

Dear Editor,

We would like to thank the reviewers for taking the time to read the manuscript and for giving constructive comments. In the revised manuscript, we have taken into consideration all the comments and suggestions made by the referees. In particular we have:

- Included a table (see Table 2) describing the location, data source, and maximum depth (as suggested by reviewer 3).
- Included annual and seasonal boxplots for nitrogen (see Fig 6 and 7), annual predicted primary production (see Fig 10), and an additional boxplot (see Fig 8) to highlight the range obtained when changing only one process at a time (as suggested by reviewer 1).
- Changed the colour scales in figure 7 and 8 (see the new Fig 11 and 12), splitting figure 8 into two figures (see Fig 12 and 13) so that the text won't get too small (as suggested by reviewer 2).
- Made the results more concise (see page 10-18 in the annotated version) and the discussion (page 18-23 in the annotated version) more explicit.

In addition below is our point by point response to all comments made by the three reviewers (reviewer comment in black, response in blue, and changes to the manuscript indicated in blue bold). Please note that all line numbers in this response refer to the annotated version attached. We have further uploaded a clean version separately.

We hope that the response would be satisfactory, and we look forward to your decision.

Kind Regards,

Authors.

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

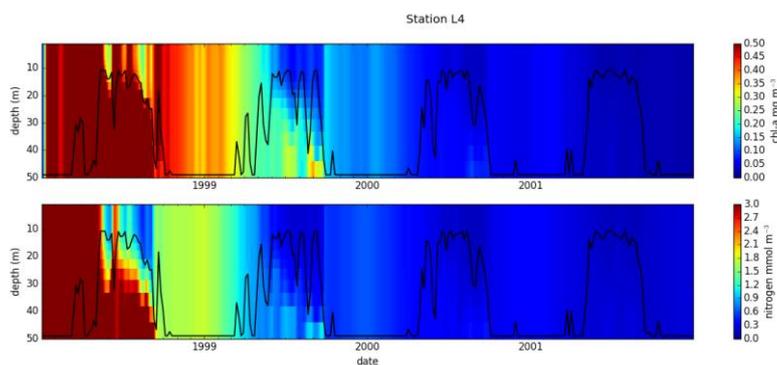


Figure 1. Chlorophyll and nitrogen concentration in the water column at station L4, when sinking rate is 3 m day^{-1}

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 respectively, but not on nitrogen, the NRR increases to 1.35) and **we have included these experiments in the supplementary material S2**. 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 have emphasised 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 improved 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 have included these results in the supplementary material S3.**

In the revised version the metrics for nitrogen and primary production (as suggested by the reviewer) have been included in Fig 6 for inter-annual variability and 7 for seasonal means. Further, predicted primary production at stations ALOHA and CARIACO have been included in Fig 10, 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 (page 6, 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 have emphasised these points in the revised text and discuss the alternative spin up method in the supplementary material section S1.**

- (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, in the revised manuscript is now in Fig. 11, 12, and 13, and also using the box plots in figure 4 and 5. **We have included a boxplot to show the range in chlorophyll annual means produced when changing only one process at a time thus better supporting the conclusions, in Fig. 8.**

- (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? could be made on the need to do sensitivity analysis 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 have include these suggestions in the discussion (page 18-25)**

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, **this have been included in the revised text with appropriate references (line 1-4 page 3)**

(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 have removed this and make the paragraph shorter (line 12-17 page 10).**

(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 have make it larger in the revised figures, **and split figure 9 into two figures (Fig. 12 and 13) to make the text clear.**

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

AR: Thank you, **this have been 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	$\mu = 0.085$	0.0126	$\mu = 0.097$	0.0085
quadratic	$\mu = 0.023$	0.0035	$\mu = 0.050$	0.0028

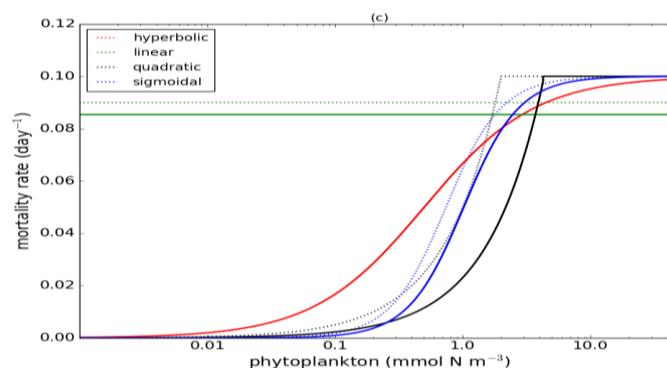


Figure 2. Mortality functional forms optimised against hyperbolic function. Dotted lines are fitting without log transform and solid lines are fitting with log transform

Specific comments:

- 1) *I1: "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 have rephrase this sentence and change 'mathematical structure' into 'mathematical equations' (page 1 line 1-2).**

- 2) *I3: "intermediately complex BGC model" -> "BGC model of intermediate complexity"*

AR: Thank you for the suggestion, **we have rephrased this sentence as suggested (page 1 line 4).**

- 3) *I9: "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 have revised this sentence please see page 1 line 9-10.**

- 4) *I11: 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 have revised this sentence in the annotated manuscript for clarity (page 1 line 12-13).**

- 5) *I14: "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 have rephrased this in the manuscript as RMSEs instead of errors (page 1 line 16).**

- 6) *I15: 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) *I7: 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 have rephrased this sentence in the annotated manuscript into ‘...address and predict the impact of climate change in the ocean ecosystems...’ (page 2 line 7-9).**

8) *I12: Even NPZ models represent "several" processes. Please be more precise.*

AR: Thank you, **we have rephrase this sentence in the annotated manuscript into ‘...More advanced biogeochemical models represent more processes and feedbacks compared to...’ (page 2 line 15-16).**

9) *I16: There can be spatial variability without iron!*

AR: We agree with this statement, **we have rephrase this sentence into ‘...such as iron, to permit phytoplankton growth limitation’ (page 2 line 18-19) for clarity.**

10) *I29: "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 have revised this sentence in the annotated manuscript into ‘... small changes in the structural process formulation often produce larger changes in the system dynamics, compared to varying parameter values alone’ (page 3 line 4-5).**

Page 3:

11) *I13: "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 have rephrased this sentence in the manuscript ‘... linear density-dependent mortality produces the biggest difference in diatoms with concentrations at mid latitudes being twice as high...’ (page 3 line 26-27).**

12) *I18: "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 have shortened these sentences in the revised manuscript for clarity (page 3 line 31-35).**

13) *I22: "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 have removed this in the revised manuscript (**page 4 line 3**).

p4:

14) I3: *"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 into '**... using possible functional form combinations within the NPZ compartments...**' (**page 4 line 18**)

15) I22: *Mention right away that Table 1 contains the equations for all functions.*

AR: Thank you, **this have been applied in the manuscript (page 5 line 2)**.

16) I29: *Mention that "T" is temperature here.*

AR: Thank you, **this have been applied in the manuscript (page 5 line 11)**.

17) I32: *"the default": Is this U_1?*

AR: Yes, this is U_1, and we **have revise this in the manuscript as U_1 instead of default (page 5 line 14)**.

p5:

18) I4: *"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 have rephrased this sentence into '**The small zooplankton, represented by the microzooplankton, graze on small phytoplankton, non-diatoms, and detritus ...**' (**page 5 line 18-19**)

19) I5: *Is "non-diatoms" referring to the "smaller phytoplankton" in the previous sentence?*

AR: Yes, **we have indicate this in the revised manuscript (page 5 line 19)**.

20) I8: *This is the third time Michaelis-Menten and Holling type II are mentioned together.*

AR: We have changed this throughout the manuscript.

21) I9: *"II" -> "III"*

AR: We have revised this in the manuscript (**page 5 line 23**)

22) I9: Why say "hereafter G₁/G₂" when "Holling type II/III" is used throughout the text?

AR: We have revised this and use G₁ and G₂ elsewhere.

23) I19: Was the shape of the curves adjusted again? If so, how?

AR: Yes, using nonlinear least squares as explained in page 4 line 31-34 in the annotated manuscript

24) I29: What is a "distinct trend" here?

AR: For clarity, we have revised this in the manuscript (page 6 line 12-13).

25) I30: 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) I31: How long is the spin-up period for the runs?

AR: See the answer to Q2 of Reviewer 1

p7:

27) I9: 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 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. **We have reduce the lengthy comment on the data assimilation in the revised manuscript, it's now on page 7 line 26-33 in the annotated manuscript.**

28) I26: 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 have now revise this description of the station at the start of section 2.5.2 (page 8 line 13-15).**

p8

29) I8: *“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) I13: *“These have been done at the five oceanographic stations which can be classified into three regional types:” This has been mentioned before.*

AR: **We have removed this in the annotated manuscript (please see the start of the Results section on page 10 line 12-17).**

31) I21: *Mention PAP.*

AR: **We have mentioned this in the annotated manuscript (page 10 line 19)**

p11:

32) I4: *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) I10: *“these members use functional combinations ...” The notation for the combinations is not clear here*

AR: We have rephrase this in the manuscript into **‘... show that ensemble members with...’ (Page 12 line 33)**

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

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

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

AR: **We have included this in the manuscript, please see Fig. 1**

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 have added more description of the x and y-axes in figure 7 and 8 in the revised manuscript (now Fig. 11-13).** 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 if possible, 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 have removed this statement in the revised version, and paraphrased it (page 3 line 3-4).** Thank you for suggesting the relevant papers also,

which have used these literatures for comparisons in our largely revised discussion section (page 23, starting line 20-25).

- 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 μM 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 have changed 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).**

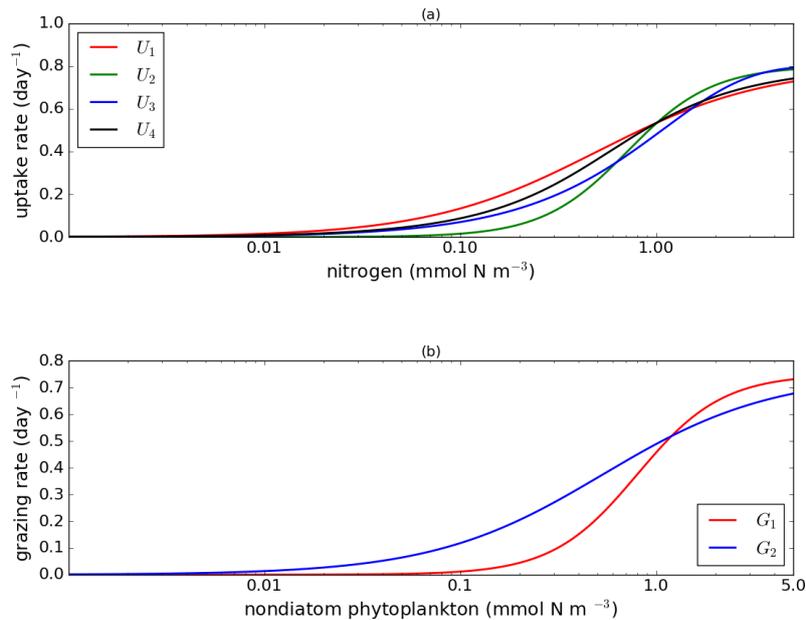


Figure 3. Uptake (a) and grazing (b) functions which have been optimised, with range of 0.001 to 5 mmol N m⁻³.

- 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 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: We have rephrased this on the main manuscript, **please see reviewer 2's answer no 9**

- 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.pocan.2007.11.001) or Pahlow, and Prowe, F. (2010), Mar. Ecol. Prog. Ser., 403, 129-144, doi:10.3354/meps08466.*

AR: We have removed this statement in the manuscript.

3) p3 line 5: "applying"

AR: We have removed this please see page 3 line 16.

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

AR: It means that biogeochemical model is likely to be structurally sensitive. We have rephrased this sentence to: 'These discrepancies from simple interaction suggest that complex biogeochemical models need to be tested by altering their default functional forms...' in the revised manuscript (page 3 line 21-23)

5) p3 line 3: "happened"

AR: We can't find happened in p3 line 3 – if this is in line 23, we have 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 have added these literatures accordingly in the revised manuscript (page 7 line 11).

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, for clarity this have been revised in the annotated manuscript, in page 8 line 8-11.

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 have included this in the new manuscript (please see table 2 in the annotated 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) P8, line 13: *"Simulations are made at 37 depth levels" - This formulation sounds as if simulations were done separately for each depth level.*

AR: This have been rephrased to **'the model is simulated at 37 depth levels...'** in the revised manuscript (page 9 line 5)

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 have been removed from the manuscript**

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.

A perturbed biogeochemistry model ensemble evaluated against in situ and satellite observations

Prima Anugerahanti¹, Shovonlal Roy^{1,2}, and Keith Haines³

¹Department of Geography and Environmental Science, University of Reading, Whiteknights, Reading, RG6 6AB, UK

²School of Agriculture, Policy, and Development, University of Reading, Whiteknights, Reading, RG6 6AR, UK

³Department of Meteorology, University of Reading, Whiteknights campus Early Gate, Reading, RG6 6BB, UK

Correspondence: Prima Anugerahanti (p.anugerahanti@pgr.reading.ac.uk), Shovonlal Roy (shovonlal.roy@reading.ac.uk)

Abstract. The dynamics of biogeochemical models are determined by the mathematical ~~structure used for equations used to describe~~ the main biological processes. Earlier studies have shown that small changes in the model formulation may lead to major changes in system dynamics, a property known as structural sensitivity. We assessed the impact of structural sensitivity in ~~an intermediately complex biogeochemical model (MEDUSA)~~ a biogeochemical model of intermediate complexity by modelling the chlorophyll and nitrogen concentrations at five different oceanographic stations spanning three different regimes: oligotrophic, coastal, and the abyssal plain, over a 10-year timescale. A 1-D MEDUSA ensemble was used with each ensemble member having a combination of tuned function parameterizations that describe the key biogeochemical processes, namely nutrient uptake, zooplankton grazing, and plankton mortalities. The impact is quantified using phytoplankton phenology (initiation, bloom time, peak height, duration, and termination of phytoplankton blooms) and ~~other statistical measures~~ statistical measures such as RMSE, mean, and range for chlorophyll and nutrients. The spread of the ensemble as a measure of uncertainty is assessed against observations using the Normalised RMSE Ratio (NRR). We found that even small perturbations in model structure can produce large ensemble spreads. The range of 10-year mean surface chlorophyll ~~concentrations are~~ concentration in the ensemble is between 0.14-3.69 mg m⁻³ at coastal stations, 0.43-1.11 mg m⁻³ on the abyssal plain, and 0.004-0.16 mg m⁻³ at the oligotrophic stations. Changing mortality and grazing functions have the largest impact on chlorophyll concentrations. The in situ measurements of bloom timings, duration, and terminations lie mostly within the ensemble range ~~and using the ensemble properties such as the~~ The RMSEs between in situ observations and the ensemble mean and median, ~~the errors~~ are mostly reduced compared to the default model output. The NRRs for monthly variability suggest that the ensemble spread is generally narrow (NRR 1.21-1.39 for nitrogen and 1.19-1.39 for chlorophyll profiles, 1.07-1.40 for surface chlorophyll, and 1.01-1.40 for depth integrated chlorophyll). Among the five stations, the most reliable ensembles are obtained for the oligotrophic station ALOHA (for the surface and integrated chlorophyll ~~10-year time series~~ and bloom peak height), for coastal station L4 (for inter-annual mean), and for the abyssal plain station PAP (for bloom peak height). Overall our ~~studies provided~~ study provides a novel way to generate ~~ensemble spread~~ a realistic ensemble of a biogeochemical model by perturbing the model ~~structure/equations/parameterizations,~~ and reliable ensemble means and spreads may be generated, ~~which will be helpful for the probabilistic predictions.~~

1 Introduction

Major changes in ocean biogeochemistry have been driven by anthropogenic activities, leading to ocean acidification, eutrophication, and increased levels of dissolved inorganic carbon (Gehlen et al., 2015; Bopp et al., 2013; Doney, 2010). To understand how the ocean ecosystem responds to these changes, marine biogeochemical models have been ~~used~~developed. The majority of these models focus on the lower trophic food-webs and explicitly represent dissolved nutrients, phytoplankton and zooplankton (NPZ). These models are then coupled with physical general circulation models to address ~~the impacts and predict the impact~~ of climate change (~~Doney et al., 2012~~) and ~~forecasting systems~~ (~~Yool et al., 2013; Butenschön et al., 2016~~)in the ocean ecosystems (~~Doney et al., 2012; Yool et al., 2013; Butenschön et al., 2016~~).

Marine biogeochemical model development began with simple NPZ models, and has become steadily more complex with increasing computing power and knowledge of ocean biogeochemistry (Anderson, 2005; Anderson et al., 2015). NPZ models consist of three compartments: nutrients as the primary resource, phytoplankton as the primary producers, and zooplankton as herbivores or grazers. Such models have been used to investigate the range of possible ecosystem behaviours before coupling them to a physical model (Franks, 2002) and seeking to represent observations at particular sites (Fasham et al., 1990; Robinson et al., 1993). More advanced biogeochemical models represent ~~several more~~ processes and feedbacks compared to the NPZ models (Raick et al., 2006), covering much more of the lower-trophic food web (Anderson, 2005). Inclusion of cell size representations (Berelson, 2002; Le Quèrè et al., 2005), different phytoplankton functional types, such as calcifiers and dimethyl sulphide producers (Le Quèrè et al., 2005), and the addition of important micronutrients, such as iron to permit ~~spatial variability in phytoplankton concentrations~~phytoplankton growth limitation (Yool et al., 2011, 2013), are now part of many biogeochemical models. Moreover, in order to investigate the effect of global climate change and anthropogenic activities in the ocean, marine biogeochemical models are now being embedded into earth system models. For example, the Model of Ecosystem Dynamics, nutrient Utilisation, Sequestration, and Acidification (MEDUSA) (Yool et al., 2011, 2013) is the chosen biogeochemical component for the UK Earth System Model, as it has high spatial correlation with patterns of pCO₂, DIC, and alkalinity (Cox and Kwiatkowski, 2013; Kwiatkowski et al., 2014).

Despite becoming more complex (Anderson, 2005), the ~~overall~~basic interactions among nutrients, phytoplankton, and zooplankton are still at the heart of all marine biogeochemical models. These interactions are governed by four primary processes, represented in the simplest NPZ models: nutrient uptake, grazing by zooplankton, phytoplankton and zooplankton mortality. These processes are functions of the state concentrations and can be parameterized by ~~more than one functional form, similar in shape but using different mathematical functions and~~different functional forms along with adjustable parameters. ~~Therefore there are two types of uncertainties that affect biogeochemical models : parametric, associated with the choice of parameter values; and structural, which relates to the underlying model equations (Hemmings and Challenor, 2012). It has been demonstrated in conventional sensitivity analyses that only small perturbations are usually produced even with large variations in parameter values, but much~~ Biogeochemical models therefore have different sources of uncertainty, such as the physical input

(Sinha et al., 2010; Doney, 1999; Hemmings and Challenor, 2012), parameters (Oschlies and Schartau, 2005; Friedrichs et al., 2006, 2007), and the model structure associated with how the ecosystem is represented, either by the number of model compartments and linkages (Friedrichs et al., 2007; Kriest et al., 2010; Ward et al., 2013), or its mathematical formulations (Anderson et al., 2010; Flora et al., 2011). Sensitivity analyses show that small changes in the structural process formulation often produce larger changes in system dynamics ~~can result from changes in the structural process formulations~~, compared to varying parameter values alone (Wood and Thomas, 1999; Fussmann and Blasius, 2005; Levin and Lubchenco, 2008; Flora et al., 2011; Adamson and Morozov, 2013; Aldebert et al., 2016), a result known as structural sensitivity (Wood and Thomas, 1999; Flora et al., 2011; Adamson and Morozov, 2013). A study by Aldebert et al. (2016) shows that parameter values are weakly correlated to food-web dynamics compared to the model formulations, as equilibrium dynamics are determined by the choice of functional forms.

Structural sensitivity may be less significant in models built on well-tested mechanisms ~~such as those as~~ in the physical sciences. ~~However,~~ ~~however~~ in biogeochemical models ~~, it is rare that a solid mechanistic basis is present, therefore it is uncertain what is the most appropriate specification of the~~ the process functional terms ~~are all gross simplifications~~. This is even more problematic if the ~~process itself is not well~~ processes are poorly understood so that ~~theoretical justification for the justification for any~~ specific representation is weak (Adamson and Morozov, 2013). Often it is difficult to implement the functional relations that are observed in the laboratory into a ~~large-scale~~ ~~large-scale~~ ecosystem with heterogeneous populations (Englund and Leonardsson, 2008). It is known from studies of simple predator-prey models that ~~applying~~ similarly shaped equations often ~~garners lead to~~ completely different stability and oscillatory model dynamics (Fussmann and Blasius, 2005; Roy and Chattopadhyay, 2007). Moreover, a specific functional form may not capture all details of the biological processes, for example, the Michaelis-Menten type function for grazing, commonly known as the ‘Holling Type II’, fails to correctly describe what happens to grazers’ movements when satiation has been reached (Flynn and Mitra, 2016). These ~~studies show that simple biological models are highly susceptible to structural sensitivity. Further, the discrepancies reported~~ ~~discrepancies~~ from simple interaction models suggest that ~~the dynamics of~~ complex biogeochemical models need to be tested by altering their default functional forms (Anderson and Mitra, 2010; Anderson et al., 2010).

~~Some~~ ~~A few~~ studies have investigated the effects of ~~different process formulations on biogeochemical models~~ ~~biogeochemical process formulations~~, e.g. Yool et al. (2011) has demonstrated ~~that~~ in an intermediately complex model, ~~that~~ linear density-dependent mortality produces the ~~most significant differences when applied to diatoms, biggest difference in diatoms, with concentrations at mid latitudes being twice as high~~, compared with sigmoidal, quadratic, or hyperbolic forms. The choice of zooplankton grazing equations ~~can also~~ affect phytoplankton concentration dramatically in a model with five plankton types, PlankTOM5.2 (Le Quèrè et al., 2005). The ~~Michaelis-Menten~~ (Holling type II) ~~grazing function produces 30% less total surface phytoplankton concentration compared to the sigmoidal (Holling type III) functions~~ ~~function~~, in the North Atlantic and North Pacific (Anderson et al., 2010). However, ~~not all processes give significantly different model outputs. Anderson et al. (2015) also~~ ~~Anderson et al. (2015)~~ shows that when two ~~similarly shaped~~ ~~similarly shaped~~ photosynthesis-irradiance curves ~~, namely,~~ (Smith and the exponential) function, were used in an NPZ-detritus model, the concentration of chlorophyll during the spring bloom was only slightly higher (0.2 mg m^{-3}) for the exponential function (~~Anderson et al., 2015~~) ~~, with little difference in phytoplankton dynamics (Anderson et al., 2015).~~

These studies suggest that when the model formulation is perturbed, each process can give rise to different effects. However, it is still unclear what will happen if formulations of all the core processes, i.e. nutrient uptake, grazing and mortality are perturbed together. Since the individual compartments of models interact with one another, any biological perturbation is likely to affect the whole ecosystem dynamics. In climate modelling, perturbed physics ensembles have been developed to investigate multiple parameter uncertainty (Murphy et al., 2007; Tinker et al., 2016), and multiple parametrization (functional) uncertainties (Subramanian and Palmer, 2017). ~~By adopting an approach similar to that in climate modelling~~ Inspired by these studies, here we attempt to generate a perturbed biogeochemical ensemble where model equations are varied by embedding different functional forms to describe the core processes, similar to the multi-parameterization ensembles in physical models. We implement this framework in the MEDUSA model (Yool et al., 2011, 2013), which is a lower trophic level model with two phytoplankton functional types, distinguished as large diatoms and small non-diatoms, two zooplankton types represented by mesozooplankton and microzooplankton, and three nutrients: silicic acid, iron, and inorganic nitrogen. Nitrogen is the primary currency of the model, similar to NPZ models, but MEDUSA allows phytoplankton to have different C:N ratios and Si:N ratios for diatoms. Diatoms utilise the silicic acid and can only be grazed by mesozooplankton. MEDUSA also includes an iron submodel developed by ~~(Parekh et al., 2005) based on (Dutkiewicz et al., 2005)~~ Parekh et al. (2005) based on Dutkiewicz et al. (2005), in which iron is separated into "free" iron and iron bound to organic ligands. Iron is removed by scavenging and added to the ocean by aeolian deposition.

We assess of the uncertainty arising from the MEDUSA model's equations from ensemble outputs generated using ~~all possible functional~~ possible functional form combinations within the NPZ compartments. For simplicity we use a 1-D version of MEDUSA-1.0 model (Yool et al., 2011; Hemmings et al., 2015), and produce results for five oceanographic stations covering abyssal plain, oligotrophic, and coastal regimes. Apart from the model outputs on concentration of nutrients and chlorophyll, we also examine the emergent properties ~~of phytoplankton~~ using phytoplankton phenology metrics. The performance of the ensemble mean, median, and the default MEDUSA run are compared with monthly and inter-annual values from in situ observations at those stations. We assessed the spread of the ensemble using the Normalised RMSE Ratio (NRR) which assesses the likelihood of the observations fitting the ensemble range. Section 2 describes the equations used and how the ensemble is run. The assessment of the uncertainty in terms of chlorophyll concentrations, phytoplankton phenology, and comparisons with the observations are described in section 3, and are further discussed in section 4.

