https://doi.org/10.5194/egusphere-2026-711
Preprint. Discussion started: 1 April 2026

SeapoPym v0.1: Implementation of the SEAPODYM low and mid trophic levels in Python with a flexible optimisation framework

Jules Victor Lehodey[1,2,3], Alexandre Mignot[4], Alexandre Ganachaud[1,2], Sarah Albernhe[4], and Simon Nicol[5]

[1]LEGOS, University of Toulouse, CNRS, IRD, CNES, UT, Toulouse, France
[2]MARBEC, University of Montpellier, CNRS, IRD, IFREMER, Sète, France
[3]Université de la Nouvelle-Calédonie, ISEA - Institut des Sciences Exactes et Appliquées (EA 7484), Nouméa, New Caledonia
[4]Mercator Ocean International, Toulouse, France
[5]Pacific Community (SPC), Noumea, New Caledonia

Correspondence: Jules Victor Lehodey (lehodey.jules@gmail.com)

Abstract. SEAPODYM-LMTL is a global advection-diffusion-reaction model that simulates age-structured zooplankton and micronekton populations driven by physical and biogeochemical forcing. This study introduces SeapoPym, a simplified version of this model that decouples biological dynamics from physical transport and incorporates a Genetic Algorithm (GA) for stochastic parameter estimation within the Python scientific ecosystem. Comparisons with SEAPODYM-LMTL show that omitting transport produces notable discrepancies in highly dynamic warm regions and cold environments with long zooplankton life cycles. However, SeapoPym remains suitable for simulating mesozooplankton across most warm- and temperate-ocean regions. Sobol sensitivity analysis identifies mortality parameters as key drivers of biomass magnitude and variability, with strong parameter interactions. Twin experiments highlight challenges in estimating recruitment-timing parameters and emphasize the importance of data from cold, contrasting environments. SeapoPym provides a flexible, low-cost framework for exploring parameter estimation, designing observational strategies, and addressing challenges in zooplankton model assessment, with the potential to integrate with circulation models or machine-learning emulators.

1 Introduction

Marine mesozooplankton, defined as organisms typically ranging from 0.2 to 2 mm in size (Sieburth et al., 1978), represent a pivotal node in pelagic food webs. While primarily composed of copepods, this diverse functional group also encompasses various planktonic larval forms (meroplankton) (Moriarty and O'Brien, 2013; Drago et al., 2022). As a key conduit of energy from lower trophic levels, mesozooplankton contribute substantially to the sustenance of micronekton—a community of larger organisms (2–20 cm) including mesopelagic fish, euphausiids, and gelatinous species (Brodeur et al., 2004; St. John et al., 2016). Given this critical role in supporting higher predators and regulating biogeochemical cycles, accurate modeling of zooplankton dynamics is essential to predict ecosystem responses to accelerating climate change, acidification, and deoxygenation (Rohr et al., 2023; Steinberg et al., 2000, 2002; Stukel et al., 2022; Siegel et al., 2023).

https://doi.org/10.5194/egusphere-2026-711
Preprint. Discussion started: 1 April 2026

A major challenge for the scientific community is the lack of consistent time-series observations spanning large scales, particularly for zooplankton and micronekton (Holland et al., 2025). There are significant gaps in both spatial and temporal coverage, as well as shortcomings in taxonomic representation (Ratnarajah et al., 2023). While traditional net trawling remains the standard method for sampling, new technologies are poised to quickly increase the number of observations (Holland et al., 2025). In the meantime, enhancing the modeling of diverse key functional groups and species, along with the development of efficient parameter-estimation methods, becomes crucial (Schartau et al., 2017). One approach to these issues is the Low and Mid-Trophic Levels modeling component of the Spatial Ecosystem and Population Dynamics Model (SEAPODYM-LMTL). SEAPODYM-LMTL mimics the distribution patterns of marine zooplankton and micronekton over space and time on a global scale, offering Essential Ocean Variables useful for modeling fish populations (Lehodey et al., 2008, 2010, 2015; Conchon, 2017; Delpech et al., 2020). The model employs a system of equations that accounts for advection, diffusion, and reaction processes, using physical data such as currents, temperature, and euphotic depth, as well as biogeochemical information, such as net primary production (NPP), derived from satellite data or other models. The water column is divided into several layers, with functional groups separated by their vertical movement strategies at different times of day. Currently, SEAPODYM-LMTL includes one group of mesozooplankton and six groups of micronekton, all of which are accessible through high-resolution global outputs available from the Copernicus Marine Service (https://data.marine.copernicus.eu).

Addressing these implementation challenges is crucial for fostering innovation in model development and for enabling more efficient exploration of biological hypotheses. The existing Advection-Diffusion-Reaction system within the SEAPODYM-LMTL model, implemented in C++, presents challenges for rapid hypothesis testing. In the current C++ implementation, biological reaction terms are tightly coupled with the transport scheme within the numerical solver; consequently, any modification to a biological mechanism requires significant changes to the resolution matrices. This structural rigidity makes code adaptation complex and time-consuming, especially as the number of simulated groups and species increases.

To overcome these limitations, we introduce a new modeling framework. This study presents SeapoPym, a simplified version of SEAPODYM-LMTL that decouples biological mechanisms from the complex transport infrastructure and leverages the Python scientific ecosystem. SeapoPym is designed as a flexible platform to address the need for rapid model development and testing. Additionally, SeapoPym incorporates a parameter-estimation module based on a Genetic Algorithm (GA), enabling stochastic optimization and maintaining independence between model design and parameter estimation.

This paper details the methodological development and validation strategy of the SeapoPym framework, with a specific focus on mesozooplankton. Numerical accuracy is demonstrated through theoretical benchmarks and comparison with the reference SEAPODYM-LMTL simulation. The model's structural behavior is then examined using a Sobol sensitivity analysis (Sobol', 2001) to assess its response to varying environmental conditions. The optimization framework's ability to recover parameters is evaluated using a twin experiment, verifying convergence to a global solution across diverse scenarios. The final section discusses model performance, optimization efficiency, and implications for future research in marine ecosystem modeling.

https://doi.org/10.5194/egusphere-2026-711
Preprint. Discussion started: 1 April 2026

2 Materials and Methods

2.1 SeapoPym model

SeapoPym is implemented as an open-source Python library released under the GPLv3 license (Lehodey, 2026). For this study, which focuses on the model's development and validation, we limit the analysis to a single zooplankton group in the epipelagic layer. However, the framework is generic and can be applied to micronekton groups by parameterizing their vertical partitioning between layers over the diel cycle (as in SEAPODYM-LMTL).

Conceptually, SeapoPym is formulated as a zero-dimensional (0D) formulation of the fundamental equations (described in Appendix A). It focuses exclusively on the biological reaction terms (i.e., recruitment, aging, and mortality), driven by environmental forcing (i.e., temperature, euphotic depth, and NPP), while explicitly neglecting spatial transport. This simplification allows the model to function independently without worrying about the complex interdependencies with the transport code.

The model adopts a modular architecture where biological processes are implemented as independent functions, enabling straightforward extension and parallel computation using the Dask library (Dask Development Team, 2016).

All data management (including input/output and intermediate states) is performed using the Xarray library (Hoyer and Hamman, 2017). This library enables the transparent use of standard data formats (e.g., NetCDF and Zarr) and deployment across heterogeneous infrastructures, from local workstations to cloud platforms, without modifying the core codebase.

A rigorous management of physical units is integrated directly into the parameterization layer using the Pint library (https://pint.readthedocs.io/), enforcing dimensional consistency and preventing semantic errors.

The framework is organized into three primary components: (i) a Configuration layer ensuring data integrity and physical units consistency; (ii) a Kernel library containing elementary biological processes, which leverage either native vectorization using the Numpy library (Harris et al., 2020) or Just-In-Time (JIT) compilation provided by the Numba library (Lam et al., 2015) to achieve execution speeds comparable to C++; and (iii) a Model orchestrator that manages the state and temporal evolution.

2.2 Parameter optimization

GAs are evolutionary optimization methods that mimic natural selection to solve global optimization problems. They evolve a population of candidate solutions through successive generations using biologically inspired operators: selection, crossover, and mutation (Goldberg, 1989). Unlike deterministic methods, GAs are stochastic and particularly effective for exploring multimodal parameter spaces where traditional techniques might get trapped in local optima (Goldberg, 1989; Falls et al., 2022). SeapoPym allows users to easily change the operators proposed by the DEAP library (Fortin et al., 2012).

Within this framework, each solution corresponds to a specific set of model parameters to be calibrated, fully described in Appendix A and categorized here into three functional groups: (i) the energy transfer coefficient E, which determines the transfer efficiency from NPP; (ii) the recruitment parameters τ_{r_0} and γ_{τ_r}, controlling the age of transition from production cohorts to recruited biomass; and (iii) the mortality parameters λ_0 and γ_λ, which define the temperature-dependent mortality rate.

https://doi.org/10.5194/egusphere-2026-711
Preprint. Discussion started: 1 April 2026

Table 1 summarizes the model parameters, their reference values, and optimization ranges. The first phase of the optimization initializes the population using a Sobol sequence to ensure comprehensive coverage of the search space.

Table 1. Model parameters with reference values and optimization ranges.

