

This document contains a point-by-point response to the reviewers, followed by a tracked changes version of the manuscript.

Anonymous Referee #1

I have read the second version of the manuscript, and I consider that the authors have responded appropriately to my original review. Well done. I only have the following minor comments, which are mostly technical corrections.

We thank the reviewer for this second reading and suggestions. We integrated all of them in the manuscript.

Abstract

1) To only focus on the main results of the study, I would suggest to remove the last sentence: “The results are in terms of...”.

Done.

1. Introduction

1) 1. 50: change “[1, 6]” by “{1, 6}”.

Done

2) 1. 62-64: I would suggest to rephrase these two sentences.

Done. Rephrased as 1 64: “For this purpose, we use Observing System Simulation Experiments (OSSEs). This method allows for simulating synthetic observations in places where real observations do not exist, and to examine how useful are the information they would provide”.

2. Method

1) 2.1 SEAPODYM-MTL and its configuration => The SEAPODYM-MTL model

Done

2) 1. 81 “models” => “simulates”

Done

3) 1.109-110: I would also suggest to remove that sentence.

Done.

4) 1.139: “This latter...” => The latter could lead to under (or over) estimate some identified regimes.

Done

5) 1.141: This sentence needs to be rephrased.

Done. We reformulated as 1 140: “This is why performing a clustering to determine the different regimes associated with the forcing fields seems a more rigorous approach here.”

6) 2.3 OSSE system configuration => your subsections (2.3.1 to 2.3.3) are very short. I would suggest here to merge them together without separating by subsections.

Done. But note that separating in subsections was intended to help clarifying the OSSE protocol we are following, in response to your previous review.

7) 1.149: “The implementation of...” => To perform realistic OSSEs, a rigorous protocol needs to be followed.

Done

8) 1.178: might change to “In the framework of OSSEs, we perform experiments using different sets of synthetic observation points ($N_e = 400$)”.

Done. We also added that N_e corresponds to the size of each set.

9) 1.183-186: a bit confusing, might need to be rephrased and further detailed (e.g. what do you mean by close enough?).

We changed “close enough” by “similar”, which is more suitable in this case. And we simplified the sentence by 1 186: “For a given group of experiments, we check that the configurations are comparable to each other by ensuring that the distribution of all secondary variables are similar (cf. marginal distribution plots in Section 3.2.1).” We do not give further details here because the idea is already widely developed in section 3.2.1. Instead, we give the reference to section 3.2.1.

3. Results

1) If you want to use “Exp.” as an abbreviation for “Experiment”, you need to explicitly say it.

Done. We added: “(usually abbreviated as Exp. in the following)”.

2) 1. 238: “are used for” => “evaluate the impacts of the bloom index...”.

Done

3) 1. 269-272: a bit confusing, might need to rephrase that paragraph and better explain what you mean by “Experiments with such cross-correlation” here.

Following reviewer 2 suggestion, we rephrased this paragraph as 1 272: “Although the distributions of the secondary variables are not always shown in the following experiments, they have been examined to ensure that the OSSE results are not biased by systematic differences in the secondary variables. Experiments with significant cross-correlation between indicator variables are not presented...”

We also introduce better the concept of cross-correlation by specifying in the previous paragraph, 1 268: “The lower velocities are coupled with lower temperatures and the higher velocities with higher temperatures. There is a cross-correlation between the velocity (primary variable) and the temperature (secondary variable).”

4) 1. 290: “The same kind of” => “Similar”.

Done.

5) 1. 311: “associated to” => “associated with”.

Done.

6) 1.331-332: might change to “...using the transects from the PIRATA cruises, and those from the British Antarctic Survey (BAS) cruises during the 2013-2015 time period (see Figure 7)”.

Done.

7) 1. 337-338: “Also, whatever the perturbation...” => need to be rephrased

We agree, we rephrased as: “regardless of the intensity of the perturbation...”.

8) 4.1 “... in term of observability” => “... in terms of observability”

Done.

9) 1. 396-398: too long, need to be split into two sentences.

Done.

4.3 Limitations and perspectives

1) 1. 410: “The realism of this approach is questionable...” sounds too dramatic, you downplay too much your approach here; I would rather say something like “Our idealized approach...”

Changed as suggested.

5. Conclusions

1) Using both “modelling” and “modeling”; you need to either stick with the UK or the US spelling.

Done, we changed every occurrence for “modelling”.

2) l. 439: “observation” => “observations”.

Done.

3) l. 441: this is not exactly the main objective of traditional OSSEs, which is to assess the impact of assimilating synthetic data; saying “...designed to correct outputs of operational models” might be misleading here. I suggest to change that sentence to something like: “...designed to estimate the impact of an observing system from the difference in the errors made by each experiment (e.g. Fujii et al., 2019)”.

Done.

4) l. 442: “to search for” => “to determine”.

Done.

5) l. 457: “in term of” => “in terms of”.

Done.

Anonymous Referee #2

We thank again the reviewer for his careful reading and useful rephrasing suggestions.

I am glad to see that the manuscript has become much easier to follow. There are only a few places where I got stuck and needed to re-read sentences multiple times to try to understand their meaning. One example is Section 4.1 which contains a lot of information (which is a good thing). Unfortunately, sentences in Section 4.1 are often still formulated in a complex or clumsy way -- and the combination of high information density and complex sentences makes it difficult to follow the text, especially for readers who do not know the model very well. I have made some suggestions below but would encourage the authors to go carefully through Section 4.1 again and revise complex sentences.

We integrated all of the reviewer suggestions (see specific comments below) to improve the clarity of section 4.1. We also simplified and cut some long sentences.

Quite a lot of emphasis is placed on DVM in the introduction and conclusions of the manuscript, for example: "Among them, the energy transfer efficiency coefficients are of great importance because they directly control the biomass of micronekton functional groups, including those that undergo DVM and contribute to the sequestration of carbon dioxide into the deep ocean ...". I agree with that statement and therefore I don't quite understand why all E_i are usually lumped together in the results section and rarely investigated separately. Are the energy transfer coefficients for micronekton undergoing DVM as well constrained by the data as others? Are there systematic differences?

In the results section, we generally investigate both all the E_i together and the different groups separately. For example:

l 254: "The estimate of the parameters for the non-migrant lower mesopelagic (lmeso) group is not sensitive to the regime of currents (Figure 3a). Conversely, the estimation is the most sensitive for the epipelagic group, whose dynamics are entirely driven by the surface currents."

l 291: "A strong stratification seems to deteriorate the estimate for all migrant groups."

l 309: "In addition, this experiment estimates the energy transfer coefficients for migrant micronekton groups with less than 1% error (Figure 3d)."

The different groups are thus not always constraint in the same way. There are differences between the different variable but no systematic differences are observed. This was however not clearly stated in the conclusion. We thus added:

l 456: "In terms of estimation performance, some functional groups are more affected by the regime variable than other. This is the case for the migrant groups that are very sensitive to the stratification or the bloom index regime for instance. However, no systematic differences between the different groups are noted".

A minor but general comment: the manuscript currently contains a mix of passive and active voice. This is not a big deal but it seems a bit inconsistent. Example: l 55ff "A method to estimate the model parameters has been developed ... While we can expect that improved estimates ...".

We made some modifications and tried to keep a same voice inside a same paragraph, as far as possible.

specific comments

l 1: I would suggest to change this from singular to plural: "Micronekton [...] are"

Done.

l 9: "Sampling regions show a variety of performances.": This is not a great sentence, I would suggest to change it to something like: "In our experiments, we obtained different results for spatially distinct sampling regions based on their prevailing ocean conditions."

Changed as suggested.

l 28: I would suggest to add units to the "1" as well.

Done.

l 36: "Progresses are" -> "Improvements in biomass estimation"

Done.

l 40: "techniques of observational estimates" -> "techniques for collecting observational estimates"

Done.

l 50: There is a leftover of the "[["-notation.

Corrected.

l 52: It would be good to describe the relationship between the 4 parameters and the 2 times of development (e.g. "(2 to parameterize recruitment, 2 for mortality)"). And I am not a biologist but I presume "stages of development" is a more suitable term than "times of development".

Done. This is indeed 2 parameters for the recruitment and 2 for the mortality. The biological terminology is really "times of development" because these are real timescales: the time it

takes for organisms to become mature (recruitment parameterization) and the time corresponding to the life expectancy of organisms (mortality parameterization).

l 59: "run" -> "model run" or "model simulation"

Done.

l 59: How can the same model run be used to generate observations and also estimate parameters? This is not a good description of twin experiments.

The point we wanted to make here is that a run with the same forcings is used to generate the simulation and to estimate the parameters (no error was introduced in the forcing fields in this study). As this point was not crucial, we removed it. "Twin experiment" is self-explanatory and less confusing.

l 65: "regarding oceanic variables" -> "characterized by different oceanic regimes"

Done. But we keep the word "variable" instead of "regime", as we introduce the concept of regimes in the method section only.

l 66: "observation gives" -> "observations provide"

Done.

l 68: "in order to simulate more realistic conditions": Add "for parameter estimation", otherwise it sounds like the error is making the ocean conditions more realistic.

Done.

l 70: "oceanic variables" -> "four oceanic variables of interest"

Done.

l 74: "set-ups" -> "set-up"

Done.

l 112: "The characterization of each observation point relies on four indicators defined from the environmental variables:": I'd suggest: "Each observation point is characterized by four indicators which are based on the following environmental variables:"

Changed as suggested.

l 123: "according to temporal variation relative to annual median threshold overshooting" -> "based on temporal variations of primary production exceeding a threshold based on its annual median"

Done.

l 128: "and regimes boundary values" -> "and regime boundary values"

Done.

l 131: remove the "a"

Done.

l 132: "values" sounds like k-means works only with 0-dimensional scalars, I would suggest to use "points" instead.

Done.

l 132: "separates N values in a given number of cluster by minimizing the distance of each value to the mean ..." -> "divides N points into a given number of clusters by minimizing the distance of each point to the mean ..."

Done.

l 133: Use n in this sentence so it is unambiguous: the number of clusters is chosen in advance. Then in the next sentence change "leads to" to "produces".

Done.

l 137: What is meant by "class" here? Please rephrase.

