Response to Reviewer #1 (RC1, egusphere-2026-711)

We thank Reviewer #1 for the review and comments. The comments on the presentation, the figures and the comparison to observations have guided much of the revision.

The reviewers' questions about the optimization and its convergence criterion led us to test CMA-ES (covariance matrix adaptation evolution strategy, Hansen and Ostermeier 2001) in place of the genetic algorithm. Our genetic algorithm gave results that depended on the random seed: the best cost and the recovered parameters differed substantially from one run to the next. CMA-ES reaches a lower cost with fewer model evaluations (Fig. 7), and it reduces the dependence on the random seed. This lowers the risk that our identifiability conclusions reflect an instability of the optimizer rather than the information carried by the data, and it lets us report the single best converged solution per experiment instead of an ensemble, which also removes the ensemble diagnostic (the HDI) that the reviewer found unclear.

Beyond this change of method, we have taken all of the reviewer's comments into account. The main revisions are a restructured manuscript with fuller, standard sections, a direct comparison of the model to in-situ observations (new Fig. 5), and reworked figures. We reply to each comment below, in the reviewer's own order, grouping closely related ones and cross-referencing where useful. Section, figure, equation and table numbers refer to the revised manuscript.

General evaluation

> *The development of a simplified model focused on zooplankton represents a valuable contribution to understanding/modelling zooplankton dynamics. That said, the manuscript would benefit from revisions to the written presentation, as well as improvements to the figures. The Methods, Results, and Discussion sections sometimes contain very short paragraphs ('one sentence') or subsections consisting of only a single paragraph, which do not fully convey the scope of the work or the significance of the results. Given the amount of work presented, these sections could be reorganized and expanded to better communicate the authors' findings.*

Author response. We agree, and the manuscript has been rewritten with this in mind. The one-sentence paragraphs have been merged into coherent paragraphs describing a single idea, and the sections now follow a standard structure that follows the logic of the study, from the reference model and its open re-implementation, through the validation and the sensitivity analysis, to the twin experiments. The Methods are now organized as the reference model (Sect. 2.1), SeapoPym and its optimization framework (Sect. 2.2), the forcing and observations (Sect. 2.3), the validation and metrics (Sect. 2.4), the Sobol analysis (Sect. 2.5) and the CMA-ES optimization (Sect. 2.6). The Results are reorganized around three messages: the model validation and the impact of transport (Sect. 3.1), the sensitivity analysis (Sect. 3.2), and the twin experiments (Sect. 3.3), and the Discussion is expanded into themed paragraphs. The specific one-sentence-paragraph comments below (L63, L87-90, the Results and the Discussion) are all covered by this rewrite.

Changes in the manuscript. Whole manuscript restructured into standard sections with consolidated paragraphs (Sects. 2 to 4 and the Conclusions).

> *The most challenging aspect of the manuscript is the lack of comparison with observational data for zooplankton biomass. Although this is a model development study, it is not sufficient to compare only two modeled zooplankton outputs. Including comparisons with observational datasets would significantly strengthen the study. This is an essential point that would also enrich both the Results and Discussion sections. I therefore encourage the authors to include more systematic comparisons, both within model results and against observations.*

Author response. We agree, and the revision strengthens both comparisons the reviewer asks for, against observations and among the model outputs. The new Fig. 5 compares SeapoPym (0D, reference parameters) and SEAPODYM-LMTL (2D) to long-term net-tow mesozooplankton biomass at HOT and BATS, the two stations with open in-situ records. Both models differ substantially from the observations, under-estimating the observed mean of 0.34 g C m^{-2} by about 30% at HOT and over-estimating the observed mean of 0.16 g C m^{-2} by roughly a factor of 2.4 at BATS. Yet the two models differ from each other far less than from the observations, the model-to-observation RMSE being 7.1 times the model-to-model RMSE at HOT and 5.5 times at BATS (Fig. 5, Table B1). Omitting transport is therefore a second-order error at these two warm, oligotrophic stations, which justifies a local 0D identifiability analysis there. This comparison concerns magnitude only, not seasonal timing. Among the model outputs, the reworked Fig. 4 maps the 0D-to-2D difference over the global ocean (mean biomass, RMSE, MAPE and bias) and Table B1 tabulates it per station (see the specific comment on Fig. 4 below).