Symbol	Description	Units	Reference	Range
E	Energy transfer coefficient	–	0.1668	[0, 0.5]
τ_{r_0}	Maximum recruitment age at T_{ref}	day	10.38	[0, 100]
γ_{τ_r}	Thermal sensitivity (recruitment)	$^\circ\text{C}^{-1}$	-0.11	$[-0.5, 0]$
λ_0	Mortality rate at T_{ref}	day^{-1}	1/150	[1/500, 1/4]
γ_λ	Thermal sensitivity (mortality)	$^\circ\text{C}^{-1}$	0.15	[0, 0.5]

It then proceeds through successive generations, evaluating individuals using the objective function and tournament selection (tournament size = 3) to select promising parents.

We use standard NSGA-II meta-parameters (Deb et al., 2002) with two-point crossover ($CXPB = 0.9$) to promote exploration of the parameter space (see Discussion).

The optimization framework is organized into modules for model instantiation, parameter specification, cost function evaluation, observation handling, and evolutionary algorithm orchestration. Additional modules support constraint handling (e.g., parameter bounds) and logging functionality for tracking optimization progress.

2.3 Forcing and reference zooplankton biomass data

The SeapoPym simulations are forced with the same temperature and NPP fields used to produce the Copernicus Marine Service product "Global Ocean Low and Mid Trophic Levels Biomass Content Hindcast" (GLOBAL_MULTIYEAR_BGC_001_033, DOI: 10.48670/moi-00020). These fields are provided alongside the LMTL variables (zooplankton and micronekton biomass) at a horizontal resolution of 1/12° (Figure 1) over the global domain. The model outputs include one zooplankton and six micronekton functional groups. The mean temperature in the epipelagic layer is computed from the global ocean model configuration GLORYS12 (Lellouche et al., 2021), and the vertically integrated NPP is derived from satellite ocean-colour observations using the VGPM algorithm (Behrenfeld and Falkowski, 1997).

To reduce computational costs associated with the global domain, a spatial subsampling is applied. The domain is discretized on a regular grid of 360° longitude × 170° latitude, spanning from 80°S to 90°N and 180°W to 179°E, with a spatial resolution of 1 °. For each grid cell, the value from the nearest corresponding cell in the original dataset is selected. The temporal coverage spans 22 years, from January 1, 1998, to December 31, 2019. The model is initialized with zero biomass and production; the first two years serve as a spin-up period for the system to reach equilibrium, and are excluded from subsequent analysis. Data were extracted from the reference simulation at six locations corresponding to regularly sampled sites (hereafter referred to as 'stations'), which were subsequently used to conduct the twin experiment and Sobol analysis. They cover a range of

https://doi.org/10.5194/egusphere-2026-711
Preprint. Discussion started: 1 April 2026

environmental conditions in temperature and NPP, spanning a gradient from cold to warm temperatures and from low to high productivity (Figure 2). The station coordinates are provided in Table 2.

Table 2. Geographic coordinates of the six oceanographic stations used for data extraction.

Station	Latitude	Longitude
Barents Sea	75.0°N	40.0°E
PAPA	50.0°N	132.0°W
Bay of Biscay	45.5°N	4.0°W
BATS	32.0°N	64.0°W
Canary	30.0°N	13.0°W
HOT	23.0°N	158.0°W

2.4 Simulation experiments

A series of simulation experiments is conducted to validate the modeling approach. The first experiment verifies the correctness of the numerical implementation and assesses model behavior in a simplified, controlled scenario. The second experiment isolates the role of transport processes to determine whether model dynamics are realistic only when physical connectivity is included or, for zooplankton, remain valid in regions with weak transport. The third experiment systematically assesses structural uncertainties through a global sensitivity analysis, identifying parameters and processes that most strongly influence model outputs and those that are poorly constrained. This analysis enhances understanding of model behavior and informs the design of the optimization procedure. The fourth experiment evaluates the optimization strategy within a controlled and transparent framework using twin experiments.

2.4.1 Validation of the numerical implementation

To validate the numerical implementation against the theoretical analytical solution ($B_{eq} = \frac{R}{\lambda}$, see Appendix A and Lehodey (2001)), we conducted a series of simulations under constant environmental forcing.

Specifically, four simulations were performed at 0, 10, 20, and 30°C to span the broad physiological thermal range of marine zooplankton. A mean value of 300 $mg \cdot m^{-2} \cdot d^{-1}$ was prescribed for NPP. In this theoretical setup, the magnitude of NPP does not affect the test outcome, as it serves solely as a scaling factor for the theoretical asymptote.

2.4.2 Impact of transport

A zooplankton simulation is performed with SeapoPym using the same forcing fields and the reference parameter values listed in Table 1, as employed in the reference LMTL simulation included in the CMEMS product (Ref. QUID, 2024). The comparison of mean zooplankton biomass distributions of LMTL and SeapoPym provides a quantitative assessment of the impact of transport removal on model predictions (see metrics below).

https://doi.org/10.5194/egusphere-2026-711
Preprint. Discussion started: 1 April 2026

2.4.3 Sobol sensitivity analysis

Following the validation and evaluation of SeapoPym's implementation above, the model's inherent structure is examined using a Sobol analysis (Sobol', 2001) to determine its identifiability limits (i.e., the ability to uniquely estimate parameter values from observations) before any optimization is performed. Sobol's global sensitivity analysis characterizes the influence of model parameters on the environment using three metrics we selected: the date of the seasonal biomass peak (argmax), the average annual biomass (mean), and the biomass variance (variance). The Python library used for this analysis is SALib (Iwanaga et al., 2022; Herman and Usher, 2017). The range of environmental conditions in temperature and NPP is represented by a set of six oceanographic stations spanning a gradient from cold to warm temperatures and from low to high productivity.

In Sobol sensitivity analysis, the first-order (S_1) and total-order (S_T) indices both measure how important the model parameters are — but they capture different kinds of effects. S_1 measures the direct, independent effect of a single parameter on the output, ignoring interactions with other parameters, i.e., how much of the output variance is explained by a given parameter alone. If $S_1 = 0.3$, then 30 % of the variance in the output is caused by that parameter by itself. S_T measures the overall contribution of a parameter, including both its S_1 (direct) effect and all interaction effects with any other parameters.

The larger the number of simulations, the lower the uncertainty in the Sobol sensitivity analysis. Saltelli (2002) and Prieur and Tarantola (2019) recommended that, for a model with P parameters, a Sobol analysis requires a minimum sample size of $(P+2) \times N$, where N typically ranges from 10^3 to 10^6 for the S_1 indices. In our case, this corresponds to a total of 7×10^3 to 7×10^6 simulations. To keep computation costs reasonable, we set the number of simulations to 10^6.

The range of values explored for the model parameters covers up to five times the reference values for zooplankton. The SALib library is used to generate the ensemble of parameter sets. For each simulation, the three metrics are computed independently for each station and stored. SALib then uses all of the resulting metrics to compute the sensitivity indices, which are subsequently summarized in a synthetic plot for the analysis.

2.4.4 Parameter optimization using the GA and twin experiments

The GA optimization was applied to the five model parameters using a twin experiment. Starting from a quasi-randomly initialized parameter set, using the Sobol sampling function provided by SALib, the twin experiment assesses the optimization algorithm's ability to recover the reference parameters by fitting the pseudo-observations. These pseudo-observations are the zooplankton biomass time series (2000-2019) created from the reference parameter set with SeapoPym at the six oceanographic stations used in the Sobol analysis (cf. section 2.4.2). The parameter search interval was restricted to a reasonable range for zooplankton (0-200 % of the reference value for each parameter). The stopping criterion for the optimization was set to 20 generations: an initial population of 14,336 individuals generated using Sobol sampling, followed by 19 generations of 10,000 individuals each.

In total, 204,336 simulations are run by experiment. Individual simulation times ranged from 250 ms to 1 s depending on the computing environment, corresponding to approximately 14–56 hours on a single core or 1–2 hours with 30 cores in parallel. All experiments were conducted using Python 3.12.

https://doi.org/10.5194/egusphere-2026-711
Preprint. Discussion started: 1 April 2026

2.5 Metrics

Two classical and complementary metrics are used to quantify differences and structural divergences in space and time. The Mean Absolute Percentage Error (MAPE) is a measure of the average relative error of the model compared to the reference. The Root Mean Square Error (RMSE) quantifies the average magnitude of differences between the two simulated fields, offering a global estimate of model accuracy while allowing spatial and temporal mapping of discrepancies.

$$\mathrm{RMSE} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2} \tag{1}$$

$$\mathrm{MAPE} = \frac{100}{n}\sum_{i=1}^{n}\left|\frac{y_i - \hat{y}_i}{y_i}\right| \tag{2}$$

For the optimization, the cost function is based on the Normalized Root Mean Square Error (NRMSE) between model predictions and pseudo-observations:

$$\mathrm{NRMSE} = \frac{\mathrm{RMSE}}{\sigma} = \frac{\sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}}{\sigma} \tag{3}$$

The use of NRMSE emphasizes the model's ability to reproduce system dynamics rather than absolute mean levels and, importantly, allows comparison across multiple stations with different variances.

The GA's ability to recover the initial parameters used to generate the observations is evaluated using the Mean Absolute Percentage Error (MAPE).

The cost function can be readily adapted to use an alternative metric (see Discussion).

