# Response to Reviewer #3 (RC3, egusphere-2026-711)

We thank Reviewer #3 for the review and comments. The points raised, together with those of the other reviewers, led to the revisions described below.

The comments on the optimization procedure 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 is the methodological basis of the revised manuscript.

We also revised the manuscript in response to the reviewer's other comments. The revised version retains the main analyses from the original submission, but presents them in a clearer structure with a more complete methodological description. The main changes are a dedicated description of SEAPODYM-LMTL and of the literature sources used to define its parameters, a rewritten Appendix A, a comparison between the model and in-situ observations, and a table of abbreviations.

We respond to each comment below. Where comments address the same methodological point, we group the responses and cross-reference the relevant sections. Section, figure, equation, and table numbers refer to the revised manuscript.

---

## General evaluation

> *This ms makes a half-baked impression at best. The authors use a whole host of terms (DEAP, NSGA, etc.) which are not explained.*

**Author response.** The terminology is now either defined or removed. DEAP and NSGA-II belonged to the genetic algorithm, which is no longer the optimization method (the optimization now uses CMA-ES, Sect. 2.6). The remaining abbreviations and acronyms are collected in a dedicated table, and the optimizer is described in plain terms in Sect. 2.6.

**Changes in the manuscript.** Optimization rewritten around CMA-ES (Sect. 2.6). New abbreviations and acronyms table (Table C1, Appendix C), referenced from the Introduction.

> *The same applies to SEAPODYM. Since SEAPOPYM is based on SEAPODYM, I think it is necessary to explain clearly the structure of SEAPODYM. All I could find in the provided references was an improved version of SEAPODYM, again without a proper explanation of the original SAPODYM.*

**Author response.** We revised Sect. 2.1 so that SEAPODYM-LMTL is introduced before SeapoPym. The new section covers the origin and purpose of SEAPODYM-LMTL, the distinction between the full SEAPODYM (upper trophic levels, tuna) and the LMTL component re-implemented here, the three euphotic-depth layers, and the day and night structure and diel vertical migration of the micronekton groups.

Section 2.1 also sets out the biological processes of the mesozooplankton group: energy intake, aging, temperature-dependent recruitment and mortality. The original formulation is attributed to Lehodey et al. (2010) and Conchon (2017), which reproduce the governing equations. The complete equations are in the rewritten Appendix A.

We also clarified the code-availability context. The operational C++ implementation used to produce the CMEMS product is not freely accessible. This access limitation is now identified in the Introduction as one motivation for the open re-implementation presented here, and the tightly coupled structure of the original C++ implementation is described in Sect. 2.1.

**Changes in the manuscript.** New Sect. 2.1 "SEAPODYM-LMTL (the reference model)". Limited accessibility of the operational C++ code named as an obstacle in the Introduction, with its tight coupling described in Sect. 2.1. Equations in Appendix A.

> *Then, SEAPOPYM is never compared directly to observations in any of the figures. We are only given the NRMSE and MAPE costs. This figures only compare SEAPOPYM and SEAPODYM, but what if both were bad?*

**Author response.** We agree that comparing two modeled outputs alone is insufficient, and we now add a direct comparison to in-situ observations. The new Fig. 5 compares SeapoPym (0D, reference parameters) and SEAPODYM-LMTL (2D) to long-term net-tow zooplankton biomass at HOT and BATS, the two stations where such records are openly available. Both models differ substantially from the observations, under-estimating at HOT and over-estimating at BATS, and reducing this discrepancy motivates the present work. The two models nonetheless differ from each other five to seven times less than either differs from the observations (Table B1), so neglecting transport is a second-order effect relative to the structural model-observation gap.

The reference model is indeed known to be imperfect, with a weak fit to the COPEPOD database ($R^2 = 0.11$, global mean biomass 1.19 versus 0.76 g C $m^{-2}$ in its Quality Information Document, Titaud et al. 2024), and this is now stated as the premise of the study in the Introduction.

This imperfection is why the study is deliberately a controlled twin (OSSE) identifiability step rather than a calibration against real data, with the fit to real observations left to a follow-up study (see urgent points 3 and 5).

**Changes in the manuscript.** New Fig. 5 and the associated text (Sects. 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.