"Class" was used in the sense of "cluster" here. We rephrased using "cluster" instead.

l 153: "The reference simulation" -> "The nature run"

Done.

l 153: "The goal is to retrieve back" I would suggest to change it to "The goal is then to retrieve"

Done.

l 165: The treatment of temperature is a bit tricky here as (1) the Celsius scale has a somewhat arbitrary zero-point and using Kelvin would probably lead to vastly different results and (2) how are negative temperature values dealt with? A little more information is needed and please mention the temperature units here.

Using a Kelvin scale would have been possible indeed, however for the sake of clarity we prefer to keep a Celsius scale (this is more classic in the literature and might allow an easier comparison with other studies). Choosing a Celsius scale instead of a Kelvin scale does not lead to different results because the way the temperature is used in the model (through the parameterization of the times of development) is unit-independent.

The existence of negative values is not related to the fact that we introduce an error. Negative temperatures already exist in the original forcing fields. The relationship between the temperature and the different times of development is exponential. As the exponential is also defined for negative values, no specific treatment is required for negative temperature values. In addition, we want to underline that negative temperature values remain close to 0 ($> -1^{\circ}\text{C}$), so there is no significant differences in terms of micronekton behaviour between the positive and negative temperatures close to zero (the zero is not a rupture point).

Following the reviewer suggestion, we added the units of the different forcing fields (including temperature) here. We also give more precision on the way the temperature is parameterized in the appendix (eq. A5 and A6 in the revised version) and state explicitly that the parameterization is also valid for negative temperatures.

l 173: I don't know if readers will immediately know what an "assimilation module" is. Maybe describe it e.g. using "used here to estimate model parameters from observations".

Done.

l 179: "in" -> "from" and "constructed as explained" -> "introduced" or "described"

Done.

l 181: "and is" -> "and it is"

Done.

l 189: "We assume 400 observations" -> "We use 400 observations in our experiments" or "We assume 400 observations were collected"

Done.

l 207: "are discussed" -> "are also discussed below"

Done.

l 230: Minor point: here section 2.2 is referred to as "subsection", previously "section" was used.

Yes, for clarity we refer all sections and subsections as "section".

l 233: I suggest to remove the "intersection", it is confusing. But do each of the 21 configurations contain less than 0.5% of the total or do they contain less than 0.5% together?

Done. We removed "intersection". Each of the 21 configurations contains less than 0.5% of the total. We tried to clarify the sentence, l 233: "21 of them are near-empty as they contain less than 0.5% of all observable points".

l 235: "The first three groups of Experiments 1a-b, 1c-d and 1e-f are meant" -> "The purpose of the first three groups of Experiments 1a-b, 1c-d and 1e-f is"

Done.

l 236: There is a sudden switch to future tense ("will").

We changed everything for present tense.

l 249: "currents" -> "current"

Done.

l 256: "is" -> "are"

Done.

l 263: Not contradictory to the experiments themselves but the conclusions that were just drawn from these experiments!

We corrected this sentence as suggested. l 264: “This result seems contradictory to the conclusions drawn from Exps. 1a-f.”

l 264: "Despite this, it has been fixed "polar regime", the temperature in configuration C' is on average lower (−0.7C) than the temperature of configuration C" (2.1C) (Figure 4)." -> "While both configurations are considered to be in the "polar regime", the temperature in configuration C' (−0.7C) is on on average lower...". Note that I moved the "(−0.7C)", otherwise the reader may think that configuration C' is on average −0.7C lower.

Done.

l 267: "conclude on" -> "assess"

Done.

l 269: "In the following, although distribution along secondary variables are not always shown, they have always been used in the analysis to check that the OSSE results are not biased by this type of difference between the distributions of randomly selected datasets."

I would suggest: "Although the distributions of the secondary variables are not always shown in the following experiments, they have been examined to ensure that the OSSE results are not biased by systematic differences in the secondary variables."

Changed as suggested.

l 271: "such" -> "significant"

Done.

l 292: "always" -> "again"

Done.

l 305: "conclusions" -> "the conclusions"

Done.

l 307: But the bloom index appears to have little effect in Exp 4a and b, so what is the message here? It would be good for the reader to more clearly state the results.

The point is that the bloom index appears to have an effect at high temperatures only. We agree that this was not clearly stated. We added l 310: “According to our results, the primary production and the regimes of the bloom index do not always play a role in the performance of the parameters estimation. A positive bloom index appears to improve the performance of the estimation at high temperatures only”. This point is further discussed in section 4.1 of the discussion.

l 350: "discussing" I'd suggest to use "examining" to avoid a double discussion.

Done.

l 352: "term" -> "terms"

Done.

Eq. 10 + 11: The notation is confusing. Is P the sum of all P_i ? Mention what P_i is. Why use different notations: $P_i(\tau=0)$ but $P_{i\{\tau=0\}}$?

P refers indeed to the sum of all P_i . We clarified the definition of P_i : “the relative amount (P_i) of P at age $\tau=0$ in each functional group i”.

l 362: There is something missing here: It should be the ratio of x and y. The second part, y, is missing in this sentence.

The ratio x/y was actually contained in the ρ notation, as defined in appendix A. However, the definition and reference to the appendix was missing here. And for the sake of clarity, we now give the complete definition of the ratio (l 365).

l 363: Use "night" and "day" in Eq 11 or switch to "n" and "d" here.

Done.

l 364: I think some of this information would be useful in the introduction to give readers who are not that familiar with this type of model the basic idea of how the model works early on.

Yes, we now better introduce what controls the micronekton biomass of the model in the introduction. This information was indeed missing and very important:

l 45: "The source and sink for the micronekton biomass are the recruitment from the potential production of micronekton at a given age and the natural mortality. ... (Equations A5 and A6)".

l 366: "Then" -> "Second" (unless that is not yet the second main mechanism)

It is indeed the second main mechanism. We changed as suggested.

l 369: "Therefore..." This sentence is difficult to understand: I don't think "will be as close as" is right here, the authors probably mean to say "will contain more information about"

Yes indeed, we changed as suggested.

l 376: "is to induce" -> "is that it induces"

Done.

l 382: "biomass. This is consistent" -> "biomass, and is consistent"

Done.

l 388: "characteristics time" -> "characteristic timescale"

Done.

l 416: "Sensitivity study" -> "Conducting a sensitivity analysis"

Done.

l 417: "on" -> "of"

Done.

Fig. 2: It would be nice to include the variable in the label for each panel, e.g. "a) temperature"

Done.

Fig. 3: It is nice to see all panels together in one Figure! I would recommend changing the maximum y-limit in panel a to 18 to make the scales comparable.

Done.

Fig. 6: "associated to" -> "associated with"

Done.

Tracked Changes Version

Influence of oceanic conditions in the energy transfer efficiency estimation of a micronekton model

Audrey Delpech^{1,2}, Anna Conchon^{2,3}, Olivier Titaud², and Patrick Lehodey²

¹Laboratoire d'Etudes Géophysiques et d'Océanographie Spatiale, LEGOS - UMR 5566 CNRS/CNES/IRD/UPS, Toulouse, France

²Collecte Localisation Satellite, CLS, Toulouse, France

³Mercator Ocean, Toulouse, France

Correspondence: Audrey Delpech (audreydelpech@wanadoo.fr)

Abstract. Micronekton – small marine pelagic organisms around 1-10 cm in size – ~~is~~are a key component of the ocean ecosystem, as it constitutes the main source of forage for all larger predators. Moreover, the mesopelagic component of micronekton that undergoes Diel Vertical Migration (DVM) likely plays a key role in the transfer and storage of CO₂ in the deep ocean: this is known as the ‘biological pump’. SEAPODYM-MTL is a spatially explicit dynamical model of micronekton. It simulates six functional groups of vertically migrant (DVM) and non-migrant (no DVM) micronekton, in the epipelagic and mesopelagic layers. Coefficients of energy transfer efficiency between primary production and each group are unknown, but they are essential as they control the production of micronekton biomass. Since these coefficients are not directly measurable, a data assimilation method is used to estimate them. In this study, Observing System Simulation Experiments (OSSEs) are used at a global scale to explore the response of oceanic regions regarding energy transfer coefficients estimation. ~~Sampling regions show a variety of performances. It appears that environmental conditions are crucial to determine the optimal observing regions~~In our experiments, we obtained different results for spatially distinct sampling regions based on their prevailing ocean conditions. According to our study, ideal sampling areas are warm and productive waters associated with weak surface currents like the eastern side of tropical Oceans. These regions are found to reduce the error of estimated coefficients by 20% compared to cold and more dynamic sampling regions. ~~The results are discussed in term of interactions between physical and biological processes.~~

1 Introduction

Micronekton organisms are at the mid-trophic level of the ocean ecosystem and have thus a central role, as prey of larger predator species such as tunas, swordfish, turtles, sea birds or marine mammals, and as a potential new resource in the blue economy (St John et al., 2016). Diel Vertical Migration (DVM) characterizes a large biomass of the mesopelagic (inhabiting the twilight zone 200-1000 m) component of micronekton of the world ocean. This migration of biomass occurs when organisms move up from a deep habitat during daytime to a shallower habitat at night. DVM is generally related to a trade-off between the need for food and predator avoidance (Benoit-Bird et al., 2009) and seem to be triggered by sunlight (Zaret and Suffern, 1976). Through these daily migrations, the mesopelagic micronekton potentially contributes to a substantial transfer of atmospheric

CO₂ to the deep ocean, after its metabolization by photosynthesis and export through the food chain (Davison et al., 2013). The understanding and quantification of this mechanism, called the ‘biological pump’, are crucial in the context of climate change (Zaret and Suffern, 1976; Volk and Hoffert, 1985; Benoit-Bird et al., 2009; Davison et al., 2013; Giering et al., 2014; Ariza et al., 2015). However, there is a lack of comprehensive datasets at global scale to properly estimate micronekton biomass and composition. The few existing estimates of global biomass of mesopelagic micronekton vary considerably between less than 1 Gt and ~ 20 Gt (Gjosaeter and Kawaguchi, 1980; Irigoien, 2014; Proud et al., 2018), so that micronekton has been compared to a "dark hole" in the studies of marine ecosystems (St John et al., 2016). Therefore, a priority is to collect observations and develop methods and models needed to simulate and quantify the dynamics and functional roles of these species’ communities.