3 Results

3.1 Theoretical validation of the numerical implementation

Figure 3 shows the temporal evolution of the modeled biomass with SeapoPym and the theoretical convergence value. As expected from the temperature-dependent definition of mortality (Eq. A.2), the time required to reach steady state from the initial condition increases as the temperature decreases. This time is only 10 days at 30°C and 30 days at 20°C, but extends to 4 and 16 months at 10°C and 0°C, respectively. Once steady state is achieved, the difference from the theoretical value is less than 0.01 % (Figure 3) and is attributed to computational accuracy.

3.2 Impact of transport

To estimate the error induced by neglecting transport, we compare the reference SEAPODYM-LMTL simulation (Fig. 1d) against the static SeapoPym output (Fig. 4a), both run with identical forcing. The resulting spatial error (RMSE) reveals that

https://doi.org/10.5194/egusphere-2026-711
Preprint. Discussion started: 1 April 2026

the impact of transport is non-uniform and driven by the interplay of current intensity, biomass gradients, and temperature (Figure 4b).

The highest discrepancies are found in dynamically active regions characterized by strong fronts, such as the Gulf Stream, the Kuroshio, and the Antarctic Circumpolar Current. In these areas, the combination of intense transport and steep biomass gradients maximizes the divergence between the static and advective models.

Temperature also plays a critical role by modulating biological time scales (Figure 3). In colder, productive regions (e.g., the Labrador and Oyashio currents), slower growth and mortality rates extend the time required for cohort development (e.g., Daase et al. (2013)). This increases the organisms' exposure to currents, thereby amplifying transport-induced errors.

Conversely, the RMSE remains minimal in subtropical gyres and tropical regions. Here, low biomass combined with rapid biological turnover (Figure 3) limits the accumulation of discrepancies, rendering the static assumption locally appropriate.

3.3 Global sensitivity analysis (Sobol)

Sobol's analysis reveals that model variability is primarily driven by parameter interactions rather than independent effects. Across all stations, the S_T indices are substantially larger than the S_1 indices, often by a factor of two to three. This large discrepancy reveals that the direct effect of any single parameter is minor compared to the variance driven by parameter interactions. However, when considering S_T, all parameters exhibit significant influence on at least one output metric. This confirms that no parameter can be neglected and that the model behaves as a highly nonlinear system in which recruitment, energy intake, and mortality processes are tightly coupled.

While all parameters are influential, they affect specific components of the zooplankton dynamics, revealing a clear functional division. First, the mortality parameters ($\lambda_0, \gamma_\lambda$) emerge as the primary direct drivers of biomass magnitude. Their S_1 indices dominate both mean biomass and variance metrics. This is physically consistent with the model structure: whereas NPP strictly bounds energy input (E), a mortality rate approaching zero can, in principle, lead to unbounded biomass growth. In sharp contrast, recruitment parameters (τ_{r_0}, γ_{τ_r}) exert negligible influence on biomass quantity ($S_T \approx 0$ for mean and variance). Their impact is almost exclusively restricted to the timing of the seasonal peak, where they exhibit high S_T. Finally, the energy parameter (E) acts as an indirect modulator. It shows no direct effect on peak timing ($S_1 \approx 0$) and influences biomass magnitude entirely through interactions with mortality parameters, effectively linking the energy source to the mortality sink.

The sensitivity of these parameters varies across the environmental gradient. Crucially, the identifiability of recruitment parameters ($\tau_{r_0}, \gamma_{\tau_r}$) depends on the clarity of the seasonal signal. In cold environments, slow development times lead to extended life cycles of 1–2 years. This temporal scale is captured by the model dynamics shown in Figure 3, where the time required to reach equilibrium at low temperatures (0°C) extends to similar durations. This is consistent with the phenology of high-latitude copepods (Tande and Båmstedt, 1985; Conover, 1988; Kosobokova, 1999). This phenology creates sharp seasonal peaks, enhancing the sensitivity of the "Peak timing" metric to recruitment parameters in these regions compared to continuous tropical production in warm waters.

In summary, accurately constraining the model requires a cost function that simultaneously captures errors in magnitude, variability, and seasonal timing. The Normalized Root Mean Square Error (NRMSE) is particularly well-suited for this purpose,

https://doi.org/10.5194/egusphere-2026-711
Preprint. Discussion started: 1 April 2026

as it inherently aggregates these three components (Taylor, 2001). Consequently, minimizing the NRMSE across contrasting environmental regimes is the most robust strategy for parameter optimization.

3.4 Twin experiment optimization using GA

Two types of experiments are conducted to evaluate the optimization approach's capabilities. The first experiment analyzes each station individually, while the second considers all stations jointly. This illustrates the approach's ability to estimate model parameters from observations collected at one or multiple stations. It also highlights its potential for conducting OSSEs to support the design of observation networks.

3.4.1 Cost function and fit to biomass

When each station is analyzed independently, the best performance is obtained in the coldest waters (BARENTS station), where the NRMSE decreases steadily over successive generations and reaches a minimum of 0.015 after 20 generations (Figure 6a). The other stations do not exhibit the same level of continuous improvement and may reach their optimal score after only a few generations. Conversely, when all stations are used jointly, all cost functions improve over the generations and reach their best values between generations 15-20, with substantially better scores than in the single-station experiments, all below 0.1 (Figure 6b). The only exception is BARENTS, for which the solution remains good but is slightly worse than the single-station result (0.035). In terms of fit to observed biomass, an $NMRSE < 0.1$ means that the coefficient of determination R^2 $(1 - NRMSE^2) > 0.99$ (assuming no bias in errors: Müller-Plath and Lüdecke (2024)).

3.4.2 Parameter estimates

The most challenging parameters to estimate are τ_{r_0} and γ_{τ_r}, which exhibit similar patterns of evolution throughout the optimisation process (Figure 7). When considering each observation station independently, only the BARENTS station achieves excellent performance for these two parameters, with MAPE values below 1 % and 10 %, respectively. However, combining all stations in the optimisation enables recovery of their initial parameter values with errors below 10 %.

In the warm, low-productivity environment of the HOT station, λ_0 cannot be accurately estimated in separate optimisation experiments. In contrast, this parameter is estimated with an error below 10 % at the other stations, with the best performance at BARENTS (<1 %). When all stations are combined, λ_0 is also recovered with an error of about 10 %. The parameter E is estimated with errors generally below or slightly above 10 %. The parameter γ_λ yields the best performance, with errors below 10 % when stations are considered separately and below 5 % when all stations are combined.

The highest density interval (HDI) is a widely used measure of uncertainty in parameter estimation (e.g., Kruschke (2015)). It represents the narrowest range of parameter values containing a specified proportion of the posterior probability mass—commonly 95 %, as in this study—derived here from the 1,000 best solutions (Figure 8). Notably, for optimisations at individual stations, the HDI of parameter τ_{r_0} decreases only for the BARENTS station. At the same time, for γ_{τ_r} the HDI remains maximal both at BARENTS and at the other stations. Even when data from all stations are combined, the HDI for γ_{τ_r} remains

https://doi.org/10.5194/egusphere-2026-711
Preprint. Discussion started: 1 April 2026

unchanged. For the other parameters, the reduction of the HDI over successive generations is clearly evident, with particularly narrow intervals achieved for λ_0 and γ_λ.

This overall performance can be explained by the complementarity of the sensitivity profiles: cold waters constrain τ_{r_0}, λ_0, and E, whereas warm waters constrain γ_λ. The overall consistency of the optimized parameters validates the multi-station approach. The non-identifiability of γ_{τ_r} is not due to a failure of the GA—which converges well for the other parameters and obtains a solution very close to the initial result in terms of NRMSE—but rather to its weak structural influence on the model outputs.

4 Discussion

This study introduces SeapoPym, a simplified version of the SEAPODYM-LMTL model that decouples biological mechanisms from transport, using the Python scientific ecosystem, and incorporates a GA for flexible, stochastic parameter estimation. The results validate SeapoPym's numerical accuracy, explore its behavior through Sobol sensitivity analysis, and evaluate its optimization framework using a twin experiment.

The influence of advection on zooplankton distribution has been recognized for a long time and is often pronounced, particularly in regions characterized by intense physical dynamics such as frontal zones, upwelling systems, and areas with strong currents (e.g., Banse, 1964; Edvardsen et al., 2003; Mackas and Coyle, 2005). Our comparison with SEAPODYM-LMTL confirms that SeapoPym remains generally suitable for simulating and estimating mesozooplankton parameters across most warm- and temperate-water regions, except in areas with strong dynamic activity. This finding aligns with theoretical expectations that physical transport becomes non-negligible when biological timescales exceed physical retention timescales. This condition is more likely in cold waters, where the generation times of zooplankton can exceed one year (Tande and Båmstedt, 1985; Conover, 1988; Kosobokova, 1999; Daase et al., 2013).

However, since cold to temperate environments are characterized by strong seasonal cycles, they provide a clear and detectable signal for parameter estimation, as demonstrated by our results. Slow rates of development in cold-water environments exert a first-order control on biomass structure, which facilitates the estimation of certain parameter values, as shown by the twin experiments.