The present paper is deliberately a controlled identifiability step rather than a calibration against real data. It first establishes, with synthetic observations, which parameters are recoverable and in which regimes, an observing system simulation experiment (OSSE) that is a prerequisite to any real-data calibration. The broader, systematic comparison the reviewer suggests, against the MAREDAT, COPEPOD and KRILLBASE databases and against biogeochemical-model outputs (CMIP), belongs to that real-data effort (see the specific point below). The reference model's documented skill against COPEPOD ($R^2 = 0.11$, global-mean biomass 1.19 versus 0.76 g C m^{-2}, Titaud et al. 2024) is now stated in the Introduction as the premise of the study.

Changes in the manuscript. New Fig. 5 and the associated text (Sects. 2.3, 2.4 and 3.1). Reference-model skill against COPEPOD quoted in the Introduction, which now frames the study as a controlled identifiability step before any real-data calibration. Per-station gaps tabulated in Table B1. The real-data calibration is named as the subject of a follow-up study in the Discussion and Conclusions.

Detailed comments

Introduction

> *Line 35: The reference is not traceable. It appears to be a general website address, but a proper, citable source is needed.*

Author response. Replaced by the citable product DOI. The Copernicus product is now cited as GLOBAL_MULTIYEAR_BGC_001_033, https://doi.org/10.48670/moi-00020, together with its Quality Information Document, Titaud et al. (2024).

Changes in the manuscript. The generic link is replaced by the product DOI and identifier, cited in Sects. 2.1 and 2.3 and in the Data Availability statement.

> *Lines 38-40: The sentence beginning with "The existing ..." requires a reference.*

Author response. The revised Introduction no longer carries this unsupported sentence, and the underlying claim is now attributed where it belongs. The advection-diffusion-reaction transport scheme of SEAPODYM-LMTL, which SeapoPym omits, is described and cited in Sect. 2.1 (Sibert et al. 1999, Lehodey et al. 2008), and the tightly coupled nature of the operational C++ implementation is stated there as well.

Changes in the manuscript. Transport scheme attributed to Sibert et al. (1999) and Lehodey et al. (2008) in Sect. 2.1, where the tight coupling is also stated.

> *Paragraphs 3-4: These paragraphs could be merged. The transition to "addressing these implementation challenges. . . " is difficult to follow, as the previous paragraph focuses on zooplankton group descriptions in SEAPODYM-LMTL and vertical layer structure in last sentences. Clarifying what "these challenges" refers to, and combining the paragraphs, would improve readability and flow.*

Author response. The Introduction has been rewritten so that each problem is followed by the response it motivates, and the vague "these challenges" is gone. It now names the three obstacles to using SEAPODYM-LMTL as a research tool: its parameters were never estimated jointly, its agreement with observations is modest, and its C++ code is not freely accessible and tightly coupled. Each is tied to the SeapoPym response, an open and modular re-implementation of the biology with a parameter-estimation framework. Because zooplankton observations are heterogeneous and sparse, whether they carry enough information to identify the parameters is an open question. The paper answers it with controlled twin experiments, in an observing system simulation experiment (OSSE) framework.

Changes in the manuscript. Introduction reorganized around the problem-to-response logic (the obstacles to using the reference model, then SeapoPym as the answer), with the disconnected paragraphs merged.

Methods

> *The Methods section contains many short subsections and dense text. A schematic figure summarizing the workflow would greatly improve clarity, especially since multiple analyses and simulations are conducted.*

Author response. We agree, and we have added a schematic overview of the workflow (Fig.~D1, Appendix~D). It shows the forcing fields (currents, temperature and net primary production), the two models (SEAPODYM-LMTL with transport in C++ and SeapoPym without transport in Python), the CMA-ES optimizer, the in-situ observations at HOT and BATS, and the four experiments, each linked to the components it uses and to the figures that report it.

Changes in the manuscript. New schematic overview figure added (Fig.~D1, Appendix~D), referenced from Sect. 2.2.

> *Line 63: There are several one-sentence paragraphs throughout the manuscript. These should either be merged with surrounding text or expanded for clarity.*

Author response. Addressed by the rewrite described in the general evaluation. This information is retained but folded into a broader paragraph rather than left standing alone.

Changes in the manuscript. Covered by the section-level restructuring (Sect. 2).

> *Table 1: Are there reference values for these parameters available in the literature? If so, please cite them.*

Author response. Yes, and they are now given with their sources. Table 1 lists, for each of the five parameters, the symbol, meaning, operational reference value with its unit, the range explored, the origin of the reference value with its literature citation, and the Appendix A equation where it appears. The recruitment and mortality temperature relationships come from the copepod development and metabolism literature (Huntley and Lopez 1992, Gillooly et al. 2002, Lehodey et al. 2010), and the energy-transfer coefficient is a maximum-likelihood estimate against the COPEPOD database (Titaud et al. 2024).

