Response to the reviewers:

Reviewer #1

Major

(1) *RC: My major criticism is that, in general, the model appears to show major discrepancies with the data, undermining the credibility of the whole exercise, including the conclusions. To be effective, the default model run should show reasonable correspondence with the data but, in several instances, it appears not to do so. Just because the MEDUSA model is already parameterised and published in this regard does not save the situation here because the work involved changing the parameterisations of sinking, maximum and grazing rates (that's rather a lot; page 6, line 7). For example, I am not convinced about the new parameterisation of sinking, namely a sinking rate of 0.1 m d-1 (page 6, line 17) which seems much too low. At PAP, the blues stars (default run) are way too high relative to the blue crosses (observations) indicating a major discrepancy for chlorophyll (Figure 4). The average chlorophyll values for the oligotrophic stations look ok, but the depth plots do not look good at all in this respect (the deep chlorophyll maxima look poorly reproduced; Figure 6). I need more convincing that the model is credible at these sites. There also seem to be large discrepancies for L4 (Figure 4). The modelled vertical concentrations of nitrate at PAP look way too high compared to the data (Figure 3). Why have box and whisker plots not been produced for nitrate, comparing model and data? And why does the appendix (supplementary material) focus only on chlorophyll, and not nitrate? Overall, I am left in doubt as to whether the model, as parameterised for the default run, is credible. The authors could help the situation by looking at some other metrics, if only for the default run. For example, what is predicted primary production at the different sites and how does this compare with data (even just comparing annual average would be highly useful)?*

AR: We agree with the reviewer that the default model does not represent the observations convincingly in many of the stations. However one of the objectives of this study was to see how far we can improve the default MEDUSA through structural perturbations in a consistent 1D set up across all stations and so we wanted to keep the model parameters unchanged or as similar as possible at every station. We changed one or two parameters of the default parameters from the literature to allow the default 1D run to be a compromise across all stations, before applying the ensemble. In particular we used 0.8 day $^{-1}$, and 0.5 day^{-1} for maximum uptake rate and zooplankton grazing respectively, similar to HadOCC model; A lower sinking rate of 0.1 m d^{-1} was needed at the coastal stations to prevent the

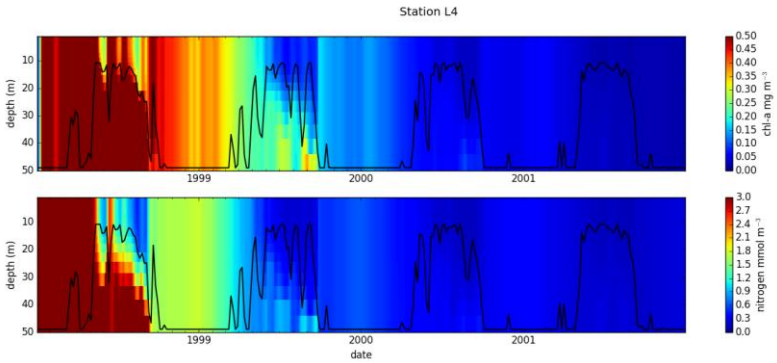

Figure 1. Chlorophyll and nitrogen concentration in the water column at station L4, when sinking rate is 3 m day^{-1}

nutrients sinking too quickly and being lost, eg. Raick et al. (2006) (a study by Ward (2013) even suggested to use 0 m d^{-1} for the optimum biogeochemical model). Considering station L4 is only 50m deep, using 3 m d^{-1} (MEDUSAs original default rate) means that all nutrients are lost from the water column after 2 years (see, the figure 1).
With the original MEDUSA default parameters the model produces too low surface chlorophyll in the oligotrophic stations, but this improves (as the reviewer observed) when the new parameters are used. But of course, the deep chlorophyll maximum it is poorly reproduced using either MEDUSA's default or the modified parameters. This also applies to station L4, where the seasonal pattern is poorly reproduced. However, the default MEDUSA parameter work better for station PAP (with NRR for surface chlorophyll and profile reduced 1.02 and 1.11, but not on nitrogen, the NRR increases to 1.35) and we will now include this in the supplementary material. Our investigations with the default parameters revealed that the large discrepancies between in situ data and the default 1D run was mostly because of the physical input data, especially the vertical velocity and vertical diffusivity coefficient as these drive the upwelling of nutrients. Since these are important to give any realistic interannual variability it is harder to tune these physical inputs in any sensible way. We will emphasise these points in the revised manuscript.