> *Another important but unanswered question is how far the parameter optimisation can compensate for the missing transport.*

**Author response.** Parameter compensation for omitted transport is a potential issue when the calibration target contains a transport signal. In the present twin experiments, this compensation cannot occur because the targets are generated by the transport-free model itself. As a result, the optimizer has no transport signal to absorb.

This question applies to calibrations against real observations. To prepare for that step, Fig. 4 and Table B1 quantify the difference between the 0D and 2D simulations by region and by station. These diagnostics identify where omitted transport contributes to the mismatch and where a 0D calibration is less affected by circulation.

At HOT and BATS, the 0D-2D transport gap is five to seven times smaller than the model-observation gap (Fig. 5, Table B1). A real-data calibration can use this diagnostic to assess where parameter changes may compensate for omitted transport (see also urgent point 3).

**Changes in the manuscript.** Explicit statement that the twin target is transport-free by construction (Sect. 2.6). Transport map and per-station table retained as the diagnostic (Fig. 4, Table B1).

> *The model equations are very difficult to follow, appear partly inconsistent, and are obviously incomplete. Several references to web-sites appear wrong or too generic. A major problem of these deficiencies is that they make it extremely hard to assess the scientific relevance of this work.*

**Author response.** Appendix A has been rewritten to address the readability, consistency, and completeness of the equations. The production dynamics are now written in the form implemented in the model, with the source at age zero and the recruitment boundary at the maximum age.

The revised Appendix A uses the same terminology throughout. The survival rate $\mu$ and the transfer-rate term have been removed, and the recruitment flux is linked to the production field. The equation-level points are addressed in the specific replies below.

**Changes in the manuscript.** Appendix A rewritten. See the specific points on Appendix A for the equation-level detail.

> *Also, the cost functions appear rather simplistic and are far behind the current state of the art (see below).*

**Author response.** This point is addressed in our reply to the specific comment on L224–225, where we discuss NRMSE and CRPS.

---

## What is urgently needed

> 1. *SEAPODYM must be explained thoroughly. Also, is the SEAPODYM code available?*

**Author response.** This point is addressed in the general evaluation above. Sect. 2.1 now describes SEAPODYM-LMTL, and the operational C++ code is not freely accessible.

> 2. *A glossary and a table of abbreviations are needed to explain the many abbreviations and concepts, as mentioned below.*

**Author response.** A table of abbreviations and acronyms has been added.

**Changes in the manuscript.** Table C1 (Appendix C), referenced from the Introduction.

> 3. *Regarding the comparison between SEAPODYM and SEAPOPYM: What is missing is a comparison of the optimised SEAPOPYM with the reference SEAPODYM. This could indicate how far the parameter optimisation may compensate for the missing circulation effects.*

**Author response.** We do not add this experiment, because it would change the scope of the manuscript from identifiability to transport compensation. The proposed experiment would use the SEAPODYM-LMTL (2D) biomass as the target for a SeapoPym optimization, so that the recovered parameters would have to absorb the missing transport, and the comparison between the optimized SeapoPym and the 2D reference would quantify that compensation.

This is the same compensation question raised in the general evaluation, where we map where transport matters (Fig. 4, Table B1). It becomes substantive only against a transport-containing target, the 2D model or real observations, which is the follow-up study.

**Changes in the manuscript.** The 0D-2D transport difference is mapped globally (Fig. 4) and tabulated per station (Table B1) as the indicator of where a 0D calibration is legitimate. The twin-experiment protocol is stated explicitly in Sect. 2.6.

> 4. *The model equations need to be made complete, e.g., the effects of DVM and interaction among different zooplankton groups, and better explained (see below).*

**Author response.** Appendix A now states the scope of the implemented equations. The present study models the single epipelagic mesozooplankton group, which does not undergo diel vertical migration. Therefore, no DVM term appears in the equations for this group.

Appendix A also states that the SEAPODYM-LMTL functional groups are not trophically coupled. Each group is driven independently by primary production, so the equations contain no inter-group interaction terms. Sect. 2.1 describes DVM only for the micronekton groups, where it applies in the reference SEAPODYM-LMTL formulation. The previous statement suggesting a straightforward extension to micronekton has been removed (see the specific point on L314-315).

**Changes in the manuscript.** Appendix A opening paragraph states the single-group scope, the absence of DVM for this group, and the absence of inter-group coupling. DVM described for micronekton in Sect. 2.1.