Observations and biomass estimations of micronekton rely traditionally on net sampling and active acoustic sampling (e.g., Handegard et al., 2009; Davison, 2011). Each method has limitations. Micronekton species can detect approaching fishing trawls and part of them can move away to avoid the net. This phenomenon leads to biomass underestimation from net trawling (Kaarvedt et al., 2012). Conversely, acoustic signal intensity may overestimate biomass due to presence of organisms with strong acoustic target strength, e.g. species that have gas inclusion inducing strong resonance (Davison, 2011; Proud et al., 2017). Progresses-Improvements in biomass estimation are expected in the coming years thanks to the combined use of different measurement techniques: multiple acoustic frequencies, traditional net sampling and optical techniques (Kloser et al., 2016; Davison et al., 2015). The accuracy of biomass estimates is predicted to benefit from this combination of techniques and from the developments of algorithms that can attribute acoustic signal to biological groups.

While these techniques of-for collecting observational estimates of biomass are progressing, new developments are also achieved in the modeling-modelling of the micronekton components of the ocean ecosystem. SEAPODYM (Spatial Ecosystem And POPulation Dynamics Model) is an eulerian ecosystem model that includes one lower- (zooplankton) and six mid-trophic (micronekton) functional groups, and detailed fish populations (Lehodey et al., 1998, 2008, 2010). Given the structural importance of DVM, the micronekton functional groups are defined based on the daily migration behavior of organisms between three broad epi- and meso-pelagic bio-acoustic layers (Lehodey et al., 2010, 2015). In addition to DVM, the horizontal dynamics of biomass in each group is driven by ocean dynamics, while a diffusion coefficient accounts for local random movements. The recruitment source and sink for the micronekton biomass are the recruitment from the potential production of micronekton at a given age and the natural mortality, respectively. The recruitment time and the natural mortality of organisms are linked to the temperature in the vertical layers inhabited by each functional group during day or night. These mechanisms are simulated with a system of advection-diffusion-reaction equations (Lehodey et al., 2008). The equations governing the model are detailed in Appendix A. Primary production is the source of energy distributed to each group according to a coefficient of energy transfer efficiency. Eleven parameters control the biological processes: a diffusion coefficient, six coefficients $(E_i^1)_{i \in \{1, \dots, 6\}}$ of energy transfer from primary production toward each mid-trophic functional group, and four parameters for the relationship between water temperature and times of development (mortality, recruitment two parameters for the life expectancy, and two parameters for the recruitment time) (Lehodey et al., 2010). The latter four parameters were estimated from a compilation of data found in the scientific literature (Lehodey et al., 2010). Therefore, the largest uncertainty remains on the energy transfer efficiency coefficients, that control the total abundance of each functional group.

A method to estimate the model parameters has been developed using a Maximum Likelihood Estimation (MLE) approach (Senina et al., 2008). A first study has shown that this method can be used to estimate the parameters E'_i using relative ratios of observed acoustic signal and predicted biomass in the three vertical layers during daytime and nighttime (Lehodey et al., 2015). However, this study was conducted for a single transect in the very idealized framework of twin experiments (~~the same run is used for observation generation and parameter estimation~~). ~~While we can expect that~~. While improved estimates of micronekton biomass ~~will be expected to~~ become available in the coming years, this will likely still require costly operations at sea. Therefore, it is important to assess realistically and more systematically how well observations can estimate parameters before deploying ~~observational systems~~. any observational system.

For this purpose, we use Observing System Simulation Experiments (OSSE OSSEs, Arnold and Dey (1986)) ~~at a global scale~~. This method allows for simulating synthetic observations in places where ~~an observing system does not exist~~ yet real observations do not exist, and to ~~see how useful the synthetic observations are for the estimation~~. examine how useful are the information they would provide. The objective of the present study is to ~~characterize and~~ identify sampling regions, ~~regarding~~ characterized by different oceanic variables, at a global scale, in which micronekton biomass ~~observation gives~~ observations provide the most useful information for the model energy transfer coefficients estimation. A set of synthetic observations is generated with SEAPODYM using a reference parameterization. Then, the set of parameter values is changed and an error is added to the forcing field in order to simulate more realistic conditions for parameter estimation. The MLE is used to estimate the set of parameters from the set of synthetic observations. The difference between the reference and estimated parameters provides a metric to select the best sampling zones. A method based on the clustering (Jain et al., 1999) of ~~oceanic variables~~ four oceanic variables of interest (temperature, currents velocity, stratification and productivity) is presented to investigate the sensitivity of the parameters estimation to the oceanographic conditions of the observation regions. This method aims at determining which conditions are the most favorable for collecting observations in order to estimate the energy transfer efficiency coefficients.

The paper is organized as follows: Section 2 describes the model ~~set-ups~~ set-up and forcings as well as the method developed to characterize regions of observations and the metrics used to evaluate the parameters estimation. Section 3 describes the outcome of the clustering method to define oceanographic regimes and synthesizes the main results of our estimation experiments. The results are then discussed in Section 4 in the light of biological and dynamical processes. Some applications and limitations of our study are also identified along with suggestions for possible future research.

2 Method

2.1 The SEAPODYM-MTL and its configuration model

SEAPODYM-MTL (Mid-Trophic Levels) ~~models~~ simulates six functional groups of micronekton in the epi- and upper and lower mesopelagic layers at a global scale. These layers encompass the upper 1000 m of the ocean. The euphotic depth (z_{eu}) is used to define the depth boundaries of the vertical layers. These boundaries are defined as follows (an approximate average depth is given in brackets): $z_1(x, y, t) = 1.5 \times z_{eu}(x, y, t)$ ($\sim 50 - 100$ m), $z_2(x, y, t) = 4.5 \times z_{eu}(x, y, t)$ ($\sim 150 - 300$

m), $z_3(x, y, t) = \min(10.5 \times z_{eu}(x, y, t), 1000)$ ($\sim 350 - 700\text{m}$), where z_{eu} is given in meters. The six functional groups are called (1) epi (for organisms inhabiting permanently the epipelagic layer); (2) umeso (for organisms inhabiting permanently the upper mesopelagic layer); (3) ummeso (for migrant-umeso, organisms inhabiting the upper mesopelagic layer at day and the epipelagic layer at night); (4) lmeso (for organisms inhabiting permanently the lower mesopelagic layer); (5) lmmeso (for migrant-lmeso, organisms inhabiting the lower mesopelagic layer at day and the upper mesopelagic layer at night) and (6) lhmmeso (for highly migrant lmeso, organisms inhabiting the lower mesopelagic layer at day and the epipelagic layer at night). The model is forced by current velocities, temperature and net primary production (see Appendix A for detailed equations). This work is based on a ten-year (2006-2015) simulation of SEAPODYM-MTL, called hereafter the nature run (NR). Eu-
100 photic depth, horizontal velocity and temperature fields come from the ocean dynamical simulation FREEGLORYS2V4 produced by Mercator-Ocean. FREEGLORYS2V4 is the global, non-assimilated version of GLORYS2V4¹ simulation that aims at generating a synthetic mean state of the ocean and its variability for oceanic variables (temperature, salinity, sea surface height, currents speed, sea-ice coverage). It is produced using the numerical model NEMO² with the ORCA025 configuration (eddy-permitting grid with 0.25° horizontal resolution and 75 vertical levels, see Barnier et al. (2006)) and forced with the
105 ERA-Interim atmospheric reanalysis from the ECMWF³. The net primary production is estimated using the Vertically Generalized Production Model (VGPM) of Behrenfeld and Falkowski (1997) with satellite derived chlorophyll-a concentration. This product is available at Ocean Productivity Home Page of the Oregon State University⁴. Due to high computational demand, the original resolution of the simulation $0.25^\circ \times \text{week}$ has been degraded to $1^\circ \times \text{month}$. Temperature, horizontal velocity and primary production fields are depth-averaged over the water column of each three layers defined by z_1, z_2 and z_3 , ending
110 with a set of three-layered forcings fields. Initial conditions of SEAPODYM-MTL come from a two-year spin-up based on a monthly climatology simulation. Reference values of SEAPODYM-MTL parameters are those published in Lehodey et al. (2010). Overall the simulation reproduces the dynamics of the ocean well, but due to the low 1° horizontal resolution, meso-scale features like eddies are not represented. The simulation captures the main temporal variability with a seasonal cycle in primary production and DVM cycle for micronekton.

115 2.2 Clustering approach to characterize potential sampling regions

~~In this section we describe the method we used to select the different observation regions for OSSE, based on environmental characteristics.~~ We define the spatio-temporal discrete observable space Ω as the set of the $1^\circ \times 1^\circ$ grid points belonging to SEAPODYM-MTL discrete domain. ~~The characterization of each observation point relies on four indicators defined from~~ the Each observation point is characterized by four indicators which are based on the following environmental variables: the
120 depth-averaged temperature \mathcal{T} , a stratification index \mathcal{S} , the surface velocity norm \mathcal{V} and a bloom index \mathcal{B} , for which different

¹<http://resources.marine.copernicus.eu/documents/QUID/CMEMS-GLO-QUID-001-025.pdf>

²<https://www.nemo-ocean.eu/>

³<https://www.ecmwf.int/>

⁴<http://www.science.oregonstate.edu/ocean.productivity/>

regimes of intensity are defined. The averaged temperature \mathcal{T} over the water-column is defined as:

$$\mathcal{T}(x, y, t) = \frac{1}{3}(T_1(x, y, t) + T_2(x, y, t) + T_3(x, y, t)), \quad (1)$$

where T_k is the depth-averaged temperature over the k^{th} trophic layer of the model. The stratification index \mathcal{S} is defined as the absolute difference of temperature between the surface and subsurface layers:

$$125 \quad \mathcal{S}(x, y, t) = |T_1(x, y, t) - T_2(x, y, t)|. \quad (2)$$

The surface velocity norm \mathcal{V} is defined as:

$$\mathcal{V}(x, y, t) = \sqrt{u_1^2(x, y, t) + v_1^2(x, y, t)}, \quad (3)$$

where u_1 and v_1 are the zonal and meridional components of the depth-averaged velocity respectively, in the first layer of the model. The phytoplankton bloom index \mathcal{B} is defined following Siegel et al. (2002) and Henson and Thomas (2007) as a Boolean: 1 for bloom regions and 0 for no bloom regions ~~according to temporal variation relative to annual median threshold overshooting~~based on temporal variations of primary production exceeding a threshold based on its annual median. More precisely, we define:

$$130 \quad \mathcal{B}(x, y) = \begin{cases} 1 & \text{if there exists } t \text{ such that } |PP(x, y, t) - \widetilde{PP}(x, y)| > 0.05 \times \widetilde{PP}(x, y), \\ 0 & \text{elsewhere.} \end{cases} \quad (4)$$

where $\widetilde{PP}(x, y)$ is the temporal median of the primary production $PP(x, y, t)$ at point (x, y) . Note that contrary to the previous indicator variables, the bloom index does not depend on time. For each indicator variable $\mathcal{G} \in \{\mathcal{T}, \mathcal{S}, \mathcal{V}, \mathcal{B}\}$ we define several ordered value-based *regimes*. The number of regimes and ~~regimes-regime~~regime boundary values are obtained by partitioning the set G_N of the values of the indicator variable \mathcal{G} at N observable locations constituting an ensemble $S_N \subset \Omega$.