For the nitrogen in station PAP, using nitrogen from the in situ as the initial condition (available from mid-2002) instead of from the test stations (described in section 2.5.2), has improve the nitrogen run and reduced the RMSE of nitrogen (from 3.16 to 2.77), and the NRR of chlorophyll (surface from 1.29 to 0.9 and profile from 1.2 to 1.07) however the nitrogen profile NRR increases (from 1.25 to 1.38). We will include this results in the supplementary material.

In the revised version the metrics for nitrogen and primary production (as suggested by the reviewer) will be included. Further, predicted primary production at stations ALOHA and CARIACO will be included, as the in situ primary production is available only at these two stations.

(2) *RC: The ensemble run at each station is initialised using in situ measurements (page6, line 31). What is needed is a stable initial condition, which will not be potentially vulnerable to initial condition instabilities. So surely what is needed is to run the first year over and over (do a spin-up) until a repeating cycle is reached, from which the run through the various years can then be undertaken.*

AR: We tried to do a spin-up run for 50 years, using first year's run and the repeating cycle of chlorophyll was achieved after 17 years of run. However, the surface nitrogen kept increasing (up to 40 mmol m^{-3}), again mainly driven by the physical model inputs, because the sum of the first year's vertical velocity is positive (upwards), continuously increasing surface nutrients with time. We decided not to use the spin up run, but instead to use in situ measurements to initialize the model. The same initialization was used for the default and ensemble run. The physical input was averaged every 5 days, controlling the biogeochemical tracers frequently. We will emphasis these points in the revised text and discuss the alternative spin up method in the supplementary material.

(3) *RC: A major conclusion of the work is (page 15, line 29) that "small perturbations in model structure can produce a wide range of results". This is a very significant conclusion and I think*

the authors can justifiably make it. For the most part, however, the results as shown in the Figures don't show this directly, because they involve various parameterisations acting simultaneously. There is plenty of text in the Results section to support their contention, focusing on individual parameters. I wonder if this conclusion could be better represented in the graphical representation of the results

AR: Thank you for suggesting the graphical representation of one of our main conclusions. We plan to show this in figure 7 and 8, and also using the box plots in figure 4 and 5. We can include a boxplot to show the range in chlorophyll annual means produced when changing only one process at a time thus better supporting the conclusions.

(4) *RC: The Introduction is generally well written, introducing the topic of model complexity nicely. The Discussion should mirror the Introduction, saying what the current study has said in context of the wider picture. Instead, the Discussion is mostly just an extended re-hash of the Results and does little to address the big picture. For example, what do the authors conclude about model sensitivity in context of complexity science and the onward drive to produce model of ever increasing complexity? A much bigger play could be made on the need ive in benefits of the ensemble analysis over previous studies that have focused more narrowly on particular parameterisations. Etc. There is plenty of scope and I would say the Discussion section needs a significant overhaul in this regard. It needs re-emphasis; a few extra lines of text will not do.*

AR: Thank you for the nice comment on the introduction, and the suggestions on the discussion. We will include these suggestions in the discussion on the revised manuscript.

Other comments:

(1) *RC: The authors articulate two types of uncertainty (page 2, line 26): "parametric, associated with the choice of parameter values; and structural, which relates to the underlying model equations". Structural uncertainty can also refer to the structure of the model itself (number of compartments, linkages, etc). This should be mentioned, stating that the authors are only looking at structural uncertainty to do with equation formulations.*

AR: Thank you for the suggestion, it will be included in the revised text with appropriate references.

