Summary
This manuscript presents an interesting combined laboratory and modelling study of the nutrient dynamics of a diatom species common in the Arctic. The laboratory component uses two experimental set-ups: 1. axenic cultures of the diatom species; 2. cultures of the diatom species that include associated bacteria. Short incubations (_2 weeks) of these cultures take them from exponential phase through to stationary phase, with the cultures sampled throughout to measure cell counts, nutrient concentrations, etc. After an initial period of diatom cell number growth (week 1) in both cultures, this stops as NOx and dSi concentrations approach limiting concentrations. However, NH4 is consistently higher in the non-axenic cultures, and the bacterial cell counts in these cultures increase exponentially during the latter period of the incubations (week 2). The authors interpret the presence of bacteria as being conducive to supplying the diatoms with regenerated nutrients. The modelling component uses a base model, G98, and an extended model based on this that includes a number of additional processes with relevance to the laboratory setting and the hypothesised role of bacterial remineralisation in supporting phytoplankton growth. The models are tuned to fit the laboratory data, with a manual phase to retain consistent parameter values between the models. The authors conclude with a discussion on the application of their results to the real Arctic and its expected future state.

I have listed a number of significant general comments below, followed by more specific and often minor comments. Overall, my assessment is that the manuscript requires major revision to clarify and amend the work described.

We want to sincerely thank the reviewer for the very thorough review and believe the suggestions helped to improve the manuscript considerably. We included all suggestions into a revised version as described below. We also changed the fixed 80% reduction term in our model to a parameter that was subject to the fitting approach and sensitivity analyses. We fitted the model again with a more automated fitting approach and reached better fits for both the G98 and extended EXT model.

General comments
Upfront, my modelling background means that I cannot comment directly on the details of the laboratory work in the study. However, I note that the experiments conducted exhibit anomalies that are not addressed in the manuscript. In Figure 1c, phosphate in bacterial cultures exhibits a strong spike upwards at day 8 that persists and shows high variability. In Figure 3d, chlorophyll in bacterial cultures shows a marked but temporary spike downward at day 8. While the latter is likely a replication or measurement issue, the former is harder to understand, and the manuscript does not discuss its scale or variability. It would be useful to know what the authors believed happened here, particularly in the case of phosphate where bottle concentrations approximately double against a backdrop of slowly declining phytoplankton and rising bacteria concentrations. The model may even be able to help on this point.

Since our study, does not focus on phosphate, we did not describe its dynamics in detail. However, we acknowledge that a short description and explanation of the anomalies is helpful for the reader to understand the overall experiment and nutrient dynamics and added some details.

The strong spike of phosphate after day 8 corresponds with the end of the exponential phase for algal growth and a spike of ammonium. At the same time bacteria abundances start increasing considerably. Thus, we explain the phosphate peak by increased bacterial regeneration (source of phosphate) and decreased algal uptake (sink of phosphate) at the same day. Due to the small bacteria biomass compared to algae, we assume limited phosphate incorporation in the bacteria biomass pool. Besides, the diatom culture may excrete additional DOM under stress, such as silicate limitation, contributing to labile DOM available for regeneration and thereby increasing the phosphate peak, which is however not part of the current extended model. We calculated the N:P ratio of the NH4 and PO4 peak at day 8, and realized that the ratio is approximately 1:1, which is different from the Redfield ratio. We see this as evidence that increased regeneration of NH4 and PO4 is not the only explanation for the PO4 peak and suggest the storage of (organic) polyphosphate in diatoms and release under stress as another potential source.

Changes in the text:

3.1) “With the onset of the stationary phase in the BAC+ experiment, PO4 and NH4 concentrations doubled within 2 to 4 days and stayed high with variations in phosphate concentrations, while they stayed low in BAC-. With depletion of NO3 in BAC+, NH4 concentrations remained high, while PO4 concentrations dropped.”
4.1) “With the start of the stationary phase, NH4 and PO4 concentrations doubled, presumably due to decreased assimilation by the silicate starved diatoms and increased regeneration by bacteria, supplied with increasing labile DOM (doubled remineralisation rate in EXT) excreted by the stressed algae. After NO3 depletion at day 15, also PO4 concentrations drop, indicating a coupling of N:P metabolism “…” Excretion of organic phosphate by diatoms is also common in cultures with surplus orthophosphate (Admiraal and Werner, 1983), which can be another explanation of the phosphate peak after silicate becomes limiting.”