$$G_N = \{g_i = \mathcal{G}(X_i) \quad X_i \in S_N\}_{1 \leq i \leq N}. \quad (5)$$

The partition of G_N is computed using ~~a~~k-mean clustering (Kanungo et al., 2002). The k -mean clustering method separates ~~N~~N ~~values-points~~value-point in a given number of cluster by minimizing the distance of each ~~value-point~~value-point to the mean (called the center) of each cluster. The number ~~n~~n of clusters is chosen according to the Elbow score (Kodinariya and Makwana, 2013; Tibshirani et al., 2001). The k -mean method ~~leads to produces~~produces ~~n~~n clusters $(\Gamma_k)_{k \in \{1..n\}}$ (called indicator variable regimes), that satisfy the following properties:

$$145 \quad \begin{cases} \bigcup_{k=1}^n \Gamma_k = G_N & \text{and} & \forall i, j \in \{1..n\}, i \neq j, \quad \Gamma_i \cap \Gamma_j = \emptyset \\ \text{and} \\ \forall i \in \{1..N\}, g_i \in \Gamma_k & \text{if } k = \underset{l \in \{1..n\}}{\operatorname{argmin}} \|g_i - \mu_l\|, \end{cases} \quad (6)$$

where μ_l is the mean of values in Γ_l . Note that Γ_k depends on the variable \mathcal{G} . In the following, we make this dependence explicit by denoting $\Gamma_k(\mathcal{G})$. The k -mean clustering allows for size-varying ~~class-clusters~~class-clusters compared to more classical statistical

analysis that would consist for example to define the regimes as the quantile of the variables distributions. ~~This~~ The latter could lead to under ~~or over estimation of some~~ (or over) estimate some identified regimes. The same kind of problem would arise from a classification defined by traditional eco-regions (Longhurst, 1995; Sutton et al., 2017), which would not account for the specificity of our forcing fields. This is why performing a clustering ~~on the set of forcing fields used to determine the different regimes associated with the forcing fields~~ seems a more rigorous approach here.

We define a *configuration* as the intersection of a selection of regimes of given indicator variables. For $i \in \{1 \dots n_{\mathcal{T}}\}$, $j \in \{1 \dots n_{\mathcal{S}}\}$, $k \in \{1 \dots n_{\mathcal{V}}\}$ and $l \in \{1 \dots n_{\mathcal{B}}\}$, the configuration C is defined as:

$$C = \mathcal{T}_i \otimes \mathcal{S}_j \otimes \mathcal{V}_k \otimes \mathcal{B}_l = \Gamma_i(\mathcal{T}) \cap \Gamma_j(\mathcal{S}) \cap \Gamma_k(\mathcal{V}) \cap \Gamma_l(\mathcal{B}), \quad (7)$$

where $n_{\mathcal{G}}$ is the number of clusters for the indicator variable \mathcal{G} . For the sake of simplicity we may also say that an observation point belongs to a configuration when the values of the indicator variables at this point belong to the corresponding regimes of the configuration. Each configuration corresponds to a subset $S_M \subset S_N$ of observable points.

2.3 OSSE system configuration

~~The implementation of OSSE requires to follow a precise protocol~~ To perform realistic OSSEs, a rigorous protocol needs to be followed (Hoffman and Atlas, 2016). Here, we describe the different steps. A scheme summarizing the OSSE methodology is given in Figure 1.

2.3.1 Nature ~~run~~ & Control runs and assimilation module

The nature run (NR) used to perform the OSSE is generated using the reference configuration of SEAPODYM-MTL described in section 2.1. The ~~reference simulation~~ NR is used to compute synthetic observations. The goal is ~~to retrieve back then to~~ retrieve the reference energy transfer coefficients of the six micronekton functional groups E_i^j by assimilating the synthetic observations into a different simulation of SEAPODYM-MTL, called the control run.

2.3.2 Control run

The control run (CR) used to perform the parameter estimate is generated using perturbed forcing fields (Figure 1). A perturbation is added to the reference forcing fields in order to consider more realistically the discrepancy between the real state of the ocean (represented here by the NR) and the simplified representation of this state by numerical models. The reference forcing fields are perturbed with a white noise whose maximal amplitude is a fraction of the averaged fields. Let F be the considered forcing field and let \bar{F} be its global average (in space and time), we define the perturbed field as:

$$\tilde{F}(x, y, t) = F(x, y, t) + \gamma(\alpha \bar{F}), \quad (8)$$

where $\alpha \in [0, 1]$ is the amplitude of the perturbation and $\gamma \in [-1, 1]$ is a uniformly distributed random number. The amplitude α is set to 0.1 for all experiments except in section 3.4 where α varies. For small values of F , this perturbation can induce a sign reversal of the forcing. This does not matter for the temperature (degree celcius, see also Equations A5 and A6) or the currents

velocities (meter per second), primary production (milimol of carbon per squared meter per day) has however been constraint to positive values. White noise has been preferred to more realistic perturbation to avoid any geographical bias pattern. The implications of this choice are further discussed in section 4.3. Its amplitude, fixed to 10% of error, is however representative of the mean error estimated for ocean circulation models (Lellouche et al., 2012; Ferry et al., 2012). The parameters E'_i are randomly sampled between 0 and 1. This *first guess* is used as initialization of the optimization scheme. We run each experiment several times with different random sampled first guess in order to ensure that the inverse model is not sensitive to the initial parameters. The set-up of the NR and CR simulations are summarized in Table 1.

2.3.2 Assimilation module

A MLE is used as an assimilation module, used here to estimate model parameters from observations. Its implementation is based on an adjoint technique (Errico, 1997) to iteratively optimize a cost function that represents the discrepancy between model outputs and observations. This approach conforms to current practices. More details about the implementation of this approach in SEAPODYM can be found in Senina et al. (2008) and Lehodey et al. (2015).

2.3.2 Synthetic observations

In the framework of OSSE, we perform estimation experiments with different sets of ~~$N_e = 400$~~ -synthetic observation points of size $N_e = 400$. The synthetic observations are sampled in from the different configurations ~~constructed as explained~~ introduced in the previous section. Let M be the number of points in a given configuration. If $M < N_e$, we consider that the configuration is too singular to be relevant for our study and it is ignored. If $M > N_e$ we randomly extract a sub-sample $S_{N_e} \subset S_M$ of observation points. In order to study the influence of one indicator at a time, we compare experiments for which the regime of the studied indicator varies and the regime of the other indicator variables remain fixed. In the following we call *primary variable* the studied indicator variable and *secondary variables* the ones whose regimes are fixed. For a given group of experiments, we check that the configurations are ~~statistically~~-comparable to each ~~others~~ other by ensuring that the distribution of all secondary variables are ~~close enough between configurations~~ similar (cf. marginal distribution plots in Section 33.2.1). If this not the case, they are not reported. A random sampling of observations within each configuration is preferred to a more realistic observation network to avoid any geographical bias. But this choice is discussed in section 3.4, where realistic networks are tested. The coverage in terms of observation numbers is however quite realistic. We ~~assume~~ use 400 observations in our experiments, which at the resolution of the model ($1^\circ \times 1$ month) corresponds for example to the deployment of six moorings during five years.

2.4 OSSE system evaluation metrics

The estimation experiments are evaluated using three metrics: (i) the performance of the estimation, (ii) its accuracy and (iii) its convergence speed.

(i) The performance is measured with the mean relative error between the estimated coefficients and the reference coefficients as defined in Lehodey et al. (2015) (Eq. 9).

$$E_r = \frac{1}{6} \sum_{i=1}^6 \left| \frac{\widehat{E}_i' - E_i'}{E_i'} \right|. \quad (9)$$

210 (ii) The accuracy is measured by the residual value of the likelihood which provides a good estimate of the discrepancy between the estimated and the observed biomass.

(iii) The convergence speed is measured by the iterations number of the optimization scheme.

The residual likelihood and iterations number metrics are provided by the Automatic Differentiation Model Builder (ADMB) 215 algorithm (Fournier et al., 2012) that is used to implement the MLE. Each metric provides different and independent information. For example, it is possible to obtain good performance and bad accuracy with an experiment that estimates correctly the energy transfer parameters for the different functional groups but over- or under-estimates the total amount of biomass. The performance is generally used to discriminate the different experiments since the aim of the study is to find the networks that better estimate energy transfer coefficients and thus directly minimize the error E_r (Eq. 9). However, the accuracy and 220 precision of the experiment are ~~discussed~~also discussed below. The convergence is necessary to ensure that the optimization problem is well defined.

3 Results

3.1 Environmental regimes clustering

The number of points per regime, obtained from the clustering (Section 2.2) and defined for each environmental variable 225 (Table 2), shows a large variability. Some regimes represent a larger amount of observable points. For instance, the tropical temperature regime covers 31% of the observable points. Almost 50% of the observable points show a weak stratification and only 10% of them have a positive bloom index or high velocities. When they are shown on a map (Figure 2) these regimes reproduce classical spatial patterns described in the scientific literature (Fieux and Webster, 2017). The regimes of the temperature variable (\mathcal{T}) show a latitudinal distribution. The polar regime (\mathcal{T}_1) is located south of the Polar front (Southern hemisphere) 230 and in the Arctic Ocean. The subpolar regime is located between the Polar front and the South Tropical front (Southern Ocean), in the Subpolar gyre region (North Atlantic) and in the Bering Sea (North Pacific). The temperate regime covers the subtropical zones of the Southern Atlantic, Indian and Pacific Oceans, located north of the South Tropical front, and extends as well in the eastern part of the Atlantic and Pacific Ocean. The tropical regime covers most of the tropical ocean and the Indian ocean. The regimes of the stratification variable (\mathcal{S}) are also structured according to the latitude, as stratification depends on 235 the temperature. The stratification decreases from the tropical oceans (where the surface waters are warm compared to the deep waters) to the pole (where the surface waters are almost as cold as the deep waters). The regimes of the surface velocity norm (\mathcal{V}) highlight the main energetic structures of the oceanic circulation. The high surface currents regime thus covers the

intense jet-structured equatorial currents, the western boundary currents (the Gulf Stream in the Atlantic and the Kuroshio in the Pacific), the Agulhas current along the South Africa coast and the Antarctic Circumpolar Current in the Southern Ocean.