(2) *RC: On page 9, line 12, there is "A selection of ensemble results are presented". A selection? On what basis?*

AR: The selection is based on the available in situ data for nitrogen and chlorophyll and some of the statistical measures we have done. We will rephrase this in the revised manuscript.

(3) *RC: Some of the text associated with the Figures is microscopically small.*

AR: Thank you for the comment, we agree that some text is too small, and we will make it larger in the revised figures, and split figure 9 into two figures to make the text clear.

(4) *RC: Be sure to cite Le Quere, not Quere without "Le".*

AR: Thank you, this will be included in the revised manuscript.

Reviewer #2

General comments:

(1) *RC: The manuscript attempts to show two aspects: (1) there is a high level of structural uncertainty in biogeochemical models and (2) the uncertainty can be exploited to better fit a range of different observations. In my opinion, the authors succeed in providing evidence for first aspect but I have doubts about the second: all comparisons of the ensemble are based on a default run that does not seem to perform very well. Other studies have shown that 1D models with the same parameter values do not perform well across multiple locations but here the same parameter values appear to be used across all stations. Have the parameters of the default run been optimized to fit the datasets used in this study? The results of the default run can have knock-on effects on the ensemble: in multiple parts of the manuscript the authors note that when there is a large bias between the model (ensemble) and the observation, that the ensemble spread is too low when really other model aspects may be to blame for the bias. In other words, problems with the parametrization, the physical model, or the 1D nature of the model cannot be explained by structural uncertainty in the biogeochemical model.*

AR: We have not formally optimised the parameter values for each stations. To allow this method to be applied in the 3D MEDUSA we kept the parameters as similar as possible at every station. Please also see response to Q1 from Reviewer 1 above.

(2) *RC: When looking at Figure 1, I noticed that the linear function in (c) provides a bad fit to the other functions and that all functions are shown on a log scale. I am wondering if a log-transformation has also been used in the function fitting exercise in Sections 2.1-2.3? If not, I would recommend that this should at least be tried as the procedure could otherwise overemphasize the fit at high tracer concentrations which may explain the slope of the linear function.*

AR: We have tried using log-transformation in the function fitting exercise, however, it does not improve the fitting - e.g., the mean absolute error between hyperbolic (the default function) and other mortality functions are larger compare to the regular nonlinear least-squares, summarised in the table 1 and figure 2 shown here. Therefore we decided to stick to a non-transformed fitting.

Table 1. Comparison between log transform and regular function fitting parameter values and its mean absolute errors.

functional form	log transform parameter	mean abs error	non log transform parameter	mean abs error
sigmoidal	k = 1.019	0.0023	k = 0.744	0.0022
linear	μ=0.085	0.0126	μ= 0.097	0.0085
quadratic	μ= 0.023	0.0035	μ= 0.050	0.0028

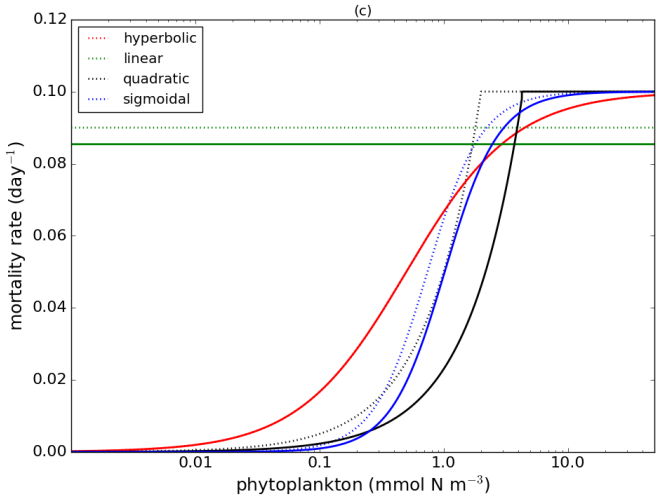

Figure 2. Mortality functional forms optimised against hyperbolic function. Dotted lines are fitting without log transform and solid lines are fitting with log transform

Specific comments:

1) *l1: "mathematical structure": What exactly does this refer to? The model formulation? I would suggest to rephrase or an improved explanation*

   AR: Yes, this means the model formulation. We will rephrase this sentence in the revised manuscript.