2 Method

~~In models, the key processes can be represented by a variety of functional forms, which are comparable in shape but different in their mathematical representation. To explore this~~ To explore structural uncertainty we first ~~attempt to~~ make the functional forms representing key processes more ~~equivalent~~ similar to each other. ~~For each process functional form we optimise by tuning~~ the shape-defining parameters ~~to make the functions equivalent to each other~~. For example, for Holling type II and Holling type III, we fix the maximum rates of each process, and implement a non-linear least squares method to optimize the half saturation coefficients so that the overall shapes are ~~as similar as possible~~. A similar similar. This approach is used for

nutrient uptake (4 functional forms), phytoplankton mortality (4 functional forms), and zooplankton mortality (4 functional forms). ~~These are described, as~~ in the subsections below. Table 1 shows the equations and parameter values.

2.1 Nutrient uptake

Alongside light, nutrient concentration ~~also~~ limits the growth of phytoplankton. In MEDUSA the standard hyperbolic monod, hereafter U_1 , function is the default. The growth of cells monotonically increases with ambient nutrient concentration, and halts when nutrients become scarce. If nutrient concentrations are high, the rate of uptake saturates. Other mathematical functions show similar properties including (i) Sigmoidal (Fennel and Neumann, 2014) ~~, similar to Holling type III,~~ U_2 , (ii) the exponential (Ivlev, 1961), U_3 , and (iv) trigonometric functions (Jassby and Platt, 1976), U_4 . All these functions include a ~~shape defining shape-defining~~ parameter, k , which for monod and sigmoidal can be interpreted as a half saturation constant, ~~with the and a~~ maximum uptake rate, V_{pT} , which is a function of temperature (Eppley, 1972): $V_{pT} = V_p 1.066^T$, where V_p is the maximum growth rate when temperature, T , is at 0°Celsius. ~~MEDUSA has silicon and iron nutrients, and two phytoplankton types: diatoms and non-diatoms.~~ The uptake function of different phytoplankton types and nutrients use similar functions but different parameter values for k , ~~as~~ summarised in Table 1. ~~Values for k are obtained from,~~ obtained by minimising the sum squared difference ~~between the default and other uptake forms~~ with U_1 . The nutrient uptake functions after optimization are shown in Fig. 1(a). The difference in shape of the optimised functional forms are more obvious at low nutrient concentrations.

2.2 Zooplankton grazing

In MEDUSA, both phytoplankton and zooplankton are grouped into "small" and "large" categories. The small ~~microzooplankton graze on smaller phytoplankton and slow sinking detritus.~~ The zooplankton, represented by the microzooplankton, graze on small phytoplankton, non-diatoms, and detritus, with the more nutrient rich, ~~and therefore~~ higher quality, non-diatoms are preferred over detritus. Larger zooplankton, represented by mesozooplankton have a broader range of prey types, including both microzooplankton and diatoms ~~which are regarded as higher,~~ which are high quality food sources. When describing multiple grazing functions, the zooplankton grazing rate is often defined using either the ~~hyperbolic Michaelis-Menten (Holling type II hereafter, G_2) or sigmoidal (Holling type II or Holling type III hereafter, G_1) expression,~~ function with maximum grazing rate g_m , and a weighted preference on the different food sources p_n (Fasham et al., 1990). ~~Since zooplankton preferences will change throughout the year, the assigned preference changes~~ The preference parameter changes through the year as a function of the food ratio. ~~Holling type II and Holling type III, G_2 and G_1~~ grazing on prey Pa are described in ~~table~~ Table 1. In MEDUSA, the default multiple grazing parameterisation is based on the sigmoid Holling type III (Ryabchenko et al., 1997) function. Apart from the weighted preference, both of these functions ~~also~~ include a half saturation constant k_x , where x is the zooplankton type.

These functions ~~have similar behaviours where grazing rates saturate and~~ both become constant at a maximum grazing rate. At low zooplankton concentrations the sigmoidal response has lower grazing rates than the hyperbolic, and therefore, the sigmoidal curve has a more rapid increase in predation rate before becoming saturated (Edwards and Yool, 2000), shown on Fig. 1(b). Preferences for food types are kept the same as MEDUSA's default parameters, with terms summarized in Table 1.

2.3 Plankton mortality

~~Apart from grazing, plankton loss is caused by mortality. MEDUSA has two mortalities:~~ MEDUSA has both density independent and density dependent mortality rates for all the phytoplankton and zooplankton types: ~~density-independent and density-dependent.~~

Density-independent loss ~~terms are~~ is modelled by a linear function representing plankton metabolic loss which was kept unchanged. Density-dependent loss includes processes such as higher-trophic grazing and disease. In MEDUSA these processes are modelled using the hyperbolic function of plankton concentration (Fasham et al., 1993). ~~As it is unclear which density dependent loss is the best choice, MEDUSA allows the option to include alternative functions to~~ Alternative functions can describe the density-dependent ~~mortalities.~~ mortality, and we use the combinations of hyperbolic (ρ_1, ξ_1), linear (ρ_2, ξ_2), quadratic (ρ_3, ξ_3), and sigmoidal (ρ_4, ξ_4) functions to describe the phytoplankton (ρ) and zooplankton (ξ) mortalities (equations and abbreviations are shown on Table 1). Similar to grazing and nutrient uptake, the functional forms have different maximum rates for each plankton type. These maximum rates ~~have~~ are made the same for all the different functions.

Of the four different mortality functions, linear and quadratic functions ~~show the most distinct trends.~~ are most different in shape, as shown on Fig. 1(b). Using the linear term is similar to a change in the value of maximum mortality rate, μ . To make the linear function ~~most~~ similar to the sigmoidal and hyperbolic functions, the maximum grazing rate is set so that the total loss integrated over the range of prey density (calculated as the area below the line function representing the total loss) is similar to that for the hyperbolic curve. The quadratic term, instead of asymptoting, continues to grow with plankton abundance. In order to keep this ~~as 'similar'~~ similar to other forms, after reaching a certain concentration, ~~the mortality~~ the function is switched to linear, so that the rate ~~reaches a plateau.~~ plateaus at high abundance. For sigmoidal mortality, the default μ are not changed but the ~~half saturation~~ half-saturation constant, k_M is optimised. The optimised mortality functions are shown in Fig. 1(c). A distinctive feature ~~in the shapes~~ of these functional forms after optimisation is that the quadratic mortality rate remains low until phytoplankton concentration reaches 10 mmol N m^{-3} , and the linear function ~~always shows constantly~~ shows consistently high plankton mortality ~~rate~~ (Fig. 1(c)).

2.4 Model Parameters

Apart from sinking rate, maximum growth, and grazing rates, parameters ~~that are~~ not listed in Table 1 are kept at their ~~respective~~ default values used in the MEDUSA model default values (Yool et al. (2011) shown on table 1-4). From a previous 3-D MEDUSA run, in the oligotrophic regions ~~MEDUSA shows~~ show a low 'background' chlorophyll concentration (Yool et al., 2011). ~~In order~~ so to raise this concentration a higher maximum growth rate and lower grazing rate has been used. ~~We chose~~ the value for The maximum uptake rate, V_p , as 0.8 day^{-1} , similar to that in the HadOCC model (Palmer and Totterdell, 2001). For zooplankton grazing, similar to ~~NPZD~~ NPZ models (Fasham et al., 1990; Fasham, 1995; Anderson et al., 2015) we use 1 day^{-1} as the ~~value for~~ maximum grazing rate, g_m . MEDUSA also contains both slow and fast detritus sinking factors. It is assumed that the latter sinks rapidly relative to the model time-step, and remineralisation of the detrital nitrogen and silicon is done implicitly. In the default model 3 m day^{-1} is used for the slow sinking detritus, however over long runs we found this leads to downward loss of nutrients from the euphotic zone. Earlier studies have used lower detrital sinking rates (Steele and

Henderson, 1981; Fasham et al., 1990; Lacroix and Gregoire, 2002; Raick et al., 2006), between 0 to 1.25 m day⁻¹. ~~Therefore we and other study have suggested to use 0 m day⁻¹ (Ward et al., 2013). We~~ chose a lower sinking rate of 0.1 m day⁻¹ to prevent depletion of state variables particularly at the shallower stations.

2.5 Running the Model and Generating the Ensemble

- 5 MEDUSA is run in the Marine Model Optimization Testbed (MarMOT-1.1) (Hemmings and Challenor, 2012; Hemmings et al., 2015), a site-based mechanistic emulator, where simulations are run in 1-D. MarMOT was developed to investigate the effect of sensitivity in plankton model simulations, especially in regard to parameter and environmental inputs (Hemmings and Challenor, 2012). Despite some uncertainties associated with the differences in physical forcing, fluxes, and initial values of biogeochemical properties, using 1-D simulations to approximate 3-D model behaviour for calibrating models based on specific sites
- 10 has improved the 3-D models' predictive skill (~~Oschlies and Garçon, 1999; Kane et al., 2011; McDonald et al., 2012). Physical and biogeochemical information are needed as input data in order to run MEDUSA. (Oschlies and Garçon, 1999; Oschlies and Schartau, 20~~
- 15 The 1-D MEDUSA is run at five oceanographic stations: PAP, ALOHA, BATS, Cariaco, and L4 shown in Fig. 2. These are chosen as they represent different oceanographic regimes: abyssal plain (PAP), oligotrophic (ALOHA, BATS), and coastal (Cariaco, L4).
- 20 At each oceanographic station, all combinations of the optimized functional forms (as described in subsection 2.1, 2.2, and 2.3), are then embedded into the 1-D MEDUSA code. The same process function is always used for both diatoms and non-diatoms, or mesozooplankton and microzooplankton. The ensemble model at each station is initialized using ~~the~~ in situ measurements ~~such as of~~ chlorophyll, nitrogen, silicic acid, and iron, and the ensemble is run over 10 years starting from January 1998. This provides a total number of 128 combinations, arising from 4 types of nutrient uptake ~~functions~~, 4 ~~types~~ phytoplankton mortalities, 2 types of zooplankton grazing, and 4 ~~types of~~ zooplankton mortalities.

2.5.1 Physical input

- Physical input files consist of gridded values of vertical velocity (m day⁻¹), vertical diffusion coefficient (m² day⁻¹), and temperature (°C), which are applied at each depth level. Additionally, time series of downwelling solar radiation (W m⁻²) and mixed layer depth (m) are also used as input. These ~~data~~ are obtained from the 5-day mean output of the Nucleus for
- 25 European Modelling of the Ocean (NEMO) model, using the Met Office Forecast Ocean Assimilation Model (FOAM), ~~which controls the physical parameters and therefore the biogeochemical tracers every 5 days.~~ The FOAM-NEMO system assimilates *in situ* satellite SST, sea-level anomaly, sea-ice concentration, temperature, and salinity profile data, in order to make the physical system more realistic (Storkey et al., 2010). However, assimilating physical data directly into a coupled physical-biogeochemical model often does not improve the simulation of the ecosystem. For example ~~a study by Ourmières et al. (2009) using the LOBSTER model, showed that although assimilating physical data improved the primary production in the Labrador Sea (due to increasing eddy activity), it does not improve the match to SeaWiFS derived chlorophyll-*a*. When when~~ assimilation
- 30 is used in the 3-D HadOCC model it overestimates the nutrient concentrations due to spurious vertical velocities (Ford et al., 2012; Ourmières et al., 2009).

~~In our work~~ To avoid overestimating surface nutrients the vertical velocities ~~taken from the assimilated from the~~ FOAM system were capped at the 90th and 10th quantiles, and the 10-year mean of the vertical velocity is also removed. This means that there is no time mean vertical velocity, ~~and these adjustments are found to give~~. These adjustments gave a better long-term vertical structure to the nutrient and other distributions. Since input data on the vertical diffusivity ~~coefficient~~ was not stored ~~from the assimilation run in FOAM~~, we used values from NEMO ORCA025-N102 output from January 1998-December 2001 and ~~from~~ ORCA0083-N01 from January 2002-December 2007, both ~~were~~ obtained from the CEDA Group workspace web (http://gws-access.ceda.ac.uk/public/nemo/#_top). ~~Similar to the other physical inputs, vertical diffusivity coefficient from these NEMO outputs~~ These physical inputs are 5-day averaged, ~~common to 3-D MEDUSA (Hemmings et al., 2015)~~, and are available at 75 depth levels (from 0.5 to 6000 m) for NEMO-FOAM and ORCA0083-N01, and 63 depth level (spanning from ~~6 to 5800 m~~) for NEMO ORCA024-N102. The level thickness increases exponentially as the depth goes deeper. Our 1D model uses these same 63 depth levels vertical resolution.

2.5.2 Biogeochemical input and validation data

The 1D MEDUSA ensemble is run at five oceanographic stations: Porcupine Abyssal Plain Sustained Observatory (PAP-SO, hereafter, PAP), A long time Oligotrophic Habitat Assessment (ALOHA), Bermuda Atlantic Time Series (BATS), Cariaco, and L4. The input for the biogeochemical environment are the initial conditions for the 11 primary tracers (state variables) including; dissolved organic nitrogen (DIN), non-diatom, diatom, silicon in diatom, silica, detritus, microzooplankton, mesozooplankton, non-diatom chlorophyll, diatom chlorophyll, and iron (mmol m^{-3}), along with the model parameter values. Initial ~~concentrations and therefore in situ data conditions for chlorophyll, silicate, iron, and nitrogen concentrations and~~ are taken from: ~~(1) station ALOHA in the Pacific ocean (22°45'N-158°00'W) part of the Hawaii Ocean Time Series (HOTS), downloaded from, (2) Bermuda Atlantic Time Series (BATS) in the subtropical North Atlantic (32°-50°N, 64°-10°W) available at, (3) the Cariaco basin (10°30'N, 64°40'W) obtained from, (4) Porcupine Abyssal Plain Sustained Observatory (PAP) in the Northeast Atlantic (?) located in 49° N, 16.5° W, taken from, and (5) station L4, part of the Western Channel Observatory located at 50°-15°N, 4°-12.3°W (Smyth et al., 2010; ?), and the data is available at~~. These stations the in situ data at the five oceanographic stations. We did not use spin up runs when initialising, as discussed in the supplementary material section S1. Location coordinate, data source, and maximum depth are summarised in Table 2 and the stations locations are shown in Fig. 2. After initialization, in situ data from these stations are used to validate the model results. For station PAP, we also use SeaWIFS-derived chlorophyll-a data with 9 km spatial resolution and 8-day ~~averaged averages~~ provided by GlobColor (<http://hermes.acri.fr/>) for validating the surface chlorophyll.

At these stations, the DIN consists of ammonia, nitrate, and nitrite, however at oligotrophic stations like ALOHA the ammonium is below the detection limit (Hawaii Ocean Time Series), and therefore DIN only consists of nitrate and nitrite. ~~At station PAP, PAP~~ we use the initial condition from one of MarMOT's test stations, located at 50°N, 20°W (Hemmings et al., 2015), since the nitrate data ~~was were~~ only collected between 30-400 m. At station L4, chlorophyll and nitrogen data were collected from the surface from 1999-2008. Therefore the initial ~~concentration concentrations~~ for chlorophyll and nitrogen are the same at every depth (total chlorophyll = 0.27 mg m^{-3} , nitrogen = 6 mmol m^{-3}). Other inputs that are not available at the websites

mentioned above, such as microzooplankton, mesozooplankton, and detritus were taken from the ~~the~~ nearest test stations. In the oligotrophic stations, 75 % of total chlorophyll was allocated initially to the non-diatom phytoplankton since these dominated the water column (Villareal et al., 2012). At the other stations, ~~half~~ of the total chlorophyll goes into the diatoms.

For validation of the model, we consider the total chlorophyll-a concentration, instead of separating diatoms and non-
5 diatoms. ~~Simulations are made~~ The model is simulated at 37 depth levels, spanning from 6-1200 m to minimise computational cost, ~~except for coastal stations where the overall depths are shallower (up apart from station L4, with maximum depth is 50 m, and Cariaco, where the maximum depth for the physical input is available down to 500 m for station Cariaco and 50 m for stations L4). The depth levels are similar to that in ORCA025-NEMO model output.~~ At the lowest level, vertical velocity and diffusion are set to zero and this level is used as the a sink for detritus. Additionally, apart from the physical input files, ~~a~~ time
10 series for soluble iron flux from dust deposition is applied, but this is ~~kept~~ constant using the average value from (Mahowald et al., 2009).

2.6 Model Metrics

We use ~~a list of statistical metrics, such as,~~ statistical metrics including correlation coefficient, root-mean squared error (RMSE), bias, ensemble range, and 10-year mean, ~~which averages the whole 10-year time-series at which both in situ data
15 and ensemble results are available for that particular time, for the~~ depth profiles of nitrogen and chlorophyll and integrated chlorophyll. For surface chlorophyll, apart from the metrics ~~mentioned above, we used inter-annual mean, which averages the chlorophyll abundance at above we use the mean chlorophyll abundance~~ each year in order to see inter-annual variability, and monthly abundance, ~~to observe the seasonal dynamics of chlorophyll, for to the seasonal variations. A similar approach is applied to nitrogen, however we use the integrated nitrogen over 200 m (integrated nitrogen / depth) to calculate
20 the inter-annual mean and monthly abundance.~~ These statistical metrics are ~~used to compare it compared~~ with in situ data. ~~Additionally, to capture the emergent properties of phytoplankton dynamics, we~~ We also consider the phenological aspects of the phytoplankton spring bloom, which are useful ecological indicators for detecting natural and anthropogenic impacts on the pelagic ecosystem (Platt and Sathyendranath, 2008). We consider seven phenology indicators as metrics ~~to investigate how structural sensitivity affects the model simulations. These indicators are centered around the phytoplankton blooms. Before the
25 blooms peak, we consider, including~~ an initiation time where the chlorophyll concentration exceeds a certain threshold, ~~in this ease at~~ half the concentration of the bloom peak. When the bloom concentration starts to diminish, we derived a termination time, where bloom concentration falls below the same threshold. The number of days when chlorophyll concentration is higher than the threshold is ~~taken as~~ the bloom duration. The concentration at the bloom peak and the date it takes place, are also included as indicators. ~~Additionally, we also noted~~ We also note the amplitude of the bloom, which is half of the peak height
30 minus the minimum chlorophyll concentration. ~~The~~ These indicators are derived using the method described in appendix A, and applied to all ensemble outputs for each year.

In an ensemble forecast system, an ensemble with good reliability is the one that is statistically consistent with the observations, such that the observation is statistically indistinguishable from the ensemble members. In order to assess the value of the ensemble probability distribution we must assess the consistency of the ensemble spread as well as the ensemble mean error

(Moradkhani and Meskele, 2010). A simple method is discussed by Anderson (2001) which takes the ratio R_a of RMSE of the ensemble mean and the mean RMSE of all the ensemble members which has the expectation value $E[R_a] = \sqrt{\frac{(n+1)}{2n}}$, where n is the number of ensemble members. This is called the Normalised RMSE Ratio (NRR= $R_a/E[R_a]$) where the desirable ensemble spread is expected to have NRR=1. If the NRR >1 then the spread is too small, and NRR <1 indicates that the ensemble spread is too large. We may expect different NRR values for different metrics and also for variability on different timescales, such as monthly or inter-annual data. This method has previously been used to set the number of ensemble members in data assimilation (Moradkhani et al., 2006; Roy et al., 2012).

3 Results

~~A selection of ensemble results are presented in order to provide a summary of the effect of perturbing model formulations of nutrient uptake, zooplankton grazing, and mortality simultaneously. These have been done at the five oceanographic stations which can be classified into three regional types: abyssal plain (PAP), oligotrophic (BATS and ALOHA), and coastal (Cariaco and L4). First the ensemble range and mean are compared with the observational fields (described in the method section), followed by the error statistics calculated for the ensemble mean/median, the default run, and the ensemble range in order to assess whether the ensemble spans the observational data. Then variability from the default run and the ensemble are compared with the in situ data, followed by comparing the NRR to assess the ensemble spread, and phytoplankton phenology, in order to investigate whether the ensemble range has captured the events that lead to phytoplankton bloom initiation and termination.~~ bloom phenology.

3.1 Abyssal Plain

In ~~this station~~ station PAP, in situ nitrate was only measured from mid 2002 to mid 2004 with a maximum depth of ~~300~~ 400 m and chlorophyll from mid 2003 to mid 2005 with maximum depth of 200 m, as in Table 2. Surface chlorophyll is derived from SeaWiFs (8-day averaged) and is available for the full 10-year time series (see supplementary Fig. S5).

Distinct seasonality ~~has been is~~ simulated by the ensemble mean. High nitrate concentrations at the surface occur during ~~the~~ winter (December-April) and decline in ~~the summer. However, below 400 m mostly continuous high (> 10 mmol m) nitrate concentration is present, shown on Fig. 3(d) summer.~~ From the seasonal mean nitrogen profile in Fig. 7(a), the ensemble inter-quartile range shows later bloom peaks compared to in situ, and mean nitrogen concentrations are high during months when both the ensemble and in situ nitrogen decline (between May to June), and show an earlier spike of nitrogen in July instead of September, and therefore underestimating the increase of nitrogen between October to December. These shared errors make the ensemble spread (NRR=1.25) still too narrow for the phenological metrics.

Chlorophyll concentration starts to decline at a depth of ~ 50 m, which also corresponds to the decline in the chlorophyll inter-quartile (between 25th and 75th percentile) range shown on Fig. 3(b). Chlorophyll also shows seasonality, similar to that ~~in nitrogen. In the in situ, of nitrogen. The in situ data show~~ high concentrations of chlorophyll ~~are recorded during May-June,~~

in the top 70 m during May-June, coinciding with the shallowing ~~of the~~ mixed layer depth. ~~However in~~ In the model this occurs earlier in spring (between end of April to May), and slightly deeper, to 100 m, as summarised in Fig. 3(a) and 3(c).

~~From table 3 chlorophyll profiles correlate better than nitrogen.~~ Chlorophyll and nitrogen profile 10-year means are also within the ensemble range, ~~although its ensemble spread is narrow as their NRR values are~~ with ensemble spread on the
5 narrow side, with NRR= 1.20 and 1.25 for chlorophyll and nitrogen respectively. ~~In terms of~~ For chlorophyll and nitrogen profiles, the ensemble median shows the highest correlation and lowest RMSE and bias, compared to ~~ensemble mean and default~~ the default or the ensemble mean. High RMSEs in nitrogen occur from ensemble members that contain U_2G_2 , U_3G_2 , ~~and the~~ U_4G_2 combinations, as shown in Fig. ~~H13~~ 13(j), which also correspond to high nitrogen mean ($< 5-9$ mmol m^{-3}), ~~apart from ensemble members that contain~~ $\rho_2\xi_3$, $\rho_1\xi_2$, $\rho_3\xi_3$, and $\rho_1\xi_4$ combinations. High chlorophyll profile RMSEs (>0.62) are
10 produced from members that combine G_1 with $\rho_1\xi_2$, $\rho_3\xi_3$, and $\rho_1\xi_4$ combinations. ~~Similar to nitrogen,~~ and this coincides with high chlorophyll mean (> 0.7 mg m^{-3}). Surface chlorophyll 10-year mean and RMSEs (>20.8) are notably high when combining U_2 with $\rho_2\xi_3$, $\rho_1\xi_3$ U_1 with $\rho_1\xi_2$, $\rho_3\xi_3$, and $\rho_1\xi_4$, as summarised in Fig. 12(e) and (j).

When compared to satellite-derived chlorophyll-a, the surface chlorophyll at this station ~~has higher correlation than shows~~ higher correlations and lower RMSEs than in other regions, especially ~~using ensemble median output, which also have lower~~ RMSEs with the ensemble median, compared to the ~~ensemble mean and default run~~. ~~In some years the default run or the~~ ensemble mean. In years when satellite-derived chlorophyll is not within the ensemble range ~~; this is~~ due to the ensemble range overestimating the satellite-derived chlorophyll (supplementary material Fig. S5). ~~Additionally, in terms of,~~ giving a slightly
15 narrow ensemble spread (NRR=1.29). There is also inter-annual ~~mean, only in decline in satellite-derived chlorophyll,~~ $(r = -0.21, p < 0.05)$. Six ensemble members capture this decline in surface chlorophyll, although with weaker correlations $(r =$
20 $-0.14 (\pm 0.06), p < 0.05)$. ~~In certain years (1998, 1999, and 2001) is satellite-derived chlorophyll the observed chlorophyll are~~ not within the ensemble inter-quartile range, although in other years they are within the full range, but outside the inter-quartile range, summarised in range, Fig. 4(a). ~~This gives a "~~ making the ensemble spread too narrow" ~~ensemble spread~~, with NRR of 1.26. ~~A decline in surface chlorophyll over the time has also been recorded in the satellite observations ($r = -0.21$ $p < 0.05$),~~ however only six ensemble member capture the decline in surface chlorophyll, with weaker correlations ($r = -0.14 (\pm 0.06)$, p
25 $< 0.05)$. ~~For monthly data, the satellite-derived surface chlorophyll concentrations are mostly within the ensemble range,~~

The range of surface chlorophyll annual mean is 0.7 mg m^{-3} . If we only allow one process function at a time to change the ensemble range reduces to 0.58 mg m^{-3} , covering 84% of the all ensemble members. If the original MEDUSA parameters are used, the interannual chlorophyll fits the ensemble slightly better, but the nitrogen fit gets worse. The results from using MEDUSA parameters and in situ nitrogen concentration as initial condition can be found in the supplementary material,
30 section S2 and closer to the ensemble median, as shown on S3. For monthly data (Fig. 5(a)-In low chlorophyll months ($<$), the ensemble shows only slightly earlier peak chlorophyll concentrations in May compared to in situ, which occur in either May or June. However, since the ensemble mean and median overestimate the satellite-derived chlorophyll during months of high chlorophyll (> 0.5 mg m^{-3}) from November-March, the satellite-derived chlorophyll is within the 75th and 25th quartiles. Although in the time series the, during peak chlorophyll in May the satellite-derived chlorophyll ~~sometimes fall outside~~

of is outside the ensemble range, ~~the overall ensemble monthly means show the highest monthly mean surface chlorophyll concentration occurring between May and June, similar to the satellite-derived chlorophyll, shown on Fig. 5(a).~~

3.2 Oligotrophic Ocean

In oligotrophic regions nutrients are ~~expected to be~~ scarce at the surface but may be plentiful at deeper depths (Dave and Lozier, 2010; Lipschultz, 2001). ~~All ensemble members represent this distribution well for ALOHA, as seen in Figure 9(d) and (e), show that the ensemble range decreases as the depth increases, with high ensemble range found at depths between 3-50 m. At ALOHA all ensemble members have similar distributions for in situ nitrogen, Fig. 9(e), with nitrogen levels > 1.0 mmol m⁻³ only at found ~ 150 m depth. However at BATS, from January 1999, for inter-annual means, Fig. 4(b), the ensemble inter-quartile range mostly overestimates the in situ observations. There is also an increasing trend of in situ nitrogen ($r = 0.69, p < 0.03$), which is not captured by the ensemble, Fig. 6(b), leading to an NRR of 1.38. This overestimation is also observed in the seasonal mean, Fig. 7(b), and in situ data is rarely within the ensemble range.~~

~~At BATS, the nitrogen concentration in the top 200 m is clearly overestimated, Fig. 9(k), with nitrogen levels > 1.0 mmol m⁻³ at ~ 10 m (with some members occasionally showing such concentrations at 3 m). Higher ensemble inter-quartile ranges are found between 3-50 m, and this range decreases with depth, shown~~ Consequently, the overall mean nitrogen concentration is overestimated, as indicated by the positive bias in the ensemble mean, in Table 3. Similar to ALOHA, nitrogen inter-annual and seasonal means are overestimated, summarised in Fig. 9(d) and 9(j). Mean nitrogen concentration is overestimated as indicated by the positive bias from the ensemble mean, as shown in Table 3. 6(c) and 7(c), respectively. This results in narrow ensemble spread (NRR= 1.40). As at ALOHA, an increasing trend in nitrogen is observed ($r = 0.67, p < 0.03$), but only 28.9% of the ensemble results, which uses G_2 as its grazing function shows similar trend.

