings as error-free ("..the considered forcings is assumed to be error-free..") and forcing uncertainty is being pushed into the observation-error covariance. This is concerning and a major conceptual mismatch, as the whole point is about calibration under uncertain physical forcing. Authors should clarify this.

Response:

As you mention, the learning-based approach directly accounts for physical uncertainties in its formulation. In this way, it coestimates different variables from noisy forcings. Conversely, the 4D-Var approach operates under the assumption of error-free forcing, as it does not explicitly account for the aforementioned uncertainty. However, explicitly representing these physical errors is computationally prohibitive in realistic 3D configurations. Furthermore, the precise accounting for the physical error is challenging due to the necessity of specifying an appropriate error covariance matrix, a task complicated by the nonlinear relationship between the physics and the BGC state variables.

We acknowledge that the phrasing may be misleading, and this will be clarified in the revised version of the manuscript.

Comment 2

Section 3.2, page 10, lines 233–240: "This weakly constrained 4Dvar scheme (Fablet et al., 2021b; Frerix et al., 2021; Tr.molet, 2007) seeks to identify an optimal set of BGC parameters.." and "This error is handled by dividing the studied time series into sub-windows.." : The authors implemented a "weak-constraint 4DVar" via sub-window consistency or a model error penalty ($\tau_{DA} = 10^-3$), which is not a standard weak-constraint 4Dvar in the true sense with explicit model error/forcing error control. This approach is okay, but the author needs to justify whether they want to call it weak constraint 4dvar in the "standard sense" or something more appropriate.

Response:

In the definition of the variational cost functions (Equations 5–7), the model error is weighted by the factor τ_{DA} relative to the observation error. We recognize that this choice is uncommon. To recover a more standard formulation, this factor will be incorporated into the model error covariance matrix B. Furthermore, we will provide an explicit expression for this matrix, namely $B = \mathbf{1}\frac{\tau_{DA}}{{\sigma_X}^2}$, where $\mathbf{1}$ is the identity matrix and σ_X denotes the standard deviation of the BGC variables across all our simulations.

Comment 3

Section 2.1, page 4, lines 100–104: "This Neumann condition guarantees a closed model with no outflows. . . " and on contrary they say in section 2.1, page 5, lines 115–125: "the sinking of detritus leads to the depletion of nitrogen from the system over time. . . an additional term is introduced. . . .": The authors state Neumann conditions gives a closed model with no outflows but then they say detritus sinking depletes nitrogen from the system and they need nitrate nudging to compensate. The author needs to clarify this. So, is the system closed or not?

Response:

The system exhibits no outflow, neither at the surface nor in the bottom layer. The latter affirmation is representative of the diffusion of the BGC components. However, the sinking of detritus is associated with a potential outflow at the base of the resolved model. Consequently, a nitrate nudging mechanism is initiated during the winter months to counteract this depletion. The term "closed model" is indeed erroneous and will be rectified in the revised version of the manuscript.

Comment 4

Section 2.2, page 7, lines 143–144: "Each simulation includes a 2-year period of spin-up with a constant nitrogen concentration as initial condition." The authors should explicitly demonstrate that the chosen spin-up period is sufficient. For example, Bianchi et al. (2023) performed a 650-year spin-up in a 1D nitrogen cycle model and reported that steady state was achieved only after about 100 years under their setup. This does not necessarily mean that the present study also requires such a long spin-up, as the model configuration is different. However, the authors should show whether the final annual cycle is repeatable and whether the vertically integrated nitrogen inventory exhibits negligible year-to-year drift by the end of spin-up. This is important because biogeochemical model skill can depend on spin-up duration and residual dependence on initial conditions, and equilibration times are sensitive to the chosen boundary and/or restoring formulation.

Response:

An experiment was previously conducted in order to assess the influence of this spin-up length on the simulated BGC variables. As demonstrated in Figure 1, the experiment exhibits a negligible impact on the integrated nitrogen after a period of two years. This will be incorporated into the supplementary material to provide a comprehensive argument for this choice.