2) *l3: "intermediately complex BGC model" -> "BGC model of intermediate complexity"*

   AR: Thank you for the suggestion, we will revise this sentence accordingly.

3) *l9: "using phytoplankton phenology (...) and other statistical measures": phytoplankton phenology is not a statistical measure.*

   AR: What we meant in this sentence is that we are using phytoplankton phenology as well as statistical measures (such as RMSE, annual mean, and bias) in order to quantify the impact of structural sensitivity in the ensemble mean, median, and other members. We will revise this sentence in the revised manuscript.

4) *l11: Is this the range found in the ensemble (as opposed to e.g. different coastal stations)? Please make this explicit.*

   AR: This is the range found in the ensemble at the coastal stations. We will revise this sentence in the revised manuscript for clarity.

5) *l14: "the errors are mostly reduced": This is not clear: model misfit with respect to the in situ obs is smaller for the ensemble mean/median than the model with standard parameters? I suggest to rephrase.*

AR: Yes, this means the model misfit with respect to the in situ observations is smaller for the ensemble mean and median, compared to the default run (using the functional forms in MEDUSA). We will rephrase this in the revised manuscript.

6) *l15: Here a narrow spread is reported, a few lines above a "large" spread was described.*

AR: What we meant was that we do produce large spread, but not wide enough to cover the observation as measured by the NRR.

Page 2

7) *l7: This reads like the forecasting systems are having an impact on ocean biogeochemistry. The climate change aspect of the sentence reads like a repeat of sentence in line 2. Please revise for clarity.*

AR: Thank you for the suggestion, we are trying to give an example of how biogeochemical models may be applied. We will rephrase this sentence in the revised manuscript.

8) *l12: Even NPZ models represent "several" processes. Please be more precise.*

AR: Thank you, we will rephrase this sentence in the revised manuscript.

9) *l16: There can be spatial variability without iron!*

AR: We agree with this statement, we will rephrase this sentence into '…such as iron, to permit phytoplankton growth limitation regionally due to the availability of micronutrients' for clarity.

10) *l29: "only small perturbations are usually produced even with large variations in parameter values" This is a very strong statement and very much depends on what a "large variation" entails. Perhaps weaken the statement and just make the point that structural uncertainty is often larger than parametric?*

AR: Thank you for the suggestion, we will revise this sentence in the revised manuscript.

Page 3:

11) l13: *"linear density-dependent mortality produces the most significant differences when applied to diatoms": What exactly does this mean? Please revise.*

AR: We meant that the difference is more apparent, we will rephrase this sentence

12) *l18: "However, not all processes give significantly different model outputs." The next sentence seems to imply that the differences maybe due to very similar inputs, can this effect thus really be attributed to the process?*

AR: In this sentence, we were trying to give an example of how changing the equations of different processes (such as grazing, mortality, and photosynthesis) may give rise to different impacts on phytoplankton dynamics. Changing the equation for photosynthesis in an NPZD model gives little change in phytoplankton dynamics. However, changing the

photosynthesis function has not been tried in our study. We will paraphrase these sentences in the revised manuscript for clarity.

