Dear Topical Editor,

We would like to thank the Editor, Referee #1 and Referee #2 for handling and reviewing our manuscript.

We responded to the referees' comments point-by-point in "Reply to the referee #1 and Reply to the referee #2". In the revised manuscript, we improved the figures and descriptions. We hope that this revision will be found acceptable for your journal.

Sincerely,
Yasuhiro Hoshiba and co-authors

Reply to the referee #1: We express our appreciation to the referee for the careful reading of, and the comments to, our paper. The below are our responses to the comments. The referee's comments are expressed in italic style and our corresponded replies are in regular style.

Having look the revised Fig. 3 and associated results, I found the parameter estimation by genetic algorithm gives a reasonable result and the manuscript is now ready for scientific discussion and review. As a whole, the manuscript was substantially improved in comparison to the one for the first submission, while there are some important points, for which more explanations/clarifications are necessary; many figures need to be improved; some part of the text should be rewritten for readability. For these reasons another round of revision is necessary.

We would like to appreciate your time to review our previous MS. According to your comments, we further revised our manuscript (MS) as detailed below:

Specific comments

1) Introduction gives concise and good summary for previous work, and points out the necessity of the current work.

Thank you.

2) line 56. ".. classical Michaelis-Menten equation .." needs reference.

The reference was added in Line 56-57 in the revised MS.

3) line 69-71. Again I have a question/concern about the use of physical field obtained from data assimilation, particularly since a 3D-Var system is used in this study. In reply to referee #1, the authors wrote "The physical field used in the offline ecosystem model (NSI-MEM) does not satisfy the mass conservation, but the passive tracers of the NSI-MEM in the offline setting do not have artificial sink/source without other boundary forcing and correction terms". I don't really understand how the authors achieve such a set-up, since the conservation of passive tracers depends on the conservation of the physical field (i.e., volume). The passive tracers in LTL models are generally represented by concentration in a grid cell. If there is an increment of SSH by assimilation, the volume of the corresponding cell changes, which automatically leads to change *of passive tracer amount (if you don't change the concentration). How did the authors handle this issue? If the authors implemented a scheme which preserves amount of passive tracers even with the volume change, then the concentration of passive tracers undergoes artificial change. How much impact do you see in this case? I'm asking this question because the observed and modeled vertical section of T (Fig. 7) exhibits difference around the inter-gyre boundary (approx. 40 degree N), implying not a small amount of SSH increment might be added in this frontal zone. I think this is an important point to be checked and discussed somewhere in the manuscript before going LTL model analyses, since a use of assimilated physical field for ecosystem modeling is (probably) one of the way to go in the future.*

We meant that temperature and salinity were not conserved by the 3D-Var system, but SSH and ecosystem tracers was actually conserved. Because the 3D-Var system used in this study only changes the temperature and salinity, and does not directly change the SSH and velocity (Fujii and Kamachi, 2003). Therefore, the amount of water mass is also conserved. We added the information at Line 72-74 in the revised MS.

4) line 82. The authors describe "iron supply was only from the dust in the model setting", while in introduction they pointed out that "The source of iron for the WNP region is not only from air-born dust but also from iron transported in the intermediate water from the Sea of Okhotsk to the Oyashio region" (line 31-33), which seems to me contradicting. A justification for the model setup or discussion on the effect of missing iron source is necessary.

Our model is not perfect and has caveats. We added a discussion on the effect of missing iron source to Line 295-298 in the revised MS.

5) line 95-102. The authors conduct a model run with SST-dependent physiological parameters, while at the same time they also stated physiological parameters (may) change with other conditions e.g., nutrient abundance, light intensity (line 303-). An explanation is necessary here, why the authors chose SST field to smooth the transition between the two assimilated stations.

The sentence (L303) quoted by the Referee was described in the context of phytoplankton physiology. On the other hand, physiological parameters estimated using the SST gradient include those of not only phytoplankton but also zooplankton. We chose SST because it directly affects physiological parameters of both phytoplankton and zooplankton whereas nutrients and light are essentially related to phytoplankton only. We added this philosophy to Line 101-104 in the revised one.

6) line 107. "The parameters values" --> "The parameter values".

Thank you. It was amended.

7) line 107-108. Why the authors selected the specific year 1998? I asked this question in my first review but the authors did not provide reasonable answer. I'm asking this question because a parameter estimation by multi-year condition gives more reliable result. If you use the data from short period, the estimated parameters may be deformed so as to make best much for specific condition. In such a case you lose generality of the estimated parameters and difficult to apply them for interpretation of plankton physiology.