240 The regimes of bloom index (\mathcal{B}) separate mostly the productive regions (North Atlantic and North Pacific, Southern Ocean, Eastern side of Tropical Atlantic, along the African coast) from the non productive regions (center of subtropical gyres mostly, as well as coastal regions of Arctic and Antarctic).

Based on these results, we construct all possible configurations, using the methodology described in [subsection Section 2.2](#).

245 Then the configurations are selected to perform the OSSEs presented in subsection 2.3. The choice of the configuration is limited by the number of observation points available in each of them. Among the 48 possible configurations, 21 of them are near-empty ~~intersection and as they~~ contain less than 0.5% of all observable points. They are thus considered as non-existent. In addition, we study the influence of the primary variable by selecting only groups of configurations whose distributions along secondary variables are similar. This leads to a selection of 7 groups of experiments (Table 3). The [purpose of the](#) first three

250 groups of Experiments 1a-b, 1c-d and 1e-f ~~are meant is~~ to study the influence of the velocity regimes \mathcal{V}_1 and \mathcal{V}_2 . The group of Experiments 2a-d ~~will be is~~ used to study the influence of the temperature regimes \mathcal{T}_1 , \mathcal{T}_2 , \mathcal{T}_3 and \mathcal{T}_4 . The group Experiments 3a-c ~~will be is~~ used to investigate the influence of the stratification index regimes \mathcal{S}_1 , \mathcal{S}_2 and \mathcal{S}_3 . Finally, Experiments 4a-b and 4c-d ~~are used for the study the influence~~ [evaluate the impact](#) of the bloom index regimes \mathcal{B}_1 and \mathcal{B}_2 .

3.2 Estimation performance with respect to environmental conditions

255 Table 3 shows the selected configurations for each experiment ~~as~~ [\(usually abbreviated as Exp. in the following\)](#) as well as their evaluation metrics. All experiments converged after 16 to 28 iterations. This confirms that the optimization problem is well defined. Since the number of iterations is partially dependent on the random initial first guess, it is not used as a criterion of discrimination between experiments.

3.2.1 Influence of the horizontal currents velocity

260 The influence of the current velocity regimes (high current velocity system or low current velocity system) on the performance of the parameters estimation is studied considering three groups of experiments (Table 3, Exp. 1a to 1f). The observation points are randomly sampled in a subset of the considered configuration for which the primary variable is the currents velocity norm \mathcal{V} .

From these sets of experiments, it appears that the performance of the parameters estimation decreases with higher ~~currents~~ [current](#) velocity at the observation points. This conclusion is valid regardless of the regime of the secondary variables: either low or high temperatures, positive or null bloom index and weak or strong stratification (Table 3). Lower velocity reduces the error on the estimated energy transfer coefficients for functional groups that are impacted by currents in the epipelagic and upper mesopelagic layers. The currents decrease with depth and are almost uniform over the different regions in the lower mesopelagic layer (not shown). Consequently, the estimate of the parameters for the non migrant lower mesopelagic (Imeso)

270 group is not sensitive to the regime of currents (Figure 3a). Conversely, the estimation is the most sensitive for the epipelagic group, whose dynamics ~~is~~ are entirely driven by the surface currents.

Note that the influence of low and high velocities is not explored for all secondary variable fixed regimes. Indeed, even within fixed regimes, the secondary variables distribution along observation points might not be statistically comparable between two experiments. This could lead to a potential bias introduced by a secondary variable, which is not the target of the study. For instance, the influence of velocity in a polar temperature regime can be investigated by comparing the configurations $C' = \mathcal{T}_1 \otimes \mathcal{S}_1 \otimes \mathcal{V}_1 \otimes \mathcal{B}_2$ (low velocity) and $C'' = \mathcal{T}_1 \otimes \mathcal{S}_1 \otimes \mathcal{V}_2 \otimes \mathcal{B}_2$ (high velocity). The corresponding estimation experiments Exp. 1' and Exp. 1'' give relative errors of 48% and 10% respectively. This result seems contradictory ~~with~~ to the conclusions drawn from Exps. 1a-f. But looking at the distributions of the observations along the secondary variables, we can notice that the temperatures are different between the two configurations. ~~Despite this, it has been fixed to~~ While both configurations are considered to be in the "polar regime", the temperature in configuration C' ~~is on average lower~~ (-0.7°C) is on on average lower than the temperature of configuration C'' (2.1°C) (Figure 4). Thus Exps. 1' and 1'' measure the combined effect of both velocity and temperature. The lower velocities are coupled with lower temperatures and the higher velocities with higher temperatures. There is a cross-correlation between the velocity (primary variable) and the temperature (secondary variable). Therefore, it is not possible to ~~conclude on~~ assess the influence of the velocity on the parameters estimation from these experiments.

285 ~~In the following, although distribution along~~ Although the distributions of the secondary variables are not always shown in the following experiments, they have ~~always been used in the analysis to check~~ been examined to ensure that the OSSE results are not biased by ~~this type of difference between the distributions of randomly selected datasets~~ systematic differences in the secondary variables. Experiments with ~~such significant~~ cross-correlation between indicator variables are not presented, this concerns 9 out of the 26 possible experiments.

290 3.2.2 Influence of temperature

In experiments 2a to 2d (Table 3), temperature is the primary variable, ranging from polar regime (Exp. 2a), to subpolar (Exp. 2b), temperate (Exp. 2c) and tropical (Exp. 2d) regimes. All other indicator variables (stratification, velocity and bloom index) are secondary variables that are set to weak, low and 1 respectively. Figure 5 shows that the distributions along the secondary variables of each configuration are close enough for the experiments to be compared, avoiding any risk of cross-correlation. The performance of the estimation increases with the temperature (Figure 3b). The mean error on the parameter estimates decreases respectively from polar (Exp. 2a; 9.1%) to subpolar (Exp. 2b; 7%), temperate (Exp. 2c; 3%) and tropical (Exp. 2d; 1.4%) configurations (Table 3).

3.2.3 Influence of stratification

The influence of stratification is first investigated with a set of three configurations combining tropical temperature regime, low velocity regime, null bloom index regime and three regimes of weak (Exp. 3a); intermediate (Exp. 3b) and strong (Exp. 3c) stratification. A marginal distribution plot of observation sets for all experiments (not shown) indicates that the three datasets differ only along the stratification variable (primary variable). The observation points display a temperature between 14°C

and 17°C, a velocity between 0 and 0.07 m s⁻¹ and a null bloom index for each experiments. The performance decreases with the intensity of stratification (Figure 3c and Table 3). The mean error is: 3.5% for a weak stratification and a vertical
305 gradient of about 0.4°C (Exp. 3a), 5.9% for an intermediate stratification with a gradient of about 5.9°C (Exp. 3b) and 8% for a strong stratification, around 11.7°C (Exp. 3c). A strong stratification seems to deteriorate the estimate for all migrant groups (Figure 3c). These results are not specific to the choice of regimes for the secondary variables. ~~The same kind of~~ Similar experiments were carried out in a temperate regime (not shown) and even though the mean error on the estimated parameters is higher on average, the result does not change: weak stratification always again leads to a better estimation than strong
310 stratification. The comparison was not fully possible in other temperature or velocity regimes because these configurations are not sufficiently represented.

3.2.4 Influence of primary production

In order to investigate the influence of primary production on the performance of the estimation, we compare the results of estimation in configurations with different bloom index regimes (primary variable). Temperature, stratification index and velocity
315 have been fixed (secondary variables) to subpolar, weak and low regimes respectively (Exp. 4a and 4b) and to tropical, strong and low for Exp. 4c and 4d. Distributions of the observation points along the secondary variables indicate that the experiments are not biased by secondary variables, as the distributions present similar modes centered at 5°C for the temperature, at 0.5°C for the stratification index and at 0.04 m s⁻¹ for the velocity (Exp. 4a and 4b) and at 15.5°C, 11°C and 0.05 m s⁻¹ respectively for Exp. 4c and 4d (not shown).

320 Both Exp. 4a and 4b result in an averaged error of 7% on the estimated parameters (Table 3). Exp. 4d (averaged error of 8%) gives a similar value as Exp. 4b. Indeed, Exp. 4d (\mathcal{T}_4 regime) has higher temperature than Exp. 4b (\mathcal{T}_2 regime) but it has also a higher stratification index (\mathcal{S}_3 regime for Exp. 4d and \mathcal{S}_1 regime for Exp. 4b). Following the conclusions from the two previous sections, better performance is achieved when temperature increases, though increasing stratification has the opposite effect. So, the two effects might compensate in this case and result in a similar estimation. However, when considering bloom regions
325 (Exp. 4c), the estimation error falls to 1.5% on average. In addition, this experiment estimates the energy transfer coefficients for migrant micronekton groups with less than 1% error (Figure 3d). According to our results, the primary production and the regimes of the bloom index do not always play a role in the performance of the parameters estimation. A positive bloom index appears to improve the performance of the estimation at high temperatures only.

3.3 Global map of parameters estimation errors

330 When considering all possible experiments, and given the fact that all these configurations are associated ~~to~~ with specific locations and times, it is possible to represent a global map of averaged estimation errors (Eq. 9). This map (Figure 6) shows that on average, the error increases from the equator towards the poles. The lowest performances (errors > 40%) are mostly found in the Arctic and Southern Ocean. Low performances are also found at some specific locations (e.g. along the main currents). The signature of the Antarctic Circumpolar Current is found in the Southern Ocean with error over 10%. Similarly,
335 the signature of the North Atlantic Drift can be seen with a patch of high errors between Canada and Ireland (Figure 2c and 6).