The spike in Chl is based on one single measurement, since the upper and lower range represent max and min values. Since chlorophyll measurements are sensible towards light, and pH, we argue that this negative spike is a measurement artifact of a single sample of the experiment.

Changes:

Figure 4, 5, B1) “…Chlorophyll a concentration in experimental cultures with a potential outlier at day 8, presumably due to photodegradation, causing a negative spike.”

The manuscript’s model description appears incomplete, with equations for terms such as those for dSi omitted. More generally, the manuscript would be improved by simply making clear which models are being run – while the text refers to model G98 and “the extended model”, the plots shown refer instead to “model + excr” and “model - excr”. What might be helpful is to have some sort of diagram of the two main models being used (G98 and Extended) to help illustrate the main connections between state variables, and make clear the differences between the two models.

We added the missing equation and double-cheked for any other incomplete model descriptions.

Changes in Table:

7a)	Silicate uptake *(Monod kinetics after Spilling et al., 2010)*	$\frac{dSi_d}{dt} = V_S^C = \left(V_{max} Si_d \frac{Si_d - S_{min}}{K_{si} S_{min}}\right) C$
7b)		$\frac{dSi_p}{dt} = \frac{-dSi_d}{dt} 14$

We also clarified, which models are run and defined abbreviations (G98 and EXT models/ BAC- and BAC+ treatments) for the different models that we kept throughout the manuscript and figures. We also added a schematic diagram of the two main models, which we agree helps clarifying the differences considerably.

Changes:

2.2) “Details regarding model equations are provided in the Appendix (Table A1) and a schematic representation of the models is given in Figure 1. We used a dynamic cell quota model by Geider et al. (1998) to describe the BAC- experiment (G98). We then extended the G98 model to represent the role of silicate limitation, bacterial regeneration of ammonium, and different kinetics for ammonium and nitrate uptake (EXT) and fitted it to the BAC+ experiment while retaining the parameter values estimated for G98.”…” For testing the importance of DON excretion we also ran the EXT model without DON excretion (EXT–excr).“