Limited computer resource was the largest reason why we chose the single year (1998); the 3D and μ-GA simulations are computationally costly (e.g. the μ-GA approach requires 2,000 "generations" to converge parameter values). We partly agree with Referee in that the generality of the estimated parameters may be increased when multi-year data is used for the parameter optimization, although such a treatment may require acclimation/adoption of organism to be well explained in the model. We incorporated this discussion in Line 311-319 in the revised MS.

8) *line 167-184. Now I found the result is reasonable and the GA optimization works well.*

Thank you.

9) line 203-209. What can we learn from this analysis? It looks to me just giving a duplicated information with Fig. 9. If the authors intend to keep this part, more explanation is needed for the purpose and necessity of this analysis.

In response to a comment from the other referee, Figure 5 was added, to show a broader horizontal assessment of our results in addition to the point assessments at KNOT and S1 stations shown in Figure 9. We revised some sentences at Line 235-237 of the revised MS.

10) line 230. The construction of the sentence seems strange.

The sentence was changed together with 9) above.

11) line 233. "maximum" --> "the maximum".

The word was eliminated, due to the revision below.

12) line 229-249. Some sentences are tedious and preventing easy-reading. For instance, line 233-236 "At St. S1, the timing of (the) maximum phytoplankton ... " can be shorten as "At St. S1, OPT case reasonably captures the timing of the phytoplankton bloom, although the amplitude is slightly overestimated." Readers already know that you compare CTRL and OPT in this section, therefore you don't have to repeat "compared to CTRL case" many times. In relation to the above comment, I suggest to use short abbreviations, e.g., control case --> CTRL, parameter-optimised case --> OPT, SST-dependent case --> T-OPT etc. for easy-reading.

We simplified the sentences in the revised MS.

13) line 263-265. This is one of the interesting results of this study (from my point of view), since you quantitatively showed how much difference occurs on the simulated biomass due to effect of horizontal advection, using 1D and 3D models which have exactly the same LTL model function. I suggest to briefly mention this in conclusion.

We briefly mentioned the difference on Line 327-328 in the revised MS.

14) line 263 and error bars in Fig. 8 and 10. Why the authors employ 0.3 degree range to define the uncertainty? Is this an autocorrelation scale of observed chlorophyll (Fig. 8) or ocean structure (Fig. 10)? An explanation or justification is needed (you can easily find such scale estimates in literature).

It is not an autocorrelation scale. Effects of advection by a mesoscale eddy would be expected within the range of ±0.3 degree (about six grids) in the physical field, because the radius scale and the lifetime of a general mesoscale eddy are O(100 km) and ⩾16 weeks, respectively (Chelton et al., 2011). We added the explanation at Line 268-270 in the revised MS.

15) line 286-287. I don't understand the meaning of this sentence. What is the verb of this sentence?

The sentence was revised (Line 294 in the new version).

16) line 322-325. I am skeptical to the statement in this paragraph. If we see the prescribed range (i.e., min and max) and estimated values of parameters in Table 2, many parameters go to its upper or lower prescribed bounds, indicating the optimization result (i.e., optimized parameter set) is strongly constrained by the prescribed bounds. In other words, the GA optimization did not search the entire parameter space freely due to the bounds. This means that the consistency between the physiological parameters and those obtained from in-situ observation (line 323) is not necessarily guaranteed by the current experiment setup (the consistency is already imposed in the experiment design). If the authors really intend to confirm the consistency, the prescribed ranges should be widen beyond the current ranges, and see whether the parameters still stay within the range of in-situ estimated values. From this point of view, I suggest to revise the concluding sentence.

Thank you for the useful comments. As suggested, we revised the concluding sentence.

17) The quality of the figures is very poor and is not suitable for publication (except Fig. 2). They should be totally redrawn.

The figures were redrawn.

-- *Fig. 3. Use a log-scale for the ordinate, otherwise it is difficult to distinguish PL lines in panel (b).*

We have tried it before, but the variations of PS lines became hard to read. As PL in the panel (b) is close to zero (i.e. PL is almost absent), we decided to keep the ordinate in linear scale. We added the information (i.e., PL is almost zero) at Line 190-191 in the revised MS.