> 5. *A figure comparing the model results (SEAPODYM and SEAPOPYM with the reference and optimised parameters both globally and per station) with the actual observations used for calibration.*

**Author response.** The revised manuscript includes a comparison of the model to in-situ observations at the reference parameters. Fig. 5 compares SeapoPym and SEAPODYM-LMTL to long-term net-tow observations at HOT and BATS, as described in the general evaluation.

The present manuscript does not optimize the model against in-situ observations. Its scope is the analysis of model structure and parameter identifiability in controlled twin experiments. The HOT and BATS records are used as an order-of-magnitude check, not as calibration targets, and SEAPODYM-LMTL enters only as the 2D reference and is not itself tuned. The manuscript therefore contains no optimized-against-observations comparison, at global scale or by station.

Within this study, the optimized model is evaluated against the synthetic target of each twin experiment, which is the recovery the reviewer asks to see. The revised CMA-ES experiments reach an NRMSE close to zero at the informative stations, and the recovery is reported in Fig. 7 and Table 2. Because the target is generated by SeapoPym itself, the biomass simulated from the recovered parameters overlaps the synthetic target in the time-series comparison shown in Fig. R1 below.

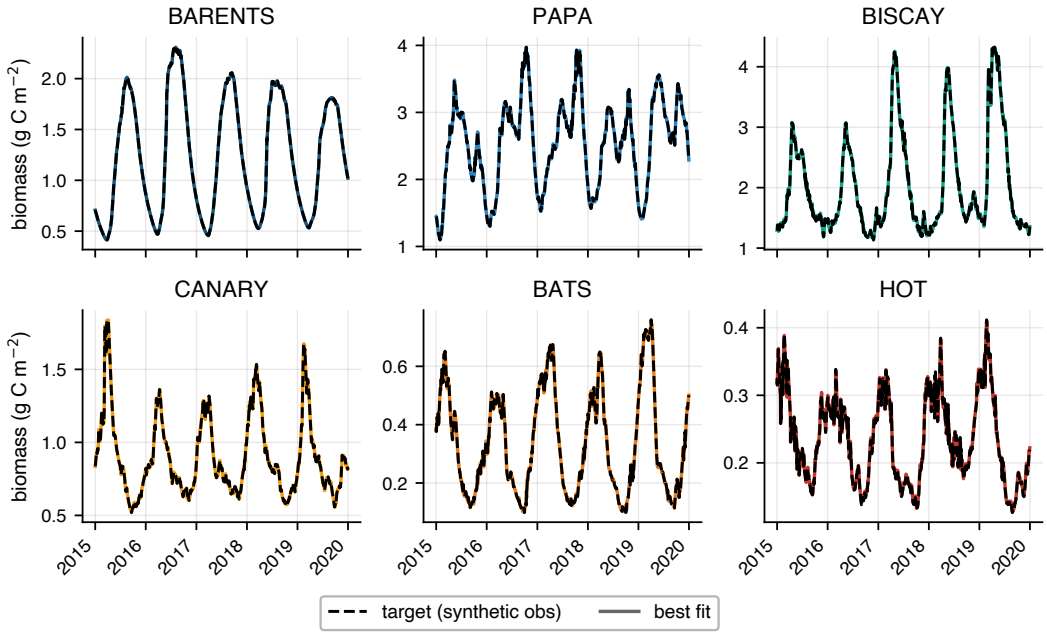


**Fig. R1.** Per-station time series of the synthetic observations used as the optimization target (dashed black, generated by SeapoPym at the reference parameters) and of the biomass simulated from the best recovered parameters (solid, station color), for the six single-station twin experiments. One panel per station ordered from cold to warm. The 2015 to 2020 window is shown for legibility, a subset of the full 2000 to 2019 analysis period. At each station, the recovered-parameter simulation overlaps the synthetic target.

**Changes in the manuscript.** Model-versus-observation comparison at reference parameters added (Fig. 5). The recovery of the optimized model against its target is reported in Fig. 7 and Table 2.

---

## Specific points for revising the ms

> *L 35: the link https://data.marine.copernicus.eu is too generic, no reference to SEAPODYM seems to exist there*

**Author response.** The generic URL has been replaced by the product DOI.

**Changes in the manuscript.** The product is now cited as GLOBAL_MULTIYEAR_BGC_001_033, https://doi.org/10.48670/moi-00020 (Titaud et al. 2024), in Sects. 2.1 and 2.3 and in the Data Availability statement.