Fig. 1)
“

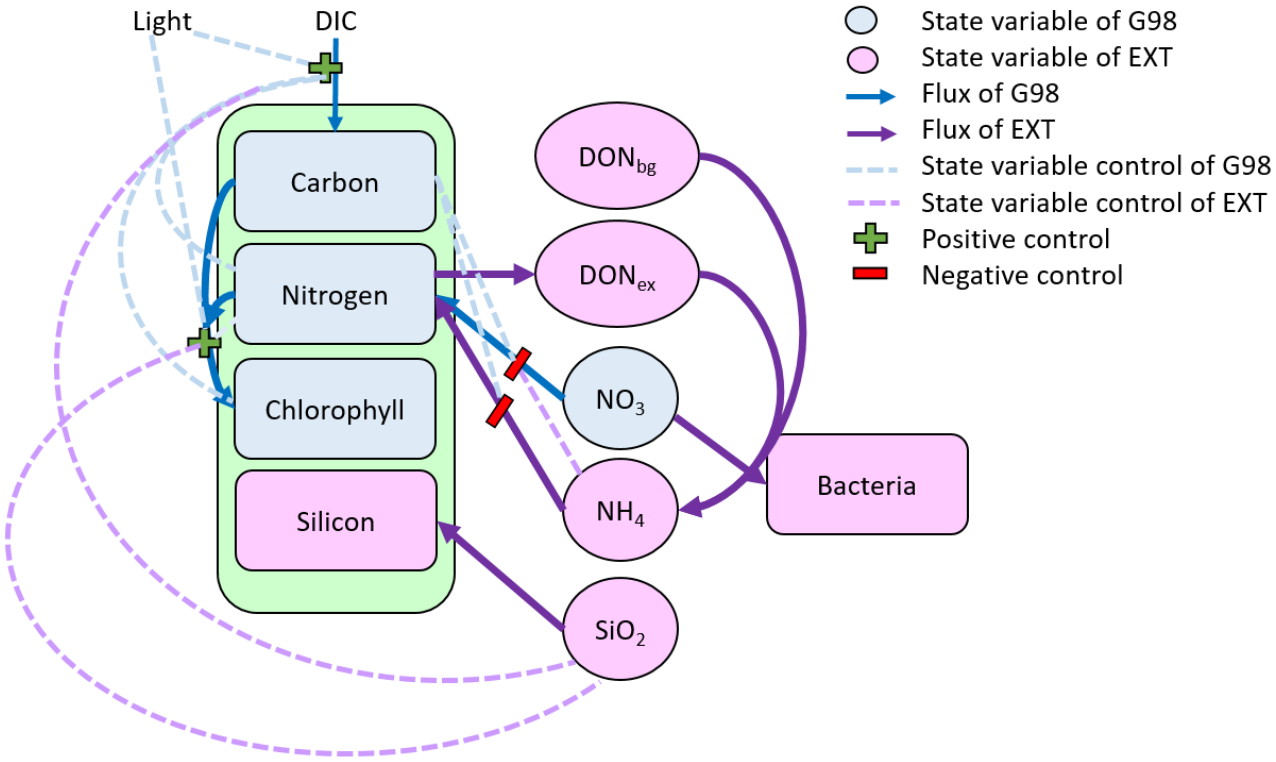

Figure 1. Schematic representation of the state variables and connections and controls in the G98 model (blue) and EXT model (purple). The EXT model has the same formulations as G98 with the additions shown in purple.”

The description of the model tuning needs to be clearer. It’s unclear why some parameters were picked for tuning while others weren’t (e.g. remineralisation parameters were not tuned), or what the rationale for picking the training data streams was (e.g. model ammonium was “loosely constrained” to observations). The text mentions several R packages used, but these are presented without any information about what they do, how they work, or what assumptions they make. For instance, is parameter space sampled by latin hypercube, genetic algorithm, or via local misfit gradient? There’s also an unclear distinction made around a “manual” component of this tuning exercise.

We added a detailed description of the model tuning, R packages, and parameter selection.

Changes:

2.3) “2.3 Model fitting
The model was written as a function of differential equations in R. All model equations are provided in the Appendix (Table A6) and the R code is available in the supplement. The differential equations were solved using the ode function of the deSolve package (Soetaert et al., 2010) with the 2nd-3rd order Runge-Kutta method with automated stepsize control. deSolve is one of the most widely used packages for solving differential equations in R.
Parameter of the G98 model were fitted to the BAC- experiment data and the EXT model was fitted to the BAC+ experiment data. The G98 parameter values were fitted first and retained without changes for the EXT model fitting. The maximum Chl:N ratio (θNmax), minimum and maximum N:C ratios (Qmin, Qmax), and irradiance (I) are given by the experimental data and needed no further fitting (Table A2). The start values and constraints for the remaining six variables (ζ, RC, αChl, n, Kno3, PCref, Table A3) were based on model fits of G98 to other diatom cultures in previous studies (Geider 1998, Ross and Geider 2009). The parameters were first fitted manually via graphical comparisons with the experimental