While these environments facilitate estimation, the S_T indices highlight indirect parameter interactions, which complicate optimization when relying on a single environment. The use of diverse observational data—especially from cold, high-contrast environments—can help constrain parameter values more effectively, as confirmed by the twin experiments. Given that these experiments were realized in an ideal framework without noise and using long, continuous time series, specific parameters will likely need to be fixed to their best-known values during model calibration in less favorable, real-world applications, especially τ_{r_0} and γ_{τ_r}, which govern recruitment timing. This strategy should help reduce the number of equally plausible solutions that fit the same observational dataset (equifinality).

The GA employed in this study is a basic, highly exploratory approach. While it entails a higher computational cost than gradient-based methods, it offers exceptional flexibility and facilitates parallel computing, enabling efficient distribution of

https://doi.org/10.5194/egusphere-2026-711
Preprint. Discussion started: 1 April 2026

the computational load (Cantú-Paz, 2001). As an evolutionary algorithm operating on discrete genotypes, the GA inherently handles both continuous and discrete variables, and applies directly without requiring any modification to the model code. For example, although the current implementation assigns a fixed Diel Vertical Migration (DVM) behavior to each group, future iterations could optimize this trait by defining it as a discrete variable. Similarly, for species-specific applications, a discrete parameterization approach would allow the algorithm to dynamically select optimal prey fields from distinct Plankton Functional Types (PFTs).

Several algorithmic enhancements could improve optimization performance: elitism to preserve the best solutions across generations (Angelova et al., 2011), Simulated Binary Crossover (SBX) to potentially accelerate convergence (Deb et al., 2002), or CMA-ES, which self-adapts to parameter correlations and search step size (Hansen and Ostermeier, 2001; Hansen, 2016).

Future model improvements will rely heavily on the availability of suitable datasets. While long-term standardized time series remain scarce, our results indicate that maintaining sampling efforts in environmentally contrasting regimes is particularly effective for parameter optimization. Beyond these fixed stations, extensive databases of discrete sampling (e.g., COPEPOD) and acoustic profiles offer immediate opportunities to refine biomass and DVM parameterizations at each grid cell (Stanton, 2000; Dmitrenko et al., 2020). Concurrently, emerging high-resolution technologies, such as automated imaging onboard autonomous profilers (UPV) and environmental DNA (eDNA), will be essential for resolving specific functional groups or species. SeapoPym is specifically designed to integrate this increasing data heterogeneity—a flexibility that extends to micronekton modeling.

Beyond parameter estimation, this framework offers a robust mechanism for observing system design. Analogous to Observing System Simulation Experiments (OSSEs) used in oceanography, our twin experiment approach allows us to quantify the potential value of new sampling stations prior to their deployment. By generating synthetic observations and forcings with realistic noise, the model can inform cost-effective monitoring strategies that maximize parameter retrieval accuracy.

Several limitations should be acknowledged. First, the omission of physical transport restricts the model's applicability to regions where biological timescales are shorter than physical retention timescales, as discussed above. Second, the twin experiments were conducted under idealized conditions—without observational noise and using long, continuous time series—meaning that real-world applications may face additional challenges in parameter retrieval. Third, the current framework focuses on a single functional group (mesozooplankton), though the modular architecture allows for straightforward extension to micronekton groups. Finally, the non-identifiability of the recruitment-timing parameter γ_{τ_r} suggests that this parameter may need to be fixed based on prior knowledge when observational constraints are limited.

5 Conclusions

This study demonstrates that SeapoPym, a Python-based reformulation of SEAPODYM-LMTL, successfully reproduces zooplankton dynamics in most oceanic regions where transport effects are secondary to biological processes. Sobol sensitivity analysis reveals mortality parameters as the primary drivers of biomass magnitude and variability, while twin experiments con-

https://doi.org/10.5194/egusphere-2026-711
Preprint. Discussion started: 1 April 2026

firm that parameter identifiability requires observations spanning contrasting thermal environments—particularly cold, high-latitude regions with pronounced seasonal cycles that facilitate the estimation of recruitment-timing parameters.

SeapoPym provides an agile 0-D modeling framework that enables non-experts to test biological hypotheses and explore parameter estimation strategies. It offers a low-cost tool for investigating how observational design and dataset composition affect parameter retrieval, without the computational burden of a complete spatial model. Future developments aim to apply the model to real-world datasets, including acoustic observations and data from emerging sensors (Everett et al., 2017). The next logical step is to address the lack of transport by coupling with an ocean circulation model (e.g., NEMO) or by using machine-learning-based emulators. Importantly, this expansion must follow strict software engineering standards to prevent the 'complexity trap' of legacy models, ensuring the code stays modular, adaptable, and accessible.

Appendix A: Underlying model Equations

A1 Biomass

The dynamics of the total biomass B are governed by the balance between recruitment and mortality:

$$\frac{\partial B}{\partial t} = R(t) - \lambda(t)B(t) \tag{A.1}$$

where:

- R is the recruitment flux $[\mathrm{g\,m^{-2}\,day^{-1}}]$;
- B is the total biomass $[\mathrm{g\,m^{-2}}]$;
- λ is the mortality rate $[\mathrm{day^{-1}}]$.

At steady state ($\partial B/\partial t = 0$), this yields the equilibrium biomass $B_{eq} = R/\lambda$.

Mortality is parameterized as a temperature-dependent function following an exponential relationship:

$$\lambda(t) = \lambda_0 \exp\left(\gamma_\lambda (T(t) - T_{\mathrm{ref}})\right) \tag{A.2}$$

where:

- λ_0 is the mortality rate at the reference temperature T_{ref} $[\mathrm{day^{-1}}]$;
- γ_λ is the thermal sensitivity coefficient $[^\circ\mathrm{C}^{-1}]$.

Within the LMTL framework, the effective temperature is constrained to positive values:

$$T(t) = \max(T_{\mathrm{env}}(t), T_{\mathrm{ref}}) \quad \text{with} \quad T_{\mathrm{ref}} = 0^\circ\mathrm{C} \tag{A.3}$$

https://doi.org/10.5194/egusphere-2026-711
Preprint. Discussion started: 1 April 2026

A2 Production

The evolution of zooplankton production $p(t,\tau)$ $[\mathrm{g\,m^{-2}\,day^{-1}}]$ as a function of time t and age τ is described by the McKendrick–Von Foerster equation (McKendrick, 1926; Von Foerster, 1959). The source term corresponds to the transfer of NPP to the first age class ($\tau = 0$), while the sink term represents the transfer from production cohorts to the biomass pool via a recruitment rate μ:

$$\frac{\partial p}{\partial t} + \frac{\partial p}{\partial \tau} = -\mu(t,\tau)\, p(t,\tau) \tag{A.4}$$

subject to the boundary condition:

$$p(t,\tau = 0) = E \times \mathrm{NPP}(t) \tag{A.5}$$

where:

- NPP is the Net Primary Production of phytoplankton $[\mathrm{g\,m^{-2}\,day^{-1}}]$;
- E is the dimensionless transfer efficiency coefficient between NPP and zooplankton production.

A3 Recruitment

The transfer rate μ activates only when the cohort age falls within a temperature-dependent recruitment window:

$$\mu(t,\tau) = \begin{cases} \alpha & \text{if } \tau \in [\tau_r(t), \tau_{r_0}] \\ 0 & \text{otherwise} \end{cases} \quad \text{with} \quad \tau_r(t) = \tau_{r_0} \exp\big(-\gamma_{\tau_r}(T(t) - T_{\mathrm{ref}})\big) \tag{A.6}$$

where:

- $\tau_r(t)$ is the minimum recruitment age at temperature T [day];
- τ_{r_0} is the maximum recruitment age at T_{ref} [day];
- γ_{τ_r} is the thermal sensitivity coefficient for recruitment age $[^\circ\mathrm{C}^{-1}]$;
- α is a constant recruitment rate $[\mathrm{day}^{-1}]$.

Assuming $\mu(t,\tau)$ is integrable, the analytical solution along the characteristic lines (for $\tau \leq t$) is:

$$p(t,\tau) = E \cdot \mathrm{NPP}(t-\tau) \cdot \exp\left(-\int_0^{\tau} \mu(t-\tau+s, s)\, \mathrm{d}s\right) \tag{A.7}$$

The exponential term represents the proportion of production retained in the cohort relative to the initial generation at $\tau = 0$. In the limiting case where $\alpha \to \infty$, total absorption occurs: all production reaching the recruitment age τ_r is instantaneously transferred to the biomass. This is equivalent to an absorbing boundary condition $p(t,\tau_r) = 0$.

https://doi.org/10.5194/egusphere-2026-711
Preprint. Discussion started: 1 April 2026

Conceptual Note: Unlike standard formulations where μ represents mortality, here it signifies a state transition flux. Conceptually, p and B represent distinct life stages (e.g., developing cohorts vs. recruited population). However, the model structure treats them differently: p is age-structured with implicit mortality (captured in the coefficient E), while B is an age-aggregated pool subject to explicit thermal mortality.

A4 Numerical Resolution

We adopt the notation X_a^n to denote the value of variable X at age $\tau = a\Delta\tau$ and time $t = n\Delta t$. To minimize numerical diffusion, we enforce the Courant-Friedrichs-Lewy (CFL) condition $\Delta t = \Delta\tau = \Delta$.