> *L 38: current C++ implementation: is the SEAPODYM code available?*

**Author response.** This point is addressed in the general evaluation. The operational C++ code is not freely accessible, and an older version is available from the authors on request only.

> *L 80: operators proposed by the DEAP library: which are these and what is the DEAP library?*

**Author response.** This point is resolved by the change in optimization method. The revised manuscript uses CMA-ES through pycma instead of the DEAP-based genetic algorithm. As a result, the DEAP operators are no longer used.

**Changes in the manuscript.** Sect. 2.6 rewritten around CMA-ES.

> *L 82: Appendix A describes the model equations, the parameters are listed in Table 1: add this info to Table 1 and refer to it here.*

**Author response.** Table 1 now provides the information needed to interpret the model parameters. For each of the five parameters it gives the symbol, meaning, reference value with unit, the range explored, the origin of the reference value with its literature citation, and a final column linking the parameter to the equation where it appears in Appendix A. The text refers to Table 1 where the parameters are introduced (Sect. 2.1).

**Changes in the manuscript.** Table 1 reworked, with columns Symbol, Meaning, Reference, Range, Origin (with literature citations), and Eq. (the Appendix A equation in which each parameter appears: A5 for $E$, A6 for the recruitment parameters, A2 for the mortality parameters). Referenced from Sect. 2.1.

> *L 86: Please define or explain what a Sobol sequence is. I don't know and could not find out from the provided references.*

**Author response.** The term Sobol sequence no longer appears in relation to the optimizer. In the submitted version, this term referred to the quasi-random sequence used to initialize the genetic algorithm. The revised manuscript uses CMA-ES with random restarts, so that initialization step has been removed.

The remaining Sobol-related analysis is the variance-based sensitivity analysis described in Sect. 2.5. The design is generated with the Saltelli sampler implemented in SALib through sample_sobol. The revised text gives the design size, $N = 8192$ base samples and $N(D + 2) = 57344$ model runs, and cites Saltelli (2002) and the SALib references.

**Changes in the manuscript.** The GA-initialization Sobol sequence is removed with the genetic algorithm. Sect. 2.5 names the Saltelli sampler (sample_sobol), gives the design size, and cites Saltelli (2002) and the SALib references.

> *L 89: What are "standard NSGA-II meta-parameters"? You need to explain briefly at least the concept.*

**Author response.** This point is resolved by the change in optimization method. NSGA-II is a multi-objective genetic algorithm and is no longer used in the revised manuscript.

Sect. 2.6 now states the CMA-ES settings used in the optimization. These settings are the default population size $\lfloor 4 + 3 \ln D \rfloor = 8$ for $D = 5$, the initial step size $\sigma_0 = 0.30$, and the standard CMA-ES convergence criteria.

**Changes in the manuscript.** NSGA-II removed. CMA-ES settings stated in Sect. 2.6.

> *L 108: Please refer to Section 2.4.3 as a reference explaining what a Sobol analysis is.*

**Author response.** The Sobol sensitivity analysis now has its own methods section. The Introduction refers to Sect. 2.5 when the Sobol analysis is first mentioned.

**Changes in the manuscript.** Cross-references to Sect. 2.5 added where the Sobol analysis is first mentioned.

> *L 128: CMEMS product (Ref. QUID, 2024): need explanation and a better reference, the URL quoted on L 395 does not show any LMTL simulation and searching for "QUID, 2024" also does not help*

**Author response.** The generic QUID reference has been replaced. The product is now cited with

its DOI and identifier, GLOBAL_MULTIYEAR_BGC_001_033, https://doi.org/10.48670/moi-00020, which points to the LMTL product page.

The Quality Information Document is now cited as Titaud et al. (2024), Copernicus Marine Service technical report CMEMS-GLO-QUID-001-033, Issue 4.2. The citation includes a direct link to the document.

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

> *L 149, 150: SALib library: please specify which functions*

**Author response.** Sect. 2.5 now specifies the SALib functions used in the analysis. The Saltelli sampler sample_sobol builds the parameter design, and sobol.analyze computes the first-order and total Sobol indices, $S_1$ and $S_T$.