13) l22: *"However, it is still unclear what will happened if formulations of all the core processes [...] are perturbed together." The preceding sentence is very general and I would say it is quite clear that the perturbations of all core processes would also "give rise to different effects". I would suggest to rephrase.*

AR: Thank you for the suggestion, we will rephrase this sentence in the revised manuscript.

p4:

14) l3: "*using all possible functional combinations": Given that there can be an infinite amount of different functional forms, I would suggest to rephrase this sentence. (Later on it becomes clear that only a few functional forms are considered.)*

AR: We have rephrased this in the revised manuscript.

15) l22: *Mention right away that Table 1 contains the equations for all functions.*

AR: Thank you, this will be applied in the manuscript.

16) l29*: Mention that "T" is temperature here.*

AR: Thank you, this will be applied in the manuscript.

17) l32*: "the default": Is this U_1?*

AR: Yes, this is U_1, and we will revise this in the manuscript as U_1 instead of default.

p5:

18) l4: *"The small microzooplankton": this makes it sound like there are small and large microzooplankton. Use something like "The small zooplankton category consists of microzooplankton..."*

AR: Thank you for the suggestion, we will rephrase this in the manuscript as 'The microzooplankton graze on non-diatoms and detritus' instead of using 'The small microzooplankton'

19) l5: *Is "non-diatoms" referring to the "smaller phytoplankton" in the previous sentence?*

AR: Yes, we will indicate this in the revised manuscript.

20) *l8: This is the third time Michaelis-Menten and Holling type II are mentioned together.*

AR: We will be more consistent in the manuscript.

21) *l9: "II" -> "III"*

AR: We will revised this in the manuscript.

22) *l9: Why say "hereafter G_1/G_2" when "Holling type II/III" is used throughout the text?*

AR: We will revise this and use G_1 and G_2 elsewhere.

23) *l19: Was the shape of the curves adjusted again? If so, how?*

AR: Yes, using nonlinear least squares as explained in P4 line 17

24) *l29: What is a "distinct trend" here?*

AR: For clarity, we will revise this in the manuscript.

25) *l30: It is not clear to me how the linear function was made to match the others. Figure 1(c) seems to suggest something went wrong. Or are large values here simply overemphasized in the fit?*

AR: Linear function describe constant removal of phytoplankton or zooplankton, therefore we set the maximum rate of the linear mortality to be similar to the total loss of integrated hyperbolic over the prey range, which resulted in 0.09 day^{-1}. We agree that the large values in the prey range may overemphasized the fit, however even after reducing the range to 10 mmol N m^{-3}, the maximum range for the linear has not changed too much (0.086 day^{-1}).

p6:

26) *l31: How long is the spin-up period for the runs?*

AR: See the answer to Q2 of Reviewer 1

p7:

27) *l9: Why this lengthy comment about physical data assimilation? Is the capping done to remove the perceived negative influence of the physical data assimilation? What about rapid shifts in mixed layer depth which is also an input of the model, may also be affected by physical data assimilation and may also drastically change nutrient concentrations in the model. It is also not quite clear how the mixed layer depth influences the 1D model.*

AR: We will reduce the lengthy comment on the data assimilation in the revised manuscript. We take the vertical velocity from the physical data assimilation. This vertical velocity is the most important physical property that determined the results. We also examined the sensitivity for mixed layer depth which is defined by the vertical diffusivity coefficient, using both model output and the mixed layer from the in situ data and we can't see much difference in the biogeochemical model results.

28) *l26: It would be good to mention these locations the first time the stations are introduced. Sec 2.5.2: Here the description is confusing, it goes from initial conditions to validation data, back to initial conditions and then to validation data.*

AR: Thank you for the suggestion, we will include this in the revised manuscript.

p8

29) l8: "one of MarMOT's test stations" What exactly is this test station?

AR: These are stations that are available within the MarMOT software, which spans from 60° - 10° N, down 20° W in the Atlantic. These stations are used to test whether the MarMOT installation has been successful. The initial conditions are taken from the MEDUSA restart files.

P9*:*

30) l13: "These have been done at the five oceanographic stations which can be classified into three regional types:" This has been mentioned before.
AR: This will be removed in the revised manuscript.