The biomass equation is solved using an explicit Euler scheme:

$$B^{n+1} = B^n + \Delta \cdot (R^n - \lambda B^n) \tag{A.8}$$

The analytical solution for p (Eq. A.7) requires knowledge of the full thermal history of each cohort, which is computationally expensive in an Eulerian framework. Therefore, p is computed iteratively from the previous time step using a survival probability S:

$$p_a^{n+1} = p_{a-1}^n \cdot S_a^{n+1} \quad \text{with} \quad S_a^{n+1} = e^{-\mu_a^{n+1}\Delta} \tag{A.9}$$

The specific recruitment flux from age class a at time $n+1$, denoted R_a^{n+1}, corresponds to the production lost by the cohort during the time step:

$$R_a^{n+1} = p_{a-1}^n - p_a^{n+1} = p_{a-1}^n(1 - S_a^{n+1}) \tag{A.10}$$

Finally, the total recruitment flux R^{n+1} feeding the biomass is the sum of contributions from all age classes within the recruitment window:

$$R^{n+1} = \sum_{a=0}^{\tau_{r_0}} R_a^{n+1} \tag{A.11}$$

Code availability. The SeapoPym v0.1 model code is open-source and distributed under the GPLv3 license. The current version of the model described in this paper is archived on Zenodo (https://doi.org/10.5281/zenodo.18348596, Lehodey 2026) and is available at https://github.com/SeapoPym/seapopym. The repository includes the source code, installation instructions, and a user manual. This documentation provides examples of simulation and optimization usage that serve as a basis for reproducing the experiments described in this paper. The specific notebooks used to generate the figures presented in this study are available upon request.

Data availability. Physical and biogeochemical forcing fields used in this study (GLORYS12, Net Primary Production) are available from the Copernicus Marine Service (CMEMS) at https://marine.copernicus.eu/ (Lellouche et al., 2021). The reference zooplankton biomass data (SEAPODYM-LMTL) used for validation and twin experiments are also distributed by CMEMS.

https://doi.org/10.5194/egusphere-2026-711
Preprint. Discussion started: 1 April 2026

Author contributions. JVL led the development of the SeapoPym software, designed the methodology, performed the simulations and sensitivity analyses, and wrote the original draft. AM and AG supervised the PhD work, provided guidance on the methodology, and contributed to the review and revision of the manuscript. SA tested new functionalities, provided feedback on the optimization framework, and contributed to the software validation. SN conceptualized the project, secured funding, supervised the overall research direction, and reviewed the manuscript. All authors approved the final version.

Competing interests. The authors declare that they have no conflict of interest.

Acknowledgements. We gratefully acknowledge the support and assistance of the Pacific Community (SPC) and the Institut de Recherche pour le Développement (IRD). We also thank the Mercator Ocean International for hosting JVL and providing computational resources, and the Copernicus Marine Service for providing the physical and biogeochemical data products. Finally, the authors are grateful to Patrick Lehodey for his careful proofreading of the manuscript and for his insightful discussions. AI-assisted tools were used for code development, code documentation, and language editing of the manuscript. All code was reviewed and validated by the authors. All scientific content, analyses, and interpretations were produced by the authors.

Financial support. Financial support for this work was provided by the Government of New Zealand via a grant from the Ministry of Foreign Affairs and Trade to the Pacific Community (WPG-0103601, DOC-4119683, ACT-0103048) to implement the Climate Science for Ensuring Pacific Tuna Access project. SA's contribution was funded by the European Union under grant agreement no. 101136748 (BioEcoOcean).

https://doi.org/10.5194/egusphere-2026-711
Preprint. Discussion started: 1 April 2026

References

Angelova, M., Tzonkov, S., and Pencheva, T.: Genetic Algorithms Based Parameter Identification of Yeast Fed-Batch Cultivation, p. 224–231, Springer Berlin Heidelberg, ISBN 9783642184666, https://doi.org/10.1007/978-3-642-18466-6_26, 2011.

Banse, K.: On the vertical distribution of Zooplankton in the sea, Prog. Oceanogr., 2, 53–125, https://doi.org/10.1016/0079-6611(64)90003-5, 1964.

Behrenfeld, M. J. and Falkowski, P. G.: Photosynthetic rates derived from satellite-based chlorophyll concentration, Limnol. Oceanogr., 42, 1–20, https://doi.org/10.4319/lo.1997.42.1.0001, 1997.

Brodeur, R. D., Seki, M. P., Pakhomov, E. A., and Suntsov, A. V.: Micronekton - What are they and why are they important?, PICES Press, 13, 7–11, 2004.

Cantú-Paz, E.: Migration Policies, Selection Pressure, and Parallel Evolutionary Algorithms, Journal of Heuristics, 7, 311–334, 2001.

Conchon, A.: Modélisation de la distribution et de la dynamique du micronecton dans l'océan global, Ph.D. thesis, Université Toulouse III - Paul Sabatier, 2017.

Conover, R. J.: Comparative life histories in the genera Calanus and Neocalanus in high latitudes of the northern hemisphere, Hydrobiologia, 167–168, 127–142, https://doi.org/10.1007/bf00026299, 1988.

Daase, M., Falk-Petersen, S., Varpe, Ø., Darnis, G., Søreide, J. E., Wold, A., Leu, E., Berge, J., Philippe, B., and Fortier, L.: Timing of reproductive events in the marine copepod Calanus glacialis: a pan-Arctic perspective, Can. J. Fish. Aquat. Sci., 70, 871–884, https://doi.org/10.1139/cjfas-2012-0401, 2013.

Dask Development Team: Dask: Library for dynamic task scheduling, https://dask.org, 2016.

Deb, K., Pratap, A., Agarwal, S., and Meyarivan, T.: A fast and elitist multiobjective genetic algorithm: NSGA-II, IEEE T. Evol. Comput., 6, 182–197, https://doi.org/10.1109/4235.996017, 2002.

Delpech, A., Conchon, A., Lehodey, P., and Senina, I.: Observation of micronekton layers in the tropical Pacific Ocean, Deep Sea Research Part I, 155, 103 173, 2020.

Dmitrenko, I. A., Petrusevich, V., Darnis, G., Kirillov, S. A., Komarov, A. S., Ehn, J. K., Forest, A., Fortier, L., Rysgaard, S., and Barber, D. G.: Sea-ice and water dynamics and moonlight impact the acoustic backscatter diurnal signal over the eastern Beaufort Sea continental slope, Ocean Sci., 16, 1261–1283, https://doi.org/10.5194/os-16-1261-2020, 2020.

Drago, L., Panaïotis, T., Irisson, J.-O., Babin, M., Biard, T., Carlotti, F., Coppola, L., Guidi, L., Hauss, H., Karp-Boss, L., Lombard, F., McDonnell, A. M. P., Picheral, M., Rogge, A., Waite, A. M., Stemmann, L., and Kiko, R.: Global Distribution of Zooplankton Biomass Estimated by In Situ Imaging and Machine Learning, Front. Mar. Sci., 9, https://doi.org/10.3389/fmars.2022.894372, 2022.

Edvardsen, A., Pedersen, J. M., Slagstad, D., Semenova, T., and Timonin, A.: Distribution of overwintering Calanus in the North Norwegian Sea, Oceanologia, 45, 433–441, 2003.

Everett, J. D., Baird, M. E., Buchanan, P., Bulman, C., Davies, C., Downie, R., Griffiths, C., Heneghan, R., Kloser, R. J., Laiolo, L., Lara-Lopez, A., Lozano-Montes, H., Matear, R. J., McEnnulty, F., Robson, B., Rochester, W., Skerratt, J., Smith, J. A., Strzelecki, J., Suthers, I. M., Swadling, K. M., van Ruth, P., and Richardson, A. J.: Modeling What We Sample and Sampling What We Model: Challenges for Zooplankton Model Assessment, Front. Mar. Sci., 4, https://doi.org/10.3389/fmars.2017.00077, 2017.

Falls, M., Bernardello, R., Castrillo, M., Acosta, M., Llort, J., and Galí, M.: Use of genetic algorithms for ocean model parameter optimisation: a case study using PISCES-v2_RC for North Atlantic particulate organic carbon, Geosci. Model Dev., 15, 5713–5737, https://doi.org/10.5194/gmd-15-5713-2022, 2022.

https://doi.org/10.5194/egusphere-2026-711
Preprint. Discussion started: 1 April 2026

Fortin, F.-A., De Rainville, F.-M., Gardner, M.-A., Parizeau, M., and Gagné, C.: DEAP: Evolutionary Algorithms Made Easy, Journal of Machine Learning Research, 13, 2171–2175, 2012.

Goldberg, D. E.: Genetic Algorithms in Search, Optimization, and Machine Learning, Addison-Wesley, 1989.

Hansen, N.: The CMA Evolution Strategy: A Tutorial, arXiv preprint arXiv:1604.00772, 2016.

Hansen, N. and Ostermeier, A.: Completely Derandomized Self-Adaptation in Evolution Strategies, Evol. Comput., 9, 159–195, https://doi.org/10.1162/106365601750190398, 2001.

Harris, C. R., Millman, K. J., van der Walt, S. J., Gommers, R., Virtanen, P., Cournapeau, D., Wieser, E., Taylor, J., Berg, S., Smith, N. J., Kern, R., Picus, M., Hoyer, S., van Kerkwijk, M. H., Brett, M., Haldane, A., del Río, J. F., Wiebe, M., Peterson, P., Gérard-Marchant, P., Sheppard, K., Reddy, T., Weckesser, W., Abbasi, H., Gohlke, C., and Oliphant, T. E.: Array programming with NumPy, Nature, 585, 357–362, https://doi.org/10.1038/s41586-020-2649-2, 2020.