**Changes in the manuscript.** Functions named in Sect. 2.5.

> *L 157: section 2.4.2: did you mean 2.4.3?*

**Author response.** The methods have been renumbered, and the faulty cross-reference has been removed. We checked the internal section references during the revision.

**Changes in the manuscript.** Section numbering and cross-references corrected throughout.

> *L 161: "by experiment" should be "for the GA" (or something similar).*

**Author response.** The sentence referred to the genetic-algorithm experiments and has been rewritten. The revised twin experiments use twenty CMA-ES restarts for each of the seven experiments described in Sect. 2.6.

**Changes in the manuscript.** Run accounting rewritten in Sect. 2.6.

> *L 169-174: please explain the symbols in the equations ($n$, $y_i$, $\hat{y}_i$, $\sigma$).*

**Author response.** Sect. 2.4 now defines the symbols used in the metrics. In the metric equations, $y_i$ is the reference value at time step $i$, $\hat{y}_i$ is the evaluated value at the same time step, and $n$ is the number of time steps.

The standard deviation $\sigma$ no longer appears in the optimization cost. The revised NRMSE is normalized by the mean of the target series, not by its standard deviation (Sect. 2.6). The text also specifies which quantities are compared in each use of the metrics: model versus reference for transport metrics, and candidate simulation versus synthetic target for optimization.

**Changes in the manuscript.** Symbols defined in Sect. 2.4 (Eqs. for bias, RMSE, MAPE). NRMSE defined as RMSE over the mean in Sect. 2.6.

> *L 179: The cost function is the metric. Not clear what you want to say here.*

**Author response.** We revised the text to distinguish the metric from the cost function. A metric measures the discrepancy between a predicted and target series, for example the NRMSE at one station.

The cost function is the scalar objective minimized by the optimizer. In a single-station experiment, the cost is that station's NRMSE. In the MERGED experiment, the cost is the mean of the

per-station NRMSE values, which assigns equal weight to each station. Thus, the cost function specifies how the metric is aggregated across the dataset.

**Changes in the manuscript.** Sect. 2.6 gives the cost as the per-station NRMSE for a single station and as the mean of the per-station NRMSE for MERGED, showing the cost as an aggregation of the metric over the dataset. Sect. 2.2 presents the cost function as a replaceable module.

> *L 198, 199: The low RMSE may be misleading in these regions. Large MAPE values could well co-occur with small RMSE, simply because of the low biomass concentrations.*

**Author response.** As the reviewer notes, the RMSE alone can be misleading in low-biomass regions, which is why the revised manuscript reports both the RMSE and the MAPE. The RMSE gives the absolute magnitude of the difference, whereas the MAPE expresses the difference relative to local biomass.

This distinction matters in low-biomass regions. In the equatorial Pacific, Indian, and Atlantic oceans, the MAPE is high even where the RMSE is low. Sect. 3.1 now interprets those regions using the MAPE rather than the RMSE alone.

**Changes in the manuscript.** MAPE panel added to Fig. 4. RMSE and MAPE contrasted in Sect. 3.1.

> *L 201–206: I suspect that this reflects a shortcoming of the Sobol' analysis, as it is a common result often seen in large ensemble simulations, when several parameters are changed simultaneously to obtain a better coverage of the parameter space. What may appear to indicate effects of parameter interaction is in reality just the result of changing several parameters at a time. If the Solbol' analysis somehow accounts for this problem, this needs to be explained.*

**Author response.** The Sobol-Saltelli design addresses the issue raised by the reviewer. In a variance-based Sobol analysis, the first-order index $S_1$ estimates the variance contribution of a parameter on its own, whereas the total index $S_T$ includes that parameter and its interactions with the other parameters. This decomposition is defined while all parameters vary simultaneously (Saltelli et al. 2002).

The concern would apply to an ensemble in which several parameters are varied at once without a variance decomposition. In the Saltelli design used here, the sampling scheme and the estimator separate the first-order and total contributions. The revised Sect. 2.5 now states this point and gives the sample size, $N = 8192$ base samples, together with bootstrap confidence intervals.

The revised results use $S_T - S_1$ as the interaction contribution associated with each parameter. For the magnitude descriptor, the parameters combine additively in $\log B$, so $S_T - S_1$ is close to zero. For the timing descriptor, recruitment age depends non-additively on its two parameters, and the interaction contribution is non-zero. This pattern is reported in Sect. 3.2.