data (POC, PON, Chl, DIN, Fig. 5 and 5), and via minimizing the model cost calculated as the root of the sum of squares normalized by dividing the squares with the variance (RMSE Eq. C2, Stow et al., 2009). The initial manual tuning approach allowed control of the model dynamics, considering potential problems with known limitations of the G98 model (e.g lag phase not modelled; Pahlow, 2005). The manual tuning also allowed obtaining good start parameters for the automated tuning approach and sensitivity/ collinearity analyses, which are sensitive to the start parameters.
After the manual tuning, an automated tuning approach was used to optimize the fits. The automated tuning was done using the FME package (Soetaert et al., 2010b), a package commonly used for fitting dynamic and inverse models based on differential equations (i.e. deSolve) to measured data. The automated analyses were based on minimizing the model cost calculated as the sum of squares of the residuals (SSR, Fitted vs measured data). The experimental data were normalized so that all normalized data were in a similar absolute range of values. This involved increasing Chl and PON values by an order of magnitude while decreasing DIN (NH4 + NO3) data by one order of magnitude The data were not weighted, assuming equal data quality and importance. Prior to the automated fitting, parameters were tested for local sensitivity (SensFun) and collinearity, or parameter identifiability (collin; e.g. Wu et al., 2014). sensFun tests for changes in output variables at each time point based on local perturbations of the model parameter. The sensitivity is calculated as L1 and L2 norms (Soetaert et al. 2009; Soetaert et al., 2010b). The sensFun output is further used as input for the collinearity, or parameter identifiability analyses. Parameters were considered collinear and not identifiable in combination with a collinearity index higher than 20 (Brun et al., 2001). In this case, only the more sensitive parameter was used for further tuning. Eventually, RC, Kno3, n, and αChl were subject to the automated tuning approach using the modfit function, based on minimizing the SSR within the given constraints. Parameters were first fitted using a Pseudorandom search algorithm (Price, 1977) to ensure a global optimum. The resulting parameters were then fine-tuned using the Nelder-Mead algorithm (Soetaert et al., 2010b) for finding a local optimum. A model run with the new parameters was then compared to the initial model via graphical comparisons of the model fit to the experimental data, and via the RMSE value.
The parameter values obtained for the G98 fit to the BAC- experiment were retained without changes or further fitting in the EXT model. The additional parameters of the EXT model were then fitted to the BAC+ experimental data (POC, Chl, PON, DIN). The model was only fitted to total DIN, due to the potential uncertainties related to ammonium immobilization in the biofilm. In fact, a test run, fitting the EXT model to NO3 and NH4 separately lead to a substantially worse overall fit (RMSE=8.79). Otherwise, the data were not weighted. Since the aim of the study was to model the effects of silicate and bacteria on algae growth and not to develop an accurate model for bacteria biomass and silicate concentrations, the parameters μbact, bactmax, Ksi, and Vmax were only fitted to the corresponding data (Bacteria, Silicate) prior to fitting the other parameters of the EXT model. Bacterial growth parameters (μbact, bactmax) were fitted to the bacterial growth curve. Silicate related parameters (Ksi, Vmax) were constrained by the study of Werner (1978) and fitted to the measured silicate concentrations. The remaining parameters were subject to the tuning approach described for G98. Ammonium related parameters (Knh4, nh4thres) were constrained by measured ammonium concentrations, and constants available for other diatom taxa described by Eppley et al. (1969). Remineralization parameters for excreted (rem) and background (remd) DOM were constrained by the data with the limitation of rem > remd, assuming that the excreted DOM is more labile. The parameters related to the effect of silicate limitation on photosynthesis and chlorophyll production (smin, SiPS) were constrained by the study of Werner (1978) and fitted as described for G98. None of the added parameters were collinear/ unidentifiable or given by the measured data and thus retained for the automated tuning approach. Eventually, the 15 parameters (Table A3) were fitted against 160 data points (Table A1).”

On a point related to tuning, I noted that the model has a key parameter for restricting phytoplankton growth (by 80%) in the absence of silicate, but that this parameter is not included in the tuning, which seems something of an omission (and, on a more presentation level, is hard-wired into the equations as a number rather than a parameter).

We agree that this parameter should be included in the tuning process since there may be variations from the study where this parameter is based on depending on the species and environment. It would also be interesting to include it in the tuning exercise to test if the 80% reduction can be confirmed after rigorous model tuning. We changed the model formulation and number of parameters in the tables accordingly and did the model fitting and sensitivity analyses again. The best fit is still an approximately 80% reduction.

Changes:

2.2) “Werner (1978) found that silicate limitation can lead to a 80% reduction in photosynthesis and a stop of chlorophyll synthesis in diatoms within a few hours. Hence, we added a parameter for the reduction of photosynthesis under silicate limitation (SiPS) and formulated a stop of chlorophyll synthesis under silicate limitations.”

3.2) “The most sensitive added parameters in EXT were the remineralisation rate of refractory DON (remd, L1=0.24), the half saturation constant for ammonium (Knh4, L1=0.1) and the inhibition of photosynthesis under Si limitation (SiPS, L1=0.07), which was comparable to other sensitive parameters of the G98 model (Qmax, RC, αChl, ζ, n, I, ΘNmax, Table A1).”
4.2) “we modelled the response of diatom growth to silicate limitation by reducing photosynthesis through a parameterized fraction (SiPS) and a stop of chlorophyll synthesis below a certain threshold, based on experimental studies (Werner, 1978; Gilpin et al., 2004) and in accordance to other ecosystem scale approaches. Automated fitting showed the same 80 % reduction of photosynthesis as described by Werner (1978).”

