This document contains a point-by-point response to the reviewer 2. The original reviewer comments are in black and our response is written below each main point in blue.

**Anonymous Referee #2**

Received and published: 8 November 2019

# General comments

Unfortunately, it is very difficult to follow parts of the manuscript because of quality of the English language. Normally I would add this to the end of my comments but in this instance, language is the main issue that I have with the manuscript – and a reason why it took me considerably longer to read through it. Many sentences are clumsily or sometimes incorrectly formulated and in certain paragraphs I had to guess the ideas the authors were trying to express. I have included a lot of corrections and questions about language in the specific comments below but I would recommend that the authors go through the manuscript carefully with the help of a native English speaker to make the manuscript more readable and easier to understand.

We thank the reviewer for his careful reading and his useful rephrasing suggestions. We took into account his many comments and hopefully improved the language quality of the manuscript. We also took advices from a native English speaker, who went through the manuscript and corrected the wrong formulations.

What I found missing in the methods section is a good description of the underlying physical model: Neither its vertical nor horizontal resolution is mentioned, it is also not clear how deep the resulting layers z_1, z_2 and z_3 are. In general, it would help to explicitly mention that SEAPODYM-MTL currents, temperature etc are based on an underlying global physical model (at least I assume so, based on l 91). At the moment this is not done and many readers may be confused how the physical transport and stratification are simulated in a model with only 3 layers.

Following the reviewer suggestions, we now give detailed precisions about the underlying physical model in section 2.1.

- We specify the physical model horizontal and vertical resolution: “ORCA025 configuration (eddy-permitting grid with 0.25° horizontal resolution and 75 vertical levels, see Barnier et al., 2006)” (lines 97-98).

- We give an approximate averaged depth for the layers: “These boundaries are defined as follows (an approximate averaged depth is given in brackets): $z\_1(x,y,t) = 1.5 \times z\_eu(x,y,t)$ (50-100 m), $z\_2(x,y,t) = 4.5 \times z\_eu(x,y,t)$ (250-300 m), $z\_3(x,y,t) = \min(10.5 \times z\_eu(x,y,t),1000)$ (400-700 m)” (lines 84-85).

- We also include a description and references for the underlying physical model: “ [forcing fields] come from the ocean dynamical simulation FREEGLORYS2V4 produced by Mercator-Ocean. FREEGLORYS is a global, non-assimilated 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 NEMO 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 Era-Interim atmospheric reanalysis from the ECMWF." (lines 93-97).

Further details and modifications are also given in the response to specific comments below.

The authors manage to identify certain prevalent ocean conditions (regimes), which are more suitable for parameter estimation in their twin experiment setup. Then they take a leap and state in multiple places throughout the manuscript that these regimes would therefore be better suited for parameter estimation outside the context of twin experiments. I am a bit skeptical about this claim, because the model's ability to simulate the ocean conditions may also be regime-dependent. There may be regimes where the model does not do a good job at simulating the ocean and model parameters do not reflect the actual energy transfer efficiency, while the model may be better suited for other regimes. This strikes me especially true for the relatively simple 3-layer model that is used in the study. Without knowledge of model error it seems difficult to make the claim that certain regions are better suited for parameter estimation than others.

This remark is completely fair. Actually, we do take into account model errors in our experiments but not inhomogeneous model errors.

First, as the reviewer 1 noted, the use of the terminology "twin experiment" was misleading in our case. Indeed, the nature run used to generate the synthetic observations has not the same forcing fields as the control run used to perform the estimation, but a "model error" was introduced (see Fig. 1 in the revised version). We then do not use the expression "twin experiment" anymore.

Second, our control run is forced with the reference forcing fields plus a perturbation that is supposed to account for model error. However, we chose a white noise perturbation that indeed does not take into account any space-dependant or regime-dependant errors. We are conscious of the implications of this choice, which are discussed in section 4.3.

In response to the reviewer comment, we develop this part of section 4.3 emphasizing the limitation of a white noise perturbation by adding:
- Lines 408-409: "The realism of this approach is questionable, as it does not take into account the possible spatial distribution of uncertainty and errors of ocean models. "
- Lines 411-412: "Indeed, we expect forcing fields to be less accurate where the ocean has strong variability."
- Lines 412-413: "However, for the purpose of our study, a spatial homogeneous error was preferable, to avoid introducing any bias."
And we also mention it in the conclusion, in response to reviewer 1 specific comments:
- Lines 450-454: "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, 2015) prevents to account for the different types of observation errors. Future studies

should include these missing components. ”

Modelling a “real” error is first quite complicated and second if we introduce a regime-dependent error, then we would not be able to highlight the main result of this study, which is the regime dependence to the performance of the estimation.