We stored first-order and total indices, not second-order indices. Therefore, $S_T - S_1$ gives the total interaction contribution associated with each parameter, but it does not identify the interacting parameter pairs. Pairwise interactions would require second-order Sobol indices.

**Changes in the manuscript.** Sect. 2.5 explains how the first-order index $S_1$ (own-effect) and the total index $S_T$ (including interactions), and their difference $S_T - S_1$, are attributed separately by the Saltelli variance decomposition even though all parameters vary at once.

> *L 224–225: The NRMSE is certainly not the best and also not the most robust cost function for parameter optimisation, since it rests on the assumption of normally distributed observations. This is also indicated by the discrepancy between the lowest MAPE and NRMSE in Fig. 7. See, e.g.,*

> *Thorarinsdottir et al., SIAM/ASA Journal on Uncertainty Quantification 1 (2013):522, for better alternatives.*

**Author response.** Proper scoring rules such as the CRPS are suited to calibration against uncertain or probabilistic observations. The present targets are deterministic twin experiments: each target is a noise-free synthetic time series generated by the model itself. Therefore, the observation-error distribution that would justify a likelihood-based or probabilistic score is not part of the present experiment.

For deterministic predictions, the CRPS reduces to the absolute error. In that setting, a proper scoring rule would provide another deterministic discrepancy measure rather than a probabilistic treatment of observation uncertainty. CRPS becomes relevant for the follow-up calibration against real, uncertain observations.

The revised optimization cost is the RMSE normalized by the mean of the target series. This normalization keeps the cost comparable across stations with different biomass levels and avoids normalization by small temporal standard deviations. The difference between the lowest MAPE and the lowest NRMSE in the submitted Fig. 7 is now interpreted as equifinality: parameters that differ from the reference can still reproduce the biomass target, as shown in Table 2.

**Changes in the manuscript.** The optimization cost is defined as the RMSE normalized by the target mean (Sect. 2.6). The noise-free deterministic nature of the twin target is stated in Sect. 2.6, the replaceable nature of the cost function in Sect. 2.2, and the option of a proper scoring rule for non-Gaussian observation noise in future real-data calibration is noted in the Discussion.

> *Section 3.4.2: Please list the optimised parameter estimates for all stations and globally, e.g., as additional columns in Table 1. This is important information for assessing the efficiency of the GA and the cost function, and also the appropriateness of the reference parameter values.*

**Author response.** Table 2 now reports the recovered parameters for each experiment. The table gives, for each parameter and each experiment, the signed relative error of the recovered value against the reference value. The reference value is printed below each parameter symbol, and the residual cost is given for the six single-station experiments and the joint MERGED experiment.

We report relative error rather than only absolute recovered values because the experiments are synthetic. In a twin experiment, the relevant diagnostic is the distance between the recovered and reference parameters. The absolute value can still be reconstructed from the printed reference value and the reported relative error.

**Changes in the manuscript.** Table 2 (recovery table) reports the per-experiment recovery as a signed relative error against the printed reference value, with the residual cost, for all seven experiments.

> *L 259–261: This seems to be a circular conclusion, since the influence on model output is gauged by the NRMSE. It could well be that a better cost function would result in a different parameter sensitivity.*

**Author response.** The revised manuscript no longer draws this conclusion from the NRMSE. Parameter influence is now diagnosed with the Sobol sensitivity analysis, whose variance decomposition does not use the optimization cost (Sect. 2.5).

The Sobol analysis measures how each parameter affects model descriptors, whereas the twin experiments measure whether the parameters can be recovered from the biomass time series. These are distinct diagnostics. Therefore, changing the optimization cost would affect the recovery experiment, but not the Sobol sensitivity indices computed from the model outputs.

The revised results also change the interpretation of recruitment. Recruitment is not described as weakly influential. Instead, the Sobol analysis shows that recruitment controls the timing of the seasonal peak. The non-recovery of recruitment parameters at warm stations is interpreted as equifinality: parameter values far from the reference can produce a biomass time series with near-zero NRMSE.

**Changes in the manuscript.** Sect. 2.5 states that the Sobol variance is independent of the optimization cost. The Discussion adds the Sobol-versus-twin distinction (influence versus recoverability). The earlier explanation by a weak influence on the outputs is removed.