31) l21: Mention PAP.

AR: This will be included in the revised manuscript.

p11:

32) l4: How well does NRR work with a significant bias?

AR: NRR depends on the ratio of the time-averaged RMSE of the ensemble mean to the mean RMSE of the ensemble members. The NRR contain the bias information from the ensemble members, as seen on Table 2.

Table 2. NRR values for Surface chlorophyll at station PAP and various NRR values for different conditions

Surface Chlorophyll	NRR
Original	1.25
Adding Error	1.30
Removing Bias	1.22

33) *l10: "these members use functional combinations ..." The notation for the combinations is not clear here*

AR: We will rephrase this sentence in the revised manuscript.

34) Table 1: It does not make sense to call \mu's the maximum rates here.

AR: In the original MEDUSA paper, the maximum loss rates are represented by \mu.

35) Fig 1: Use "U_1" etc. here.

AR: Thank you, the figure will be revised in the manuscript.

36) *Fig 7: A better description of the x and y axes are needed. Why do b,d,f and h have no y-axis? Use the same color scale across all stations. Same comment applies to Fig. 8 where the font becomes too small.*

AR: Thank you for the suggestions. We will add more description of the x and y-axes in figure 7 and 8 in the revised manuscript. Figure 7 b, d, and f have the same y-tick labels as a, c, and e, therefore in order to maximise the space, we decided not to put the y-tick label. In terms of colour scale, we are not quite sure whether using the same scale across all stations would be a good idea, due to the range of values between different stations and regions. For example, the chlorophyll profile RMSE at station ALOHA and BATS are on different range (ALOHA is between 0.08 and 0.15, and BATS is between 0.3 and 0.35). Therefore we will keep the colour scale on the nitrogen and chlorophyll concentrations between regions similar, and if possible also in the RMSEs.

Reviewer #3

Major comments:

1) Firstly, in the introduction (page three, line 29) the authors state that "It has been demonstrated in conventional sensitivity analyses that only small perturbations are usually produced even with large variations in parameter values, but much larger changes in system dynamics can result from changes in the structural process formulations". I am not quite sure what "conventional" means, but I do think that this statement is misleading, as it neglects previous works that indicate a large sensitivity of marine biogeochemical models to their parameters, when compared to structural sensitivity. These studies have been carried out at a local scale, across different oceanic regimes, or in 3D (see, e.g., Friedrichs et al., 2007, Jour. Geophys. Res., 112, C08001, doi:10.1029/2006JC003852; Ward et al., 2013, Prog. Oceanog.116,49–65, or Kriest et al., 2012, Glob. Biogeochem. Cyc. 26, GB2029, doi:10.1029/2011GB004072, to name just a few examples). Some of them even address the role of different functional forms, or have been applied to the BATS site (e.g., study by Ward et al., 2013). They may be helpful for presenting and discussing this current work in a wider context. Thus, more exploration about what has been found for marine biogeochemical models and their structural and parametric uncertainty can help to improve the discussion, which is currently somehow repetitive, lacks a critical discussion of the results, and how they might relate to other uncertainties (structural, parametric, physical, ...).

AR: Here "conventional sensitivity analysis" was referring to parameter sensitivity analysis, but not the structural sensitivity. We will clarify this in the revised version. Thank you for suggesting the relevant papers also, which will use for comparisons in our largely revised discussion section.