Another feature of the oligotrophic ocean is a deep chlorophyll maximum (DCM) that occurs below the mixed layer (Fennel and Boss, 2003). In Fig. 9(b) and 9(h), high chlorophyll concentrations are simulated by the ensemble mean between 70-90 m in BATS and up to 150 m in ALOHA. A DCM occurs when lower chlorophyll is detected at the surface, which roughly matches with the in situ profiles at ALOHA (see Fig. 9(c) and Letelier et al. (2004)) and BATS (Fig. 9(i)) although the depth of the ~~DCM is slightly ensemble DCM at both stations is 10-20 m shallower than in situ (down to 150 m).~~ The high subsurface chlorophyll coincides with a higher ensemble range, with the range decreasing with depth. However neither BATS nor ALOHA show ~~a continuous DCM as the continuous DCM~~ seen in the in situ profiles, Fig. 9(c) and 9(i).

The majority of ensemble members underestimate in situ 10-year mean chlorophyll profile concentrations, especially at BATS where all ensemble members show positive bias ~~towards for~~ both surface and integrated chlorophyll profiles. ~~This in turn results in NRR > 1, showing that the ensemble spread is too narrow. At station ALOHA, in situ chlorophyll 10-year means (surface, profile, and integrated) are always within the ensemble range. In contrast, the modelled 10-year mean nitrogen from the ensemble mean and median are more than twice the in situ observations, also leading to a narrow ensemble spread with the ALOHA NRR value for nitrogen being the largest, summarised in Table 3. At BATS some members show a very low chlorophyll mean Figure 11(a) and (b) show that ensemble members with $G_2, \rho_2\xi_2, \rho_2\xi_3, \rho_2\xi_4, \rho_3\xi_1,$ and U_3 produce lowest chlorophyll concentrations (< 0.015-0.08 mg m⁻³) and high nitrogen concentrations (> 0.34 mmol m), see, which then~~

coincide with high RMSE, shown in Fig. 11(b), 11(e) and (f), 12(l), and 12(q), resulting in high RMSEs for both variables. Most of these members use functional combinations G_2 , $\rho_2\xi_2$, $\rho_2\xi_3$, $\rho_2\xi_4$, $\rho_3\xi_1$, and U_3 . The low chlorophyll concentrations, coinciding with high RMSEs, also come from the same ensemble members as for station ALOHA, except for U_3 . The magnitude and range of RMSEs at BATS are highest, and the default run has lower RMSEs than ensemble mean and median, summarised in Table 3. Ensemble members that use U_1G_1 and U_4G_1 show higher profile 10-year mean concentration concentrations of chlorophyll at both stations, Fig. 11, although when paired with $\rho_3\xi_3$ and $\rho_1\xi_4$, the RMSEs increase. High nitrogen concentrations are almost always observed when U_3 and U_2 were used in the oligotrophic regions, these oligotrophic regions, summarised in Fig. 13(a) and (b).

Surface As for profiles, surface chlorophyll at ALOHA (supplementary Fig. S2) has lower RMSEs and higher 10-year mean concentration compared to those at concentrations compared to BATS, summarised in Table 3. Low chlorophyll with high RMSEs (> 0.1) have not been observed in station ALOHA. Ensemble members with lower low surface chlorophyll concentrations were similar to the observation profiles, and are the same as those with low chlorophyll profiles, although high surface chlorophyll RMSEs also coincide with high surface concentrations, summarised in Fig. 12(a) and (f). The low RMSEs for surface chlorophyll at ALOHA is are also reflected in the NRR, with a value (NRR = 1.07) close to unity, although slightly narrow, and the ensemble almost always encompasses the in situ observations. During low chlorophyll, (see supplementary material Fig. S2). During summer months (June-September), most of the ensemble members still underestimate the in situ monthly mean means, summarised in Fig. 5(b), and not all peaks are peak values are not always covered by the ensemble.

Figure 4(b) and (c), shows that there is show no distinct inter-annual variability at either ALOHA or BATS. Figure 5(b) and (c) shows that the highest mean of in situ chlorophyll concentration is usually found in December (0.13 mg m⁻³) and April (0.28 mg m⁻³), at station ALOHA and BATS respectively, and these are within the ensemble range. At station, and in most years the default run is closer to in situ. Model output at BATS have a lower 10-year mean surface chlorophyll than in situ data and most ensemble members underestimate the surface inter-annual means, making the ensemble spread appear narrow both in the 10-year and inter-annual means, shown on Table 3 and 4. Changing the functional forms one at a time produces an ensemble range of 0.11 and NRR = 1.39, slightly wider than the whole ensemble, summarised in Fig. 8. At BATS in 2004, a high in situ chlorophyll mean was recorded (0.65 mg m⁻³) that was not captured by all and none of the ensemble members captured this high mean, see the supplementary material Fig. S1 and 4(c). Since model outputs at BATS have a lower 10-year mean chlorophyll than in situ data, most of the ensemble members underestimate the surface, which therefore narrows the NRR value for annual mean at BATS. At ALOHA the range for inter-annual means is too wide, with NRR = 0.84. This is broader than the overall time series mean, as the in situ inter-annual means, therefore making the ensemble spread too narrow both in the 10-year mean and are mostly within the 75th quartile, making the inter-annual mean, shown on Table 3 and 4. This is also reflected in the monthly mean, whereby in the mean RMSE of the ensemble higher than the RMSE of the ensemble mean (0.043 and 0.025 respectively). However, when only one process is perturbed at a time, the NRR becomes narrow (1.17), and the in situ is no longer within the inter-quartile range, as shown on Fig. 8 and summarised in Table 4. In 1999, the ensemble mean and median is higher than the in situ, which is even clearer in the inter-annual mean of the primary production.

Unlike at PAP, there is no distinct seasonality in chlorophyll. At ALOHA during months of low concentration in the ensemble and in situ chlorophyll ($< 0.1 \text{ mg m}^{-3}$, occurring in July-October), the in situ concentrations are above the ensemble range. At ALOHA, the inter-annual mean, the in situ data are mostly within the 75th and 25th quartiles. This is also shown in the monthly means, especially during high in situ (> 0.1) surface chlorophyll months from November-January most ensemble members underestimate surface chlorophyll. At BATS, in situ concentrations are clearly underestimated during the same months, summarised in Fig. 5(b). However, the range of the ensemble for inter-annual mean at station ALOHA is seen to be too wide, with NRR of 0.84, as the in situ inter-annual means are mostly closer to the 75th quartile, making the mean RMSE of the ensemble higher than the ensemble mean's RMSE (0.043 and 0.025 respectively). At ALOHA, between December to May, when in situ chlorophyll seasonal means are $> 0.1 \text{ mg m}^{-3}$, the in situ data are within the inter-quartile range, but at BATS this only happens when in situ chlorophyll means decrease from 0.20 to 0.08 mg m^{-3} in May.

At station ALOHA the ensemble mean and median produce smaller errors RMSEs for both chlorophyll and nitrogen. In although in the depth profiles, bias compared to BATS. Bias in the default run is still smaller than for the ensemble mean and median. However the for surface and integrated chlorophyll show that the ensemble mean and median produce have lower bias than the default concentrations. This is the opposite for BATS where both RMSEs and. At BATS low RMSEs and bias with high correlation coefficient are higher for produced by the default run compared to ensemble mean and median, as well as for the biases. At both stations, integrated chlorophyll from ensemble mean and median shows have smaller RMSEs and a better correlation coefficient, compared to the default run. At ALOHA, NRR for the integrated chlorophyll is closer to 1 compared to either than for the surface and chlorophyll profiles. However the default run in oligotrophic regions generally produces higher chlorophyll and lower nitrogen concentrations compared to the ensemble mean and median. This also matches better with in situ patterns data as the correlation coefficient, r is higher. This is because using Using U_1G_1 gives is seen to give rise to higher chlorophyll concentrations.

3.3 Coastal

In the coastal stations, in situ observations show strong seasonality, shown on Fig. 14(c), (f), (g), and (h). In terms of the inter-annual mean, the ensemble range at stations Cariaco and L4 always includes the observations (Fig. 4(c) and (d)), despite the ensemble spread mostly being quite narrow, as described by the NRR values in Table 3.

The in situ profiles at Cariaco show high chlorophyll concentrations ($> 1 \text{ mg m}^{-3}$) within the upper 30 m occurring between January to February (see This has not been captured by the ensemble mean in station Cariaco, shown in Fig. 14(e)). This coincides with the rise of nitrogen from deeper depths to $\sim 30 \text{ m}$, as seen on Fig. 14(f), increasing the nitrogen concentration to $\sim 5 \text{ mmol m}^{-3}$. However this is not captured by the ensemble mean, a, with chlorophyll concentration almost constant above 0.7 mg m^{-3} in the upper 30 m, as shown in 14(a) and the surface (supplementary Fig. S3), shown in Fig. 5(d). Since the range of seasonal chlorophyll is wide, apart from in August and November, in situ concentrations are within the ensemble range. A decline of chlorophyll has been recorded in station at Cariaco from 2004 (Taylor et al., 2012), and this has been is captured by the ensemble mean, median, and default ($r=r=-0.72$, $p<p<0.05$, $r=r=-0.66$, $p<p<0.05$, and $r=r=-0.35$, $p<p<0.05$ respectively). For nitrogen the seasonal upwelling is not captured, although in Similar to chlorophyll, nitrogen

from the ensemble also shows no seasonality, see Fig. 7(d). Nonetheless, in situ concentrations are still mostly within the ensemble range, apart from November, where in situ nitrogen decreases to 5.32 mg m^{-3} . However in 2001, and between 2005-2006, inter-annual downwelling of nutrients are reproduced, summarised in Fig. 14(d). Figure 14(e) shows the inter-quartile range for nitrogen increasing below $\sim 40\text{m}$ and then decreasing again at $\sim 100 \text{ m}$. Similarly on Fig. 14(b) chlorophyll interquartile range is high at depths where chlorophyll is plentiful. Despite the lack of seasonality, annual means of chlorophyll and nitrogen, are mostly within the ensemble range, Fig. 6(f) and 4(f), with the NRR 0.78 and 1.15 for chlorophyll and nitrogen respectively.

At station L4 the in situ and ensemble means both show seasonality of nitrogen with high concentration ($> 8 \text{ mmol m}^{-3}$) occurred during November to February, and close to zero ($> 0.1 \text{ mmol m}^{-3}$), during summer months, consistent with the observation from Smyth et al. (2010). For chlorophyll Figure 14(g) shows sharp peaks in spring time, sharp peaks are observed during spring (March-April) and fall (September) for in situ data, and the ensemble means peak around one month later (May-June), without a distinct secondary peak, similar to the typical North Atlantic spring bloom (??). Observed chlorophyll concentrations generally range from coinciding with the sharp decline of nitrogen in spring. However this has not been represented in the model, where only one peak is simulated between May-June, summarised in Fig. 5(e). If only diatom chlorophyll concentration is shown, the two bloom events are clearer, especially in the default run (see supplementary material Fig. S4). The ensemble mostly overestimates the in situ during non-bloom periods (in situ range = $0.09\text{-}2 \text{ mg m}^{-3}$, apart from the sharp increases during bloom events (up to 6.41 mg m^{-3}), yet during non bloom period, ensemble range from $0.28\text{-}3.13 \text{ mg m}^{-3}$ and during bloom events, the highest peak is 5.95 mg m^{-3} , therefore the surface chlorophyll is not fully captured by the ensemble. This is reflected by the high NRR value of 1.31, and therefore NRR value is 1.21, indicating a too narrow spread.

Both stations show weak positive correlations of surface chlorophyll from the ensemble mean, summarised in Table 3. For the inter-annual mean, the ensemble range always includes the observations (Fig. 4(c) and (d)), the NRR values are given in Table 3. The ensemble mean and median show better correlation and smaller RMSEs compared to the default run, apart from nitrogen at station L4. Chlorophyll is biased at both stations, for the ensemble mean at Cariaco, and for the ensemble median at L4. Integrated Similar to the oligotrophic stations, the integrated chlorophyll shows better correlation with in situ observations at station Cariaco, compared to both surface and chlorophyll profiles. Nonetheless, compared to other oceanic regions, Cariaco still has the highest RMSE for both chlorophyll profile and surface values. At L4, the ensemble mean shows high RMSE for surface nitrogen, but low RMSE for surface chlorophyll, see summarised in Table 3.

Although from Table 3, in situ surface chlorophyll concentrations are slightly overestimated by the ensemble mean, most of the ensemble outputs at Cariaco are underestimated other ensemble outputs are underestimated at Cariaco, except for ensemble members that use the combinations $\rho_2\xi_3$, $\rho_1\xi_2$, $\rho_3\xi_3$, and $\rho_1\xi_4$. This in turn makes the ensemble spread narrow, as indicated by the NRR value appear too narrow in the NRR. Unlike the oligotrophic regions, these high chlorophyll concentrations also

coincide with higher RMSE (> 1.7). Higher nitrogen concentrations ($> 1.2 \text{ mmol m}^{-3}$) with high RMSEs (> 1.5) are also associated with the same ensemble members. ~~Despite this, these members, summarised in Fig 13(c). These same ensemble members however~~ show relatively low nitrogen concentration ($> 7.5 \text{ mmol m}^{-3}$) at station L4. The chlorophyll mean at L4 shows that high concentrations ($> 0.9\text{--}0.2 \text{ mg m}^{-3}$) are produced when ~~G_1 is paired with the model uses $\rho_2\xi_3, \rho_1\xi_2, \rho_3\xi_3,$ and $\rho_1\xi_4$.~~ ~~These also combinations. These~~ coincide with high RMSEs, especially in members which pair ~~U_1 and U_2 and $\rho_2\xi_3, \rho_1\xi_2, \rho_3\xi_3,$ and $\rho_1\xi_4$.~~ ~~Low chlorophyll concentrations and RMSEs at the coastal stations are produced from U_2G_2 and $U_3G_2,$ and additionally U_3G_1 in station L4.~~ High nitrogen concentrations ($> 9 \text{ mmol m}^{-3}$) are produced by U_4G_2 , with correspondingly high RMSE.

Surface chlorophyll at ~~coastal stations~~ these coastal stations also has a higher relative range than other stations, with L4 ~~showing the higher range compared to Cariaco, summarised in higher than Cariaco, see~~ Table 3. ~~The wider spread for annual means compared to the monthly data over 10-years have been observed at all the stations, including the coastal stations.~~ Despite having lower range than L4 in terms of surface 10-year ~~mean, in the annual mean means, for the inter-annual means~~ (Fig 4(d) and (g)), the NRR value for Cariaco is too small (0.78), indicating the ensemble spread is wider than necessary.

~~At station L4, the in situ, which is also observed for~~ inter-annual ~~means are closer to the ensemble median, indicated by the smaller bias and RMSE compared to both the default and ensemble mean, shown on Table 3 and Fig. 4(e).~~ ~~Despite the narrow ensemble range in the overall mean of surface chlorophyll, the spread for the overall primary production, shown in Fig. 10(b) and (c). However, if processes are perturbed one at a time, the NRR is closer to the ideal ensemble range (NRR= 0.90). On the other hand, at L4 the in situ annual mean is reliable (NRR=1.001) and in situ means are almost always, since the in situ chlorophyll is~~ close to the ensemble ~~mean and median.~~ ~~In the monthly means, shown in Fig5 from September–April, in situ observations are within the ensemble range, however, in the summer months when chlorophyll starts to decline (May–August) due to the exhaustion of nutrients, in situ monthly means are below the ensemble range. This in turn indicates that seasonally, the ensemble does not always cover the in situ observations, making the spread for the overall 10-year mean too narrow. The highest in situ monthly mean chlorophyll concentration occurs in April, yet median (see Fig. 4(e)). However if the ensemble is reduced by only perturbing one process at a time the NRR increases to 1.36, and the in situ data is no longer within~~ the ensemble ~~mean and median show this peak in June, and the default run in May. There are also two peaks that occur in the in situ monthly means, one in April, and the other in September. If only diatom chlorophyll concentrations is shown, the two bloom events are shown better, especially in the default run (see supplementary material Fig. S4).~~ ~~shown on Fig. 8, despite the range still covering 86% of the full ensemble (2.14).~~

3.4 Phytoplankton Phenology

At most stations, the phenology metrics are covered by the ensemble range. There are differences in the timing of phenological events between the ensemble mean, median, and default run, ranging from a couple of days to a couple of weeks, as shown in Table 4. The timing of initiation, bloom peak, and termination show wide interquartile ranges for all stations and can lie between ~ 20 and 100 days earlier than the in situ timing, apart from stations PAP and ALOHA, see Fig. 15(b). At stations PAP and ALOHA the inter-quartile range is at least ~ 40 days too early. However, the ensemble mean and median at station L4 and

Cariaco are later than in situ timings. For initiation both stations are two months late ~~and are but are stil~~ within the ensemble range. In terms of the timing of the bloom peak and termination, ~~they these~~ are up to 3 months ~~late and~~ 120 ~~days too day~~ late respectively.

BATS has the largest range of phenological timings, especially in termination time. ~~In terms of For~~ initiation, the in situ timing is within the interquartile range and only three days earlier than the ensemble median. However, ~~in at~~ ALOHA the initiation time shows more inter-annual variability (supplementary Fig. S6) ~~eg. in some years bloom initiation and~~ may occur in June, August and October, as well as ~~in~~ December and January. This causes the mean observed initiation time to ~~become~~ ~~end up~~ in May. From Fig. 15(a), the ensemble run shows a mean initiation time between late January and April instead and so the observations fall outside the ensemble range. Due to this variable initiation, although ~~peak~~ bloom time is within the ~~full~~ ensemble range at ALOHA, the timing is outside the 75th and 25th percentile range, making the ensemble spread too narrow (NRR=1.35). The peak chlorophyll at ALOHA shown in Fig. 5(b), where high ($> 0.1 \text{ mg m}^{-3}$) chlorophyll monthly means are recorded in June, August, and September as well as December and January, yet the ensemble mean and median show highest concentrations only in January and February, also placing the bloom timing outside the inter-quartile range, ~~summarised in~~ ~~see~~ Fig. 15(b). At BATS the earliest initiation ~~is mid January~~ in the ensemble ~~is mid January~~, but the earliest in situ initiation occurs in February. Therefore, peak bloom time from the ensemble at BATS are usually later than in situ. However, ensemble estimates of bloom peaks for 30°N, where BATS is located, agree with a study by Racault et al. (2012), who identify early April as the peak time. ~~Although the range of peak bloom time in BATS is very high (174 days), the NRR suggests that this range is still narrow (1.17), because most ensemble members produce blooms between April and May, and the in situ timing occurs on 29 March and so is still outside the interquartile range. The large full range is caused by some ensemble members blooming much later. However, since the in situ timing is earlier, it is not within most of the ensemble range, so the overall ensemble spread appears narrow.~~

Both coastal stations show in situ initiation typically happens in mid-March, ~~and these are which is usually~~ within the ensemble range, which spans ~~for~~ 100 days (between the end of February and late June). The ensemble means show later initiation, with the 75th and 25th spanning mid April to end of May for Cariaco, and between early and mid May for L4. This later timing is also clear in peak bloom times, shown on Fig. 15(b). Figure 5(e) shows the in situ bloom at L4 is one to two months overestimated by the ensemble. Cariaco is the only station with peak bloom time, duration, and termination outside the ensemble range, due to the lack of chlorophyll seasonality, as ~~explained noted~~ in section 3.3. ~~This results in the timing of initiation, bloom peak, duration, and termination having high, also resulting in higher~~ NRR values.

Initiation timing is captured best at station PAP, with the ensemble median's initiation ~~only averaging averaging only~~ eight days earlier than for the satellite-derived chlorophyll, resulting in NRR ~~for initiation closest to one (1.14) = 1.14 for initiation, closer to one~~ compared to other stations. A typical North Atlantic bloom happens during spring (Raymont, 1980), however most blooms at PAP occur in late May-early June, as shown in Table 4 and Fig. 5. Later blooms are recorded from satellite-derived chlorophyll-a in 2005, three months later than the average and much later than the ensemble mean and median, although the bloom timing is still within the ensemble range, although the ~~range itself is still narrow, according to the NRR value (NRR value is 1.31)~~. At L4, also in the North Atlantic, the spring bloom is in April, but most ensemble members show later initiation

and peak bloom time, mostly in June. Due to this delay the NRR values at L4 indicate that the ensemble range is too narrow, although still within the full ensemble range. Ensemble mean and median at PAP show good agreement with in situ termination date. ~~Although, and although~~ other station termination times are also within the ensemble range, most are later than the inter-quartile range. However, at ALOHA, located at 22°N, the ensemble median for termination at the end of August ~~falls close to~~
5 ~~agrees with~~ the observations from Racault et al. (2012).

Compared to running only the default MEDUSA, where only a single ~~mean peak value bloom peak~~ is produced, the ensemble range mostly encompasses the in situ peak amplitudes, shown on Fig. 15(c). Only at BATS are ~~the~~ in situ peak height and amplitude outside the ensemble range, ~~resulting from the narrow ensemble range seen from the NRR value~~. This is expected since most of the ensemble members underestimate in situ chlorophyll. At Cariaco, in situ peak heights are within the ensemble
10 range, but observed peaks are higher (mean= 3.5 mg m⁻³, maximum peak= 7.7 mg m⁻³), and the ensemble reaches less than half of the in situ peak (mean= 1.2 mg m⁻³, maximum height= 5 mg m⁻³). This ~~underestimates underestimate of~~ the peak and ~~consequently also the amplitude, resulting bloom amplitude, results in~~ NRR of 1.40 and 1.39 respectively. Ensemble members with higher peak and amplitudes are also those with higher chlorophyll biases. Despite the narrow ensemble range, at L4 chlorophyll peaks are within the 75th and 25th range box, and ~~its amplitude is amplitudes are~~ within the full spread. ~~In contrast~~
15 ~~stations~~ ALOHA and PAP have reliable ensemble spreads according to their NRR values for peak height (see Table 4).

Similar to peak heights, the bloom durations at most stations are within the ensemble range, apart from station Cariaco, which shows the narrowest ensemble spread according to its NRR ~~value~~. The duration at Cariaco is overestimated because the peak is very wide (up to 143 bloom days). This, along with the late initiation of the bloom, results in a three month late termination. At ALOHA, duration is outside the 75th and 25th quartile box, since the peak is also much broader compared to in
20 situ blooms. This results in too narrow ensemble mean according to the NRR ~~value~~. The opposite is true at BATS where in situ peaks are generally broader, and the ensemble members with lower chlorophyll concentration showing narrower peaks, and a greater range in bloom durations, which consequently lowers the NRR value.

4 Summary and Discussion

In this paper we have investigated structural sensitivity ~~of~~, associated with the mathematical formulation of the processes in an
25 intermediately complex biogeochemical model by generating its ensemble outputs of chlorophyll and nitrogen and comparing ~~them~~ with a single default run, and with in situ observations at five oceanographic stations. The ensemble consists of 128 ensemble members, each with different process function combinations. ~~Following the work of Fussmann and Blasius (2005) In~~ order to maintain phenomenological similarity, these functions ~~have been previously calibrated, are calibrated~~ using non-linear least squares, ~~and while~~ keeping the maximum process rates fixed ~~in order to maintain phenomenological similarity~~. We have
30 chosen nutrient uptake, zooplankton grazing, and plankton mortalities to vary, as these are the core processes of every marine biogeochemical model, from the simplest to the most complex.