Table A7, 1b)
$\frac{dC}{dt} = Si_{PS}(P^C - \zeta V_N^C - R^C - xf)C = \mu C$

Finally, the authors identify three central hypotheses in their study:

1. Bacterial regeneration of ammonium will extend a phytoplankton growth;
2. Silicate or nitrogen limitations have different physiological responses;
3. A simple experiment can adequately represent Arctic spring bloom dynamics.

On the first, the model has a very poor performance replicating the time history of ammonium concentrations. On the second, this study would be more convincing if the concentrations of Si and N had been experimentally manipulated to enhance / diminish limitation of each. On the third, the model’s inconsistent fit with observations, and its omission of significant real world factors (e.g. zooplankton) make it difficult to evaluate whether this is true. And because the model is only being run for the short incubation period (i.e. rather than beyond the incubation period, or in some mode investigating more realistic or extrapolated settings), it’s not clear how it behaves when “unleashed”.

Overall, I very much like the combined laboratory and modelling approach, but judge that the modelling component in particular needs to be made much clearer, and evaluated more critically.

We agree that the hypotheses are not perfectly addressed with the data and model, due to the resons mentioned and reformulated the hypotheses in the following way:

We hypothesize that: I) Bacterial regeneration extends a phytoplankton growth period and gross carbon fixation; II) Diatoms continue photosynthesis under silicate limitation at a reduced rate if DIN is available; III) Cultivation experiments are powerful for understanding the major spring bloom dynamics.

1. Bacterial regeneration extends a phytoplankton growth period and gross carbon fixation
2. Diatoms continue photosynthesis under silicate limitation at a reduced rate if DIN is available
3. Cultivation experiments are powerful for understanding the major spring bloom dynamics

Each hypothesis can be tested by the cultivation experiment and can be discussed and evaluated in more detail with the modelling approach.

Concerning hypothesis 1 we suggest that the poor fit to ammonium is mainly related to the measurements rather than the model. Ammonium is most likely immobilized in the biofilm via adsorption to the EPS and accumulation in pockets unavailable to diatoms (See response to Referee #3). These immobile NH4 pools are still part of the measured data. With the model assuming all NH4 being available for algae growth, this is a problem. Hence, we did not put a strong weighting on ammonium for the fitting

routines but fitted the parameter to DIN instead. We did try to fit the model with heavy weighting on ammonium, but could still not reproduce the high ammonium concentrations in the stationary phase, while having a substantially worse fit for the other measured variables (RMSE=8.8).

Response to Referee #3:

“This could, in particular, explain the high values of measured NH4 compared to the model results as shown in Figure 5c. In addition we could add a small discussion of a potential pH dependence of NH4+ adsorption to the EPS in terms of the pKa values of NH4+ and carboxylic groups, which belongs to the acidic polysaccharides as a fraction of EPS:

- Carboxylic groups have a pKa < 5, i.e. far away from seawater pH ~ 8, which means that they are always in the deprotonized negatively charged form R-COO- in seawater.
- NH4+ has a pKa ~9 closer to seawater pH.
- Thus, the NH4+/NH3 ratio will be higher in more acidic microenvironments (pH ~7.5-8).
- Thus, a lower pH due to bacterial respiration would increase the concentration of NH4+ in comparison to the bulk medium, which results in a higher immobile NH4 pool due to adsorption to the EPS.
- This could explain the higher discrepancy between modelled and measured NH4+ values in the experiments with bacteria (as seen in Figure 5c).”

We also included a model run going beyond the experimental time frame in the supplementary material. Overall, the model reaches stable state after approx. 20 days when all nutrients are used up. Bacterial regeneration can keep some levels of N and C assimilation going beyond the loss for excretion and maintenance respiration, but they do not build substantially more biomass, which would be expected in the environment, where, however, sinking and grazing would lead to an additional export leading to a net loss. In order to keep the keep the manuscript streamlined, would prefer not adding these plots to the main manuscript, but to the Supplement instead. In the main manuscript, we suggest adding a short statement of the models stability if run longer (stable state after all nutrients are used up).