An attempt in modelling an inhomogeneous error is to take an error proportional to the deviation of a field to its climatology, as we also explain it in section 4.3. This approach has in fact been tested; it was the subject of previous studies:

- Delpech, Audrey (2017). *Sensitivity study of SEAPODYM parameters to physical and biogeochemical forcing fields in the framework of Observing System Simulation Experiment*. Master Thesis from the Ecole Nationale Supérieure de Techniques Avancées (Available upon request).

- Lehodey, Patrick, Titaud, Olivier, Delpech, Audrey and Conchon, Anna (2018) *Optimal design of ecosystem module*. AtlantOS Deliverable, D5.5. AtlantOS, 29 pp. DOI 10.3289/atlantos_d5.5.
(Available at : https://www.atlantos-h2020.eu/download/deliverables/AtlantOS_D5.5.pdf)

Section 3.2 (linear perturbation) and 4.1.2 of the AtlantOS report cited above show how our results are robust to the introducing of an inhomogeneous error on the forcing fields. In particular, figure 6 of this same report shows that, even if the perturbation is higher in the tropical regions (warm regime), these regions remain the best locations for parameters estimation.

Note that this study was conducted in a slightly different framework since a different physical simulation was used, and the synthetic observations were taken on the tracks of real ship transects. The results can however be interpreted in a similar fashion.

# Specific comments

l 5: "migrant and non-migrant micronekton": Does the "migrant" refer to DMV? It would be good to be explicit here.
Yes it does. We now specify “vertically migrant” and we added “DVM” in brackets.

l 15: The "all" is too general.
We removed the “all” and give concrete examples of micronekton predators instead: tunas, swordfishes, turtles, seabirds and marine mammals.

l 16: "Migrations" -> "Migration"
Done.

l 16: It would be good to briefly summarize DMV.
We added (Lines 19-22): “This migration of biomass occur when organisms move up from a deep habitat during to a shallower habitat at night. DVM is generally related to a trade-off between the need for food and predator avoidance (Benoit-Bird, 2009) and seem to be triggered by sunlight (Zaret, 1976)”

l 17: Mesopelagic already implies "inhabiting the twilight zone (200-1000 m)", I would rephrase to "the mesopelagic (inhabiting the twilight zone from 200m to 1000m depth) component of micronekton"
Done.

l 25: Does "develop the datasets" mean to collect observations?
We changed for “collect observations”, “develop” was in fact for “methods and models”.

l 33: "making this component strongly underestimated" -> "leading to an underestimation bias for this group of micronekton"
This sentence does not exist anymore in the revised version (following referee 1 short comments, we shortened this paragraph).

l 34: "acoustic frequencies associated to traditional net sampling": I am not sure what this means.
The idea was that progresses are expected from the association of many different sampling technics: multi-frequency acoustics, optical measurements and traditional net samplings. We rephrased that sentence as:
“Progresses are expected in the coming years thanks to the combined use of different measurement techniques: multiple acoustic frequencies, traditional net sampling and optical techniques” (Lines 36-38).

l 35: "More accurate biomass estimates should benefit from" -> "The accuracy of biomass estimates is predicted to benefit from"
Done.

l 40: What are "target fish", I would suggest to remove "target" or rephrase.
We removed it.

l 41: Does "the functional groups" refer to the micronekton? If so, please include this information.
Yes, it does refer to micronekton. We added this information.

l 42: "The spatial dynamics of biomass in each group...": Add "In addition to DMV"
Done. We also added: “spatial *horizontal* dynamics” which was missing.

l 43: Which processes are included in "The time of development"?
The times of development include the recruitment time and the mortality. We clarified this sentence (lines 46-47 in the revised version): “The recruitment time and natural mortality of organisms …”.

l 57: Make sure to include information here about what is new in this study compared to Lehodey et al. (2015).
What is new compared to Lehodey et al. (2015) is first the goal of the study. While they intended to validate that parameter estimation method, we investigate the sensitivity of the parameter estimation to the environmental conditions by using a clustering approach. We thus perform

OSSE at a global scale and in a more realistic framework by introducing an error on the forcing field.
We now highlight it better in the text.