~~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. For physical models, perturbed parameter~~

ensembles have been explored and utilized to quantify climate change uncertainties in a probabilistic sense (Murphy et al., 2007; Tinker et al. 2011). Here, through this approach, we provide a perturbed biology ensemble conditioned upon process structural uncertainties. Applying structural sensitivity in the 1-D framework has also allowed a large parameter space of concurrent variations to be explored for several different oceanographic regions, and with minimal computational cost. From these assessments, we find that small perturbations in model structure can produce a wide range of results, particularly regarding chlorophyll and nutrient concentration as well as phytoplankton phenology. Apart from the assessment of uncertainties arising from the structural sensitivity and the reliability of the ensemble spread, we have also compared the RMSEs against observations for the ensemble mean and median, and for the deterministic model default run.

Our findings reveal that in all regions, the Holling Type II (G_2) grazing function decreases the chlorophyll concentration, and pairing it with linear phytoplankton mortality (ρ_2) lowers the concentration even further. The nutrients respond in the opposite direction with enhanced nitrogen concentrations. This is expected as at low concentrations, using the G_2 function would graze more phytoplankton, as shown on Fig. 1(b). Pairing G_2 with the linear (ρ_2) mortality of phytoplankton, which constantly removed the phytoplankton regardless of the phytoplankton concentrations, will reduce the chlorophyll concentration even further; but the opposite will happen when G_2 is paired with linear zooplankton mortality. Yool et al. (2011) has similarly shown that using a linear mortality causes the biggest changes, and Anderson et al. (2010) show that type III or sigmoidal grazing depletes less phytoplankton at low concentrations compared with hyperbolic grazing. It is therefore consistent that the lowest chlorophyll concentrations are observed from the combination of these functional forms. We found that default phytoplankton mortality in phytoplankton concentrations compared to quadratic and sigmoidal. In contrast, the default phytoplankton (ρ_1) and sigmoidal zooplankton mortality (ξ_4) produce the highest chlorophyll concentrations in all regions, similar to the experiment from Yool et al. (2011). Linear zooplankton mortality (ξ_2) produces enhanced chlorophyll concentrations due to higher zooplankton mortalities and lower phytoplankton mortalities, especially when combined with the default hyperbolic phytoplankton mortality (ρ_1). In terms of nutrient uptake, the exponential (U_3) and sigmoidal (U_2) functions show inefficient uptake, as they produce low chlorophyll and especially high nitrogen concentrations, especially in the oligotrophic regions. Figure 1 shows that in low nitrogen regions, uptake rates using U_3 and U_2 are lower than those using the default Michaelis-Menten (U_1), or trigonometric (U_4), functions. Yet, the differences when using U_3 and U_2 compared to using U_4 and U_1 are not as large as using type II grazing and linear mortality on chlorophyll concentrations at the oligotrophic stations. Another example is found at station L4; when pairing U_4 uptake and G_2 , the ensemble produces high nitrogen concentrations but low chlorophyll concentration is not seen, Fig. 12 (as shown on Fig. 12(a), (b), 13(a), and (b), especially in the oligotrophic region. Even though the functional forms have been optimised, the most deviations are observed when nitrogen is $< 1 \text{ mmol N m}^{-3}$ shown in Fig. 1(a). This is because when phytoplankton concentration is $> 1 \text{ mg m}^{-3}$, G_2 depletes less phytoplankton compared to U_3 and U_2 , which uptake less nitrogen in low nitrogen concentrations, produce high nitrogen and low chlorophyll. However, the effect is not as noticeable compared to using G_1 , and combining this with U_4 lowers the uptake rate, as summarised in Fig. 1, leaving higher nitrogen or G_2 .

As an additional metric for the ensemble spread we computed the Normalised RMSE Ratio (NRR) (Anderson, 2001) to measure whether the ensemble has a reasonable spread, which could then be regarded as an uncertainty when using the results of the model simulations. The NRR values for the five oceanographic stations indicate that the spread is usually too narrow, especially for nitrogen profiles, with the smallest NRR station L4, and the largest at station ALOHA, shown in Table 3. Nitrogen in situ mean concentrations are generally overestimated by ensemble especially in the oligotrophic region. But the model well represents days when concentrations are high, and for chlorophyll at the oligotrophic stations, only very low chlorophyll concentrations fall within the ensemble range (see supplementary material Fig. S1 and S2).

Overall, station Cariaco shows the highest chlorophyll RMSEs. This station also produces model outputs in which high chlorophyll concentrations coincide with high nitrogen. Consistently high chlorophyll concentrations are seen throughout the time series, while the observations only show a phytoplankton bloom once per year, leading to the higher chlorophyll RMSE than at other stations. The observed timing of the bloom, duration, and termination are outside the ensemble range, and consequently the ensemble range is assessed as too narrow according to its NRR value (>1.39). At this station, chlorophyll is mostly driven by the upwelling of nutrients. These disparities between the ensemble members that are caused by the structural differences therefore garner some range. Stations that have produced high chlorophyll concentrations also have high ensemble range. For example, in station Cariaco where chlorophyll concentration is high, despite the discrepancy between the in situ seasonal pattern and the ensemble, the range still covers the in situ concentrations. The mismatch between the observation and ensemble is mostly caused by the trade winds (?). However, the upwelling is not captured well physical dynamics, in which the upwelling of nutrients that feeds the phytoplankton is not well simulated by the assimilated vertical velocity we used. Instead of chlorophyll peaks occurring between December-January throughout the time-series, the ensemble produces a constantly high chlorophyll concentration, summarised in Fig. 5(d). Despite these, This emphasises that despite using the ensemble approach, a biogeochemical model is only as good as its physical model (Doney, 1999). Even though there are discrepancies, the chlorophyll profile at Cariaco has an NRR value closest to 1, but not for the surface annual mean. Although the in situ inter-annual mean concentrations are almost always within the inter-quartile range (see due to the large range in the ensemble. However, in the annual mean of chlorophyll and primary production (Fig. 4(d)), some of the means are either closer to the upper quartile or the lower quartile. This in turn widens the ensemble spread (NRR=0.78), and 10(b)) the large range makes the ensemble spread too large. The ensemble range is considerable even if the equation of only one process is changed at a time, which accounts for at least 80% of the full ensemble range. This has been observed at all of the stations, summarised in Table 4 and Fig. 8, emphasising that perturbing functional forms will produce a large range of model results. In some cases, this reduced range may be statistically more meaningful than the full range. For example, compared with the full ensemble, the reduced ensemble range for Cariaco's annual-mean chlorophyll gives an NRR closer to unity. Therefore, it may be possible through a further study to systematically reduce the number of ensemble members, whilst retaining a realistic ensemble range, which will reduce computational costs.

The NRR for chlorophyll (mean, surface, and integrated) at station BATS is

At most of the stations, the ensemble mean produced lower RMSE compared to the default run, suggesting that the structural ensemble with a wide range of predictions covering the in situ observations, is likely to produce a mean field closer to the

observation, than a single-structure model. Even in stations such as BATS, where the in situ chlorophyll is underestimated by most ensemble members, the highest and BATS is the only station at which the default run performs better than the ensemble mean for RMSEs of chlorophyll and nitrogen concentration. The low concentrations modelled at BATS give the high NRR values for surface and profile integrated chlorophyll, indicating too narrow ensemble spread. Only the lower or 'background' chlorophyll are inside the ensemble range, but not the higher bloom peaks. Peak height and amplitude in the phenology are therefore underestimated. However in the monthly means (Fig. 5(b)) from December to April, the months where modelled chlorophyll means are high, in situ means are c)) of the in situ values (during months of high chlorophyll) are within the ensemble range. This is also observed at ALOHA, showing that in oligotrophic regions during higher chlorophyll concentrations, the ensemble range increases to encompass the in situ concentrations. The higher in situ chlorophyll concentration at BATS compared to ALOHA may be because the station is in the mode-water region of the subtropical gyre and has deeper mixed layers allowing deep nutrients to reach the surface and supply the growth of phytoplankton. In the ensemble run, the mean chlorophyll profile concentrations are very similar between ALOHA and BATS, shown on Table 3. At the surface, ALOHA shows higher mean chlorophyll than BATS, perhaps due to high concentrations ($\sim 0.24 \text{ mg m}^{-3}$). Similar monthly-mean chlorophyll patterns have also been observed in 1999, simulated at the surface at ALOHA, contributing to the overall mean, while at BATS steady low chlorophyll concentrations are found throughout the ensemble. Lower PAP, ALOHA, Cariaco and L4 (with some exceptions in summer month) (Fig. 5(b)), whereby the in situ chlorophyll is generally within the ensemble range. We further note that, considerable model bias such as lower modelled concentrations of chlorophyll, compared to the in situ, in the subtropical gyre have also been observed in the 3-D default data, has been observed for the default 3-D MEDUSA model itself (Yool et al., 2011).

Although also oligotrophic, station ALOHA shows the lowest RMSE for both nitrogen and chlorophyll, compared to other stations, especially for surface chlorophyll. At this station, the surface and integrated chlorophyll, peak height, and amplitude have the NRR close to 1 (see Table 3 and 4). However for bloom initiation ALOHA in situ timing is outside the ensemble range. In some years, the observed bloom initiation has been recorded from June to August and these patterns are captured by the ensemble. However in most years, modelled bloom initiations are between December and January. The summer chlorophyll bloom that occurs in the north pacific subtropical gyre consists of picoplankton (White et al., 2015) and may be caused by the addition of nitrogen from nitrogen-fixing organisms (Dore et al., 2008) which have not been explicitly represented in MEDUSA. This may explain the discrepancy in bloom timing. At BATS in the Atlantic subtropical gyre, phytoplankton surface blooms occur in the spring time. The initiation, bloom time, and duration are within the ensemble range, summarised on Table 4. The large range is due to, e.g., in the initiation time varying strongly with concentration, therefore also affecting the duration. For example, high concentration chlorophyll produces earlier initiation times (February-April), and years with lower concentrations show later initiation times (May-June).

Station PAP shows the best match between the observed bloom and the ensemble. At PAP, seasonality is very well defined with both nitrogen and chlorophyll concentrations affected greatly by the mixed layer depth. Phytoplankton dynamics in regions like the North Atlantic is dictated by the mixed layer deepening and subtropical gyre (Yool et al., 2011). This may be due to the absence of nitrogen fixers and picoplankton in MEDUSA, which cause the increase of plankton concentration in the

enrichment of nutrients in the surface layer (?). In terms of peak height and amplitude, although within the ensemble range with NRR values close to unity, they are still larger than 1, indicating a slightly narrow spread (see Table 4). In some years (eg. 2002 and 2006), the peak height is overestimated by most the ensemble, and in the monthly means, the month where the ensemble concentrations peak, is later, summarised in Fig. 5. summer (White et al., 2015), or due to the fact that phytoplankton uptake equation in MEDUSA does not allow phytoplankton to acclimatise in the oligotrophic region through optimum uptake kinetics (Smith et al., 2009; Yool et al., 2011).

At station L4, Apart from the model's state variables such as chlorophyll and nutrient concentrations, we have looked into the model-derived phytoplankton phenology because of its importance to marine ecosystems e.g. importance of the timing of phytoplankton blooms for the survival of zooplankton and fish larvae (Cushing, 1990). The timing of the blooms has also been shown to control the variability of pCO₂ in the sub-polar region (Bennington et al., 2009). Despite having a reliable spread in the annual mean of surface chlorophyll has a reliable spread annual mean, in terms of phytoplankton phenology, stations such as L4 show some mismatch with the observation. In situ initiation, bloom timing, and duration in L4 are earlier than in most of the ensemble members, although still lying within the ensemble range, ~~despite this being narrow by the NRR. Some and some~~ ensemble mean timings (termination and peak bloom time) are similar to the satellite observations at this latitude (Racault et al., 2012), ~~such as termination and peak bloom time.~~ When in situ chlorophyll is fitted with a smooth curve, the highest peak mostly occurs during spring (March-April). But model metrics, including ensemble mean and median, are noisy, and peaks mostly fall in the summer (May-July), ~~although the peak height is usually within the ensemble range. At~~. Moreover, at L4, distinct phytoplankton blooms occur twice a year: first in spring and the second in fall-autumn (Smyth et al., 2010). These blooms are sometimes well simulated, eg. e.g. in Fig. 14(g) and 5(d), but are not as distinct as in situ measurements because of due to the variability of the model. However, the difference in peak timing does not affect the duration of Some of these discrepancies may also be caused by the way zooplankton select their prey in MEDUSA. In a study by Sailley et al. (2014) grazing selection based on total prey concentration would result in rapid nutrient turn-over, which results in a single peak event, but if the selection is based on the stoichiometry of C:N, the blooms, and the in situ duration is well within the ensemble inter-quartile range.

At some stations, we have observed different NRR value for the inter-annual means (shown on Table 4 and Fig. 4) and the monthly data (shown on table 3 and Fig. 5). The discrepancy may be because in the monthly data, not all in situ peaks and troughs are being covered by the ensemble spread, but nutrients would regenerate slower, and therefore result in two chlorophyll peaks. However, the inter-annual in situ means are almost always within the ensemble range. The most notable difference is in the coastal stations where for station Cariaco, the inter-annual in situ means are mostly in the upper quartile, summarised in Fig. 4(d). The addition of the ensemble members that produced lower than average concentration therefore widens the overall ensemble spread, reducing the NRR. On the other hand, at station L4, the NRR for annual mean is close to unity because most of the year, the in situ data is very close to the ensemble median, see Fig. 4(e), with R_a value of 0.711. In terms of phytoplankton phenology, BATS ensemble has the highest range of peak bloom time (174 days), however the NRR suggests too narrow range (1.17). This is because most of the ensemble members produce blooms between April and May difference in peak timing does not affect the duration of the blooms, and the in situ timing occurred in the 29 March and so the in situ

mean bloom time is still outside the interquartile range. The large range is caused by some ensemble members blooming much later. However, since the in situ timing is earlier, it is not within most of the ensemble range, so the overall ensemble spread is deemed narrow.

We have chosen phytoplankton phenology to define model metrics because of its importance to marine ecosystem productivity; eg. Cushing (1990) show that the survival of zooplankton and fish larvae is affected by the timing of phytoplankton blooms. The timing of the blooms have also been shown to control the variability of pCO₂ in the sub-polar region (Bennington et al., 2009). Despite the importance of bloom timing, duration is well within the ensemble inter-quartile range. More generally, discrepancies in predicting bloom timing by large-scale biogeochemical models ~~is~~ are also reported in many studies, e.g., Henson et al. (2017) and Kostadinov et al. (2017). Henson et al. (2017) shows that ~~the~~ compared with the satellite data, the 3-D MEDUSA 2.0 (Yool et al., 2013) model ~~initiates~~ estimates spring bloom start ~~dates~~ date ~50 days ~~early~~ late, and in ~~the southern hemisphere it~~ initiates them ~~southern hemisphere, model estimates subtropical bloom start date~~ ~50 days ~~late~~ earlier. By generating an ensemble of 7 CMIP5 models, Kostadinov et al. (2017) highlighted that the difference in bloom timing between the model ensemble and satellite-derived chlorophyll can be more than one month over most of the ocean. This agrees with our study (see, Table 4), as most of our ensemble members have earlier bloom initiation dates, and the difference between the ensemble mean and in situ timing of bloom, ~~eg-e.g.~~ PAP and L4, are more than one month. ~~However, the use of the~~ Additionally, the whole ensemble range ~~, produced by this study~~ can help to provide an uncertainty range for the timing of phytoplankton blooms. By utilising the ensemble, start date differences may be reduced. The ensemble range almost always encompasses the observed annual mean, peak height, and amplitude. Therefore it may be suitable to use the ensemble model in order to forecast these phenological aspects. Further, it may also be possible to improve the accuracy of the ensemble range, by systematically removing certain ensemble members in a future study.

Finally, the unresolved discrepancy between in situ observations and large number of process ensemble results, such as in the oligotrophic stations for nitrogen and L4 for phytoplankton peak timings, emphasise that the inclusion of some missing processes, such as active prey selection, and species would improve the performance of the model (Friedrichs et al., 2007; Kriest et al., 2010) and functional forms which describe chemostat experiments, such as the droop function or the active prey selection (Sailley et al., 2014) are not as structurally sensitive as the logistic equations (Aldebert et al., 2018). Additionally, MEDUSA uses logistic functions such as Monod and Holling type III equations to describe its processes and are well known to be structurally sensitive (Aldebert et al., 2018). We did not include equations that allow such selection or species, as in this paper we tried to ensure that all the equations have similar properties to the default MEDUSA, in order to show that perturbing the structure of the model equations would result in different plankton and nutrient dynamics. Therefore, comparing the performance of model complexity and the ensemble method was beyond the scope of this study.

5 Conclusions

Our study highlights that it is important to conduct structural sensitivity analyses in addition to parameter sensitivity analyses ~~It~~ and it is crucial to include mathematical functions that can capture sufficient information of the key biogeochemical processes

known from experimental studies. However, none of the deterministic functions can capture all details of these processes (Anderson et al., 2010), therefore we have introduced a method whereby instead of having only one default model output, we have an ensemble generating a range of possible outcomes arising from alternative model structures. We have explored the structural sensitivity of the 1-D version of MEDUSA, the ocean biogeochemistry component of UK-ESM1, ~~to reduce the errors between in situ and model outcome~~which is becoming widely used in the community. This study emphasises that small ~~perturbation~~perturbations in MEDUSA process ~~structure~~equations can produce very different model results.

~~Linear phytoplankton mortality and hyperbolic (Holling type II) grazing generally produces lower chlorophyll concentrations, thereby reducing the ensemble mean and median chlorophyll concentrations. In regions of low nitrogen, sigmoidal and exponential uptake produce high nitrogen concentrations. The spread of the ensemble for chlorophyll and nitrogen profiles is the widest at coastal stations (NRR between 1.19-1.31), and narrowest at the oligotrophic stations (NRR 1.29-1.40). However, a reasonable range of ensemble is produced at oligotrophic station ALOHA, with NRR values of 1.01, and 1.07, for 10-years of integrated and surface chlorophyll time series respectively, and L4 for surface inter-annual means, with NRR value of 1.001. The ensemble range mostly covers the ensemble of perturbations generally encompasses the in situ observations (particularly the surface overall annual mean) of chlorophyll and nutrients at all stations. At oligotrophic stations, the ensemble mean and median tend to simulate lower chlorophyll concentrations and higher nitrogen concentrations compared with the default run and the in situ observations. However, at coastal and abyssal stations, the ensemble mean and median tend to overestimate chlorophyll, particularly during low chlorophyll months (June to September) in the North Atlantic. This in turns means that the spreads are wider for the inter-annual means at most of the stations, since the in situ inter-annual means are within the ensemble range, whereas in the monthly data the ensemble often doesn't capture the peak, making the NRR narrow. For phytoplankton phenology, the observed bloom initiation, peak bloom time, and termination, are all within the ensemble range at most stations. Although the ensemble members mostly show late blooms compared with in situ, the bloom durations are still within the 75th and 25th percentiles of the ensemble.~~

~~Our~~. ~~Therefore, our~~ study shows promise that ~~an ensemble~~the ensemble of a single biogeochemical model resulting from perturbing the model structure, ~~can produce a meaningful range of chlorophyll and nitrogen~~may produce meaningful prediction ranges of its state variables. However, our study is based on 1-D simulation, and further study with a 3-D biogeochemical model would help extend results to the global ocean. It may also be possible to further minimise the computational costs by systematically reducing the number of ensemble members whilst retaining a realistic ensemble range. Further studies could include varying the weighting of ensemble members, or reducing the number of model combinations to improve the ensemble range and to assess properly different plankton functional types and dissolved inorganic carbon. Such a perturbed biology ensemble may also be ~~used~~useful for data assimilation ~~eg.e.g.~~ with satellite-derived chlorophyll.

Data availability. The raw model outputs will be available at Pangaea after the manuscript has been published and upon request from the authors (p.anugerahanti@pgr.reading.ac.uk, shovonlal.roy@reading.ac.uk)

Appendix A: Determining phytoplankton phenology

Before determining the initiation time, bloom timing must be identified. This is done by taking the ten years of surface chlorophyll output and breaking it down into individual years. These are then rearranged into two datasets: January-December and June-May, and the date of maximum chlorophyll concentration in each year is determined. If the peak timing occurs mostly towards the end or the beginning of the year, June to May datasets are used instead of the former. The timing is then adjusted if the calendar year has changed.

The initiation is determined by the day that chlorophyll concentration exceeds a given threshold. However, since in situ chlorophyll has some data gaps and modelled chlorophyll is not smooth, some studies have fitted a function or model to the datasets to make the chlorophyll data smoother (Platt et al., 2009; Sapiano et al., 2012; Brody et al., 2013). Here we use a 5th order polynomial curve to get a smooth fit of the bloom peaks in the data (Fig. A1), from which phenology metrics are calculated. After being fitted, a threshold of half the bloom peak concentration is chosen. To find the **peak time**, the date at which maximum chlorophyll concentration is achieved in the fitted curve is determined, and this date is used as a reference to calculate other metrics. **Amplitude** is then calculated as half of the highest peak minus the minimum concentration. **Initiation** is the day when chlorophyll concentration goes just above the threshold towards the maximum (Brody et al., 2013). **Termination** of the bloom is defined when concentration falls below the threshold (Racault et al., 2012). If two peaks are detected the termination of the spring bloom is determined when the first bloom reduces to its minimum, just before the second bloom starts (in the first valley). **Duration** of the bloom is simply the total number of days on which chlorophyll concentration is above the threshold or termination minus initiation.

This phenology is useful to see how the bloom develops and terminates, whether the concentration increases rapidly and decreases slowly or vice versa. The phenology is summarised in Fig. A1. The curve fitting method is only applied if the data shows potential outliers especially in higher concentrations. If there is only one prominent bloom each year, as at stations ALOHA and BATS, and the data is smooth, the regular threshold method (when the concentration is above 50% of the maximum bloom, and the associated initiation and termination times), without fitting the data with a curve is applied. To avoid results being affected by how bloom phenology is determined, the same method is used for determining the metrics from both in situ and model output.

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

Acknowledgements. This study was funded by the Bakrie Center Foundation (grant no. 1307/BCF-SK/RSCH/VII/2015), Indonesia. The authors would like to thank Kevin White, for his advise on this study and John Hemmings, for providing the most recent version of the MarMOT code. We would like to thank Ruth Airs and Denise Cummings for providing the L4 chlorophyll and nutrient measurements. This study uses various oceanographic station data and we would like to thank all crews and scientists involved in collecting, processing, and making the data publicly available. [Three anonymous reviewers are acknowledged for their valuable comments and suggestions.](#)

References

- Adamson, M. W. and Morozov, A. Y.: When can we trust our model predictions? Unearthing structural sensitivity in biological systems, *Proceedings of the Royal Society of London A: Mathematical, Physical and Engineering Sciences*, 469, 20120500, <https://doi.org/10.1098/rspa.2012.0500>, 2013.
- 5 Aldebert, C., Nerini, D., Gauduchon, M., and Poggiale, J. C.: Does structural sensitivity alter ~~complexity~~ “stability-complexity-stability” relationships?, *Ecological Complexity*, 28, 104–112, <https://doi.org/10.1016/j.ecocom.2016.07.004>, 2016.
- [Aldebert, C., Kooi, B. W., Nerini, D., and Poggiale, J. C.: Is structural sensitivity a problem of oversimplified biological models? Insights from nested Dynamic Energy Budget models, *Journal of Theoretical Biology*, 448, 1–8, <https://doi.org/10.1016/j.jtbi.2018.03.019>, 2018.](https://doi.org/10.1016/j.jtbi.2018.03.019)
- Anderson, J. L.: An Ensemble Adjustment Kalman Filter for Data Assimilation, *Monthly Weather Review*, 129, 2884–2903, [https://doi.org/10.1175/1520-0493\(2001\)129<2884:AEAKFF>2.0.CO;2](https://doi.org/10.1175/1520-0493(2001)129<2884:AEAKFF>2.0.CO;2), 2001.
- Anderson, T. R.: Plankton functional type modelling: Running before we can walk?, *Journal of Plankton Research*, 27, 1073–1081, <https://doi.org/10.1093/plankt/fbi076>, 2005.
- Anderson, T. R. and Mitra, A.: Dysfunctionality in ecosystem models: An underrated pitfall?, *Progress in Oceanography*, 84, 66–68, 2010.
- Anderson, T. R., Gentleman, W. C., and Sinha, B.: Influence of grazing formulations on the emergent properties of a complex ecosystem model in a global ocean general circulation model, *Progress in Oceanography*, 87, 201–213, <https://doi.org/10.1016/j.pocean.2010.06.003>, 2010.
- Anderson, T. R., Gentleman, W. C., and Yool, A.: EMPOWER-1.0: An Efficient Model of Planktonic ecosystems Written in R, *Geoscientific Model Development*, 8, 2231–2262, <https://doi.org/10.5194/gmd-8-2231-2015>, 2015.
- ~~Behrenfeld, M. J. and Boss, E. S.: Resurrecting the ecological underpinnings of ocean plankton blooms., *Annual review of marine science*, 6, 167–94., 2014.~~
- ~~Behrenfeld, M. J., Doney, S. C., Lima, I., Boss, E. S., and Siegel, D. A.: Annual cycles of ecological disturbance and recovery underlying the subarctic Atlantic spring plankton bloom, *Global Biogeochemical Cycles*, 27, 1291–1293., 2013.~~
- Bennington, V., McKinley, G. A., Dutkiewicz, S., and Ullman, D.: What does chlorophyll variability tell us about export and air-sea CO₂ flux variability in the North Atlantic?, *Deep-Sea Research II*, 23, 1–11, 2009.
- 25 Berelson, W.: Particle settling rates increase with depth in the ocean, *Deep-Sea Research II*, 49, 237–251, 2002.
- Bopp, L., Resplandy, L., Orr, J. C., Doney, S. C., Dunne, J. P., Gehlen, M., Halloran, P., Heinze, C., Ilyina, T., Séférian, R., Tjiputra, J., and Vichi, M.: Multiple stressors of ocean ecosystems in the 21st century: Projections with CMIP5 models, *Biogeosciences*, 10, 6225–6245, <https://doi.org/10.5194/bg-10-6225-2013>, 2013.
- Brody, S. R., Lozier, M. S., and Dunne, J. P.: A comparison of methods to determine phytoplankton bloom initiation, *Journal of Geophysical Research: Oceans*, 118, 2345–2357, <https://doi.org/10.1002/jgrc.20167>, 2013.
- ~~Butenschön~~
- [Butenschön, M., Clark, J., Aldridge, J. N., Icarus Allen, J., Artioli, Y., Blackford, J., Bruggeman, J., Cazenave, P., Ciavatta, S., Kay, S., Lessin, G., Van Leeuwen, S., Van Der Molen, J., De Mora, L., Polimene, L., Sailley, S., Stephens, N., and Torres, R.: ERSEM 15.06: A generic model for marine biogeochemistry and the ecosystem dynamics of the lower trophic levels, *Geoscientific Model Development*, 9, 1293–1339, <https://doi.org/10.5194/gmd-9-1293-2016>, 2016.](https://doi.org/10.5194/gmd-9-1293-2016)
- 30 Cox, P. M. and Kwiatkowski, L.: Assessment of the iMarNet Ocean Biogeochemical Models, Tech. rep., University of Exeter, Exeter, http://imarnet.org/pdf/iMarNet_Model_Evaluation_-_CoxKwiatkowski, 2013.