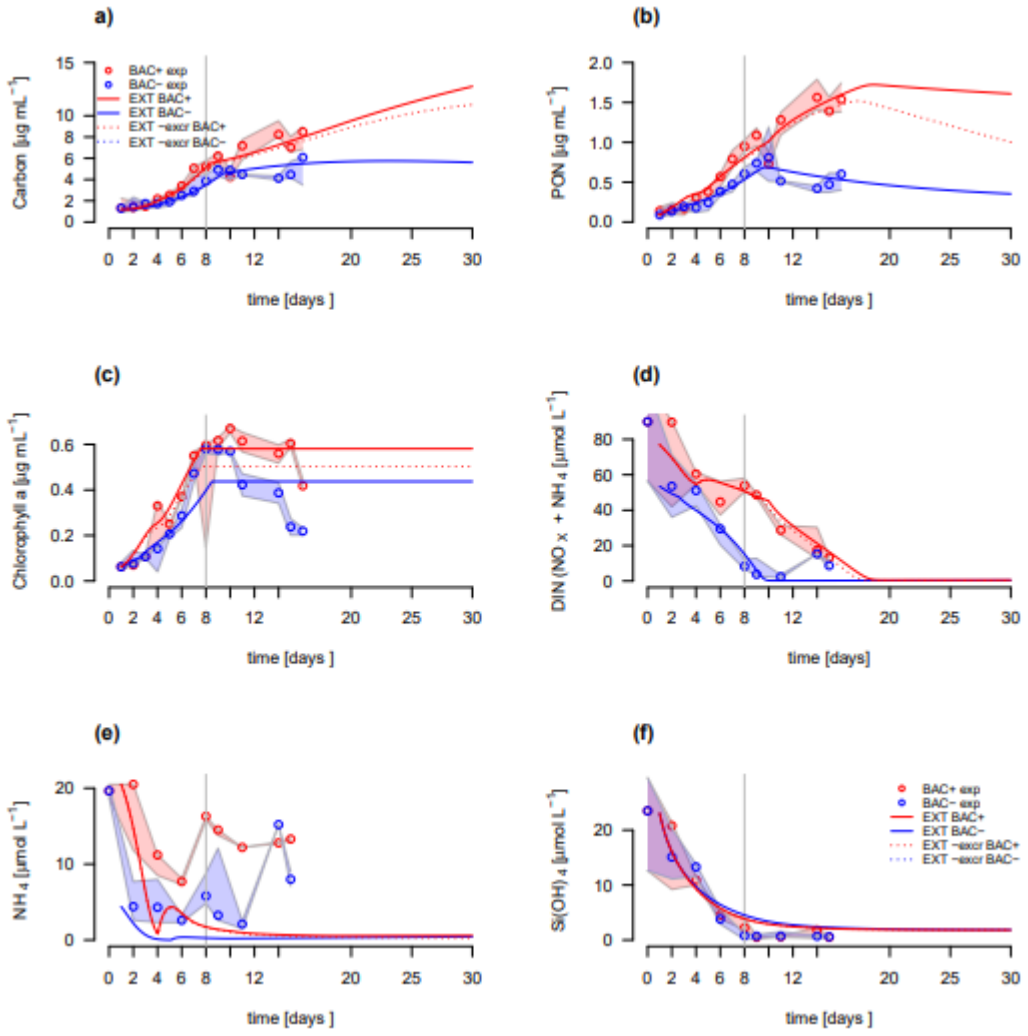

Figure S1: Model fit of the EXT model to the BAC- (blue) and BAC+ (red) experiment. Circles show median values and the colored polygons show the minimum and maximum of the measured data (n=3). Solid lines show the model outputs extended to 30 days of a) POC, b) PON, c) Chl (including outlier at day 8 in BAC+), d) NOX, e) NH_4, and f) Silicate. The dotted line show the output of EXT without excretion.

Specific comments

Pg. 1, ln. 20: "neglect" or "simplify"?; the distinction is important

We changed it to the term “simplify” since a general regeneration component is common in most models. When using the term “neglect” we were mostly focusing on a regeneration component dependent on bacteria biomass which is typically neglected in favor of a general purely substrate dependent remineralization formulation.

Pg. 1, ln. 23: surplus "and"

We removed the surplus “and”

Pg. 1, ln. 25: regarding the importance of organic matter excretion, was this based on observational evidence?