Comment 5

Section 3.3, page 12, line 305-310: The authors state the input tensor has size (Nbatch*Nch*NT*NT) where NT=240, but the linkage between the 120-day analysis period, 35 vertical levels and this 240*240 represenattion is not clear to me and needs more explanation.

Response:

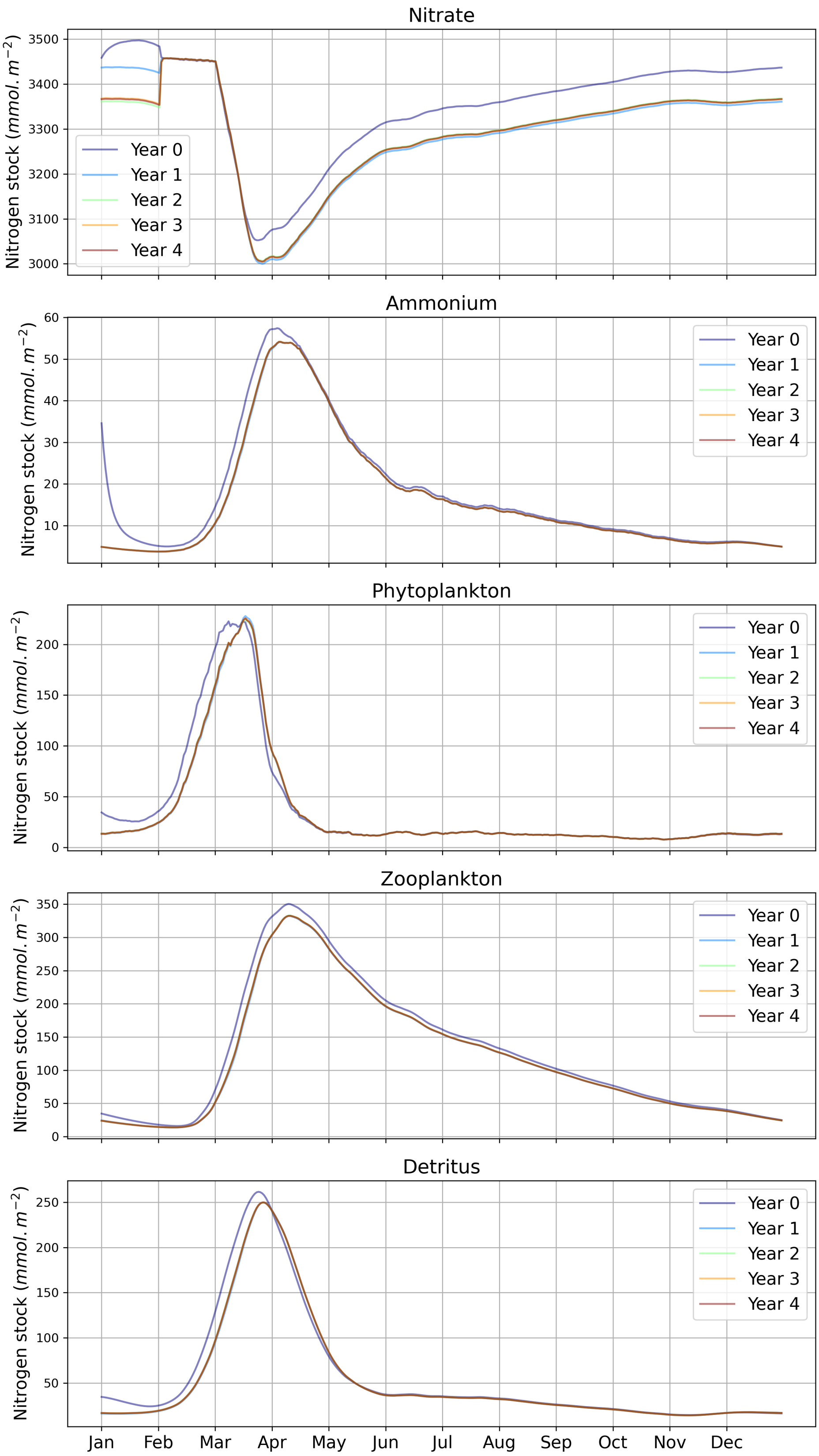

Figure 1: **Simulated nitrogen stocks,** averaged over one hundred samples across five consecutive years. The stocks are shown for nitrate, ammonium, phytoplankton, zooplankton, and detritus (from top to bottom). The one hundred samples constitute the test dataset and differ in their sets of BGC parameters and physical forcing.

Thank you for pointing out this error. The input tensor has dimensions $(N_{batch} \times (N_{ch} \times (N_z + 1)) \times 2N_T)$, where N_{batch} is the number of samples per batch, $N_{ch} = 7$ is the number of channels, $N_z = 34$ is the number of layers at which tracers are resolved (vertical diffusion is computed at the interfaces, i.e., at $N_z + 1 = 35$ depths), and N_T is the time-window length in days (since the forcings are provided at 12-hour intervals). This will be corrected and clarified in the revised version of the manuscript.

Comment 6

Section 2.3, page 7, lines 151–165: The authors represented uncertainty as spatial shifting of forcing, but in a real scenario, reanalysis error includes amplitude bias, timing bias, and vertical profile error due to mixing. The author should discuss these limitations properly, which will strengthen the paper.

Response:

In the present study, the physical errors under consideration are represented as a time-varying spatial shift. This spatio-temporal uncertainty is responsible for the mislocated submesoscale structures observed in contemporary physical reanalyses. The considered error scale represents the finer simulated scale with regard to current small-scale physical reanalyses [1, 2]. A pertinent question to pose would be an evaluation of the impact of larger-scale biases, including amplitude and timing biases, on these calibration approaches. These errors are common in global simulations, such as reported in [3].