> *L 290: How is the DVM implemented? It is not mentioned in the model equations. Also, it is unclear to me how this is done in a 0-D modelling framework (e.g., L 323).*

**Author response.** DVM is not implemented for the group simulated in this study. The implemented group is epipelagic mesozooplankton, which does not migrate vertically in SEAPODYM-LMTL. Therefore, DVM does not enter the equations in Appendix A.

DVM applies to the migrating micronekton groups of SEAPODYM-LMTL, which are described in Sect. 2.1 but not simulated here. In the reference formulation, DVM enters through day-and-night layer averaging of the temperature and currents experienced by each migrating group. The revised manuscript removes the previous wording that could imply that DVM was active in the present 0D experiments.

**Changes in the manuscript.** DVM described for micronekton in Sect. 2.1. Appendix A states it does not apply to the modeled group. Ambiguous mentions removed.

> *L 294-297: But what about different cost functions (see above)?*

**Author response.** The revised Discussion now includes the cost function as a possible extension. The paragraph that previously listed only optimizer changes now states that the modular cost function can be replaced.

Examples are given for future real-data calibration. A proper scoring rule such as the CRPS could be used for uncertain or probabilistic observations, and a phase-sensitive metric could be used to constrain seasonal timing. This point is linked to the reply on L224–225.

**Changes in the manuscript.** Cost-function lever added to the future-work paragraph of the Discussion.

> *L 304: "SeapoPym is specifically designed to integrate this increasing data heterogeneity": How is this achieved? This remains completely unclear from the model description in this ms.*

**Author response.** We revised this statement to match what the implementation demonstrates. The framework separates the observation component from the cost function, so different data sources could be assigned different metrics in a future calibration.

The present study uses one biomass time series as the target. Therefore, the revised manuscript presents heterogeneous-data integration as a design property of the framework, not as a capability demonstrated by the experiments in this paper.

**Changes in the manuscript.** Claim rephrased to a modular-design statement (Sect. 2.2).

> *L 314, 315: "the modular architecture allows for straightforward extension to micronekton groups": Again, how is this implemented remains unclear from the presented equations.*

**Author response.** We removed this claim. The revised text states that the SEAPODYM-LMTL functional groups share the same formulation and differ in their vertical-layer occupancy.

The scope of the present implementation is also stated. The manuscript implements and tests the single epipelagic mesozooplankton group only.

**Changes in the manuscript.** "straightforward extension to micronekton" removed. Scope stated factually (Sect. 2.2).

> *Appendix A: The equations are apparently only for a single zooplankton group. The main text stresses the ability of SEAPOPYM to accommodate multiple groups, but then the equations should somehow allow for interactions among these groups, e.g., if one eats the other etc.*

**Author response.** The Appendix now states the single-group scope of the equations. The present study models the epipelagic mesozooplankton group only.

The functional groups in SEAPODYM-LMTL are not trophically coupled. Each group is driven independently by primary production, so the equations contain no terms in which one group consumes another. The Appendix also states that DVM does not apply to the epipelagic mesozooplankton group.

**Changes in the manuscript.** Appendix A opening paragraph.

> *It remains also unclear how many age classes are resolved.*

**Author response.** The revised manuscript states the number of age classes. With a daily time step, the model resolves $\tau_{r_0}/\Delta$ age classes, which gives about eleven daily classes at the reference temperature. This information is given in Sect. 2.1 and in the Numerical solution part of Appendix A.

**Changes in the manuscript.** Number of age classes stated in Appendix A and Sect. 2.1.

> *Eq. (A.4): I don't understand this equation. L 348–349 mention source and sink terms but (A.4) has only one (product) term on the right. The left-hand-side looks like a partial differential equation but the term $\partial p/\partial\tau$ seems to make no sense, since $\tau$ (age class) is a discrete variable (according to L 349) and hence $\partial\tau$ appears meaningless. This needs to be explained clearly (or corrected). Also, $\mu$ is termed recruitment rate on L 349 but transfer rate on L 358.*

**Author response.** Appendix A has been rewritten to clarify the age-structured production equation. The production equation is now written as the loss-free McKendrick–Von Foerster transport equation. In this formulation, production ages at unit velocity and is conserved while it ages.

The source and sink are boundary conditions. The source is the injection at age zero, $p(t, \tau = 0) = E\,\mathrm{NPP}(t)$. The sink is the absorbing boundary at the recruitment age. Therefore, the equation has no product term on the right-hand side.