- Cushing, D.: Plankton production and year-class strength in fish populations - an update of the match mismatch hypothesis., *Advances in Marine Biology*, 26, 249–293, <http://www.sciencedirect.com/science/article/pii/S0065288108602023>, 1990.
- Dave, A. C. and Lozier, M. S.: Local stratification control of marine productivity in the subtropical North Pacific, *Journal of Geophysical Research: Oceans*, 115, 1–16, <https://doi.org/10.1029/2010JC006507>, 2010.
- 5 [Doney, C.: Major challenges confronting marine biogeochemical modeling, *Global Biogeochemical Cycles*, 13, 705–714, 1999.](#)
- Doney, S. C.: The growing human footprint on the planet, *Science*, 328, 1512–1516, <https://doi.org/10.1126/science.1185198>, 2010.
- Doney, S. C., Ruckelshaus, M., Emmett Duffy, J., Barry, J. P., Chan, F., English, C. A., Galindo, H. M., Grebmeier, J. M., Hollowed, A. B., Knowlton, N., Polovina, J., Rabalais, N. N., Sydeman, W. J., and Talley, L. D.: Climate Change Impacts on Marine Ecosystems, *Annual Review of Marine Science*, 4, 11–37, <https://doi.org/10.1146/annurev-marine-041911-111611>, 2012.
- 10 Dore, J. E., Letelier, R. M., Church, M. J., Lukas, R., and Karl, D. M.: Summer phytoplankton blooms in the oligotrophic North Pacific Subtropical Gyre: Historical perspective and recent observations, *Progress in Oceanography*, 76, 2–38, <https://doi.org/10.1016/j.pocean.2007.10.002>, 2008.
- Dutkiewicz, S., Follows, M. J., and Parekh, P.: Interactions of the iron and phosphorus cycles: A three-dimensional model study, *Global Biogeochemical Cycles*, 19, 1–22, <https://doi.org/10.1029/2004GB002342>, 2005.
- 15 Edwards, A. and Yool, A.: The role of higher predation in plankton population models, *Journal of Plankton Research*, 22, 1085–1112, <https://doi.org/10.1093/plankt/22.6.1085>, 2000.
- Englund, G. and Leonardsson, K.: Scaling up the functional response for spatially heterogeneous systems, *Ecology Letters*, 11, 440–449, <https://doi.org/10.1111/j.1461-0248.2008.01159.x>, 2008.
- Eppley, R. W.: Temperature and phytoplankton growth in the sea, *Fishery bulletin*, 70, 1063–1085, <https://doi.org/163346>, 1972.
- 20 Fasham, M. J. R.: Variations in the seasonal cycle of biological production in subarctic oceans: A model sensitivity analysis, *Deep-Sea Research Part I*, 42, 1111–1149, [https://doi.org/10.1016/0967-0637\(95\)00054-A](https://doi.org/10.1016/0967-0637(95)00054-A), 1995.
- Fasham, M. J. R., Ducklow, H. W., and McKelvie, S. M.: A nitrogen-based model of plankton dynamics in the ocean mixed layer, *Journal of Marine Research*, 48, 591–639, <https://doi.org/10.1357/002224090784984678>, 1990.
- Fasham, M. J. R., Sarmiento, J. L., Slater, R. D., Ducklow, H. W., and Williams, R.: Ecosystem behaviour at Bermuda station "S" and ocean
25 weather station "India": A general circulation model and observational analysis, *Global Biogeochemical Cycles*, 7, 379–415, 1993.
- Fennel, K. and Boss, E.: Subsurface maxima of phytoplankton and chlorophyll: Steady-state solutions from a simple model, *Limnology and Oceanography*, 48, 1521–1534, <https://doi.org/10.4319/lo.2003.48.4.1521>, 2003.
- Fennel, W. and Neumann, T.: Introduction to the Modelling of Marine Ecosystem, Elsevier Science, second edn., 2014.
- Flora, C., David, N., Mathias, G., Andrew, M., and Jean-Christophe Poggiale, P.: Structural sensitivity of biological models revisited, *Journal
30 of Theoretical Biology*, 283, 82–91, <https://doi.org/10.1016/j.jtbi.2011.05.021>, 2011.
- Flynn, K. J. and Mitra, A.: Why Plankton Modelers Should Reconsider Using Rectangular Hyperbolic (Michaelis-Menten, Monod) Descriptions of Predator-Prey Interactions, *Frontiers in Marine Science*, 3, [165, 1–17](#) <https://doi.org/10.3389/fmars.2016.00165>, 2016.
- Ford, D. A., Edwards, K. P., Lea, D., Barciela, R. M., Martin, M. J., and Demaria, J.: Assimilating GlobColour ocean colour data into a pre-operational physical-biogeochemical model, *Ocean Science*, 8, 751–771, <https://doi.org/10.5194/os-8-751-2012>, 2012.
- 35 Franks, P. J. S.: NPZ models of plankton dynamics: Their construction, coupling to physics, and application, *Journal of Oceanography*, 58, 379–387, <https://doi.org/10.1023/A:1015874028196>, 2002.
- [Friedrichs, M. A., Dusenberry, J. A., Anderson, L. A., Armstrong, R. A., Chai, F., Christian, J. R., Doney, S. C., Dunne, J., Fujii, M., Hood, R., McGillicuddy, D. J., Moore, J. K., Schartau, M., Spitz, Y. H., and Wiggert, J. D.: Assessment of skill and portability in](#)

- [regional marine biogeochemical models: Role of multiple planktonic groups](https://doi.org/10.1029/2006JC003852), *Journal of Geophysical Research: Oceans*, 112, 1–22, <https://doi.org/10.1029/2006JC003852>, 2007.
- [Friedrichs, M. A. M., Hood, R. R., and Wiggert, J. D.: Ecosystem model complexity versus physical forcing: Quantification of their relative impact with assimilated Arabian Sea data](https://doi.org/10.1016/j.dsr2.2006.01.026), *Deep-Sea Research Part II: Topical Studies in Oceanography*, 53, 576–600, <https://doi.org/10.1016/j.dsr2.2006.01.026>, 2006.
- 5 Fussmann, G. F. and Blasius, B.: Community response to enrichment is highly sensitive to model structure., *Biology letters*, 1, 9–12, <https://doi.org/10.1098/rsbl.2004.0246>, 2005.
- Gehlen, M., Barciela, R., Bertino, L., Brasseur, P., Butenschön, M., Chai, F., Crise, A., Drillet, Y., Ford, D., Lavoie, D., Lehodey, P., Perruche, C., Samuelsen, A., and Simon, E.: Building the capacity for forecasting marine biogeochemistry and ecosystems: recent advances and future developments, *Journal of Operational Oceanography*, 8, s168–s187, <https://doi.org/10.1080/1755876X.2015.1022350>, 2015.
- 10 [Hartman, S. E., Jiang, Z., Turk, D., Lampitt, R. S., Frigstad, H., Ostle, C., and Schuster, U.: Biogeochemical variations at the Porcupine Abyssal Plain sustained Observatory in the northeast Atlantic Ocean, from weekly to inter-annual timescales, pp. 845–853, 2015.](#)
- Hawaii Ocean Time Series: Analytical methods and results, <http://hahana.soest.hawaii.edu/hot/methods/inuts.html>.
- Hemmings, J. C. P. and Challenor, P. G.: Addressing the impact of environmental uncertainty in plankton model calibration with a dedicated software system: The marine model optimization testbed (MarMOT 1.1 alpha), *Geoscientific Model Development*, 5, 471–498, <https://doi.org/10.5194/gmd-5-471-2012>, 2012.
- 15 Hemmings, J. C. P., Challenor, P. G., and Yool, A.: Mechanistic site-based emulation of a global ocean biogeochemical model (MEDUSA 1.0) for parametric analysis and calibration: An application of the Marine Model Optimization Testbed (MarMOT 1.1), *Geoscientific Model Development*, 8, 697–731, <https://doi.org/10.5194/gmd-8-697-2015>, 2015.
- 20 Henson, S. A., Yool, A., Cole, H. S., Hopkins, J., and Martin, A. P.: Detection of climate change-driven trends in phytoplankton phenology, <https://doi.org/10.1111/gcb.13886>, *Global Change Biology*, 00, 1–11, 2017.
- Ivlev, V.: *Experimental Ecology of The Feeding of The Fishes*, Yale University Press, New Haven, CT, 1961.
- Jassby, A. D. and Platt, T.: Mathematical Formulation of the Relationship Between Photosynthesis and Light for Phytoplankton, *Limnology and Oceanography*, 21, 540–547, <https://doi.org/10.4319/lo.1976.21.4.0540>, 1976.
- 25 Kane, A., Moulin, C., Thiria, S., Bopp, L., Berrada, M., Tagliabue, A., Crépon, M., Aumont, O., and Badran, F.: Improving the parameters of a global ocean biogeochemical model via variational assimilation of in situ data at five time series stations, *Journal of Geophysical Research: Oceans*, 116, 1–14, <https://doi.org/10.1029/2009JC006005>, 2011.
- Kostadinov, T. S., Cabré, A., Vedantham, H., Marinov, I., Bracher, A., Brewin, R. J., Bricaud, A., Hirata, T., Hirawake, T., Hardman-Mountford, N. J., Mouw, C., Roy, S., and Uitz, J.: Inter-comparison of phytoplankton functional type phenology metrics derived from ocean color algorithms and Earth System Models, *Remote Sensing of Environment*, 190, 162–177, <https://doi.org/10.1016/j.rse.2016.11.014>, 2017.
- 30 [Kriest, I., Khatiwala, S., and Oschlies, A.: Towards an assessment of simple global marine biogeochemical models of different complexity](https://doi.org/10.1016/j.pocean.2010.05.002), *Progress in Oceanography*, 86, 337–360, <https://doi.org/10.1016/j.pocean.2010.05.002>, 2010.
- Kwiatkowski, L., Yool, A., Allen, J. I., Anderson, T. R., Barciela, R., Buitenhuis, E. T., Butenschön, M., Enright, C., Halloran, P. R., Le [Quere](#) [Quèrè](#), C., De Mora, L., Racault, M. F., Sinha, B., Totterdell, I. J., and Cox, P. M.: IMarNet: An ocean biogeochemistry model inter-comparison project within a common physical ocean modelling framework, *Biogeosciences*, 11, 7291–7304, <https://doi.org/10.5194/bg-11-7291-2014>, 2014.
- 35

- Lacroix, G. and Gregoire, M.: Revisited ecosystem model (MODECOGeL) of the Ligurian Sea: seasonal and interannual variability due to atmospheric forcing, *Journal of Marine Systems*, 37, 229–258, [https://doi.org/http://dx.doi.org/10.1016/S0924-7963\(02\)00190-2](https://doi.org/http://dx.doi.org/10.1016/S0924-7963(02)00190-2), 2002.
- [Le Quèrè, C., Harrison, S. P., Prentice, I. C., Buitenhuis, E. T., Aumont, O., Bopp, L., Claustre, H., Leticia Cotrim Da Cunha, R. G., Giraud, X., Klaas, C., Kohfeld, K. E., Legendre, L., Manizza, M., Platt, T., Rivkin, R. B., Sathyendranath, S., Uitz, J., Watson, A. J., and Wolf-Gladrow, D.: Ecosystem dynamics based on plankton functional types for global ocean biogeochemistry models, *Global Change in Biology*, 11, 2016–2040, <https://doi.org/10.1111/j.1365-2486.2005.01004.x>, 2005.](#)
- Letelier, R. M., Karl, D. M., Abbott, M. R., and Bidigare, R. R.: Light driven seasonal patterns of chlorophyll and nitrate in the lower euphotic zone of the North Pacific Subtropical Gyre, *Limnol. Oceanogr.*, 49, 508–519, <https://doi.org/10.4319/lo.2004.49.2.0508>, 2004.
- Levin, S. and Lubchenco, J.: Resilience, robustness, and marine ecosystem-based management, *BioScience*, 58, 27–32, <https://doi.org/10.1641/B580107>, 2008.
- Lipschultz, F.: A time-series assessment of the nitrogen cycle at BATS, 48, 1897–1924, 2001.
- ~~Lorenzoni, L., Taylor, G. T., Benitez-Nelson, C., Hansell, D. A., Montes, E., Masserini, R., Fanning, K., Varela, R., Astor, Y., Guzmán, L., and Muller-Karger, F. E.: Spatial and seasonal variability of dissolved organic matter in the Cariaco Basin, *Journal of Geophysical Research: Biogeosciences*, 118, 951–962, , 2013.~~
- 15 Mahowald, N. M., Engelstaedter, S., Luo, C., Sealy, A., Artaxo, P., Benitez-Nelson, C., Bonnet, S., Chen, Y., Chuang, P. Y., Cohen, D. D., Dulac, F., Herut, B., Johansen, A. M., Kubilay, N., Losno, R., Maenhaut, W., Paytan, A., Prospero, J. M., Shank, L. M., and Siefert, R. L.: Atmospheric iron deposition: global distribution, variability, and human perturbations., *Annual review of marine science*, pp-245–278, <https://doi.org/10.1146/annurev.marine.010908.163727>, 2009.
- McDonald, C., Bennington, V., Urban, N., and McKinley, G.: 1-D test-bed calibration of a 3-D Lake Superior biogeochemical model, *Ecological Modelling*, 225, 115–126, <https://doi.org/10.1016/j.ecolmodel.2011.11.021>, 2012.
- Moradkhani, H. and Meskele, T. T.: Probabilistic Assessment of the Satellite Retrieval Error Translation to Hydrologic Response, in: *Satellite Rainfall Applications for Surface Hydrology*, edited by Gebremichael, M. and Hossain, F., chap. II, pp. 235–241, Springer, Netherlands, 2010.
- Moradkhani, H., Hsu, K., Hong, Y., and Sorooshian, S.: Investigating the impact of remotely sensed precipitation and hydrologic model 25 uncertainties on the ensemble streamflow forecasting, *Geophysical Research Letters*, 33, 1–5, <https://doi.org/10.1029/2006GL026855>, 2006.
- Murphy, J. M., Booth, B. B. B., Collins, M., Harris, G. R., Sexton, D. M. H., and Webb, M. J.: A methodology for probabilistic predictions of regional climate change from perturbed physics ensembles., *Philosophical transactions. Series A, Mathematical, physical, and engineering sciences*, 365, 1993–2028, <https://doi.org/10.1098/rsta.2007.2077>, 2007.
- 30 Oschlies, A. and Garçon, V.: An eddy-permitting coupled physical-biological model of the North Atlantic. 1. Sensitivity to advection numerics and mixed layer physics, *Global Biogeochemical Cycles*, 13, 135–160, <https://doi.org/10.1029/98GB02811>, 1999.
- [Oschlies, A. and Schartau, M.: Basin-scale performance of a locally optimized marine ecosystem model, *Journal of Marine Systems*, 63, 335–358, 2005.](#)
- Ourmières, Y., Brasseur, P., Lévy, M., Brankart, J.-m., and Verron, J.: On the key role of nutrient data to constrain a coupled 35 ~~physical-physical-~~ biogeochemical assimilative model of the North Atlantic Ocean, *Journal of Marine Systems*, 75, 100–115, <https://doi.org/10.1016/j.jmarsys.2008.08.003>, 2009.
- Palmer, J. and Totterdell, I.: Production and export in a global ecosystem model, *Deep-Sea Research I*, 48, 1169–1198, [https://doi.org/10.1016/S0967-0637\(00\)00080-7](https://doi.org/10.1016/S0967-0637(00)00080-7), 2001.

- Parekh, P., Follows, M. J., and Boyle, E. A.: Decoupling of iron and phosphate in the global ocean, *Global Biogeochemical Cycles*, 19, 1–16, <https://doi.org/10.1029/2004GB002280>, 2005.
- Platt, T. and Sathyendranath, S.: Ecological indicators for the pelagic zone of the ocean from remote sensing, *Remote Sensing of Environment*, 112, 3426–3436, <https://doi.org/10.1016/j.rse.2007.10.016>, 2008.
- 5 Platt, T., White, G. N., Zhai, L., Sathyendranath, S., and Roy, S.: The phenology of phytoplankton blooms: Ecosystem indicators from remote sensing, *Ecological Modelling*, 220, 3057–3069, <https://doi.org/10.1016/j.ecolmodel.2008.11.022>, 2009.
- ~~Quèrè, C. L., Harrison, S. P., Prentice, I. C., Buitenhuis, E. T., Aumont, O., Bopp, L., Claustre, H., Leticia-Cotrim-Da-Cunha, R. G., Giraud, X., Klaas, C., Kohfeld, K. E., Legendre, L., Manizza, M., Platt, T., Rivkin, R. B., Sathyendranath, S., Uitz, J., Watson, A. J., and Wolf-Gladrow, D.: Ecosystem dynamics based on plankton functional types for global ocean biogeochemistry models, *Global Change in Biology*, 11, 2016–2040, 2005.~~
- 10 Racault, M.-F., Le Quere ~~Quèrè~~, C., Buitenhuis, E., Sathyendranath, S., and Platt, T.: Phytoplankton phenology in the global ocean, *Ecological Indicators*, 14, 152–163, <https://doi.org/10.1016/j.ecolind.2011.07.010>, 2012.
- 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.
- 15 Raymont, J.: *Plankton and Productivity in the Oceans*, vol. 1, Pergamon Press, ~~Oxford, 2nd edn., London~~, 1980.
- Robinson, C. L. K., Ware, D. M., and Parsons, T. R.: Simulated annual plankton production in the northeastern Pacific Coastal upwelling Domain, *Journal of Plankton Research*, 15, 161–183, <https://doi.org/10.1093/plankt/15.2.161>, 1993.
- Roy, S. and Chattopadhyay, J.: Enrichment and stability: A phenomenological coupling of energy value and carrying capacity, *BioSystems*, [90, 371–378](https://doi.org/10.1016/j.biosystems.2006.10.001), <https://doi.org/10.1016/j.biosystems.2006.10.001>, 2007.
- 20 Roy, S., Broomhead, D. S., Platt, T., Sathyendranath, S., and Ciavatta, S.: Sequential variations of phytoplankton growth and mortality in an NPZ model: A remote-sensing-based assessment, *Journal of Marine Systems*, 92, 16–29, <https://doi.org/10.1016/j.jmarsys.2011.10.001>, 2012.
- Ryabchenko, V. A., Fasham, M. J. R., Kagan, B. A., and Popova, E. E.: What causes short-term oscillations in ecosystem models of the ocean mixed layer?, *Journal of Marine Systems*, 13, 33–50, [https://doi.org/10.1016/S0924-7963\(96\)00110-8](https://doi.org/10.1016/S0924-7963(96)00110-8), 1997.
- 25 [Sailley, S. F., Polimene, L., Mitra, A., Atkinson, A., and Allen, J. I.: Impact of zooplankton food selectivity on plankton dynamics and nutrient cycling, *Journal of Plankton Research*, 37, 519–529, <https://doi.org/10.1093/plankt/fbv020>, 2014.](https://doi.org/10.1093/plankt/fbv020)
- Sapiano, M. R. P., Brown, C. W., Schollaert Uz, S., and Vargas, M.: Establishing a global climatology of marine phytoplankton phenological characteristics, *Journal of Geophysical Research: Oceans*, 117, 1–16, <https://doi.org/10.1029/2012JC007958>, 2012.
- ~~Siegel, D. A., Doney, S.~~
- 30 ~~Sinha, B., Buitenhuis, E. T., Le Quere, C., and Yoder, J. A. Anderson, T. R.: The North Atlantic spring phytoplankton bloom and Sverdrup's critical depth hypothesis. Progress in Oceanography Comparison of the emergent behavior of a complex ecosystem model in two ocean general circulation models, *Progress in Oceanography*, 84, *Science*, 296, 730–3, 2002. 204–224, <https://doi.org/10.1016/j.pocean.2009.10.003>, 2010.~~
- [Smith, S., Yamanaka, Y., Pahlow, M., and Oschlies, A.: Optimal uptake kinetics: physiological acclimation explains the pattern of nitrate uptake by phytoplankton in the ocean, *Marine Ecology Progress Series*, 384, 1–12, <https://doi.org/10.3354/meps08022>, 2009.](https://doi.org/10.3354/meps08022)
- 35 Smyth, T. J., Fishwick, J. R., Al-moosawi, L., Cummings, D. G., and Harris, C.: A broad spatio-temporal view of the Western English Channel observatory, 32, 585–601, <https://doi.org/10.1093/plankt/fbp128>, 2010.

- ~~Smyth, T. J., Atkinson, A., Widdicombe, S., Allen, J. I., Queiros, A., Sims, D., and Brange, M.: The Western Channel Observatory, *Progress in Oceanography*, 137, 335–341, 2015.~~
- Steele, J. H. and Henderson, E. W.: A Simple Plankton Model, *The American Naturalist*, 117, 676–691, 1981.
- Storkey, D., Blockley, E., Furner, R., Guiavarc’h, C., Lea, D., Martin, M., Barciela, R. M., Hines, A., Hyder, P., and Siddorn, J.: Forecasting
5 the ocean state using NEMO:The new FOAM system, <https://doi.org/10.1080/1755876X.2010.11020109>, 2010.
- Subramanian, A. C. and Palmer, T. N.: Ensemble superparameterization versus stochastic parameterization: A comparison of model uncertainty representation in tropical weather prediction, *Journal of Advances in Modeling Earth Systems*, 9, 1231–1250, <https://doi.org/10.1002/2016MS000857>, 2017.
- Taylor, G. T., Muller-karger, F. E., Thunell, R. C., Scranton, M. I., Astor, Y., and Varela, R.: Ecosystem responses in the southern Caribbean
10 Sea to global climate change, 109, 19 315–19 320, <https://doi.org/10.1073/pnas.1207514109>, 2012.
- Tinker, J., Lowe, J., Pardaens, A., Holt, J., and Barciela, R.: Uncertainty in climate projections for the 21st century northwest European shelf seas, *Progress in Oceanography*, 148, 56–73, <https://doi.org/10.1016/j.pocean.2016.09.003>, 2016.
- Villareal, T. A., Brown, C. G., Brzezinski, M. A., Krause, J. W., and Wilson, C.: Summer Diatom Blooms in the North Pacific Subtropical Gyre : 2008 ~~â€“~~ 2009, 7, 4, [e33109](https://doi.org/10.1371/journal.pone.0033109), <https://doi.org/10.1371/journal.pone.0033109>, 2012.
- 15 [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.](https://doi.org/10.1016/j.pocean.2013.06.002)
- White, A., Letelier, R. M., Whitmire, A. L., Barone, B., Bidigare, R. R., Church, M. J., and Karl, D. M.: Phenology of particle size distributions and primary productivity in the North Pacific subtropical gyre (Station ALOHA), *Journal of Geophysical Research Oceans*, ~~pp-~~[120, 7381–7399](https://doi.org/10.1002/2015JC010897), <https://doi.org/10.1002/2015JC010897>, 2015.
- 20 Wood, S. and Thomas, M.: Super-sensitivity to structure in biological models, *The Royal Society*, 266, 565–570, <https://doi.org/10.1098/rspb.1999.0673>, 1999.
- Yool, A., Popova, E. E., and Anderson, T. R.: MEDUSA-1.0: A new intermediate complexity plankton ecosystem model for the global domain, *Geoscientific Model Development*, 4, 381–417, <https://doi.org/10.5194/gmd-4-381-2011>, 2011.
- 25 Yool, A., Popova, E. E., and Anderson, T. R.: MEDUSA-2.0: An intermediate complexity biogeochemical model of the marine carbon cycle for climate change and ocean acidification studies, *Geoscientific Model Development*, 6, 1767–1811, <https://doi.org/10.5194/gmd-6-1767-2013>, 2013.

Tables

Table 1. Parameter values for resource uptake (U), zooplankton grazing (G), and plankton mortalities (ρ and ξ for phytoplankton and zooplankton respectively), described using similar functional forms (shown in Fig. 1). In grazing equation, g_m represents maximum grazing rate, P_a is the prey, and p_a denotes the grazing preference. Starred equations are the default functional responses in MEDUSA.