Herman, J. and Usher, W.: SALib: An open-source Python library for Sensitivity Analysis, J. Open Source Softw., 2, 97, https://doi.org/10.21105/joss.00097, 2017.

Holland, M. M., Artigas, L. F., Atkinson, A., Best, M., Bresnan, E., Devlin, M., Eerkes-Medrano, D., Johansen, M., Johns, D. G., Machairopoulou, M., Pitois, S., Scott, J., Schilder, J., Stern, R., Tait, K., Whyte, C., Widdicombe, C., and McQuatters-Gollop, A.: Mind the gap - The need to integrate novel plankton methods alongside ongoing long-term monitoring, Ocean & Coastal Management, 262, 107 542, https://doi.org/10.1016/j.ocecoaman.2025.107542, 2025.

Hoyer, S. and Hamman, J.: xarray: N-D labeled Arrays and Datasets in Python, J. Open Res. Softw., 5, 10, https://doi.org/10.5334/jors.148, 2017.

Iwanaga, T., Usher, W., and Herman, J.: Toward SALib 2.0: Advancing the accessibility and interpretability of global sensitivity analyses, Socio-Environ. Syst. Model., 4, 18 155, https://doi.org/10.18174/sesmo.18155, 2022.

Kosobokova, K. N.: The reproductive cycle and life history of the Arctic copepod Calanus glacialis in the White Sea, Polar Biol., 22, 254–263, https://doi.org/10.1007/s003000050418, 1999.

Kruschke, J. K.: JAGS, p. 193–219, Elsevier, ISBN 9780124058880, https://doi.org/10.1016/b978-0-12-405888-0.00008-8, 2015.

Lam, S. K., Pitrou, A., and Seibert, S.: Numba: a LLVM-based Python JIT compiler, in: Proceedings of the Second Workshop on the LLVM Compiler Infrastructure in HPC, SC15, p. 1–6, ACM, https://doi.org/10.1145/2833157.2833162, 2015.

Lehodey, J. V.: SeapoPym/seapopym: SeapoPym v0.1, Zenodo, https://doi.org/10.5281/zenodo.18348596, software, 2026.

Lehodey, P.: The pelagic ecosystem of the tropical Pacific Ocean: dynamic spatial modelling and biological consequences of ENSO, Prog. Oceanogr., 49, 439–468, https://doi.org/10.1016/s0079-6611(01)00035-0, 2001.

Lehodey, P., Senina, I., and Murtugudde, R.: A spatial ecosystem and populations dynamics model (SEAPODYM) – Modeling of tuna and tuna-like populations, Prog. Oceanogr., 78, 304–318, https://doi.org/10.1016/j.pocean.2008.06.004, 2008.

Lehodey, P., Senina, I., Sibert, J., Bopp, L., Calmettes, B., Hampton, J., and Murtugudde, R.: Preliminary forecasts of Pacific bigeye tuna population trends under the A2 IPCC scenario, Prog. Oceanogr., 86, 302–315, https://doi.org/10.1016/j.pocean.2010.04.021, 2010.

Lehodey, P., Senina, I., Nicol, S., and Hampton, J.: Modelling the impact of climate change on South Pacific albacore tuna, Deep-Sea Res. Pt. II, 113, 246–259, https://doi.org/10.1016/j.dsr2.2014.10.028, 2015.

Lellouche, J.-M., Greiner, E., Bourdallé-Badie, R., et al.: The Copernicus Global 1/12° Oceanic and Sea Ice GLORYS12 Reanalysis, Frontiers in Earth Science, 9, 698 876, https://doi.org/10.3389/feart.2021.698876, 2021.

Mackas, D. and Coyle, K.: Shelf–offshore exchange processes, and their effects on mesozooplankton biomass and community composition patterns in the northeast Pacific, Deep-Sea Res. Pt. II, 52, 707–725, https://doi.org/10.1016/j.dsr2.2004.12.020, 2005.

https://doi.org/10.5194/egusphere-2026-711
Preprint. Discussion started: 1 April 2026

McKendrick, A. G.: Applications of Mathematics to Medical Problems, Proceedings of the Edinburgh Mathematical Society, 44, 98–130, 1926.

Moriarty, R. and O'Brien, T. D.: Distribution of mesozooplankton biomass in the global ocean, Earth Syst. Sci. Data, 5, 45–55, https://doi.org/10.5194/essd-5-45-2013, 2013.

Müller-Plath, G. and Lüdecke, H. J.: Normalized coefficients of prediction accuracy for comparative forecast verification and modeling, Res. Stat., 2, https://doi.org/10.1080/27684520.2024.2317172, 2024.

Prieur, C. and Tarantola, S.: Variance-based sensitivity analysis: Theory and estimation algorithms, Handbook of Uncertainty Quantification, pp. 1217–1239, 2019.

Ratnarajah, L., Abu-Alhaija, R., Atkinson, A., Batten, S., Bax, N. J., Bernard, K. S., Canonico, G., Cornils, A., Everett, J. D., Grigoratou, M., Ishak, N. H. A., Johns, D., Lombard, F., Muxagata, E., Ostle, C., Pitois, S., Richardson, A. J., Schmidt, K., Stemmann, L., Swadling, K. M., Yang, G., and Yebra, L.: Monitoring and modelling marine zooplankton in a changing climate, Nat. Commun., 14, https://doi.org/10.1038/s41467-023-36241-5, 2023.

Rohr, T., Richardson, A. J., Lenton, A., Chamberlain, M. A., and Shadwick, E. H.: Zooplankton grazing is the largest source of uncertainty for marine carbon cycling in CMIP6 models, Communications Earth & Environment, 4, https://doi.org/10.1038/s43247-023-00871-w, 2023.

Saltelli, A.: Making best use of model evaluations to compute sensitivity indices, Comput. Phys. Commun., 145, 280–297, https://doi.org/10.1016/s0010-4655(02)00280-1, 2002.

Schartau, M., Wallhead, P., Hemmings, J., Löptien, U., Kriest, I., Krishna, S., Ward, B. A., Slawig, T., and Oschlies, A.: Reviews and syntheses: parameter identification in marine planktonic ecosystem modelling, Biogeosciences, 14, 1647–1701, https://doi.org/10.5194/bg-14-1647-2017, 2017.

Sieburth, J. M., Smetacek, V., and Lenz, J.: Pelagic ecosystem structure: Heterotrophic compartments of the plankton and their relationship to plankton size fractions 1, Limnol. Oceanogr., 23, 1256–1263, https://doi.org/10.4319/lo.1978.23.6.1256, 1978.

Siegel, D. A., DeVries, T., Cetinić, I., and Bisson, K. M.: Quantifying the Ocean's Biological Pump and Its Carbon Cycle Impacts on Global Scales, Annu. Rev. Mar. Sci., 15, 329–356, https://doi.org/10.1146/annurev-marine-040722-115226, 2023.

Sobol', I.: Global sensitivity indices for nonlinear mathematical models and their Monte Carlo estimates, Math. Comput. Simul., 55, 271–280, https://doi.org/10.1016/s0378-4754(00)00270-6, 2001.

St. John, M. A., Borja, A., Chust, G., Heath, M., Grigorov, I., Mariani, P., Martin, A. P., and Santos, R. S.: A Dark Hole in Our Understanding of Marine Ecosystems and Their Services: Perspectives from the Mesopelagic Community, Front. Mar. Sci., 3, https://doi.org/10.3389/fmars.2016.00031, 2016.

Stanton, T.: Review and recommendations for the modelling of acoustic scattering by fluid-like elongated zooplankton: euphausiids and copepods, ICES J. Mar. Sci., 57, 793–807, https://doi.org/10.1006/jmsc.1999.0517, 2000.

Steinberg, D. K., Carlson, C. A., Bates, N. R., Goldthwait, S. A., Madin, L. P., and Michaels, A. F.: Zooplankton vertical migration and the active transport of dissolved organic and inorganic carbon in the Sargasso Sea, Deep-Sea Res. Pt. I, 47, 137–158, https://doi.org/10.1016/s0967-0637(99)00052-7, 2000.

Steinberg, D. K., Goldthwait, S. A., and Hansell, D. A.: Zooplankton vertical migration and the active transport of dissolved organic and inorganic nitrogen in the Sargasso Sea, Deep-Sea Res. Pt. I, 49, 1445–1461, https://doi.org/10.1016/s0967-0637(02)00037-7, 2002.

Stukel, M. R., Décima, M., and Landry, M. R.: Quantifying biological carbon pump pathways with a data-constrained mechanistic model ensemble approach, Biogeosciences, 19, 3595–3624, https://doi.org/10.5194/bg-19-3595-2022, 2022.

https://doi.org/10.5194/egusphere-2026-711
Preprint. Discussion started: 1 April 2026

Tande, K. S. and Båmstedt, U.: Grazing rates of the copepods Calanus glacialis and C. finmarchicus in arctic waters of the Barents Sea, Mar. Biol., 87, 251–258, https://doi.org/10.1007/bf00397802, 1985.

Taylor, K. E.: Summarizing multiple aspects of model performance in a single diagram, J. Geophys. Res. Atmos., 106, 7183–7192, https://doi.org/10.1029/2000jd900719, 2001.