Comment 7

Section 4.4, paragraph 1: page 20, around lines 423–429: The authors introduced a hybrid method here, which, as a reader, appears quite late. I suggest the exact role of the hybrid relative to the two other methods should be stated earlier in the methods, especially what is and what is not re-estimated in the DA stage.

Response:

The novel approach presented here offers an additional method for evaluating the relevance of the corrected physical forcing produced by the UNet. This additional method is intended to refine the states while maintaining the UNet calibrated parameters constant. A thorough description of this method, accompanied by its justification, will be provided in the methodology section.

Comment 8

(1) Table 1: lambda is 0.05 d^{-1}, whereas in the model description (Section 2.1, paragraph 2: page 5, around lines 120–123), lambda is set to 1 d^{-1} during Feb and 0 otherwise. One of them is wrong and the other is correct, I suppose. Clarifications needed.

(2) Figure 5 caption: "leaning-based scheme" should be learning-based scheme.

(3) Fig. 7 caption: "30-day state sampling" appears twice, in both red and brown distributions.

(4) Section 3.2, page 11, line 273: "the considered forcings are assumed" should be "...forcings are assumed".

Response:

The value of lambda is indeed 0.05 d^{-1} during February and 0 otherwise. These and the other technical aspects will be addressed in the revised version of the manuscript.

References

[1] Romain Escudier et al. "A high resolution reanalysis for the Mediterranean Sea". In: *Frontiers in Earth Science* 9 (2021), p. 702285. DOI: `10.3389/feart.2021.702285`.

[2] Andrea Storto et al. “Ocean reanalyses: recent advances and unsolved challenges”. In: *Frontiers in Marine Science* 6 (2019), p. 418. DOI: 10.3389/fmars.2019.00418.

[3] Olivier Boucher et al. “Presentation and evaluation of the IPSL-CM6A-LR climate model”. In: *Journal of Advances in Modeling Earth Systems* 12.7 (2020), e2019MS002010. DOI: 10.1029/2019MS002010.