Changes in the manuscript. Table 1 given an Origin column with literature citations and an Eq. column. Prior partial calibrations described and referenced in Sect. 2.1, pointing to Table 1.

> *Lines 87–90: Another one-sentence paragraph, consider revising.*

Author response. Addressed by the rewrite, as for L63.

Changes in the manuscript. Covered by the restructuring (Sect. 2).

> *Figure 2: The light blue dots (39,853) in the caption. What are these?*

Author response. The figure and its caption have been redrawn, and the obsolete reference to "39,853 light blue dots", left over from an earlier version, is removed. The background is now described for what it is, a two-dimensional histogram of the ocean grid cells binned by their mean temperature and mean net primary production, the color giving the number of cells per bin.

Changes in the manuscript. Fig. 2 caption rewritten to describe the 2D histogram, obsolete sentence removed.

> *Table 2: This table only lists latitude and longitude of stations that are not analyzed in detail. Its inclusion as a separate table may not be necessary. This information can be also given somewhere else. Instead, you can consider giving a table summarizing the simulation that have been conducted.*

Author response. Rather than remove the table, we have expanded it into a more informative one. The former latitude-longitude table becomes the per-station summary Table B1, which keeps the station positions and adds, for each station, the mean temperature, the recruitment age at the mean and warmest points, the 0D-to-2D difference (RMSE and MAPE), and, where in-situ records exist, the model-observation gap and its ratio. The coordinates are also stated inline where the stations are introduced (Sect. 2.3). The reviewer's alternative suggestion, a table summarizing the simulations conducted, is met by the new workflow schematic (Fig. D1, Appendix D), which lays out the experiments and the data each one uses.

Changes in the manuscript. The coordinate table is expanded into the per-station summary Table B1 (positions, environmental conditions, recruitment ages, 0D-to-2D errors and the model-observation ratio). Coordinates also given inline (Sect. 2.3). The experiments are summarized in the new workflow schematic (Fig. D1, Appendix D).

> *Line 128: The reference "REF. QUID, 20224" cannot be found in the reference list. Please check and correct.*

Author response. The broken pointer is replaced. The product is now cited by its DOI and identifier (GLOBAL_MULTIYEAR_BGC_001_033, https://doi.org/10.48670/moi-00020) and its Quality Information Document as Titaud et al. (2024), a Copernicus Marine Service technical report (CMEMS-GLO-QUID-001-033) with a direct link to the document, so both are traceable and appear in the reference list.

Changes in the manuscript. The product is cited by DOI and ID, and the QUID as Titaud et al. (2024), in Sects. 2.1, 2.3 and the Data Availability statement.

> *Subsection 2.5: This section includes three equations but minimal written explanation, along with several very short paragraphs. If these equations are important enough to include, they should be properly explained in the text.*

Author response. We have kept the equations and now explain them, as a second reviewer also requested. The metrics are introduced together in Sect. 2.4 with all symbols defined, y_i and $\hat{y}_i$ being the reference and evaluated values at time step i among the n steps of the series. Each is given a one-line interpretation: the bias is the mean difference $\hat{y}_i - y_i$, so positive and negative errors cancel, whereas the RMSE measures the magnitude of the difference and the MAPE the mean absolute difference relative to the local biomass. The optimization cost is defined separately in Sect. 2.6 as the RMSE normalized by the mean of the target series (the NRMSE), so that stations of different biomass enter on a comparable scale. We write the equations out rather than only cite them because their exact form fixes the sign convention: $\hat{y}_i$ minus y_i places the evaluated value first, so a positive bias reads directly as an over-estimation of the reference.

Changes in the manuscript. Metrics consolidated and explained in Sect. 2.4 (equations for bias, RMSE and MAPE, with all symbols defined). The optimization cost is defined as the mean-normalized RMSE in Sect. 2.6.

Results

> *Similar to the Methods section, there are many short paragraphs and some subsections consisting of only one paragraph. Consider reorganizing the Results around the main messages for better readability. Some sentences are difficult to connect with the corresponding figures.*

Author response. The Results are reorganized around their three messages, as described in the general evaluation, and each claim is now tied to the figure that supports it. The validation and impact of transport are gathered in Sect. 3.1, the sensitivity analysis in Sect. 3.2, and the twin experiments in Sect. 3.3, with explicit pointers to Figs. 3 to 7, Table 2 and Table B1.

Changes in the manuscript. Results reorganized into three message-driven subsections (Sects. 3.1 to 3.3) with explicit figure references.