Von Foerster, H.: Some Remarks on Changing Populations, in: The Kinetics of Cellular Proliferation, edited by Stohlman, J. F., pp. 382–407, Grune and Stratton, New York, 1959.

https://doi.org/10.5194/egusphere-2026-711
Preprint. Discussion started: 1 April 2026

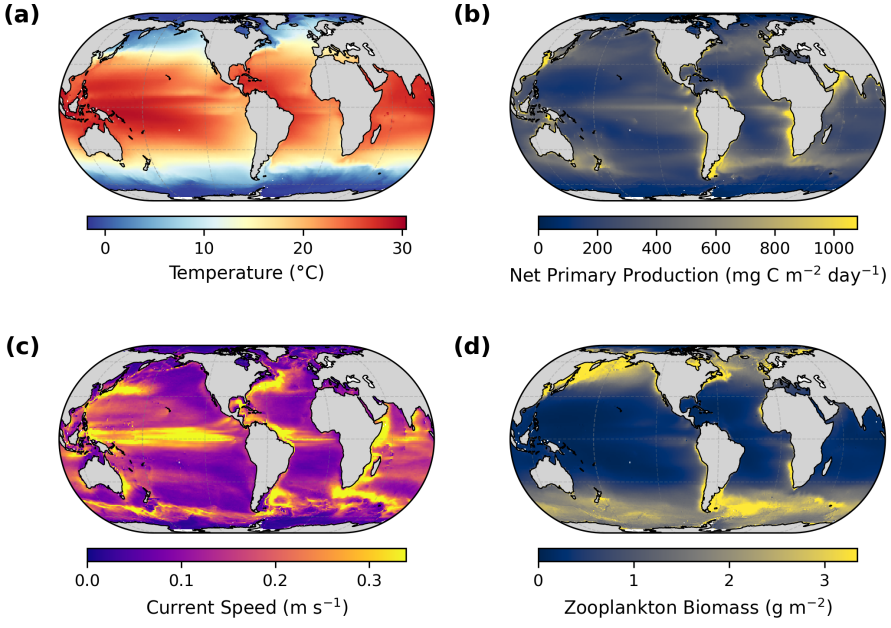

Figure 1. Average map distributions (2000-2020) of a) temperature in the epipelagic layer and b) vertically integrated NPP used as forcing for the SeapoPym simulations. c) Mean current norm and d) zooplankton biomass modeled by LMTL are shown for further comparison and analysis.

https://doi.org/10.5194/egusphere-2026-711
Preprint. Discussion started: 1 April 2026

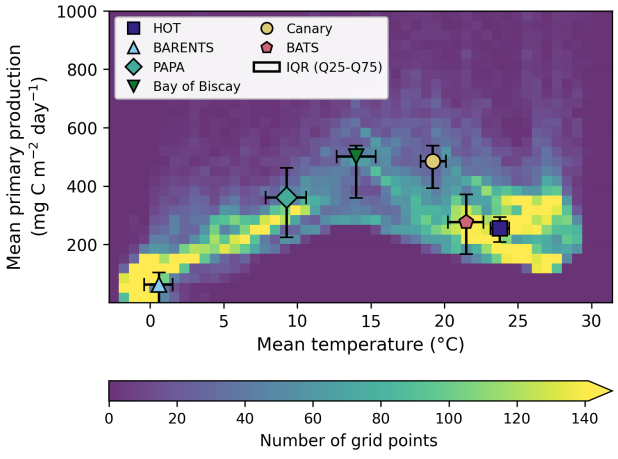

Figure 2. Global distribution of mean temperature versus mean net primary production from 1998 to 2020. Light blue dots (39,853) represent the global forcing dataset. The six selected oceanographic stations are highlighted with distinct symbols that indicate their mean positions within this environmental space. Error bars indicate the 25th and 75th quantiles over the time series.

https://doi.org/10.5194/egusphere-2026-711
Preprint. Discussion started: 1 April 2026

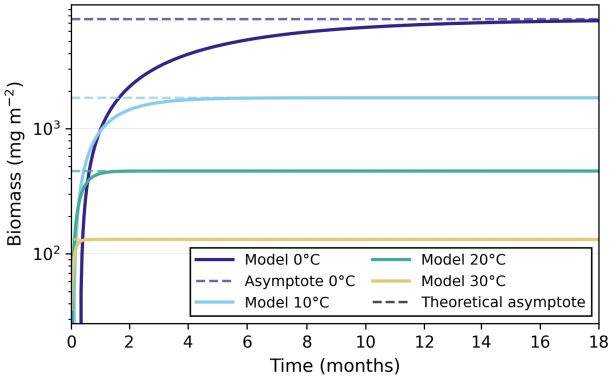

Figure 3. Results from the theoretical benchmark with constant mortality and recruitment rates at four target temperatures for the SeapoPym model. The theoretical asymptote line is shown as a dashed line.

https://doi.org/10.5194/egusphere-2026-711
Preprint. Discussion started: 1 April 2026

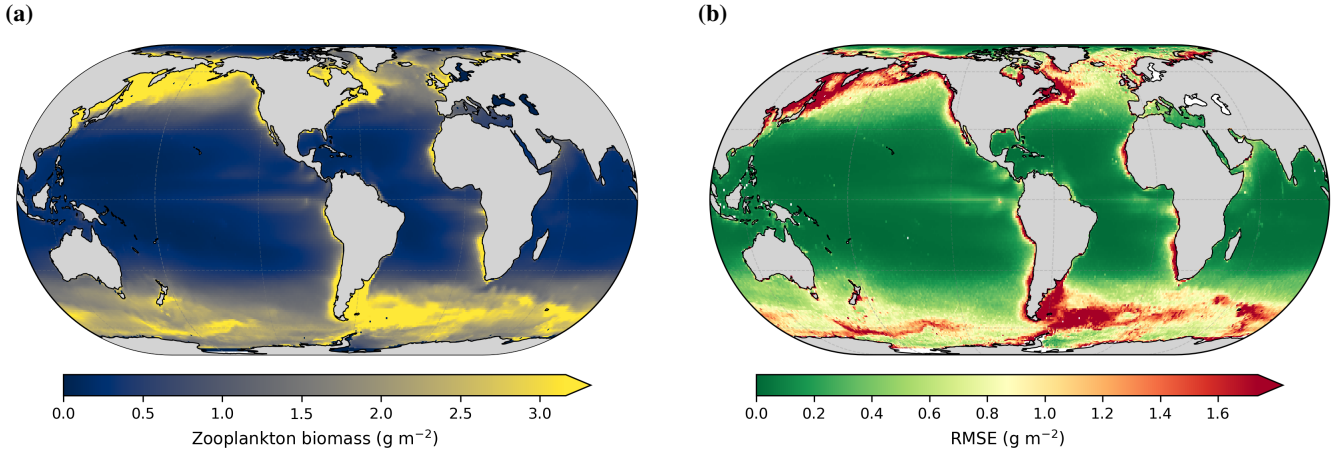

Figure 4. (a) Mean biomass distribution of zooplankton predicted by SeapoPym; (b) Root Mean Square Error (RMSE) between SeapoPym and LMTL.

https://doi.org/10.5194/egusphere-2026-711
Preprint. Discussion started: 1 April 2026

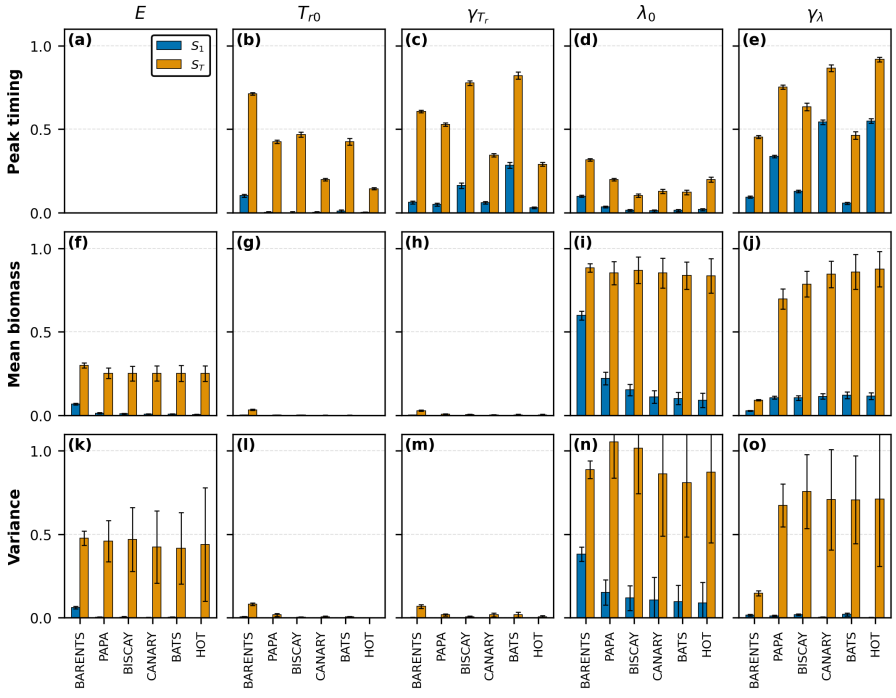

Figure 5. Sobol analysis for three metrics (rows) and five parameters (columns) at six stations (x-axis), totaling 1,190,700 simulations (N=99,225 samples). The first-order index S_1 is shown with blue bars and the total order index S_T with orange bars. The error bars represent bootstrap 95 % confidence intervals. Note that stations are sorted along the x-axis from coldest to warmest conditions.