Process/ Plankton type	Symbol	Meaning	Parameter value (mmol m ⁻³)			
Nutrient Uptake (U)			Monod* (U_1) $\frac{n}{n+k}$	Sigmoidal (U_2) $\frac{n^2}{n^2+k^2}$	Exponential (U_3) $1 - \exp(-\frac{n}{k})$	Trigonometric (U_4) $\frac{2}{\pi} \arctan(\frac{n}{k})$
Non-diatom	kN_{nd}	shape defining constant for nitrogen	0.5	0.74	1.12	0.60
	kFe_{nd}	shape defining constant for iron	0.33 $\times 10^{-3}$	0.49 $\times 10^{-3}$	0.74 $\times 10^{-3}$	0.40 $\times 10^{-3}$
Diatom	kN_d	shape defining constant for nitrogen	0.75	1.12	1.68	0.91
	kSi_d	shape defining constant for silicon	0.75	1.12	1.68	0.91
	kFe_d	shape defining constant for iron	0.67 $\times 10^{-3}$	0.99 $\times 10^{-3}$	1.50 $\times 10^{-3}$	0.81 $\times 10^{-3}$
Grazing (G)			Holling type III* (G_1) $\frac{g_m P_a P_a^2}{k_g^2 + p_a P_a^2 + p_b P_a^2}$	Holling type II (G_2) $\frac{g_m P_a^2}{k_g(p_a P_a + p_b P_b) + p_a P_a^2 + p_b P_b^2}$		
Microzooplankton	k_{mi}	half saturation constant	0.80	0.46		
	pmi_{nd}	grazing preference for non-diatom	0.75	0.75		
	pmi_{det}	grazing preference for detritus	0.25	0.25		
Mesozooplankton	k_{me}	half saturation constant	0.30	0.17		
	pme_{nd}	grazing preference for non-diatom	0.15	0.15		
	pme_{det}	grazing preference for detritus	0.15	0.15		
	pme_d	grazing preference for diatoms	0.35	0.35		
	pme_{mi}	grazing preference for microzooplankton	0.35	0.35		
Mortality (ρ, ξ)			Hyperbolic* (ρ_1, ξ_1) $\mu \frac{P}{P+k_M} P$	Linear (ρ_2, ξ_2) μP	Quadratic (ρ_3, ξ_3) μP^2	Sigmoidal (ρ_4, ξ_4) $\mu \frac{P^2}{P^2+k_M^2} P$
Non-diatom	μ_{nd}	maximum rate (day ⁻¹)	0.10	0.10	0.05	0.10
	k_{Mnd}	half saturation constant	0.50	-	-	0.74
Diatom	μ_d	maximum rate (day ⁻¹)	0.10	0.10	0.05	0.1
	k_{Md}	half saturation constant	0.50	-	-	0.74
Microzooplankton	μ_{mi}	maximum rate (day ⁻¹)	0.10	0.10	0.05	0.10
	k_{Mmi}	half saturation constant	0.50	-	-	0.74
Microzooplankton	μ_{mi}	maximum rate (day ⁻¹)	0.20	0.20	0.07	0.20
	k_{Mmi}	half saturation constant	0.75	-	-	1.12

Table 2. [Location, data source, and available depth range for the five oceanographic stations](#)

Station	Location	Source	depth range
ALOHA	22°45'N, 158°00'W	http://hahana.soest.hawaii.edu/hot/hot-dogs/interface.html	5-5000 m
BATS	32°50'N, 64°10'W	http://bats.bios.edu/	4-4000 m
Cariaco	10°30'N, 64°40'W	http://imars.marine.usf.edu/cariaco	1-1310 m
L4	50°15'N, 4°12.3'W	http://www.westernchannelobservatory.org.uk/data.php (available upon request)	surface
PAP	49°N, 16.5°W	http://projects.noc.ac.uk/pap/data	7-400m

Table 3. Error statistics, 10-year mean, and NRR of chlorophyll (mg m^{-3}) and nitrogen (mmol m^{-3}) concentration at five stations for the default run, ensemble mean, ensemble median, and the ensemble range (ensemble maximum - ensemble minimum). These are calculated from surface to 200 m depth, starting from January 1998 to December 2007. Bias is (model output) – (in situ observation). Bold text indicate the smallest RMSE. At Station L4 error statistics and mean are taken from the surface and starts from January 1999 for chlorophyll and June 2000 for nitrogen. For station PAP, error statistics are taken from 2002-2004 since in situ data is only available during that time.

Stations	Nitrogen profile				Chlorophyll profile				Surface chlorophyll				Integrated chlorophyll				
	r	RMSE	Bias	Mean	r	RMSE	Bias	Mean	r	RMSE	Bias	Mean	r	RMSE	Bias	Mean	
PAP	Ens mean	0.23	3.26	0.61	6.59	0.42	0.32	0.06	0.48	0.45	0.51	0.22	0.66				
		(±0.07)	(±2.57)	(±5.13)	(±5.24)	(±0.37)	(±0.73)	(±0.68)	(±0.75)	(±0.38)	(±0.73)	(±0.68)	(±0.76)				
	Ens median	0.23	3.16	0.54	6.38	0.49	0.29	0.003	0.42	0.54	0.46	0.15	0.60				
	Default run	0.21	3.32	-0.20	5.64	0.28	0.40	0.18	0.59	0.36	0.57	0.30	0.74				
	In situ				5.83				0.42				0.44				
NRR		1.25				1.20				1.29							
ALOHA	Ens mean	0.77	1.06	0.67	1.20	0.22	0.10	-0.06	0.06	0.22	0.05	-0.01	0.10	0.69	2.73	-0.72	3.80
		(±0.03)	(±0.19)	(±0.39)	(±0.39)	(±0.49)	(±0.04)	(±0.11)	(±0.11)	(±0.47)	(±0.09)	(±0.13)	(±0.14)	(±0.60)	(±5.49)	(±7.09)	(±10)
	Ens median	0.77	1.06	0.68	1.18	0.14	0.11	-0.07	0.05	0.13	0.05	-0.01	0.07	0.56	3.3	-1.17	3.34
	Default run	0.77	1.09	0.61	1.10	0.28	0.10	-0.03	0.09	0.27	0.07	0.03	0.11	0.70	4.71	1.25	5.77
	In situ				0.50				0.12				0.08				4.52
NRR		1.39				1.29				1.07				1.01			
BATS	Ens mean	0.56	1.39	1.16	1.77	0.19	0.33	-0.12	0.05	0.22	0.33	-0.12	0.05	0.39	52.13	-19.39	6.18
		(±0.38)	(±0.84)	(±1.00)	(±1.01)	(±0.37)	(±0.05)	(±0.16)	(±0.16)	(±0.58)	(±0.15)	(±0.05)	(±0.15)	(±0.54)	(±9.40)	(±21)	(±14)
	Ens median	0.55	1.39	1.16	1.77	0.11	0.33	-0.12	0.05	0.06	0.34	-0.12	0.05	0.27	23.30	-17.71	4.51
	Default run	0.58	0.73	0.62	1.35	0.23	0.31	-0.07	0.10	0.28	0.31	-0.07	0.09	0.43	48.58	-10.77	13.14
	In situ				0.98				0.17				0.15				23.90
NRR		1.38				1.39				1.40				1.40			
Cariaco	Ens mean	0.78	2.97	0.61	5.39	0.29	0.83	-0.02	0.49	0.13	1.23	0.02	0.77	0.41	17.73	-1.05	11.47
		(±0.08)	(±0.49)	(±2.54)	(±2.54)	(±0.34)	(±0.42)	(±0.93)	(±0.93)	(±0.22)	(±0.33)	(±1.90)	(±0.57)	(±0.40)	(±7.90)	(±17)	(±17)
	Ens median	0.76	3.24	0.51	5.29	0.20	0.88	-0.18	0.32	0.072	1.29	-0.29	0.46	0.29	19.46	-5.51	7.00
	Default run	0.76	3.29	0.59	5.37	0.22	0.87	-0.09	0.42	0.11	1.27	-0.18	0.57	0.34	18.71	-3.86	8.65
	In situ				4.78				0.51				0.76				12.52
NRR		1.25				1.19				1.21				1.17			
L4	Ens mean	0.70	2.94	1.56	4.52					0.25	1.05	0.42	1.76				
		(±0.14)	(±2.13)	(±4.06)	(±4.06)					(±0.33)	(±1.67)	(±2.61)	(±2.61)				
	Ens median	0.68	3.10	1.73	4.69					0.21	1.02	0.27	1.61				
	Default run	0.52	2.67	1.12	4.08					0.31	1.13	0.83	2.17				
	In situ				2.96								1.34				
NRR		1.31								1.21							

Table 4. Surface annual mean and phytoplankton phenology from in situ, ensemble mean, median, and default run. [The range and NRR in the bracket are the values for changing the functional form one process at a time \(shown on Fig. 8\).](#)

Stations		Annual Mean (mg m ⁻³)	Initiation Time	Bloom (mg m ⁻³)	Peak Height (mg m ⁻³)	Amplitude (mg m ⁻³)	Duration	Termination
PAP	Ens mean	0.61	01 Apr	07 May	2.07	0.96	95	26 Jul
	Range	±0.70(0.58)	±51	±45	±2.98	±1.63	±99	±124
	NRR	1.26 (1.37)	1.14	1.31	1.08	1.09	1.42	1.60
	Ens med	0.55	12 Apr	15 May	2.03	0.95	87	22 Jul
	Default run	0.71	03 Apr	05 May	2.1	0.96	99	21 Aug
	In situ	0.44	20 Apr	03 Jun	1.52	0.44	95	24 Jul
ALOHA	Ens mean	0.07	21 Mar	21 Apr	0.14	0.047	62	15 Aug
	Range	±0.13(0.11)	±89	±119	±0.28	±0.11	±95	±119
	NRR	0.84 (1.17)	1.35	1.29	0.97	1.19	1.56	1.28
	Ens med	0.063	26 Mar	02 May	0.14	0.05	85	24 Aug
	Default run	0.10	14 Mar	18 Apr	0.25	0.096	66	10 Aug
	In situ	0.084	08 May	26 May	0.14	0.048	47	23 Jun
BATS	Ens mean	0.047	02 Mar	12 Apr	0.1	0.043	89	06 Jul
	Range	±0.14(0.11)	±187	±174	±0.42	±0.19	±116	±198
	NRR	1.40 (1.39)	1.18	1.17	1.42	1.42	1.08	1.20
	Ens med	0.038	28 Feb	06 Apr	0.08	0.033	95	02 Aug
	Default run	0.091	06 Mar	25 Apr	0.29	0.13	65	19 Jun
	In situ	0.17	25 Feb	29 Mar	0.58	0.27	93	28 May
Cariaco	Ens mean	0.61	20 May	22 Jul	1.09	0.38	133	30 Sep
	Range	±1.53(1.29)	±101	±66	±2.61	±0.86	±63	±61
	NRR	0.78 (0.90)	1.48	1.40	1.39	1.42	1.88	1.55
	Ens med	0.37	22 May	14 Jul	0.83	0.34	110	06 Sep
	Default run	0.46	21 May	22 Jul	0.98	0.39	122	19 Sep
	In situ	0.61	16 Mar	21 Apr	2.39	1.15	76	01 Jun
L4	Ens mean	1.65	13 May	06 Jun	3.25	1.13	64	17 Aug
	Range	±2.48(2.14)	±100	±82	±3.12	±1.50	±78	±167
	NRR	1.00 (1.36)	1.49	1.42	1.32	1.48	1.22	1.19
	Ens med	1.49	18 May	07 Jun	3.09	1.13	70	18 Sep
	Default run	2.03	19 Apr	08 Jun	3.73	1.3	94	11 Aug
	In situ	1.20	09 Mar	11 Apr	3.58	1.64	80	28 May

: Figure Captions

- Figure 1. Nearly identical curves which describes resource uptake (a), zooplankton grazing (b), and phytoplankton mortality (c). Figure (a) shows four uptake functions, which have been optimised to the default uptake function, monod (U_1). Figure (b) shows two grazing functional forms, the holling type III (G_1) and type II (G_2) functions. Four phytoplankton mortality functions are shown on figure (c), whereby hyperbolic is the default function. The optimisation method is describe in section 2.1, 2.2, and 2.3. Table 1 describes the function's equations and parameters.
- Figure 2. SeaWIFs-derived mean 1998 chlorophyll-*a* (mg m^{-3}) overlain with the 5 oceanographic stations time series site (Red dots). These stations are located in different oceanic regions: oligotrophic (ALOHA and BATS), coastal (L4 and Cariaco), and abyssal plain (PAP).
- 10 Figure 3. Chlorophyll and nitrogen profiles from ensemble mean ((a) and (d) respectively), in situ observations ((c) and (f) for chlorophyll and nitrogen respectively), and 75th and 25th quartile range of concentrations at each depth ((b) for chlorophyll and (e) for nitrogen) at station PAP. The range are obtained by averaging the concentrations from all ensemble members for 10 years at each depths. Black dots in the second column show the mean concentration of the ensemble mean over the time series (from January 1998-December 2007). White solid line in (a) shows mixed layer depth.
- 15 Figure 4. Inter-annual mean of surface chlorophyll from all the study sites ((a)-(e)) and the 10-year annual mean (g), all measured in mg m^{-3} . The boxplots show the ensemble annual means. Blue cross is the in situ observation, red open circle, black dot, and blue stars are the ensemble mean, median, and the default run respectively. The blue box is the 75th (top) and 25th (bottom) quartiles. Red line is the median. The whiskers are the ensemble minimum and maximum mean of surface chlorophyll. Annual mean values and NRR are described in Table 4.
- 20 Figure 5. 10-year monthly mean surface chlorophyll from all the study sites ((a)-(e)), showing the seasonal dynamics of surface chlorophyll (mg m^{-3}). The boxplots show the ensemble ~~annual~~ seasonal means. Blue cross is the in situ observation, red open circle, black dot, and blue stars are the ensemble mean, median, and the default run respectively. The blue box is the 75th (top) and 25th (bottom) quartiles. The red line is the median. The whiskers are the ensemble minimum and maximum mean of surface chlorophyll. In station PAP, in situ data for December is not available due to low light and high cloud cover.
- 25 Figure 6. Inter-annual variability of averaged 200 m integrated nitrogen, from all the study sites ((a)-(e)), and the annual mean (f). Since the in situ data for PAP does not always cover the first 200m, the overall mean nitrogen concentration from all depth

is used instead. For station L4, in situ nitrogen is only collected on the surface. Blue cross is the in situ observation, red open circle, black dot, and blue stars are the ensemble mean, median, and default run respectively. The blue box is the 75th(top) and 25th(bottom) quartiles. Red line is the median, and the whiskers are the ensemble minimum and maximum of the integrated nitrogen. In station L4 and PAP data for nitrogen is only available from 2000-2007 and 2002-2004 respectively.

- 5 Figure 7. 10-year monthly mean of averaged 200 m integrated nitrogen from all the study sites ((a)-(e)), showing the seasonal dynamics of nitrogen (mmol m^{-3}). For station PAP, the nitrogen shown is the overall profile, and in L4, the in situ nitrogen concentration is only available at the surface. The boxplot shows the ensemble monthly means. Blue cross is the in situ observation, red open circle, black dot, and blue stars are the ensemble mean, median, and the default run respectively. The blue box is the 75th (top) and 25th (bottom) quartiles. The red line is the median. The whiskers are the ensemble minimum and maximum mean of integrated nitrogen. In station PAP, the in situ data is only collected from 2002-2004 and L4 from 2000-2007.

Figure 8. Annual mean of surface chlorophyll when changing only one process at a time (blue box), overlain with annual mean of all ensemble members (green box) at five oceanographic stations. Ensemble mean and median plotted in the figure (shown in red open circle and black closed circle), are the from the 128 ensemble members.

- 15 Figure 9. Time series (from January 1998-December 2007) of ensemble mean and in situ, and range of chlorophyll and nitrogen concentrations at oligotrophic stations. Station ALOHA is shown on (a)-(f) and BATS is shown on (g)-(l). White solid line in (b) and (g) represents mixed layer depth. (b), (d), (h), and (j) are the 75th and 25th percentile range of chlorophyll ((b) for ALOHA and (h) for BATS) and nitrogen ((d) for ALOHA and (j) BATS) over the depth. The range is obtained by averaging the chlorophyll and nitrogen concentrations of each ensemble members over the time series at each depth. Black dots in (b), 20 (d), (h), and (j) are the mean of the ensemble. Ensemble mean chlorophyll profiles (shown on (a) and (g)) and nitrogen ((e) and (k)) are obtained from all of the ensemble members. *In situ* chlorophyll are shown in (c) and (i), and nitrogen are shown in (g) and (l), for ALOHA and BATS respectively.

Figure 10. Mean integrated primary production averaged over 200m that are available in (a) ALOHA and (b) Cariaco, and (c) the annual mean. The NRR for ALOHA and Cariaco are 1.12 and 0.80 respectively.

- 25 Figure 11. Chlorophyll profile 10-year means ((a)-(d)) and its RMSEs ((e)-(f)) at four oceanographic station from all of the ensemble members. Station L4 is not included as chlorophyll data is only taken at the surface. These are arranged by the lowest chlorophyll (top left) mean to the highest (bottom right), depending on the oceanographic regions.

Figure 12. 10-year mean and RMSE of surface chlorophyll (mg m^{-3}) and nitrogen (mmol m^{-3}) at five stations from all ensemble members. The first panel ((a)-(e)) shows surface chlorophyll mean and the third panel ((k)-(o)) shows nitrogen mean. RMSEs are shown on the second panel ((f)-(j)) and fourth ((p)-(t)) for surface chlorophyll, and nitrogen respectively. Concentrations and RMSEs are arranged by the lowest chlorophyll (top left) mean to the highest (bottom right), depending on the oceanographic regions. For station PAP, the sequence is sorted based on coastal station. The y-axis shows combination of uptake (U_1, U_2, U_3 , and U_4) and grazing (G_1 and G_2), and x-axis shows combinations of phytoplankton (ρ) and zooplankton (ξ) mortalities.

Figure 13. 10-year mean and RMSE of nitrogen (mmol m^{-3}), at five stations from all ensemble members. The first panel ((a)-(e)) shows nitrogen mean and RMSEs are shown on the second panel ((f)-(j)). Concentrations and RMSEs are arranged by the lowest chlorophyll (top left) mean to the highest (bottom right), depending on the oceanographic regions. For station PAP, the sequence is sorted based on coastal station. The y-axis shows combination of uptake (U_1, U_2, U_3 , and U_4) and grazing (G_1 and G_2), and x-axis shows combinations of phytoplankton (ρ) and zooplankton (ξ) mortalities.

Figure 14. Time series of chlorophyll and nitrogen profile of ensemble mean, their range, and in situ concentrations at the coastal stations Cariaco (a-f) and L4 (g-h) from January 1998-December 2007. (a) and (d) show chlorophyll and nitrogen ensemble mean at Cariaco respectively. White solid line in (a) is the mixed layer depth. (b) and (e) shows the 75th and 25th percentile of chlorophyll and nitrogen concentrations at each depth. The black dots are the mean of the ensemble. These range are obtained from the 10-year mean concentrations at each depth. Since in situ chlorophyll and nitrogen were taken at the surface in station L4, only surface time series were shown in (g-h). The grey shades on chlorophyll, shown in (g), and nitrogen, shown in (h) time series show 75th and 25th percentile of the range. Blue and red dots are in situ concentrations for chlorophyll and nitrogen respectively.

Figure 15. Phytoplankton phenology metrics at the five stations. Blue cross is the in situ, red, black, and blue dots are the ensemble mean, median, and the default run respectively. The timings and concentrations are averaged annually from January 1998 to December 2007.

Figure A1. Determining phenology using a combination of threshold method and curve fit at station L4, here the initiation is when the fitted curve is above 50% of the maximum peak, however the termination is on the first valley.

: Figures

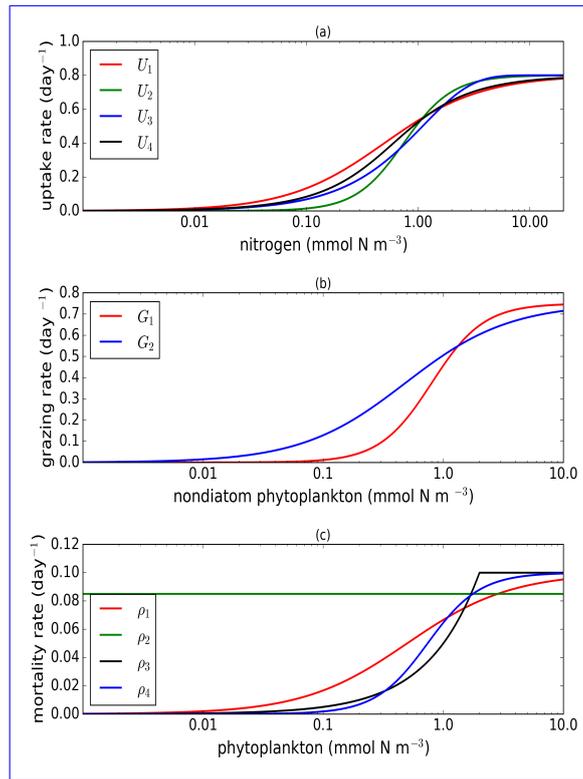


Figure 1. Nearly identical curves which describes resource uptake (a), zooplankton grazing (b), and phytoplankton mortality (c). Figure (a) shows four uptake functions, which have been optimised to the default uptake function, monod (U_1). Figure (b) shows two grazing functional forms, the holling type III (G_1) and type II (G_2) functions. Four phytoplankton mortality functions are shown on figure (c), whereby hyperbolic is the default function. The optimisation method is describe in section 2.1, 2.2, and 2.3. Table 1 describes the function's equations and parameters.

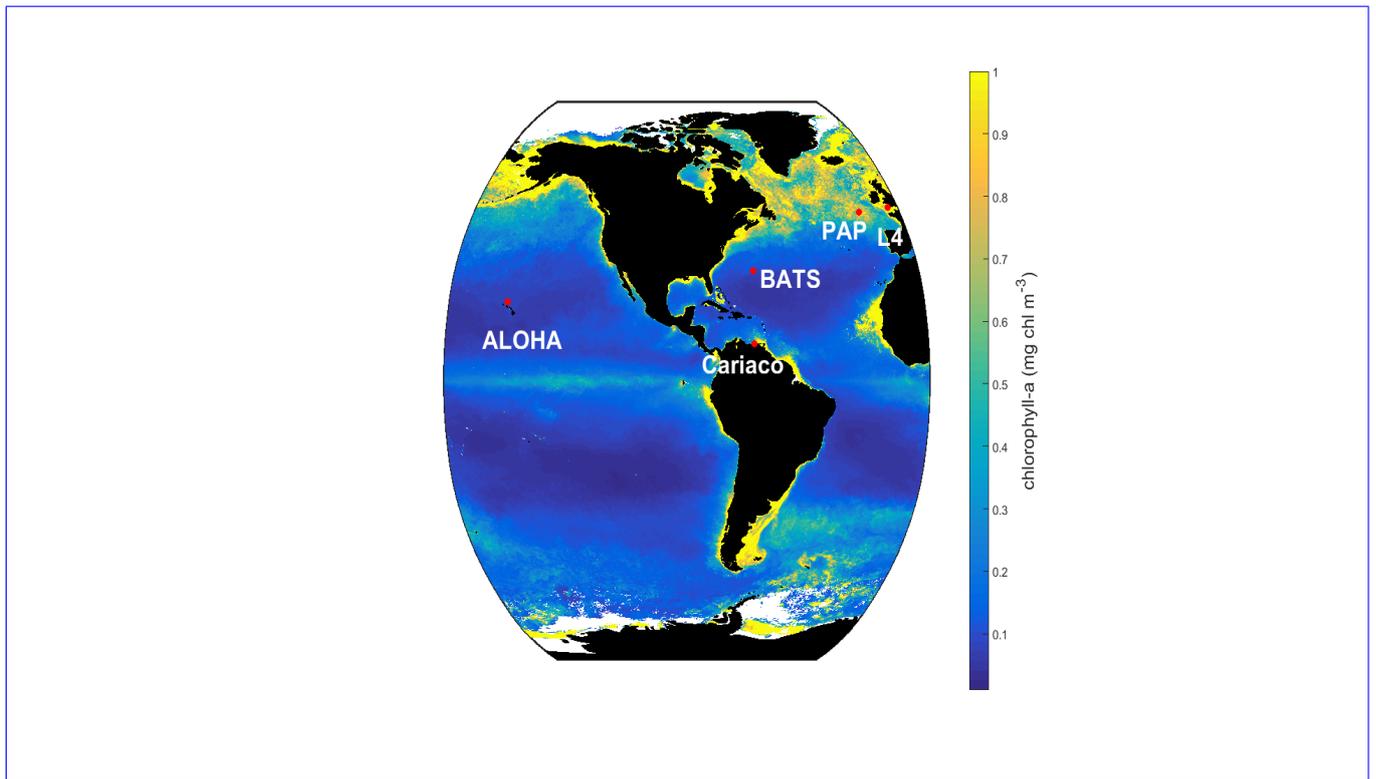


Figure 2. SeaWIFs-derived mean 1998 chlorophyll-*a* (mg m^{-3}) overlain with the 5 oceanographic stations time series site (Red dots). These stations are located in different oceanic regions: oligotrophic (ALOHA and BATS), coastal (L4 and Cariaco), and abyssal plain (PAP).