The age coordinate $\tau$ is continuous in the continuous formulation, so $\partial p/\partial\tau$ is defined there. Age is discretized only in the numerical solution, where the model represents age classes.

The rate $\mu$ has been removed. There is no survival term during aging: losses before recruitment are included in the efficiency $E$, and mortality $\lambda$ acts only on the biomass pool $B$ after recruitment. This revision removes the earlier naming inconsistency.

**Changes in the manuscript.** Appendix A rewritten: loss-free transport equation, boundary-condition source and sink, continuous $\tau$ discretized only later, $\mu$ removed.

> *L 369: p was defined as production on L 347. Now you say here that for total absorption, $p(t, \tau_r) = 0$. But if nothing is produced, nothing could be transferred, which seems illogical. I may missing something here but please explain this clearly.*

**Author response.** Appendix A now defines $p(t, \tau)$ as production density per unit age and states how recruitment transfers this production into biomass. The production field has a source at age zero, $p(t, \tau = 0) = E\,\mathrm{NPP}(t)$, and an absorbing boundary at the recruitment age $\tau_r$.

The condition $p(t, \tau_r) = 0$ means that production reaching the recruitment age has been transferred into the biomass pool. It does not mean that no production exists. The recruitment flux is the production reaching that boundary, $R(t) = p(t, \tau_r(t))$, and this flux feeds the biomass pool $B$.

The survival exponential and the associated partial-transfer explanation have been removed because the implemented model uses complete transfer at recruitment.

**Changes in the manuscript.** Appendix A (Recruitment): complete-transfer scheme stated, $R(t) = p(t, \tau_r(t))$, "transfer not a loss" clarified, survival exponential removed.

> *L 374: Did you mean Numerical Solution?*

**Author response.** Yes. The heading has been corrected to “Numerical solution”.

**Changes in the manuscript.** Section title set to "Numerical solution" (Appendix A).

> *L 376: What is the Courant-Friedrichs-Lewy (CFL) condition? What is $\Delta\tau$ ($\tau$ being a discrete variable)?*

**Author response.** Appendix A now explains the age discretization in plain terms. Production ages at unit velocity, $d\tau/dt = 1$. With age classes of width $\Delta_\tau$ and time steps $\Delta_t$, the implementation sets $\Delta_t = \Delta_\tau = \Delta = 1$ day, a Courant number of one.

With this choice, one time step advances a cohort by exactly one age class, so the age transport is exact, with no numerical diffusion, and the production field needs no integration scheme. This is the plain statement that replaces the CFL term.

The step $\Delta_\tau$ is the age discretization, introduced only in the numerical solution: the continuous age coordinate $\tau$ of the transport equation is discretized at the implementation stage. The number of resolved age classes is $\tau_{r_0}/\Delta$.

**Changes in the manuscript.** Appendix A (Numerical solution): Courant-one condition explained with $\Delta = 1$ day, age discretization and class count stated, CFL jargon removed.

> *Fig. 3: The figure shows at least 3 dashed lines but the legend only 2, Asymptote 0°C and Theoretical asymptote (whose color doesn't appear to match any of the dashed lines).*

**Author response.** Fig. 3 has been redrawn. Each solid line shows the simulated biomass at one of four constant temperatures, and each matching dashed line gives the analytical equilibrium $B = R/\lambda$ for that temperature.

The legend now corresponds to the plotted lines.

**Changes in the manuscript.** Fig. 3 redrawn. The legend now has one colored entry per constant temperature for the solid simulated curves, plus a single black dashed entry labeled $B = R/\lambda$ for the analytical steady state, with both axes logarithmic. Each legend entry corresponds to a line in the plot.

> *Fig. 4: Please add a panel showing the percentage error MAPE. The RMSE may be more relevant for assessing the model output as input (e.g., food) for higher-trophic-level models, but the MAPE also provides essential information on model performance, particularly in oligotrophic regions.*

**Author response.** Fig. 4 now includes a MAPE panel. The figure contains four panels: SeapoPym mean biomass, RMSE, MAPE, and bias.

The added MAPE panel shows relative error in low-biomass regions, including the oligotrophic gyres and the equatorial Pacific, Indian, and Atlantic oceans.

**Changes in the manuscript.** Fig. 4 now includes the MAPE panel (Sect. 3.1).

---