Lines 58-59: “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)”

Lines 62-64: “For this purpose, we use Observing System Simulation Experiments (OSSE) at a global scale. This method allows for simulating synthetic observations in places where an observing system does not exist yet, and to see how useful the synthetic observations are for the estimation. The purpose of the present study …”

l 58: "Therefore, it is useful ...": This sentence is confusing, please rephrase.
Done. We reformulated the whole paragraph, in response to your previous comment.

l 78: So more information about the physical model would be useful here. Is it truly a 3-layer model or are these subdivided into more layers? How is the model divided in the horizontal?
Some information about the physical model was indeed missing, as also noted by reviewer 1. The physical model comes from a simulation of the numerical model NEMO that have a 0.25° horizontal resolution and 75 vertical levels). The physical fields are then horizontally degraded to 1° horizontal resolution and depth-averaged in 3 vertical layers defined by the value of the euphotic depth (line 84). We give these precisions in the revised version (lines 96-101)

l 83: Why is "migrant-umeso" not abbreviated as "mumeso" and why is "meso" part of the name when they also migrate to the epipelagic? I think it would be beneficial to the reader to rethink the names. For example, given that the layers have just been introduced as z_1, z_2 and z_3, the names could include the indices of the layers they inhabit (the use of "nekton" is just a suggestion here): (1) epi -> nekton_1 (2) umeso -> nekton_2 (3) ummeso -> nekton_12 (4) lmeso -> nekton_3 (5) lmmeso -> nekton_23 (6) lhmmeso -> nekton_13
The name and abbreviations have been chosen to follow the official denomination from the CMEMS (Copernicus Marine Environment Monitoring Service). This service provides indeed a global reanalysis of mid-trophic levels, where the name and abbreviations are already “ummeso” …etc for the different functional groups. Here is the reference documentation:
http://resources.marine.copernicus.eu/documents/PUM/CMEMS-GLO-PUM-001-033.pdf.

l 90: Is this equivalent to reduction in the model resolution? Is the horizontal resolution that is used 1 degree?
We run SEAPODYM-MTL with 1°x1° forcing fields (temperature, currents velocity, primary production). All the forcing fields were previously degraded or interpolated if needed (this is particularly the case of the FREEGLORYS 0.25° ocean model).
We reformulated and better specified the model description section of the manuscript in response to your second general comment and to reviewer 1 specific comments.

l 107: It would be useful to know the height of each layer here. What if layer 1 is well stratified but deep enough that its average temperature does not differ much from T_2?

The height of each layer depends linearly on the euphotic depth, so it changes with space and time (line 84). But we now give an approximate averaged depth for each layer to get a feeling of the numbers (lines 84-85).
In practice as the first layer is generally no deeper than 50-100 m, its averaged temperature differs from the underlying layer. But even if the layer 1 would be deep and well stratified, it would not be a problem either because from the model construction, the micronekton is impacted by the vertically-averaged temperature of each layer. So, what really matters is the difference of *vertically-averaged* temperature between the different pelagic layers, regardless of the stratification in each layer.

Eq 6: '[1,N]' denotes an interval that contain real numbers, '{1,...,N}' would be the correct way to denote integers (see my comment about '[[' below). Furthermore, shouldn't the intersect of any two Gamma_k be empty? The current equation only states the weaker condition that the intersect of all Gamma_k is empty.
We indeed intended to mention an integer interval; we corrected it with "{1…N}" and leaved out the notation "[[1, N]]". The second comment is right too, this is indeed the intersection of any two Gamma_k that is empty and not only the intersection of all of them. We corrected the text accordingly.

l 124: It would not be difficult to express the results of a k-means clustering in words. I would recommend that the authors do that, so that some one without good knowledge of mathematical notation can understand the results.
We added a definition of the k-mean clustering in words: "The k-mean clustering method separate N values in a given number of cluster by minimizing the distance of each value to the mean (called the center) of each cluster." (Lines 130-131).

l 126: I don't think the double-bracket notation '[[ ]]' for integer ranges is very common (I have found it on the French wikipedia page but not the English, compare https://en.wikipedia.org/wiki/List_of_mathematical_symbols to https://fr.wikipedia.org/wiki/Table_de_symboles_math%C3%A9matiques.) I would suggest to either change it or briefly explain it.
We changed for the notation "{…}" instead.

l 124: "we explicit this dependence" -> "we make this dependence explicit"
Done

l 132: "the inverse model": What is the inverse model, this is the first time it is mentioned? Better explanation is required.
Due to reorganisation of section 2.3 in response to reviewer 1 specific comments, this mention does not appear anymore. Note that we now detail this section and added a figure (Figure 1) to further explain the OSSE configuration. The "inverse model" is the model that uses the MLE (cf. control run in Figure 1 in the revised version), but it would have been indeed more accurate in the sentence you mentioned to use "MLE" directly.