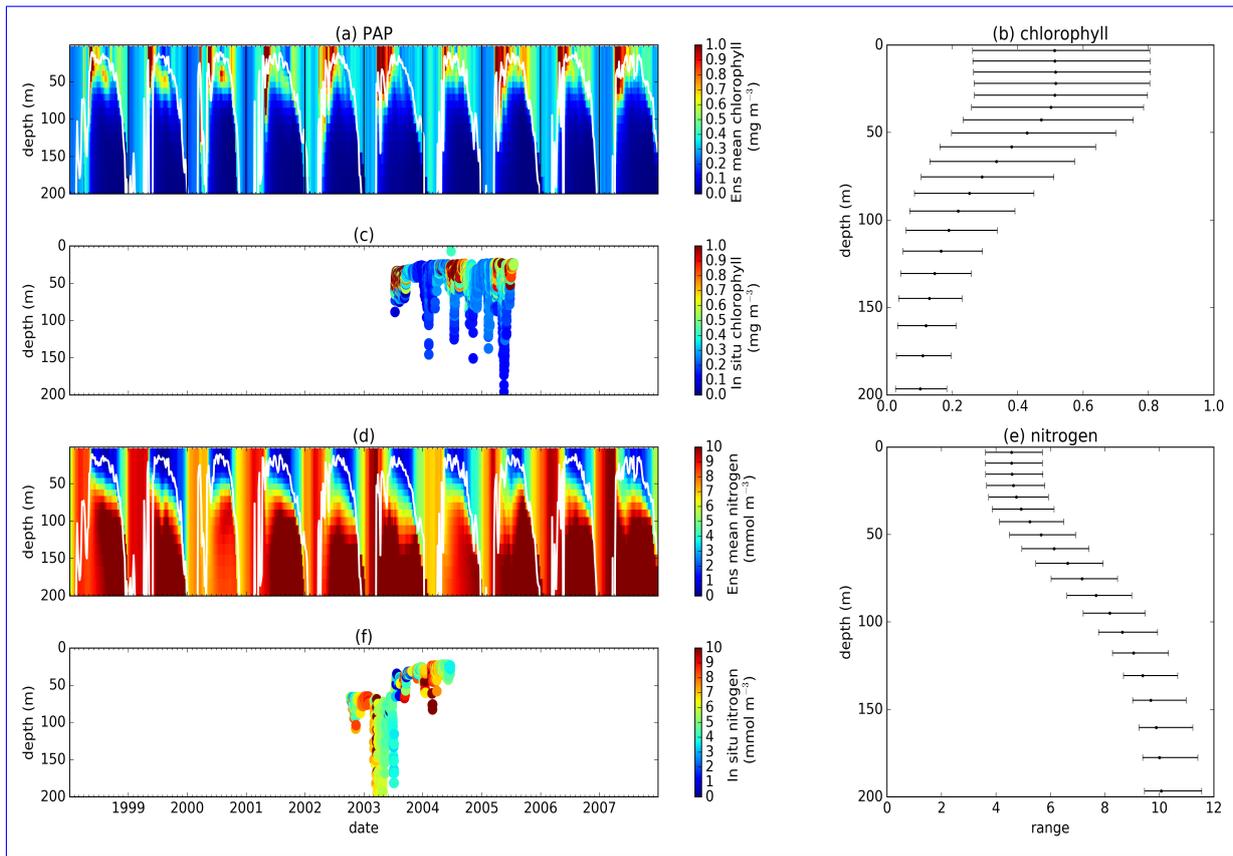


Figure 3. Chlorophyll and nitrogen profiles from ensemble mean ((a) and (d) respectively), in situ observations ((c) and (f) for chlorophyll and nitrogen respectively), and 75th and 25th quartile range of concentrations at each depth ((b) for chlorophyll and (e) for nitrogen) at station PAP. The range are obtained by averaging the concentrations from all ensemble members for 10 years at each depths. Black dots in the second column show the mean concentration of the ensemble mean over the time series (from January 1998-December 2007). White solid line in (a) shows mixed layer depth.

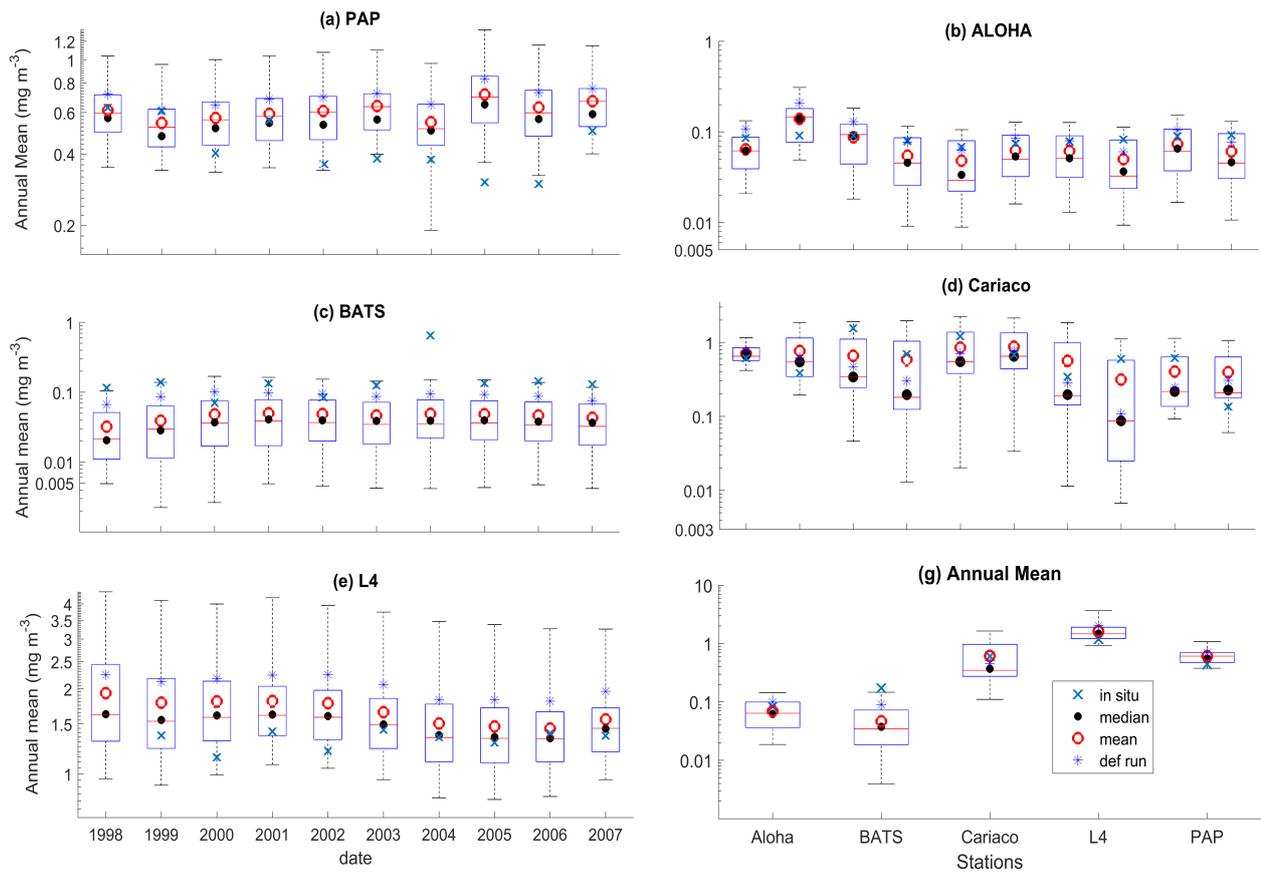


Figure 4. Inter-annual mean of surface chlorophyll from all the study sites ((a)-(e)) and the 10-year annual mean (g), all measured in mg m^{-3} . The boxplots show the ensemble annual means. Blue cross is the in situ observation, red open circle, black dot, and blue stars are the ensemble mean, median, and the default run respectively. The blue box is the 75th (top) and 25th (bottom) quartiles. Red line is the median. The whiskers are the ensemble minimum and maximum mean of surface chlorophyll. Annual mean values and NRR are described in Table 4.

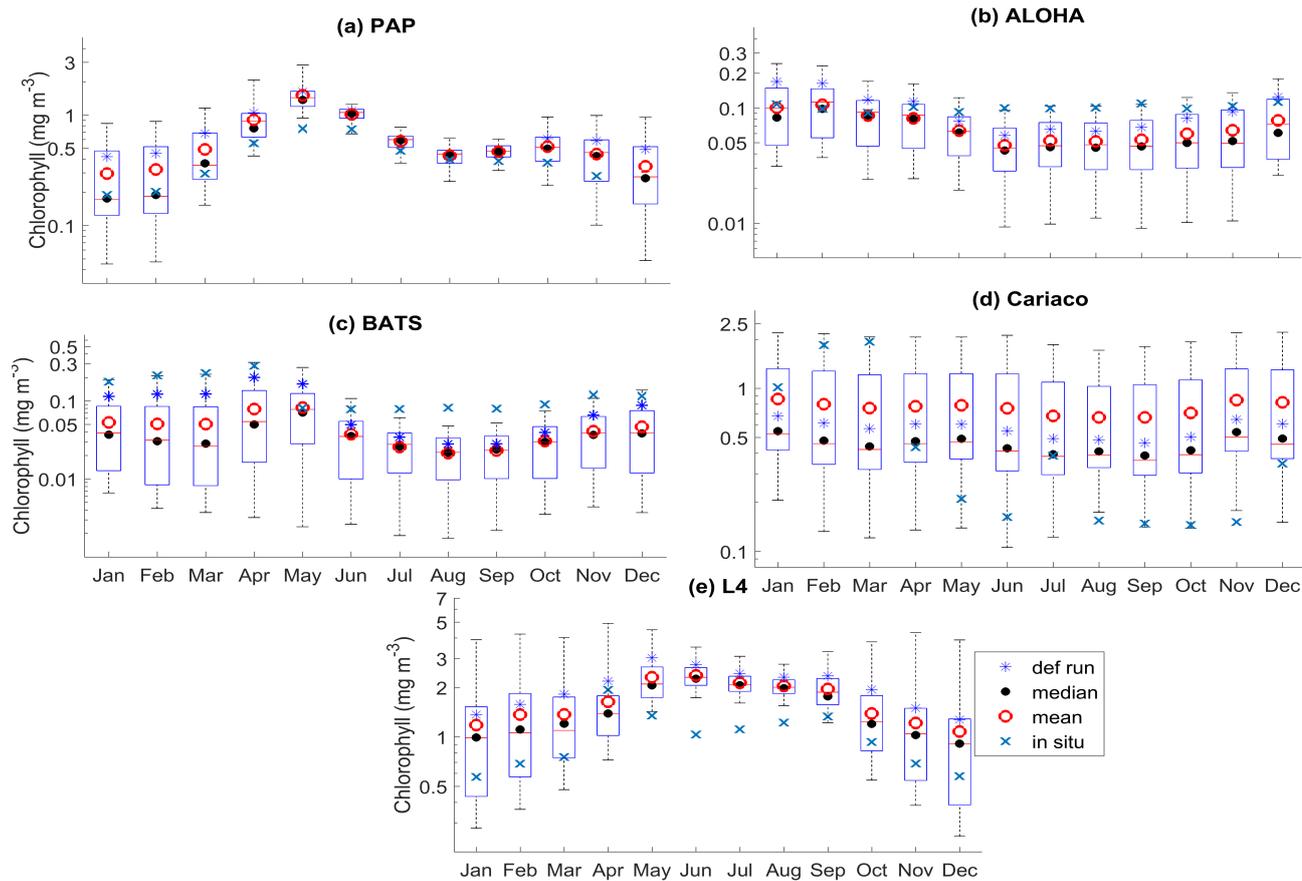


Figure 5. 10-year monthly mean surface chlorophyll from all the study sites ((a)-(e)), showing the seasonal dynamics of surface chlorophyll (mg m⁻³). The boxplots show the ensemble annual seasonal means. Blue cross is the in situ observation, red open circle, black dot, and blue stars are the ensemble mean, median, and the default run respectively. The blue box is the 75th (top) and 25th (bottom) quartiles. The red line is the median. The whiskers are the ensemble minimum and maximum mean of surface chlorophyll. In station PAP, in situ data for December is not available due to low light and high cloud cover.

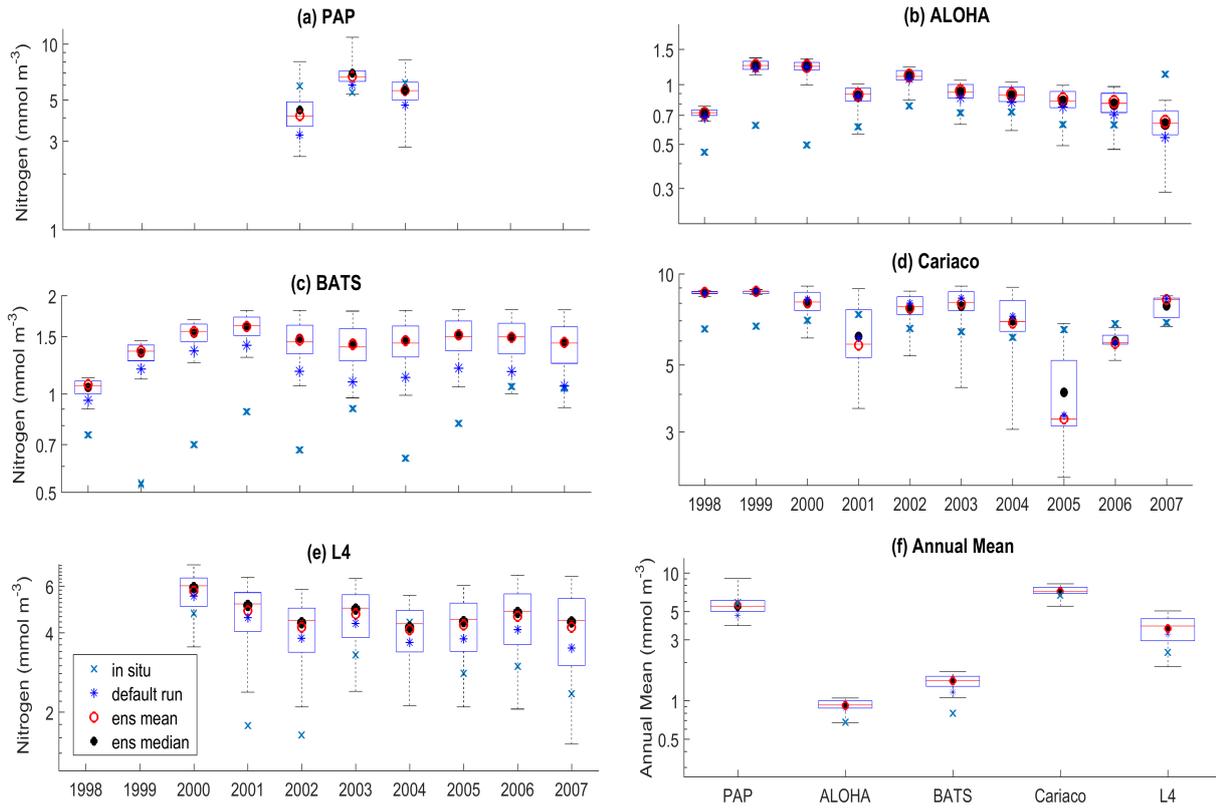


Figure 6. Inter-annual variability of averaged 200 m integrated nitrogen, from all the study sites ((a)-(e)), and the annual mean (f). Since the in situ data for PAP does not always cover the first 200m, the overall mean nitrogen concentration from all depth is used instead. For station L4, in situ nitrogen is only collected on the surface. Blue cross is the in situ observation, red open circle, black dot, and blue stars are the ensemble mean, median, and default run respectively. The blue box is the 75th(top) and 25th(bottom) quartiles. Red line is the median, and the whiskers are the ensemble minimum and maximum of the integrated nitrogen. In station L4 and PAP data for nitrogen is only available from 2000-2007 and 2002-2004 respectively.

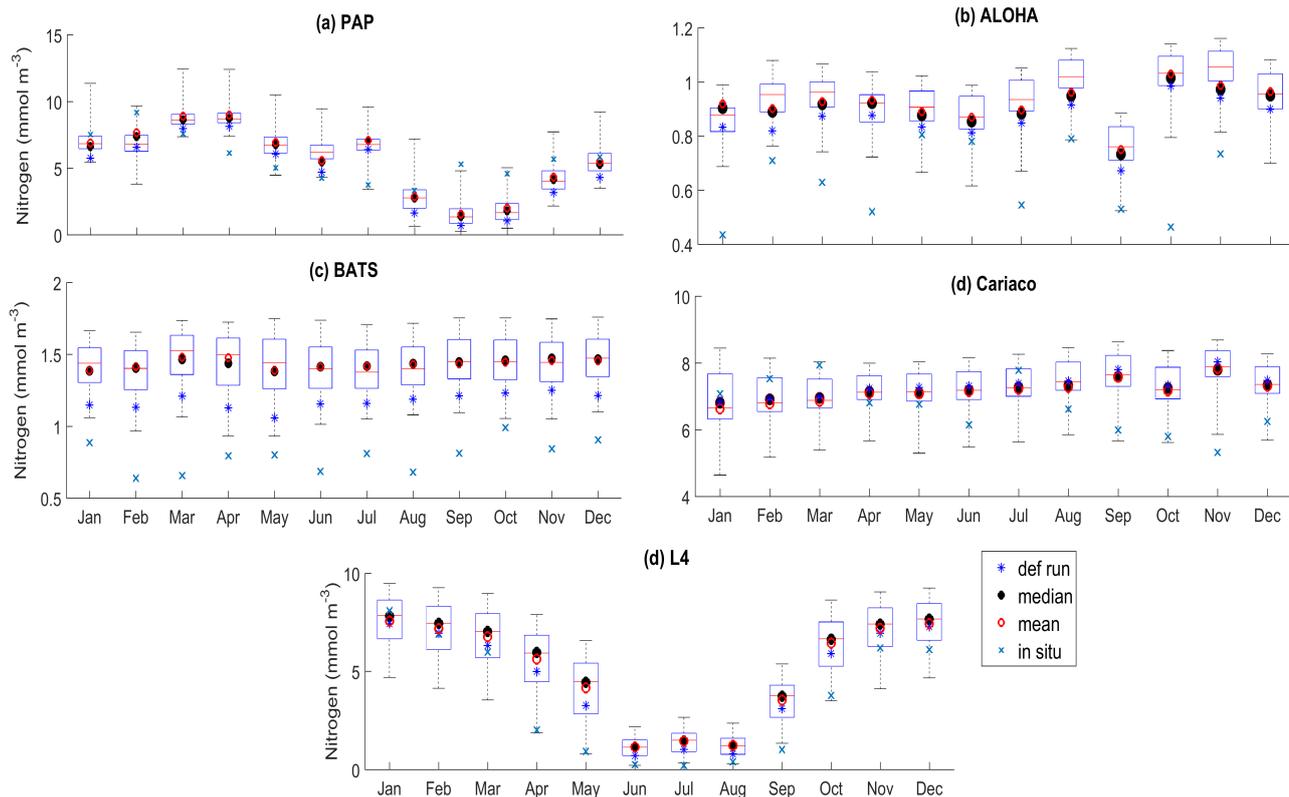


Figure 7. 10-year monthly mean of averaged 200 m integrated nitrogen from all the study sites ((a)-(e)), showing the seasonal dynamics of nitrogen (mmol m⁻³). For station PAP the nitrogen shown is the overall profile, and in L4, the in situ nitrogen concentration is only available at the surface. The boxplot shows the ensemble monthly means. Blue cross is the in situ observation, red open circle, black dot, and blue stars are the ensemble mean, median, and the default run respectively. The blue box is the 75th (top) and 25th (bottom) quartiles. The red line is the median. The whiskers are the ensemble minimum and maximum mean of integrated nitrogen. In station PAP, the in situ data is only collected from 2002-2004 and L4 from 2000-2007.

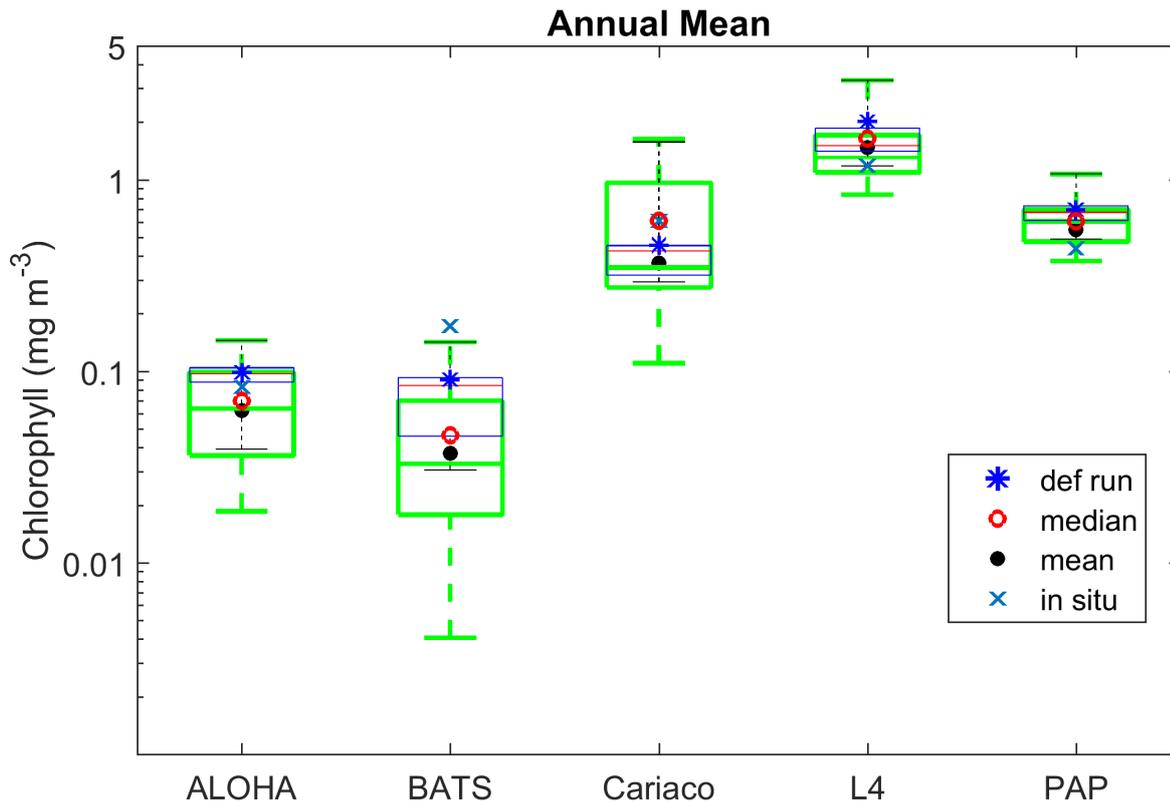


Figure 8. Annual mean of surface chlorophyll when changing only one process at a time (blue box), overlain with annual mean of all ensemble members (green box) at five oceanographic stations. Ensemble mean and median plotted in the figure (shown in red open circle and black closed circle), are the from the 128 ensemble members.

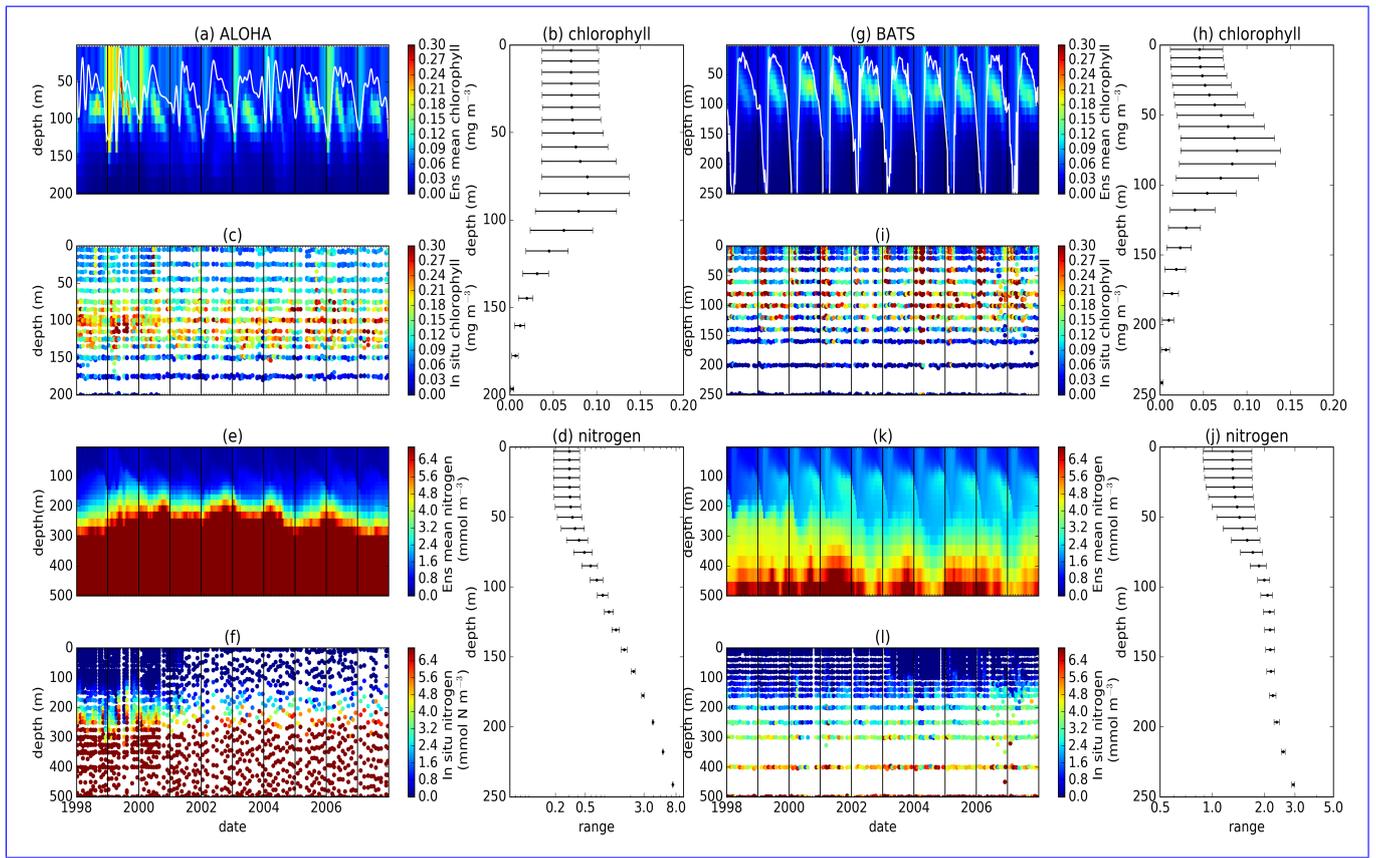


Figure 9. Time series (from January 1998-December 2007) of ensemble mean and in situ, and range of chlorophyll and nitrogen concentrations at oligotrophic stations. Station ALOHA is shown on (a)-(f) and BATS is shown on (g)-(l). White solid line in (b) and (g) represents mixed layer depth. (b), (d), (h), and (j) are the 75th and 25th percentile range of chlorophyll ((b) for ALOHA and (h) for BATS) and nitrogen ((d) for ALOHA and (j) BATS) over the depth. The range is obtained by averaging the chlorophyll and nitrogen concentrations of each ensemble members over the time series at each depth. Black dots in (b), (d), (h), and (j) are the mean of the ensemble. Ensemble mean chlorophyll profiles (shown on (a) and (g)) and nitrogen ((e) and (k)) are obtained from all of the ensemble members. *In situ* chlorophyll are shown in (c) and (i), and nitrogen are shown in (f) and (l), for ALOHA and BATS respectively.