> *Line 187 ("Impact of transport"): This section compares zooplankton biomass from SeapoPYm and the original model (SEAPODYM-LMTL), but the comparison is difficult to interpret. It can be considered: (1) Presenting both results in the same figure (Figure 4), (2) Showing absolute differences. In the same figure RMSE sub-figure can be kept as it is.*

Author response. We have followed both suggestions and added a third panel. Fig. 4 is now a four-panel figure that keeps the RMSE panel as the reviewer proposes, adds the SeapoPym mean biomass on the same color scale as the reference biomass of Fig. 1a so the two fields can be read side by side, adds the signed difference (bias) of SeapoPym minus SEAPODYM-LMTL to show where and in which direction the two models differ, and adds the MAPE, the difference taken relative to the local biomass, which a second reviewer also requested to bring out the low-biomass regions. The text reads the panels together (Sect. 3.1), the RMSE locating the large absolute differences along the strong currents and at high latitudes, the bias showing where transport displaces biomass, and the MAPE bringing out the low-biomass equatorial band.

Changes in the manuscript. Fig. 4 turned into a four-panel figure (SeapoPym biomass, RMSE, MAPE, bias), read jointly in Sect. 3.1.

> *Additionally, comparing model results with zooplankton datasets (e.g., MAREDAT, COPEPOD, KRILLBASE) would strengthen the manuscript. A comparison with outputs from global biogeochemical models (e.g., CMIP datasets) would also provide useful context for discussion.*

Author response. We added an order-of-magnitude comparison of both models against the HOT and BATS in-situ series (Fig. 5, general evaluation above). For the COPEPOD database, the reference model's documented skill ($R^2 = 0.11$) is now quoted in the Introduction. The broader, systematic comparison against the databases the reviewer lists and against biogeochemical-model outputs belongs to the real-data calibration we are preparing as a separate paper, rather than a partial version folded into this identifiability study.

Changes in the manuscript. As in the general evaluation (Fig. 5, COPEPOD skill in the Introduction, systematic comparison deferred to the follow-up study).

> *Line 226 (Subsection 3.4): This paragraph seems more appropriate for the Methods section.*

Author response. We agree. This introductory paragraph no longer appears, and the Results section now opens directly on the findings.

Changes in the manuscript. Introductory paragraph removed from the Results (Sect. 3.3), which now opens on the results.

> *Figure 7: The caption should be improved for clarity.*

Author response. The figure and its caption have been remade. Figure 7 now shows the optimization convergence, the best-so-far cost against the number of model evaluations for the seven experiments. The cost is the mean-normalized NRMSE, and the caption explains the restart selection and how to read the curves. The parameter recovery, previously hard to read, is now a separate shaded table (Table 2) in which each entry gives the signed relative error of a recovered parameter against its reference.

Changes in the manuscript. Fig. 7 replaced by the CMA-ES convergence figure with a full caption. Parameter recovery shown separately as the shaded recovery Table 2.

> *Line 250: The term "HDI" is introduced without explanation. This should be defined earlier, likely in the Methods section.*

Author response. The HDI no longer appears. It was a property of the ensemble of solutions, and the revised analysis reports only the single best restart of each experiment rather than an ensemble (a simplification a second reviewer also pointed toward). The uncertainty discussion now rests on the recovery table (Table 2) and on the notion of equifinality, both defined in the text, so there is no undefined term left.

Changes in the manuscript. HDI and the ensemble analysis removed. Recovery reported through Table 2.

Discussion

> *The Discussion would benefit from further development. In particular, there is little comparison between model results and observational data. Including such comparisons would significantly strengthen the manuscript and its impact. There is a one-sentence paragraph, these should be revised for better flow.*

Author response. The Discussion has been expanded into themed paragraphs and the one-sentence paragraph is gone. It now develops the lessons of the twin experiments for calibration, the divergence between the Sobol sensitivity and the twin recovery, what the experiments leave open, the interplay between transport error and identifiability across regimes, the implications for observing-system design, and the limitations and the way forward. The comparison to observations runs through it, with the model-observation gap of Fig. 5 used to argue that neglecting transport is a second-order error where data exist, and the move to real, uncertain observations named as the next step.

Changes in the manuscript. Discussion expanded into themed paragraphs (Sect. 4), the observation comparison woven through it and the real-data calibration set as future work.

> *Line 303: "UPV" likely refers to "UVP"*

Author response. Thank you, the typo no longer appears. The rewritten text refers to the emerging observation technologies (automated imaging on autonomous platforms, environmental DNA, acoustic profiling) without the erroneous acronym.

Changes in the manuscript. The sentence carrying the "UPV" typo has been rewritten (Introduction and Discussion).