Eq 8: If alpha is constant here, I would suggest to abandon it and let gamma be uniformly distributed in '[-0.1,0.1]'.

Alpha is constant for most experiments except the ones in section 3.4.
We now mention it: “The amplitude alpha is set to 0.1 for all experiments except in section 3.4 where alpha varies” (line 163).

Eq 8: For a particularly small F, this could become negative.
Yes, this is generally not a problem (the current velocity can become negative and the current simply reverse direction, or the temperature can also become negative in extreme polar regions, as it is already the case for the native forcing fields). But we indeed ensured that the primary production is always positive.
We now mention it in the text: “For small values of F, this perturbation can induce a sign reversal of the forcing. This does not matter for the temperature or the currents velocities, primary production has however been constraint to positive values.” (Lines 163-165 in the revised version).

l 230: "The temperature shows the presence of a strong bias is". Bias with regard to what? I would suggest to change to something like: "Temperatures are different between the two configurations."
We changed as suggested, note that this paragraph has also been further simplified.

l 233: "Therefore, it seems here that the difference observed in the temperature values of the two datasets has a stronger impact on the parameter estimation than the regime of currents." So far it has been demonstrated that both temperature and velocity differ between the two experiments/configurations. What is the evidence that differences in temperature have a stronger impact?
We show that low velocities give better estimations than high velocities. We show also that high temperatures give better estimation than low temperatures. But what if we now compare a configuration of low velocities and low temperatures with a configuration of high velocities and high temperatures? This is what experiment 1’ and 1’’ are showing. The configuration with high temperatures and high velocities give better estimation than the configuration with low temperature and low velocities. So it is as if the temperature governed the performance more than the velocities. This suggests thus that differences in temperature have a stronger impact.

However, this information was slightly beyond the scope of the point we wanted to make here. We thus removed it.

l 237: It would be good to get a feel for the numbers, do significant cross-correlations occur often, what is their proportion w.r.t. the total number of experiments?
Among the 26 possible experiments, 9 had a significant cross-correlation. We now specify it in the text (lines 269-270).

l 242: Fig 5 is referenced before Fig 4.
This is not the case anymore since we combined figures 2, 4, 6 and 7 in one figure 3.

l 258: "if the mean error on the estimated parameters were higher in average, the result does not change". I am not sure what is meant here, is the mean error higher on average? Please rephrase.
We rephrased the sentence as: “The same kind of 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 leads to a better estimation than strong stratification." (Lines 289-290 in the revised version).

l 267: Do you mean "Exp. 1a and 1b" here or 4a and 4b?
We meant Exp. 4a and 4b and corrected it in the text.

l 270: "Indeed, not only the temperature is higher but also the vertical gradient of temperature.": It is not fully clear what this sentence is referring to.
We rephrased it as: "Indeed, Exp. 4d (T_4 regime) has higher temperature than Exp. 4b (T_2 regime) but it has also a higher stratification index (S_3 regime for Exp. 4d and S_1 regime for Exp. 4b)" (Lines 302-303).

Eq 10: What is E_pp?
E_pp is the total energy transfer from the primary production to the mid-trophic level, all functional groups together. Its definition is given in the Appendix A, but we also recall it below equation 10 in the revised version.

Table 3: It would be nice to add the number of samples/observable points for each configuration.
We included this information in the Table 3. We also added a note in the caption to explain that even if the number of observable point differs from one configuration to another, the experiment were conducted with a constant number of observation points (400 synthetic observations).

Fig 1: Reference Table 2 in caption.
Done

Fig 1: I understand the intent of using transparency to indicate uncertainty but I doubt it is done correctly here. It appears like the colors are plotted on top each other in a predefined order, so even light orange colors in (b) will appear more orange than blue because the orange is plotted on top of the blue.
The idea was that even if orange is plotted on top of the blue, transparency would make the resulting colour neither orange nor blue but something in between. However, given the resolution of the figure and the very few point that are concerned, this information may be not relevant. We removed it.

Fig 2,4,6,7 can be combined into one, which would allow a nice comparison between the different configurations.
We agree and it is done so in the revised version (Figure 3).