2) *Secondly, I miss some discussion about the way the different functional forms have been made "equivalent to each other." (p4 line 17). As it seems, the parameters of the different equations (e.g., half saturation-constants) were fitted against the default function "so that the overall shapes are as similar as possible." (p 4, line 19), by "minimising the sum squared difference between the default and other uptake forms" (line 32ff). Obviously, when looking at Fig 1, this happened across a very wide range of potential nutrient or chlorophyll (in case*

of zooplankton grazing) concentrations. The upper limits are far outside the range of values for most stations simulated in this study (up to 100 uM nitrate or phytoplankton N will likely never be found at BATS or ALOHA). Thus, it seems that the different functional forms were homogenised for a range that, at many stations, is outside the expected and/or observed range. On the other hand, the functions deviate most strongly when nutrients or phytoplankton are scarce (Fig 1a and 1b), and more representative for the simulated regimes. What would have happened, if the test functions (e.g., sigmoidal or Holling III) were made equivalent to the default functions at lower substrate levels, representative for more oligotrophic regimes? Could it be that the effects of switching to alternative forms becomes less important? Again, the paper to my opinion would benefit a lot from a more critical discussion.

AR: We agree that from looking at figure 1a and 1b, the functions deviate mostly when the nutrient or phytoplankton are scarce, and overfitting may occur due to the large value of nitrogen and phytoplankton. However, we are trying to capture the whole range of nutrient and phytoplankton at all the different region, and optimise the functions when both are the closest to each other (when phytoplankton and nutrient are plentiful) and within the nitrogen and chlorophyll range of all the stations. (See also response to Q2 and 25 of reviewer 2) Suppose we are optimising the nutrient uptake on the similar range of station BATS and ALOHA (with maximum nitrogen and phytoplankton concentration of 5 mmol N m^{-3}, shown on Figure 3, although at stations like Cariaco, PAP, and L4, we may see nitrogen larger than 5 mmol N m^{-3}), the functions still deviate at low nitrogen and phytoplankton concentration. Additionally, the value of half saturation constant have not changed much (for nutrient, the half saturation constant for sigmoidal, exponential, and trigonometric are 0.71, 1.10, and 0.58 respectively, and for grazing the half saturation constant for Holling type II is 0.48). Therefore, the effects of switching to alternative forms will still generate a range of different model outputs. We will change Figure 1 in the manuscript to only use the range that are available in the model (between 0 – 20 mmol N m^{-3} for nitrogen and 0 – 10 mmol N m^{-3} for phytoplankton).

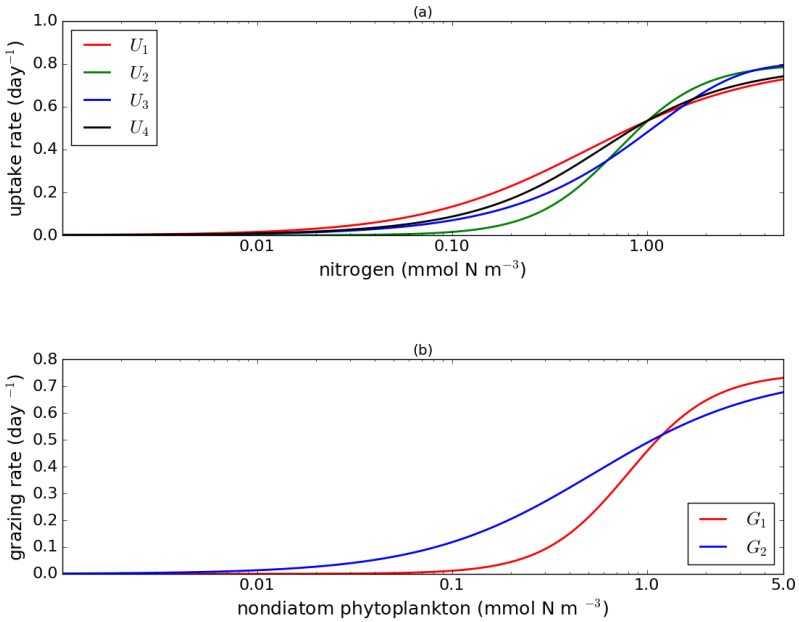

Figure 3. Uptake (a) and grazing (b) functions which have been optimised, with range of 0.001 to 5 mmol N m^{-3}.