-- *Fig. 6, 7, 10 and 11. Use the same vertical range for consistency*

-- *Fig. 3, 8, 10, 11. Apply a consistent manner for color and line type in these figures, e.g., dotted-line for 1D case, solid-line for 3D case; red line for CTRL, blue line for OPT, green line for T-OPT, black line for OBS. etc. The different rules between different figures makes readers confused.*

-- *Fig 8 and 10. Use color shade (with transparency) instead of the error bars. In some cases the error bars are stacked each other and not distinguishable.*

-- *Fig 11. "(/day)" --> "[day^{-1}]".*

-- *Fig. 3, 5, 6, 7, 8, 10, 12. Divisions and labels for abscissa, ordinate and color bars should be reconsidered for easy-reading.*

-- *Fig. 3, 8, 9, 10, 11. Put appropriate legend for line type and color with examples, in empty space in panels.*

-- *Fig. 1. Use the same panel size for (a), (b) and (c), since the practical information in each panel is nearly the same.*

Reply to the referee #2: We express our appreciation to Referee #2 for constructive and positive comments to our manuscript. Below are our response to the comments. The original referee's comments are written in italic style and our corresponding replies are in regular style.

I have read the revised version of the Hoshiba et al manuscript as well as their answers to my criticisms. The authors significantly improve the manuscript: they add some text that clarify a lot of issues, produce new figures in order to reinforce the performances of the model with parameters estimations and most importantly provide a new version of Figure 3.
Even if I am still not fully convinced by the significance of the improvement of the data assimilation experiment on the quality of 3D model results (I would have liked to see more error skills like bias, RMS), I think that the authors did a lot of efforts for improving their work and answering our criticisms. I still have some comments.

We would like to appreciate your understanding of revision of our manuscript (MS).

Specific comments

1) Line 220: what do you mean by effective nutrient? Limiting nutrient?

Yes, we meant limiting nutrient. We clarified it in the revised MS.

2) Line 271: please clarify what is the mean iron growth rate and give the units. Iron uptake rate? So, the production increases because you change the parameters describing the iron limitation which results that your phytoplankton is not anymore limted by iron. How realistic are your calibrated parameters? (I am not an iron expert and so I cannot realize if the iron uptake efficiency is overestimated or not)
Please clarify how the parameters are set at the interface between the "KNOT" and S1 regions? Are you imposing a sharp gradient or are you using some smoothing? From the figures it seems that it is the first option and this results in sharp transition of the simulated biogeochemical fields (e.g. figure 6c). You mention in section 3.5 that the version of the parameters that is smoothed with SST does gives worse results compared to the other one. DO you know why?

In the present context, the word "the mean iron growth rate" was used to express an average value of the growth rate limited by iron. The limited growth rate (molN/m^3/day) was standardized for phytoplankton group (i.e. it was mathematically divided by PS or PL biomass (molN/m^3)), hence its unit is day^{-1}. The limited growth rate is a limitation function of growth rate by either nitrogen, silicate or dissolved iron. The smallest rate among them indicates the largest limitation to the rate of phytoplankton's photosynthesis. We added the unit and the information at Line 275-281 in the revised MS.

It is difficult for us, all modelers to know how realistic the parameters used in ecosystem models are. So, we conducted the parameter estimation. We improved the expression about the future task of the estimated parameters in Conclusions in revised MS.

We imposed a sharp gap between the subarctic (KNOT) and subtropical (S1) regions in Parameter-optimised case, and a smoothed boundary in the SST-dependent case. In the performance, the SST-dependent case was slightly worse than the parameter-optimised at the grids of KNOT and S1, potentially due to advection and diffusion from the grids surrounding. However, the SST-dependent method is useful for eliminating the boundary gap and improving performances in some regions, as in Section 3.5.

3) Figure 10: 1D model results are not represented here in Figure 10 b and 10c.

We show 1D model results only in Figure 10 (a), because our main focus is phytoplankton's biomass and variation. We revised the caption of Figure 10.

*4) Figure 11: Please clarify what is represented in this Figure: maximum growth rate * limitation function by either nitrogen, silicate and dissolved iron? Uptake rate of nitrogen, iron?*

These represent the growth rates limited by nitrogen, silicate and iron. The smallest rate among the three rates most limits phytoplankton's photosynthesis. The growth rate limited by iron is smallest in this situation, therefore it dominates the rate of photosynthesis. We added the information to the caption of Figure 11 and at Line 275-281 in the revised MS.