The patch of high errors in the North Pacific Ocean, however, is difficult to interpret. The equatorial regions show interesting patterns that are similar across the three oceans. In the vicinity of the equator, good performances are observed (mean error 2%). On both northern and southern sides of this low error band, the performance is decreased with errors reaching about 8%. The equatorial regions are characterized by strong currents and warm surface waters. As described above, these environmental features have opposite effects on the performance of the estimation. Therefore, a possible explanation of this distribution of errors is that water temperature is high enough to overcome the effect of currents in the equatorial band, but when moving poleward, the temperature decreases cannot compensate anymore for the negative effect of currents which is still quite strong. It is to note that the map presented in Figure 6 has been obtained for a given set of forcing fields (temperature, velocity, primary production). It is thus dependent on the simulation that is used. The regime-dependence of the estimation performance is however independent of the simulation.

3.4 Testing realistic networks

The above experiments are based on random selection of observation points within a large subset. This technique was chosen to avoid any bias related to the temporal or spatial potential auto-correlation of observation networks. However, sampling at sea is rarely randomly distributed and can generate correlations. To relax this strong assumption, we perform experiments based on positions from real acoustic transects (underways ship measurements). Two regions are compared using ~~positions data of the the transects from the~~ PIRATA cruises in the Equatorial Atlantic ~~Ocean (PIRATA) and cruises of (?) and those from the~~ British Antarctic Survey ~~in Antartec peninsula region~~ (BAS) ~~close to the Antarctic peninsula~~ (Figure 7).

The same forcing, method and initial parameterization were used with a random noise amplitude (α) increasing from 0 to 0.2. Subsets of $N_e = 400$ observations were selected along the transects to run the experiments. The resulting averaged relative error on the coefficients is shown as a function of the amplitude of perturbation (Figure 8a) for both networks. It appears that the estimation error increases with the amplitude of the error introduced on the forcing field. Also, ~~whatever the~~ ~~regardless of the intensity of the~~ perturbation, the estimation error is always lower when using PIRATA observation networks than BAS observation networks. These results are fully consistent with the previous results indicating that networks located in tropical warm waters, as for PIRATA, give better estimates than the ones located in cold waters, as for the BAS (Figure 8b). This should give confidence in the fact that our results are robust when the "random sampling" hypothesis used in the previous section is relaxed and that more realistic sampling designs are considered. Here in particular, the temporal auto-correlation of the different samplings is very strong since PIRATA and BAS are both underway ship measurements taken from 2-month cruises, repeated annually. The results seem much less dependent to the exact design of the samplings and the seasonality of the measurements than to their actual geographical location. Oceanic conditions of the observations (correlated to their geographical location) are the first order of sensitivity. In this sense, the PIRATA network is thus a very promising observatory for the micronekton, especially since it already includes a complete set of various physical and biogeochemical parameters measurements (Foltz et al., 2019).

4 Discussion

In the following, we will discuss a possible theoretical interpretation of the outcome of the estimation experiments (section 4.1) and a potential application of our results (section 4.2). Section 4.3 closes this discussion ~~discussing~~ examining the particular framework used to conduct this study and opening some perspectives for future work.

4.1 An interpretation of the performance in ~~term~~ terms of observability

The differences in the performance of parameter estimation can be interpreted in the light of the characteristic ~~times~~ timescales of physical and biological processes. The parameters we want to estimate (E'_i) control the energy transfer efficiency between the primary production (PP) and micronekton production (P) (Eq. A3; Appendix A). These parameters are thus directly related to the relative amount (P_i) of P ~~at age $\tau = 0$~~ in each functional group i ~~at age $\tau = 0$~~ and we have:

$$E'_i = \frac{P_i(\tau = 0)}{cE_{pp} \int PP dz} \quad (10)$$

where E_{pp} is the total energy transfer from the primary production to the mid-trophic level ~~;(all functional groups together)~~ and c a conversion coefficient (see Appendix A). It is possible to rewrite the initial condition (Eq. A3) as a system of six equations involving the energy transfer coefficients.

$$\begin{cases} \rho_{1,d}(P|_{\tau=0}) = E'_1 \\ \rho_{1,n}(P|_{\tau=0}) = E'_1 + E'_3 + E'_6 \\ \rho_{2,d}(P|_{\tau=0}) = E'_2 + E'_3 \\ \rho_{2,n}(P|_{\tau=0}) = E'_2 + E'_4 \\ \rho_{3,d}(P|_{\tau=0}) = E'_4 + E'_5 + E'_6 \\ \rho_{3,n}(P|_{\tau=0}) = E'_4 \end{cases} \quad (11)$$

where $\rho_{K,\omega}(P|_{\tau=0})$, ~~with ρ defined as $\rho_{K,\omega}(P) = \frac{\sum_{i|K(i,\omega)=K} P_i}{\sum_{i=1}^6 P_i}$~~ is the ratio of age 0 potential micronekton production in the layer $K \in \{1, 2, 3\}$, at the time of the day $\omega \in \{\text{day, night}\}$ ~~$\omega \in \{d, n\}$ (for day and night).~~

The predicted micronekton biomass at a given time and location (grid cell) results from two main mechanisms. First, the potential production (P) evolves in time from age $\tau = 0$, and is redistributed by advection and diffusion until the recruitment time τ_r , when it is transferred into biomass (B). ~~Then~~ Second, the biomass is built by the accumulation of recruitment over time in each grid cell and is lost due to a temperature-dependent mortality rate, while the currents redistribute the biomass spatially. The observations ~~are~~ correspond to the relative amount of biomass in each layer, i.e. the ratios of biomass $\rho_{K,\omega}(B|_{t=t^o})$ (Eq. A7), where t^o is the time at which the observation is collected. Therefore, the observation will ~~be as close as contain~~ more information about the energy transfer parameters we want to estimate if $\rho_{K,\omega}(B|_{t=t^o})$ is close to $\rho_{K,\omega}(P|_{\tau=0})$ ~~;(Eq. 11).~~ This requires that the integrated mixing and redistribution of biomass during the elapsed time between the age 0 of potential production and the time of observation (i.e. at least the recruitment time) is as weak as possible. This can be achieved in

two ways: (i) either the currents are weak so that the ~~advective mixing~~ advection of biomass is also weak (but the ~~diffusive mixing~~ diffusion will still remain); (ii) Or the temperature is high, leading to a short recruitment time with reduced period of transport ~~and biomass redistribution~~, mixing and redistribution of biomass (Eq. A5). These two mechanisms can explain why warm temperatures and weak currents were found to improve the estimations compared to cold temperatures and high velocities (Sections 3.2.1 and 3.2.2). An additional effect of warm temperature is ~~to induce~~ that it induces a higher mortality rate ~~(Eq. A6)~~. When warm waters are combined with high primary production (e.g. the equatorial upwelling region), there is a rapid turnover of biomass and the relative ratios of biomass by layer are closer to the initial ratio of production and thus to the energy transfer efficiency coefficients. Conversely, at cold temperature, the mortality rate is lower; biomass is accumulated from recruitment events and carries with it the integrated mixing and the perturbed ratio structures. This can explain why, at warm temperature, high productivity is needed for a better estimation (section 3.2.4). A side effect is that if temperature is not homogeneous across layers, then the mortality rate λ will differ for each functional group, depending on the layers it inhabits. This will be an additional driver of perturbation on the observed ratios of biomass compared to the initial ratios of potential production. This is consistent with the result that a strong thermal stratification degrades the performance of estimation (section 3.2.3).

An observation will thus be the most effective for the estimation of parameters if it carries the information of the initial distribution of primary production into functional groups. This is the case if the biomass is renewed quickly enough compared to the time it takes for the currents and diffusive coefficient to mix it. This condition can be seen in terms of equilibrium between the biological processes (production, recruitment and mortality) and the physical processes (advection and diffusion). For an observation to be the most useful to the parameter estimation, it is necessary that the ~~characteristics time~~ characteristic timescales governing biological processes (τ_β) is shorter than the one governing physical processes (τ_ϕ) at the location of the observation : $\tau_\beta \ll \tau_\phi$.

This interpretation highlights the problem of observability of the parameters E'_i from the measurements $\rho_{K,\Omega}(B)$. The parameters are directly observable at the age $\tau = 0$ of the ~~primary~~ production, but the measurements and the information we can get on the system are available only after a time τ_r . The observability will then be the better if the observable variables have not changed too much during the time τ_r (short τ_r , slow ocean dynamics). This is intrinsically linked to governing equations of the system (Eq. A1-A3) and therefore should not be dependent of the framework of the study.

4.2 Towards eco-regionalization ?

The clustering approach we propose allowed identifying oceanic regions that provide optimal oceanic characteristics for our parameters estimation ~~by separating~~. It separates regions where the distribution of biomass is driven by physical processes from regions where it is driven by biological processes. ~~It gives essential information about the optimal regions for implementing observational networks~~. This could be seen as a new definition of eco-regions based on similar ecosystem structuring dynamics. The definition of ocean eco-regions has been proposed based on various criteria (Emery, 1986; Longhurst, 1995; Spalding et al., 2012; Fay and McKinley, 2014; Sutton et al., 2017; Proud et al., 2017). A convergence of these different approaches to identify regions characterized by homogeneous mesopelagic species communities would be of great interest to facilitate the

~~modeling~~ modelling and biomass estimate of the mesopelagic components. Acoustic observation models could be developed and validated at the scale of these regions. Then, the observation models integrated to ecosystem and micronekton models as the one used here, would serve to convert their predicted biomass into acoustic signal to be directly compared to all acoustic observations collected in the selected region. This approach would allow to account for (and estimate) the sources of biases and errors linked to acoustic observations directly in the data assimilation scheme.