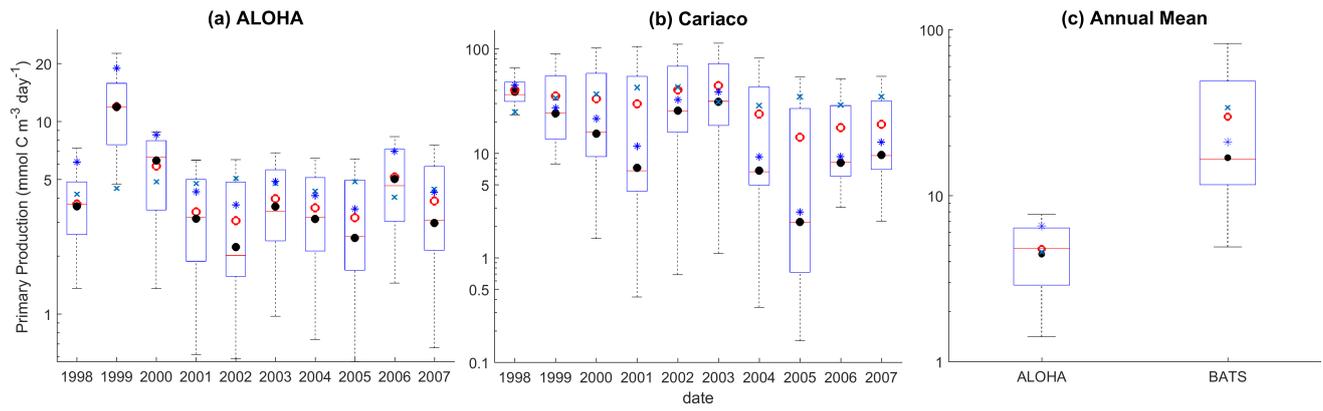


Figure 10. Mean integrated primary production averaged over 200m that are available in (a) ALOHA and (b) Cariaco, and (c) the annual mean. The NRR for ALOHA and Cariaco are 1.12 and 0.80 respectively.

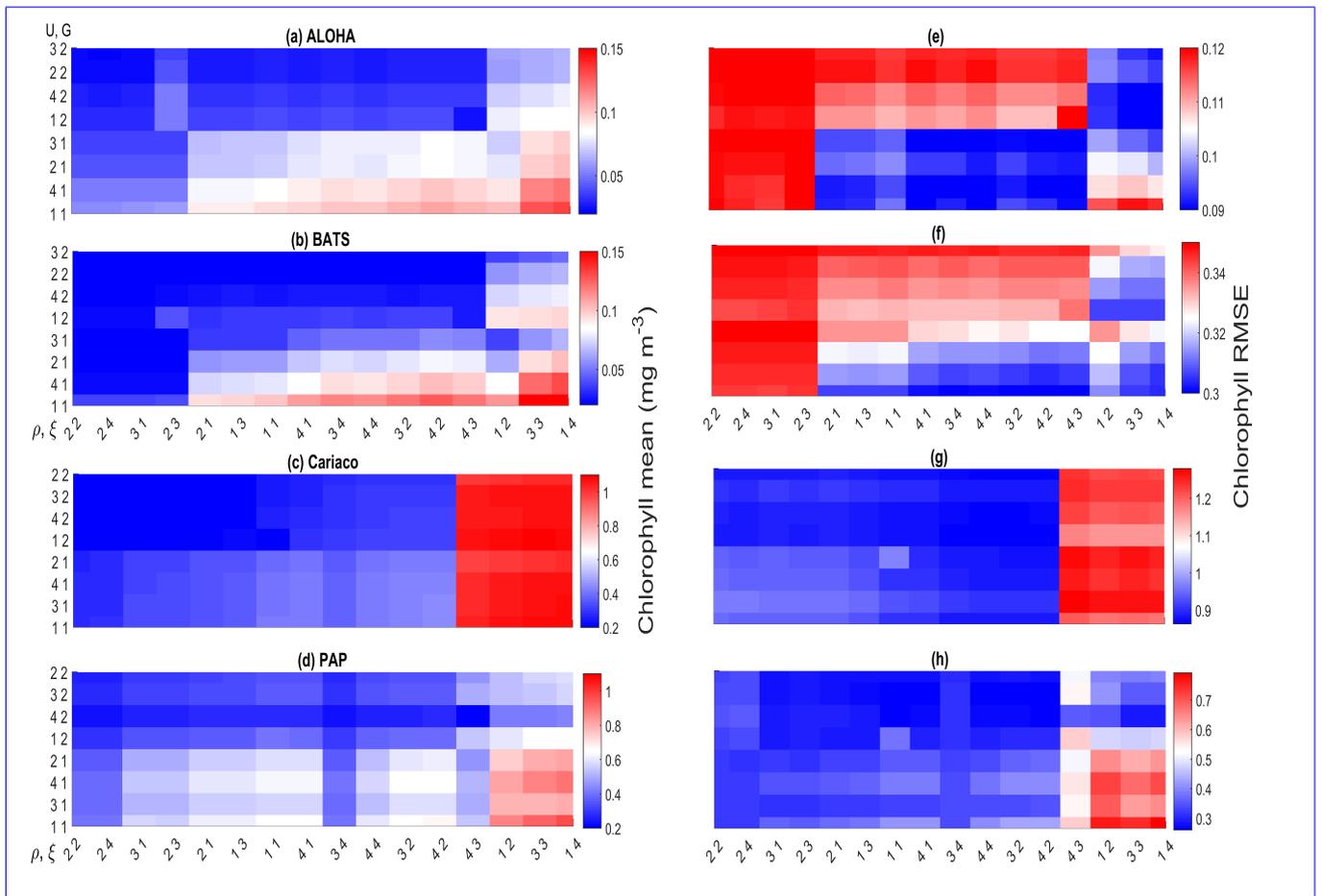


Figure 11. Chlorophyll profile 10-year means ((a)-(d)) and its RMSEs ((e)-(f)) at four oceanographic station from all of the ensemble members. Station L4 is not included as chlorophyll data is only taken at the surface. These are arranged by the lowest chlorophyll (top left) mean to the highest (bottom right), depending on the oceanographic regions.

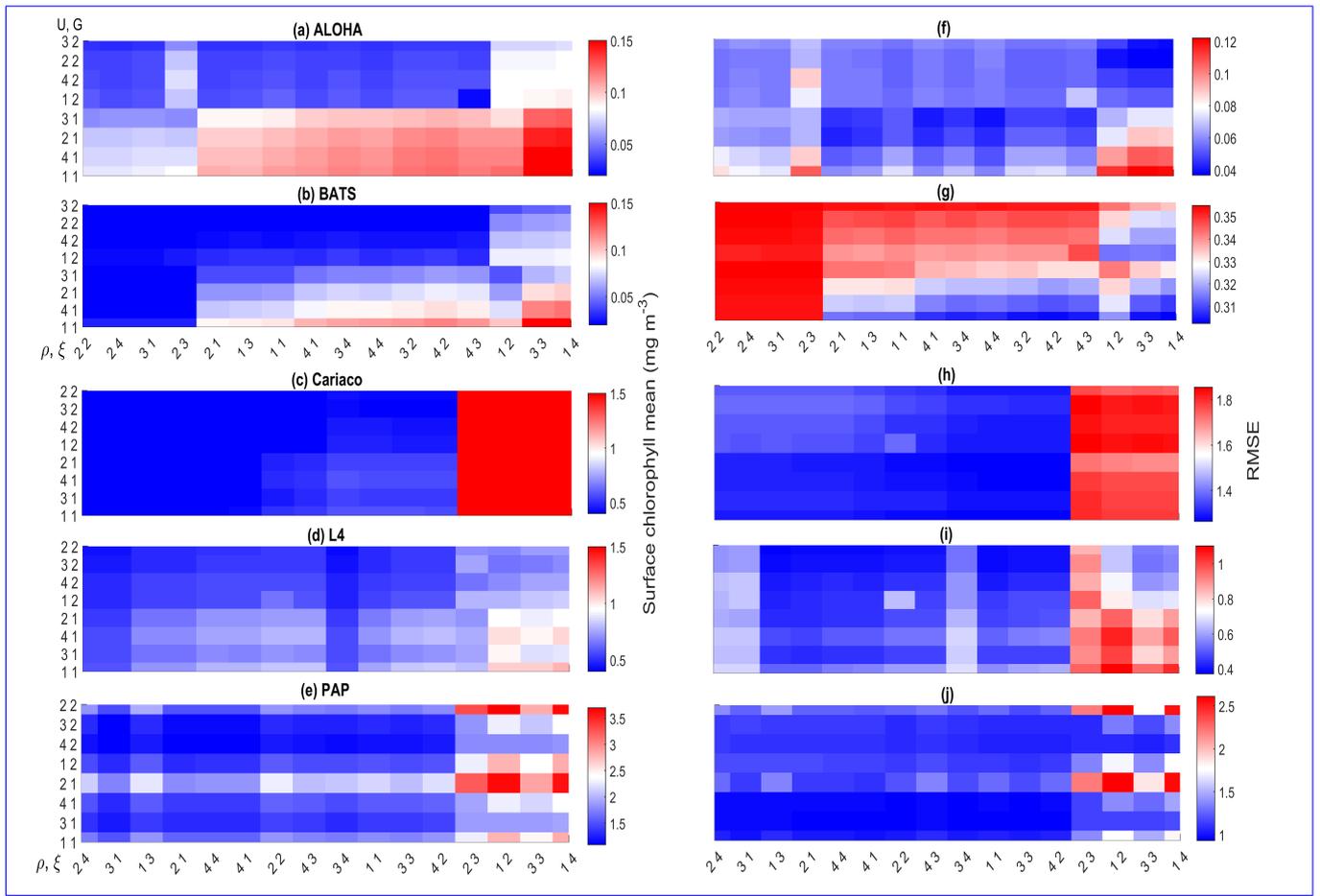


Figure 12. 10-year mean and RMSE of surface chlorophyll (mg m^{-3}), and nitrogen (mmol m^{-3}), at five stations from all ensemble members. The first panel ((a)-(e)) shows surface chlorophyll mean, and the second panel ((f)-(j)) shows RMSEs. Concentrations and RMSEs are arranged by the lowest chlorophyll (top left) mean to the highest (bottom right), depending on the oceanographic regions. For station PAP, the sequence is sorted based on coastal station. The y-axis shows combination of uptake (U_1, U_2, U_3 , and U_4) and grazing (G_1 and G_2), and x-axis shows combinations of phytoplankton (ρ) and zooplankton (ξ) mortalities.

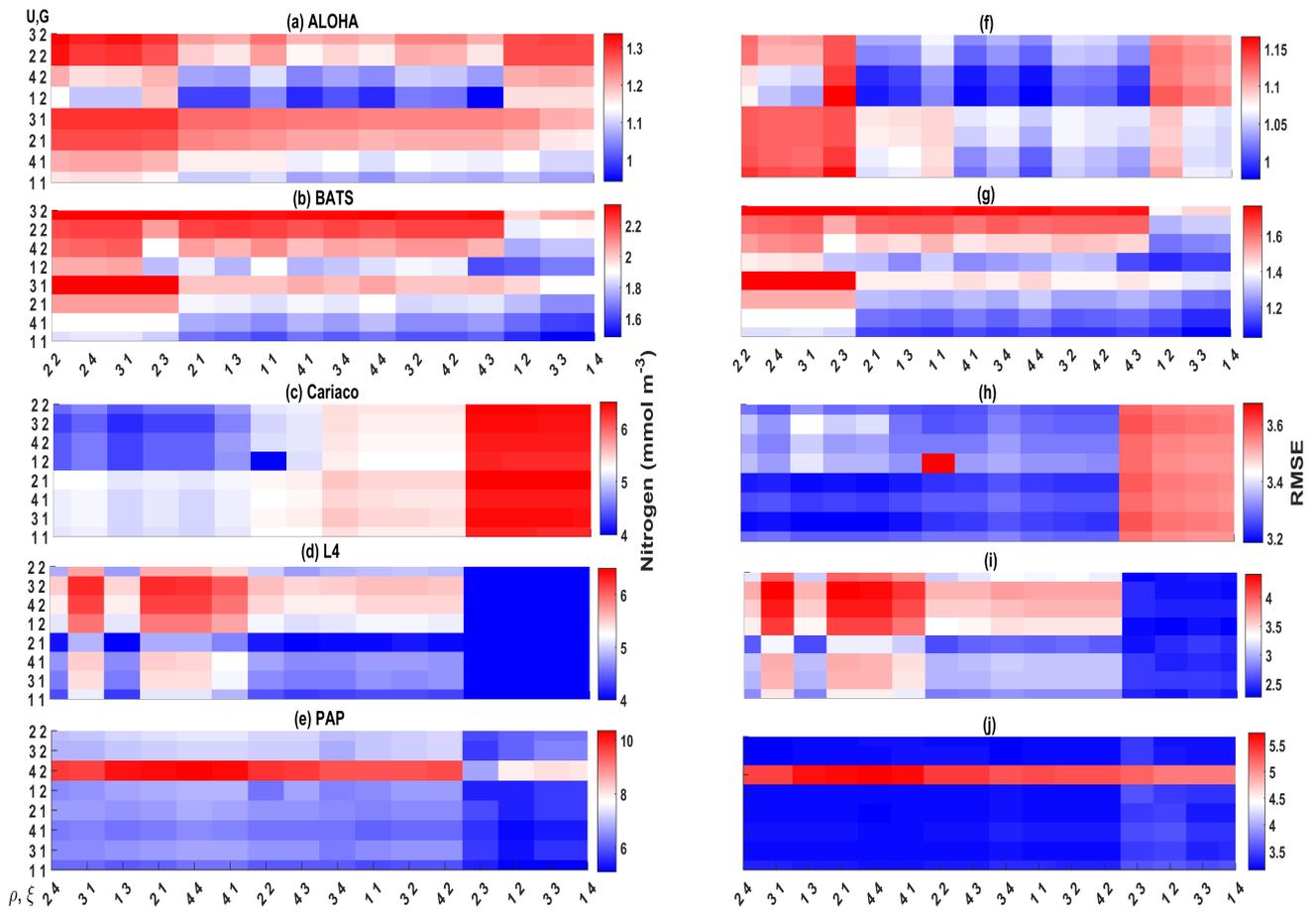


Figure 13. 10-year mean and RMSE of nitrogen (mmol m^{-3}), at five stations from all ensemble members. The first panel ((a)-(e)) shows nitrogen mean and RMSEs are shown on the second panel ((f)-(j)). Concentrations and RMSEs are arranged by the lowest chlorophyll (top left) mean to the highest (bottom right), depending on the oceanographic regions. For station PAP, the sequence is sorted based on coastal station. The y-axis shows combination of uptake (U_1, U_2, U_3 , and U_4) and grazing (G_1 and G_2), and x-axis shows combinations of phytoplankton (ρ) and zooplankton (ξ) mortalities.

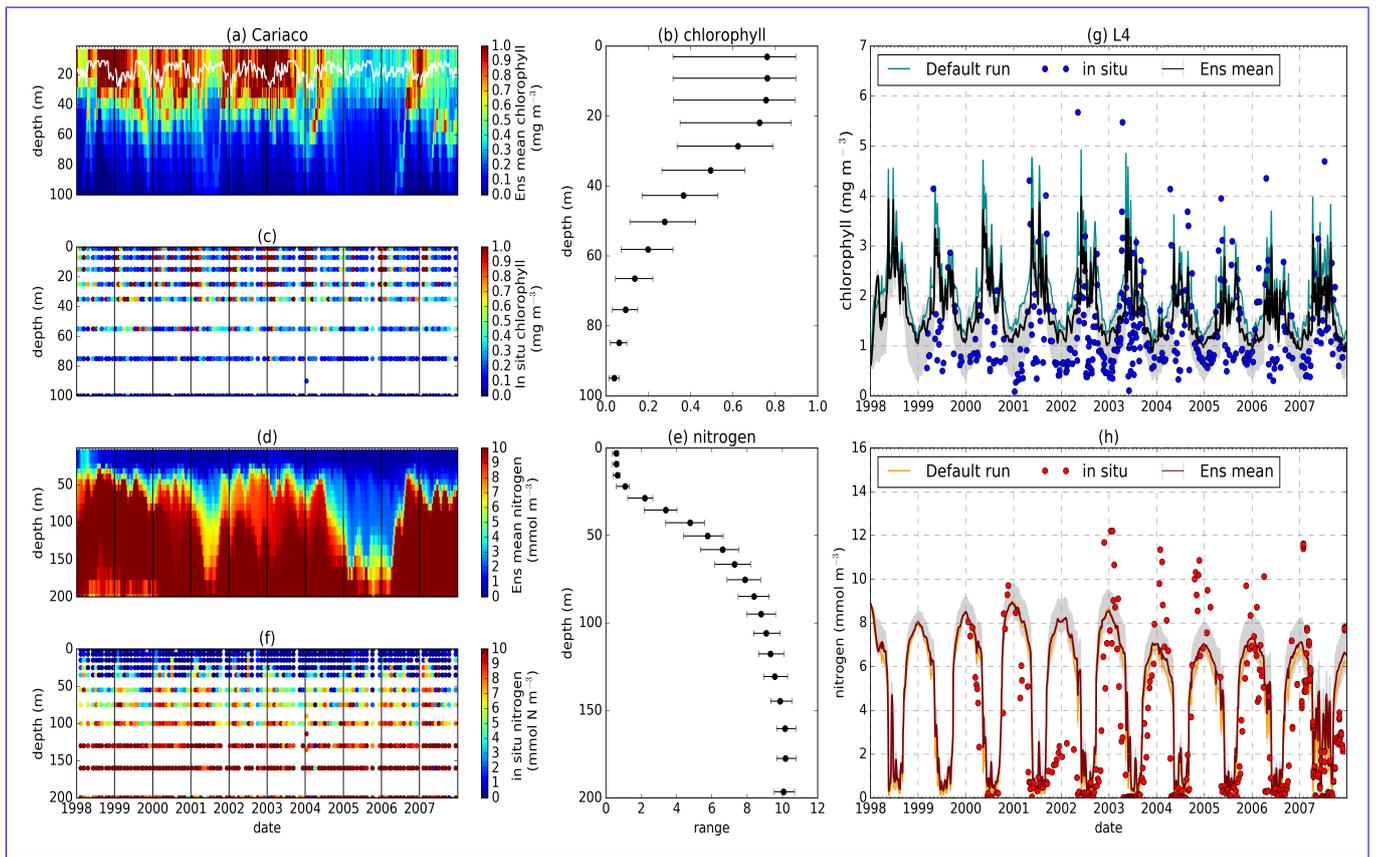


Figure 14. Time series of chlorophyll and nitrogen profile of ensemble mean, their range, and in situ concentrations at the coastal stations Cariaco (a-f) and L4 (g-h) from January 1998-December 2007. (a) and (d) show chlorophyll and nitrogen ensemble mean at Cariaco respectively. White solid line in (a) is the mixed layer depth. (b) and (e) shows the 75th and 25th percentile of chlorophyll and nitrogen concentrations at each depth. The black dots are the mean of the ensemble. These range are obtained from the 10-year mean concentrations at each depth. Since in situ chlorophyll and nitrogen were taken at the surface in station L4, only surface time series were shown in (g-h). The grey shades on chlorophyll, shown in (g), and nitrogen, shown in (h) time series show 75th and 25th percentile of the range. Blue and red dots are in situ concentrations for chlorophyll and nitrogen respectively.

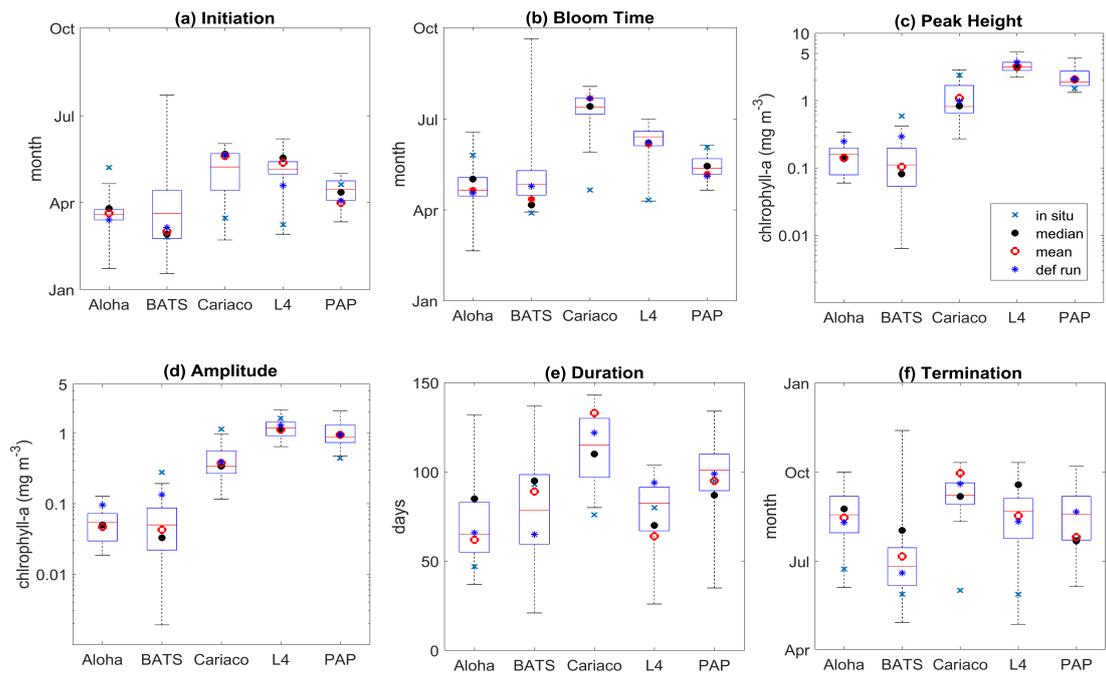


Figure 15. Phytoplankton phenology metrics at the five stations. Blue cross is the in situ, red, black, and blue dots are the ensemble mean, median, and the default run respectively. The timings and concentrations are averaged annually from January 1998 to December 2007.

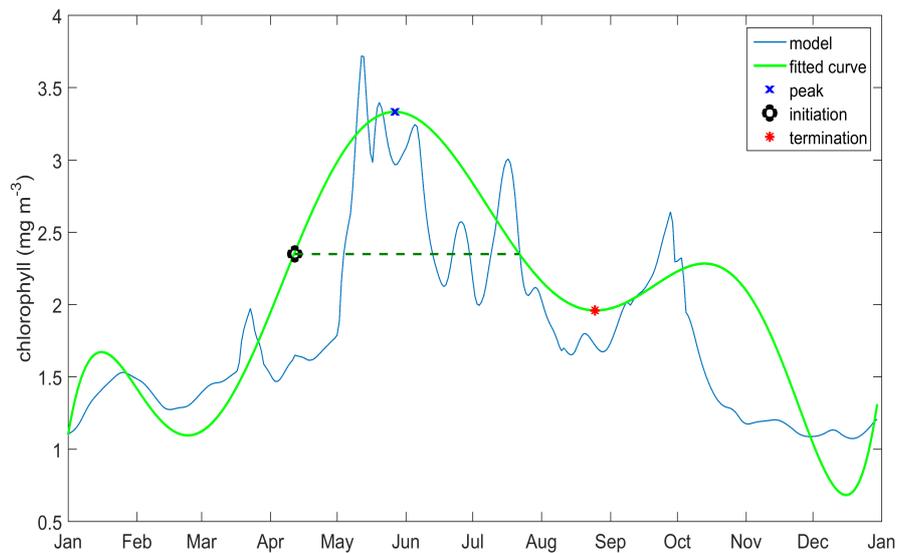


Figure A1. Determining phenology using a combination of threshold method and curve fit at station L4, here the initiation is when the fitted curve is above 50% of the maximum peak, however the termination is on the first valley.