https://doi.org/10.5194/egusphere-2026-711
Preprint. Discussion started: 1 April 2026

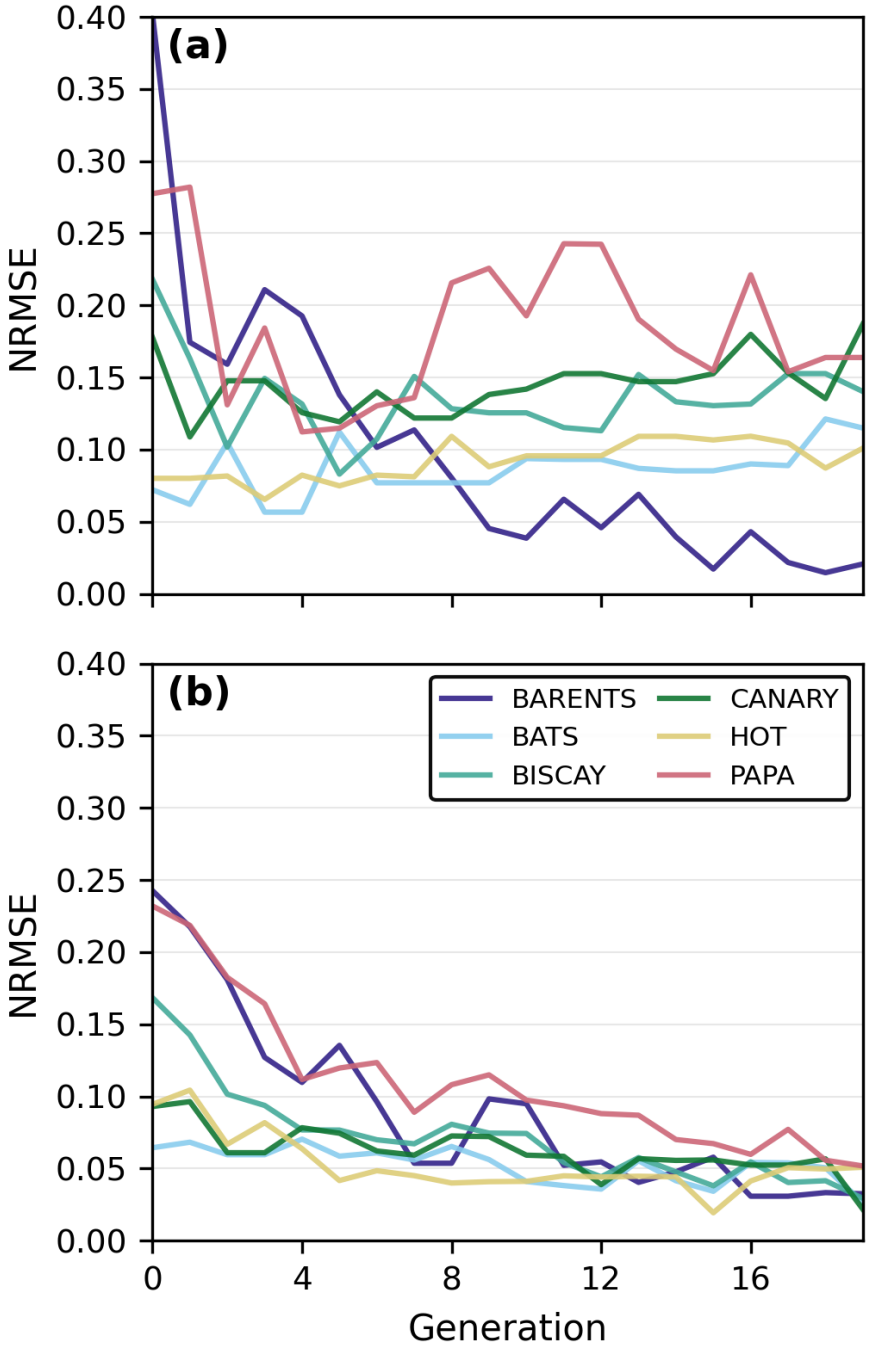

Figure 6. Evolution of the NRMSE for the predicted biomass at each station for optimization experiments considering each station separately (top) or together (bottom).

https://doi.org/10.5194/egusphere-2026-711
Preprint. Discussion started: 1 April 2026

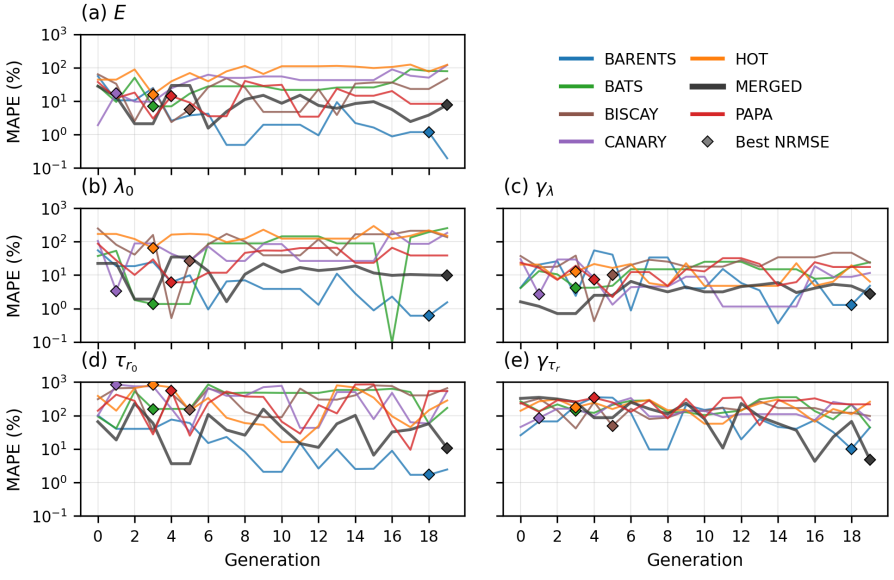

Figure 7. Evolution through the optimization process of the Mean Absolute Percentage Error (MAPE) for each parameter for twin experiments, with an independent optimization for each simulated observation time series (BARENTS, HOT, BATS, BISCAY, CANARY, PAPA) or a joint optimization (MERGED).

https://doi.org/10.5194/egusphere-2026-711
Preprint. Discussion started: 1 April 2026

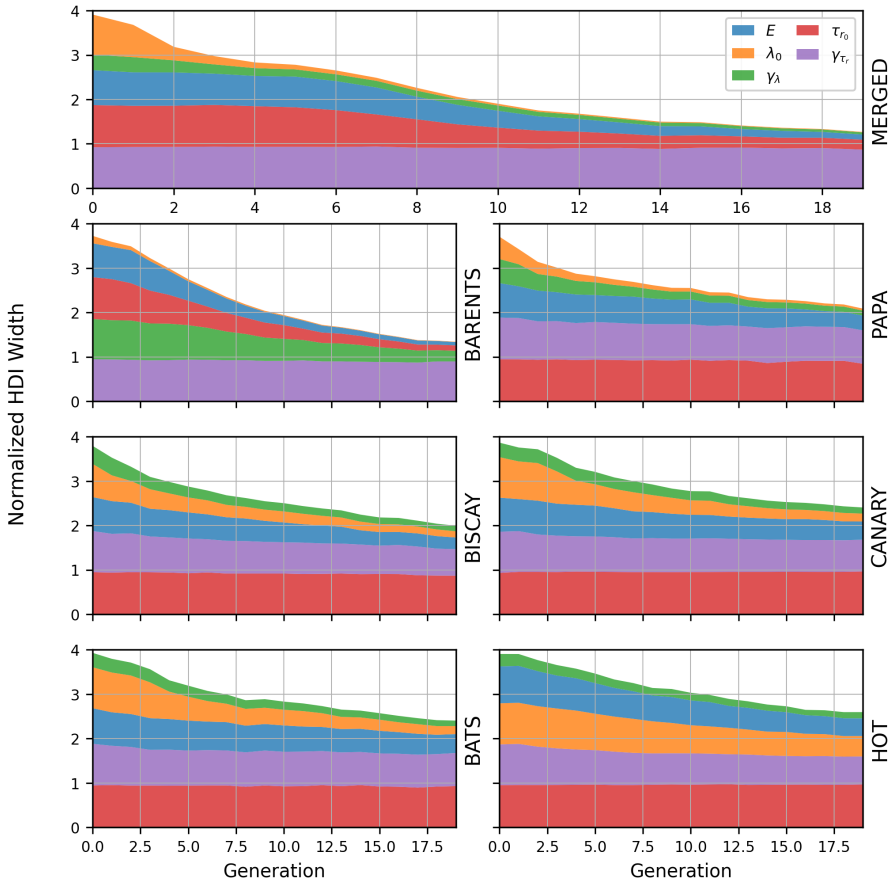

Figure 8. Highest density interval (HDI) evolution of the 1,000 best individuals per generation for each parameter estimate in twin simulation experiments using all (MERGED) or separate stations.