4.3 Limitations and perspectives

We have chosen to model the error between the true state of the ocean and the modelled state by adding a white noise perturbation to the forcings of the NR as input of the CR. ~~The realism of this approach is questionable, as it~~ Our idealized approach does not take into account the possible spatial distribution of uncertainty and errors of ocean models, and other approaches would be interesting to explore. For instance, implementing an error proportional to the deviation of the climatological field should be more realistic because it would be based on the natural and intrinsic variability of the ocean. Indeed, we expect forcing fields to be less accurate where the ocean has strong variability. However, for the purpose of our study, a spatial homogeneous error was preferable to avoid introducing any bias. Random noise ensures that the results obtained in different locations are directly comparable. ~~Sensitivity study~~ Conducting a sensitivity analysis with respect to the choice of forcing errors modelling was beyond the scope of this study. In addition to the uncertainty ~~on~~ of ocean models outputs, other sources of uncertainties remain to be explored to progress toward more realistic estimation experiments. For instance, we considered that the observation operator (Eq. A7) is perfect but field observations are always tainted by errors. The micronekton biomass estimates at sea require a chain of extrapolation and corrections to account for the sampling gear selectivity and the portion of water layer sampled. For acoustic data, many factors need to be considered sources of potential error: the correction with depth, the target strength of species, the intercalibration between instruments and the signal processing methods (Handegard et al., 2009, 2012; Kaartvedt et al., 2012; Proud et al., 2018). This is an important research domain that requires to combine multiple observation systems, including new emerging technologies as broadband acoustic, optical imagery and environmental DNA to reduce overall bias in estimates of micronekton biomass (e.g., Kloser et al., 2016) and use those estimates to assess, initiate and assimilate into ecosystem models. Finally, the results of the clustering approach need to be confirmed with other ocean circulation model outputs, especially at higher resolution to check the impact of the mesoscale activity on the definition of optimal regions for energy transfer efficiency estimation. In a future study, in addition to test the impact of introducing noises in the observations, the same approach could be used to directly estimate also the model parameters that control the relationship between the water temperature and the time of development of micronekton organisms. Other perspectives may include a study of the sensitivity to the design of the samplings (the impact of moored instruments in comparison with underway measurements), in the continuity of the work of Lehodey et al. (2015).

5 Conclusions

Understanding and modelling marine ecosystem dynamics is considerably challenging. It generally requires sophisticated models relying on a certain number of parameterized physical and biological processes. SEAPODYM-MTL provides a parsimonious approach with only a few parameters and an MLE to estimate these parameters from observations. Among them, the energy transfer efficiency coefficients are of great importance because they directly control the biomass of micronekton functional groups, including those that undergo DVM and contribute to the sequestration of carbon dioxide into the deep ocean (Davison et al., 2013; Giering et al., 2014; Ariza et al., 2015). Therefore, a correct assessment of energy transfer coefficients is crucial for climate studies. Given the high cost of ~~observation~~ observations at sea, the design of optimal observational networks through simulation experiments (OSSEs) is a valuable approach before the deployment of such platforms. Our objective was different from most OSSEs studies designed to ~~correct outputs of operational models, e.g., for weather and physical oceanography forecast systems~~ estimate the impact of an observing system from the difference in the errors made by each experiment (Fujii et al., 2019). Here the objective was to ~~search for~~ determine the optimal observations to estimate the set of invariant fundamental parameters of the model. This study provides insights for implementing such observations, based on the definition of oceanic regions using only four variables: the depth-averaged temperature, a thermal stratification index, the surface current velocity norm and a bloom index. Experiments that were conducted in these regions with random sampling or based on realistic existing networks have shown that the quality of the MLE for the energy transfer efficiency coefficients is mainly linked to environmental conditions. We found that observations from warm temperature regions (such as temperate or tropical regions) were more effective than those from cold regions. The presence of a bloom at the location of observation also improves the performance of the estimation (especially in warm environment). Conversely, high temperature stratification and high intensity of currents are both found to deteriorate the estimate. Thus, an optimal combination of environmental factors is found at a global scale for productive, warm and moderately stratified waters, with weak dynamics, such as the eastern side of the tropical Oceans. In terms of estimation performance, some functional groups are more affected by the regime variable than other. This is the case for the migrant groups that are very sensitive to the stratification or the bloom index regime for instance. However, no systematic differences between the different groups are noted. The main limitation in this study is certainly the absence of realistic modelling of the different sources of errors: the error between the modelled and the true state of the ocean have been modelled with a white noise perturbation that does not allow for spatially inhomogeneous errors. And the observations have been assumed to be directly proportional to biomass. The absence of a realistic observation model converting the acoustic signal into biomass (Jech et al., 2015) prevents to account for the different types of observation errors. Future studies should include these missing components. An interpretation of the results in ~~term~~ terms of balance between characteristic times of biological and physical processes has been proposed, pointing out a mathematical problem of observability. Hopefully this study will help in the next development of observing networks for micronekton and more generally will provide a useful methodology for future research aiming at investigating the influence of environmental conditions on the observability of some parameters. In any cases, we believe it is a next step in the ~~modeling~~ modelling of mid-trophic ecosystems and its implications

Appendix A: SEAPODYM-MTL underlying equations

SEAPODYM-MTL is based on a system of advection-diffusion-reaction equations for each functional group i , $i \in \{1, 6\}$, $i \in \{1 \dots 6\}$, involving two state variables: the potential production P_i (expressed in gram of wet weight by squared meters by day, gWWm⁻²d⁻¹) and the biomass B_i (expressed in gramm of wet weight by squared meters, gWWm⁻²):

$$\frac{\partial B_i}{\partial t} = - \left(\frac{\partial}{\partial x}(uB_i) + \frac{\partial}{\partial y}(vB_i) \right) + D \left(\frac{\partial^2 B_i}{\partial x^2} + \frac{\partial^2 B_i}{\partial y^2} \right) - \lambda(T)B_i + P_i(\tau_r(T)), \quad (\text{A1})$$

$$\frac{\partial P_i}{\partial t} = - \left(\frac{\partial}{\partial x}(uP_i) + \frac{\partial}{\partial y}(vP_i) \right) + D \left(\frac{\partial^2 P_i}{\partial x^2} + \frac{\partial^2 P_i}{\partial y^2} \right) - \frac{\partial P_i}{\partial \tau}, \quad (\text{A2})$$

where x, y, t and τ are the variables for space, time and age respectively. u, v (ms⁻¹) and T (°C) are the currents velocities and temperature respectively. These variables are integrated over each layer K , $K \in \{1, 3\}$, $K \in \{1 \dots 3\}$ and weighted by the time each functional group i spends in the layer. D is the diffusion coefficient accounting for both the physical diffusion and the ability of micronekton organisms to swim short distances. τ_r (days) is the recruitment coefficient corresponding to the age for which the potential production converts into biomass of micronekton. λ (days⁻¹) is the mortality coefficient which accounts for natural mortality. Note that these two last parameters depend on the temperature.

The initial conditions for this system are :

$$B_i(t=0) = B_0, \quad P_i(t=0) = P_0, \quad (\text{A3})$$

$$P_i(\tau=0) = cE'_i E_{pp} \int_{z_3}^0 PP dz, \quad (\text{A4})$$

where B_0 and P_0 are obtained by spinup, PP (in milimol of carbon per cubic meters per day, mmolCm⁻³d⁻¹) is the net primary production, E_{pp} (adimensional) is the total energy transfer from the primary production to the mid-trophic level, E'_i (adimensional) is the distribution of this energy into the different functional groups, c is the conversion coefficient between mmolC and gWW and $z_3 = \min(10.5 \times z_{eu}, 1000)$, z_{eu} the euphotic depth (in meters).

Following Lehodey et al. (2010), the recruitment and mortality coefficients are parameterized as:

$$\tau_r(T) = \tau_{r0} e^{\tau_{rc} T} \quad (\text{A5})$$

with $\tau_{r0} = 527$ days and $\tau_{rc} = -0.125$ days⁻¹.

$$\lambda(T) = \lambda_0 e^{\lambda_c T} \quad (\text{A6})$$

with $\lambda_0 = 5 \times 10^{-4} \text{ days}^{-1}$ and $\lambda_c = 0.125 \text{ days}^{-1}$

Note that these coefficient are also defined for negative temperature values.

A module estimates SEAPODYM-MTL parameters by a variational data assimilation method : a Maximum Likelihood
520 Estimation (MLE) (Senina et al., 2008). This method minimizes a cost function (the likelihood) that measures the distance
between the biomass predicted by the model and the observed biomass. As the model outputs and the observations are not
directly comparable, they are transformed with an observation model operator \mathcal{H} . \mathcal{H} is defined for each layer K as:

$$\mathcal{H} : B \mapsto \rho_{K,\omega} = \frac{\sum_{i|K(i,\omega)=K} B_i}{\sum_{i=1}^6 B_i} \quad (\text{A7})$$

where $K(i,\omega)$ denotes the layer that the functional group number i occupies at the time of the day ω . \mathcal{H} gives for each layer
525 the relative amount of biomass that we call *ratio* (Lehodey et al., 2015).

The gradient of the likelihood function is computed using the adjoint state method. The parameters are then estimated using
a quasi-Newton algorithm implemented by the Automatic Differentiation Model Builder (ADMB) algorithm (Fournier et al.,
2012). SEAPODYM-MTL and the exact formulation of the cost function are described in detail in Lehodey et al. (2015).

Author contributions. All authors contributed to the design of the study. AD developed the method, conducted the experiments, analyzed
530 the results and wrote the original manuscript. AC and OT contributed to the development of the parameter estimation component of
SEAPODYM-MTL. OT prepared the forcing fields and contributed to the revision of the manuscript. PL coordinated the AtlantOS activity
at CLS and contributed to the analysis of results and the revision of the manuscript.

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

Acknowledgements. This work has been supported by the European Union's Horizon 2020 research and innovation project AtlantOS
535 (633211). The authors thank the Groupe Mission Mercator Coriolis (Mercator Ocean) for providing the ocean general circulation model
FREEGLORYS2V4 simulation and Jacques Stum and Benoit Tranchant at Collecte Localisation Satellite for processing satellite pri-
mary production and ocean reanalysis data. We also thank Bernard Bourlès and Jérémie Habasque from the Institut de Recherche pour
le Développement and Sophie Fielding from the British Antarctic Survey for making the PIRATA (<http://www.brest.ird.fr/pirata/pirata>)
and BAS (<https://www.bas.ac.uk/project/poets-wcb>) cruise trajectories available. The authors are also grateful to Susanna Michael and two
540 anonymous reviewers whose comments and suggestions helped improving the manuscript.