3) *Thirdly, as recommended by the second referee, I suggest that the authors read through the manuscript again carefully, revise some sections for clarity, and correct spelling and grammar. The results section already contains a lot of detail, which is partly repeated in the discussion. I would suggest to to shorten and streamline the presentation of results, highlighting those that are common among stations (or differ), as well as the effects of different parameter combinations, and use the discussion to clarify and discuss some of the aspects mentioned above.*

AR: Thank you for the suggestions, and also the addition of literatures which you have suggested. We have indeed revised and streamlined the new paper.

Some detailed comments:

1) p2, line 14ff: "Inclusion of ..." - As mentioned by the other referee, even the spatial variability of light, nutrient availability and mixing already induce a spatial variability of plankton concentrations.

   AR: This will be rephrased in the revised manuscript

2) p2, line 34ff: "However, in biogeochemical models, it is rare that a solid mechanistic basis is present, ..." But see e.g., more recent developments of adaptive models based on mechanistic approaches, such as Pahlow, et al. (2008, Prog.Oceanog., 76 (2), 151-191, doi:10.1016/j.pocean.2007.11.001) or Pahlow, and Prowe, F. (2010), Mar. Ecol. Prog. Ser., 403, 129-144, doi:10.3354/meps08466.

   AR: We will remove this statement in the manuscript

3) p3 line 5: "applying"

   AR: We will rephrase this to ‘applied’ in the manuscript

4) p3 line 9: "highly susceptible" - What does this mean?

   AR: It means that biogeochemical model is likely to be structurally sensitive. We will rephrase this sentence.

5) p3 line 3: "happened"

   AR: We can’t find happened in p3 line 3 – if this is in line 23, we will rephrase this sentence as mentioned by reviewer #2

6) p6 line 25: "Oschlies and Garcon, 1999" - a follow-up study by Oschlies and Schartau (2005, Jour. Mar. Res., 63, 335–358) highlighted this even more; see also the study by Friedrichs et al. mentioned above.

AR: Thank you for the suggestions, we will include these literatures accordingly.

7) p7, section 2.5.1: Physical input: please indicate the vertical grid on which this model was run, including its maximum depth.

   AR: This has been stated in the biogeochemical input but this will be revised in the manuscript.

8) p7 section 2.5.2: Biogeochemical input and validation data: I would suggest to list all the details of the different stations (location, max depth, data source, data assimilated) in a table.

   AR: Thank you for the suggestion, we will include this in the new manuscript, however we do not assimilate any data into our model

9) p7 section 2.5.2: Do I understand correctly, that the observations were used for initialisation as well as for model validation? If so, then the model is not validated against fully independent data (at least not at depth, given a short simulation time of just 10 years), and I would suggest to mention it here.

   AR: Indeed, we are using the observation to initialise the model (using in situ chlorophyll, nitrogen, iron, and silicate data from January 1998), but we do not use the later in situ data to force the model, so the validation data is independent.

10) p7, line 13: "Simulations are made at 37 depth levels" - This formulation sounds as if simulations were done separately for each depth level.

    AR: This will be rephrased in the revised manuscript.

11) p15 line 24: "Most current biogeochemical models are run in a deterministic, rather than a probabilistic, manner, even though data from observations contain many uncertainties, eg. in satellite-derived chlorophyll." - I think I can guess what you want to say, but in the current form this sentence is not clear.

    AR: This will be rephrased in the revised manuscript for clarity.

    Literatures cited:

Raick, C., Soetaert, K., and Grégoire, M.: Model complexity and performance: How far can we simplify? Progress in Oceanography, 70, 27–57, https://doi.org/10.1016/j.pocean.2006.03.001, 2006.

Ward, B. A., Schartau, M., Oschlies, A., Martin, A. P., Follows, M. J., and Anderson, T. R.: When is a biogeochemical model too complex? Objective model reduction and selection for North Atlantic time-series sites, Progress in Oceanography, 116, 49–65, https://doi.org/10.1016/j.pocean.2013.06.002, 2013.