References

- Ariza, A., Garijo, J., Landeira, J., Bordes, F., and Hernández-León, S.: Migrant biomass and respiratory carbon flux by zooplankton and micronekton in the subtropical northeast Atlantic Ocean (Canary Islands), *Progress in Oceanography*, 134, 330–342, 2015.
- Arnold, C. P. and Dey, C. H.: Observing-systems simulation experiments: Past, present, and future, *Bulletin of the American Meteorological Society*, 67, 687–695, 1986.
- 545 Barnier, B., Madec, G., Penduff, T., Molines, J.-M., Tréguier, A.-M., Le Sommer, J., Beckmann, A., Biastoch, A., Böning, C. W., Dengg, J., Derval, C., Durand, E., Gulev, S., Rémy, E., Talandier, C., Theetten, S., Maltrud, M. E., McClean, J., and De Cuevas, B.: Impact of partial steps and momentum advection schemes in a global ocean circulation model at eddy-permitting resolution, *Ocean Dyn.*, 56, 543–567, <https://doi.org/http://dx.doi.org/10.1007/s10236-006-0082-1>, 2006.
- 550 Behrenfeld, M. and Falkowski, P.: Photosynthetic rates derived from satellite-based chlorophyll concentration, *Limnology and Oceanography*, 42, 1–20, 1997.
- Benoit-Bird, K., Au, W., and Wisdom, D.: Nocturnal light and lunar cycle effects on diel migration of micronekton, *Limnology and Oceanography*, 54, 1789–1800, 2009.
- Davison, P.: The specific gravity of mesopelagic fish from the northeastern pacific ocean and its implications for acoustic backscatter., *Journal of Marine Sciences*, 68, 2064–2074, 2011.
- 555 Davison, P., Checkley Jr, D., Koslow, J., and Barlow, J.: Carbon export mediated by mesopelagic fishes in the northeast Pacific Ocean, *Progress in Oceanography*, 116, 14–30, 2013.
- Davison, P. C., Koslow, J. A., and Kloser, R. J.: Acoustic biomass estimation of mesopelagic fish: backscattering from individuals, populations, and communities, *ICES Journal of Marine Science*, 72, 1413–1424, 2015.
- 560 Emery, W. J.: Global water masses: summary and review, *Oceanologica acta*, 9, 383–391, 1986.
- Errico, R. M.: What is an adjoint model?, *Bulletin of the American Meteorological Society*, 78, 2577–2592, 1997.
- Fay, A. and McKinley, G.: Global open-ocean biomes: mean and temporal variability, *Earth System Science Data*, 6, 273–284, 2014.
- Ferry, N., Parent, L., Garric, G., Drevillon, M., Desportes, C., Bricaud, C., and Hernandez, F.: Scientific validation report (ScVR) for reprocessed analysis and reanalysis., 2012.
- 565 Fieux, M. and Webster, F.: The planetary ocean, *Current natural sciences*, EDP sciences, 2017.
- Foltz, G. R., Brandt, P., Richter, I., Rodríguez-Fonseca, B., Hernandez, F., Dengler, M., Rodrigues, R. R., Schmidt, J. O., Yu, L., Lefevre, N., et al.: The tropical atlantic observing system, *Frontiers in Marine Science*, 6, 2019.
- Fournier, D. A., Skaug, H. J., Ancheta, J., Ianelli, J., Magnusson, A., Maunder, M. N., Nielsen, A., and Sibert, J.: AD Model Builder: using automatic differentiation for statistical inference of highly parameterized complex nonlinear models, *Optimization Methods and Software*, 27, 233–249, 2012.
- 570 Fujii, Y., Remy, E., Zuo, H., Oke, P. R., Halliwell, G. R., Gasparin, F., Benkiran, M., Loose, N., Cummings, J., Xie, J., et al.: Observing system evaluation based on ocean data assimilation and prediction systems: on-going challenges and future vision for designing/supporting ocean observational networks, *Frontiers in Marine Science*, 6, 417, 2019.
- Giering, S., Sanders, R., Lampitt, R., Anderson, T., Tamburini, C., and Boutif, M.: Reconciliation of the carbon budget in the ocean's twilight zone., *Nature*, 507, 480–483, 2014.
- 575 Gjosaeter, J. and Kawaguchi, K.: A review of the world resources of mesopelagic fishes., *Food Agriculture Org*, pp. 193–199, 1980.

- Handegard, N., Du Buisson, L., Brehmer, P., Chalmers, S., De Robertis, A., Huse, G., and Kloser, R.: Acoustic estimates of mesopelagic fish: as clear as day and night?, *Journal of Marine Sciences*, 66, 1310–1317, 2009.
- Handegard, N., Du Buisson, L., Brehmer, P., Chalmers, S., De Robertis, A., Huse, G., and Kloser, R.: Towards an acoustic-based coupled observation and modelling system for monitoring and predicting ecosystem dynamics of the open ocean., *Fish and Fisheries*, 2012.
- 580 Henson, S. A. and Thomas, A. C.: Interannual variability in timing of bloom initiation in the California Current System, *Journal of Geophysical Research: Oceans*, 112, 2007.
- Hoffman, R. N. and Atlas, R.: Future observing system simulation experiments, *Bulletin of the American Meteorological Society*, 97, 1601–1616, 2016.
- 585 Irigoien, X.: Large mesopelagic fishes biomass and trophic efficiency in the open ocean., *Nature Communication*, 5, 2014.
- Jain, A. K., Murty, M. N., and Flynn, P. J.: Data clustering: a review, *ACM computing surveys (CSUR)*, 31, 264–323, 1999.
- Jech, J. M., Horne, J. K., Chu, D., Demer, D. A., Francis, D. T., Gorska, N., Jones, B., Lavery, A. C., Stanton, T. K., Macaulay, G. J., et al.: Comparisons among ten models of acoustic backscattering used in aquatic ecosystem research, *The Journal of the Acoustical Society of America*, 138, 3742–3764, 2015.
- 590 Kaartvedt, S., Staby, A., and Aksnes, D. L.: Efficient trawl avoidance by mesopelagic fishes causes large underestimation of their biomass, *Marine Ecology Progress Series*, 456, 1–6, 2012.
- Kanungo, T., Mount, D. M., Netanyahu, N. S., Piatko, C. D., Silverman, R., and Wu, A. Y.: An efficient k-means clustering algorithm: Analysis and implementation, *IEEE Transactions on Pattern Analysis & Machine Intelligence*, pp. 881–892, 2002.
- Kloser, R. J., Ryan, T. E., Keith, G., and Gershwin, L.: Deep-scattering layer, gas-bladder density, and size estimates using a two-frequency acoustic and optical probe, *ICES Journal of Marine Science*, 73, 2037–2048, 2016.
- 595 Kodinariya, T. M. and Makwana, P. R.: Review on determining number of Cluster in K-Means Clustering, *International Journal*, 1, 90–95, 2013.
- Lehodey, P., Andre, J.-M., Bertignac, M., Hampton, J., Stoens, A., Menkès, C., Mémery, L., and Grima, N.: Predicting skipjack tuna forage distributions in the equatorial Pacific using a coupled dynamical bio-geochemical model, *Fisheries Oceanography*, 7, 317–325, 1998.
- 600 Lehodey, P., Sennina, I., and Murtugudde: A spatial ecosystem and population dynamics model - modeling of tuna and tuna-like population., *Progress in Oceanography*, 78, 304–318, 2008.
- Lehodey, P., Murtugudde, R., and Senina, I.: Bridging the gap from ocean models to population dynamics of large marine predators : A model of mid-trophic functional groups., *Progress in Oceanography*, 84, 69–84, 2010.
- Lehodey, P., Conchon, A., Senina, I., Domokos, R., Calmettes, B., Jouano, J., Hernandez, O., and Kloser, R.: Optimization of a micronekton model with acoustic data, *Journal of Marine Science*, 2015.
- 605 Lellouche, J.-M., Galloudec, O. L., Dré villon, M., Régnier, C., Greiner, E., Garric, G., and Ferry, N.: Evaluation of real time and future global monitoring and forecasting systems at Mercator Océan., *Ocean Science Discussions*, 2012.
- Longhurst, A.: Seasonal cycles of pelagic production and consumption, *Progress in oceanography*, 36, 77–167, 1995.
- Proud, R., Cox, M. J., and Brierley, A. S.: Biogeography of the global ocean’s mesopelagic zone, *Current Biology*, 27, 113–119, 2017.
- 610 Proud, R., Handegard, N. O., Kloser, R. J., Cox, M. J., and Brierley, A. S.: From siphonophores to deep scattering layers: uncertainty ranges for the estimation of global mesopelagic fish biomass, *ICES Journal of Marine science*, 76, 718–733, 2018.
- Senina, I., Silbert, J., and Lehodey, P.: Parameter estimation for basin-scale ecosystem-linked population models of large pelagic predators : Application to skipjack tuna., *Progress in Oceanography*, 2008.

- 615 Siegel, D., Doney, S., and Yoder, J.: The North Atlantic spring phytoplankton bloom and Sverdrup's critical depth hypothesis, *science*, 296, 730–733, 2002.
- Silverman, B. W.: *Density estimation for statistics and data analysis*, Routledge, 2018.
- Spalding, M. D., Agostini, V. N., Rice, J., and Grant, S. M.: Pelagic provinces of the world: a biogeographic classification of the world's surface pelagic waters, *Ocean & Coastal Management*, 60, 19–30, 2012.
- 620 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, *Frontiers in Marine Science*, 3, 31, 2016.
- Sutton, T. T., Clark, M. R., Dunn, D. C., Halpin, P. N., Rogers, A. D., Guinotte, J., Bograd, S. J., Angel, M. V., Perez, J. A. A., Wishner, K., et al.: A global biogeographic classification of the mesopelagic zone, *Deep Sea Research Part I: Oceanographic Research Papers*, 126, 85–102, 2017.
- 625 Tibshirani, R., Walther, G., and Hastie, T.: Estimating the number of clusters in a data set via the gap statistic, *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 63, 411–423, 2001.
- Volk, T. and Hoffert, M. I.: Ocean carbon pumps: Analysis of relative strengths and efficiencies in ocean-driven atmospheric CO₂ changes, *The carbon cycle and atmospheric CO₂: natural variations Archean to present*, 32, 99–110, 1985.
- Zaret, T. and Suffern, J.: Vertical migration in zooplankton as a predator avoidance mechanism, *Limnology and Oceanography*, 21, 804–816, 1976.