Yes, the excretion is based on the observation of algae aggregation in the stationary phase. However, since this is not the strong part of the model, we removed this statement from the abstract. We also added a more detailed discussion of the biofilm formation and implications for the model as described below and in more detail in the response to reviewer 3.

Pg. 1, ln. 26: “model complexity is comparable to other ecosystem models” – this is misleading as the model here is really an incubator model and not an ecosystem

model; it's missing most of the components that such models include (e.g. detritus, zooplankton)

When comparing model complexity (or number of parameters), we only compare the phytoplankton growth compartment within the ecosystem scale models, which are comparable to our extended model. We omit the complexity in ecosystem scale models not part of our model (e.g. Zooplankton, detritus). We clarified this by following change:

"Overall, model complexity (number of parameters) is comparable to the phytoplankton growth and nutrient biogeochemistry formulations in common ecosystem models used ..."

Pg. 2: maybe be a little clearer on the distinction between autotrophic and heterotrophic bacteria throughout; cyanobacteria, for instance, are unlikely to play the role that's described as "bacterial" here

We distinguished heterotrophic and autotrophic bacteria clearer by adding the term heterotrophic to the bacteria, we are discussing for remineralization. We also agree that the term phototroph for cyanobacteria and heterotroph for the cyanobacteria associated bacteria is especially important to avoid confusion. However, we suggest adding the term heterotrophic only to the first occurrence of bacterial regeneration to keep the text readable.

Pg. 2: also, you should probably say something about the role of zooplankton; they graze phytoplankton and excrete some of the nitrogen they acquire; how quantitatively important is this here?; (I've added a cite to a paper that hints that they might not be all that important)

We added a section about zooplankton with a statement that their importance is low for regenerated production, compared to bacteria regeneration citing additional literature including the suggested paper doi:10.1016/j.dsr.2012.10.003

Change:

"Zooplankton grazing is typically of low importance for terminating blooms (e.g. Saiz et al., 2013), while inorganic nutrients are considered driving bloom termination (Krause et al. 2019, Mills et al. 2018)."…"Zooplankton may also release some ammonium after feeding on phytoplankton, but we suggest that this process is likely far less important than bacterial regeneration (e.g. Saiz et al., 2013). Previously measured ammonium excretion of Arctic mesozooplankton is typically low compared to bacterial remineralization (Conover and Gustavson, 1999), with the exception for one study in summer in a more open ocean setting (Alcaraz et al., 2010). "

Pg. 2, ln. 36: "marine phytoplankton *are*"?

We corrected the term accordingly

Pg. 2, ln. 41: predictions of what?

Corrected to: "predicitons of primary production with climate change"

Pg. 2, ln. 47: you might want to cite something like doi:10.1016/j.dsr.2012.10.003 as evidence of the reduced role of mesozooplankton in controlling / terminating blooms

As described above, we added following statement:

"Zooplankton grazing is typically of low importance for terminating blooms (e.g. Saiz et al., 2013), while inorganic nutrients are considered driving bloom termination (Krause et al. 2019, Mills et al. 2018)."

Pg. 2, ln. 48: remineralisation of what?; a bit of clarity would be helpful; dead diatoms, TEP, faecal material, etc.?

Clarified in the following way: "remineralisation of organic matter"

Pg. 2, ln. 57: heterotrophic bacterioplankton?

Yes, we clarified it by using the term "heterotrophic bacterioplankton" (See also response above).

Pg. 2, ln. 60: regarding “neglected”, do you mean omitted or simplified?; most models include remineralisation of detrital material, and this implicitly bacterial

We are mainly referring to culture based experiments such as Ross and Geider 2009, and Flynn 2001, where remineralisation is omitted. We changed the term to “simplified or omitted”.

Pg. 2, ln. 63: cultivation experiments normally provide parameter values for things like maximum rates of processes, half-saturations, etc., so it’s not clear this is problematic; if model tuning is using cultivation experiments at equilibrium then this might be more of an issue

Problematic is that obtaining axenic algae cultures is challenging, not possible for most species, and does usually not last long (See our response to bacteria growing in the axenic treatment below). In previous cultivation experiments, no efforts for obtaining axenic cultures were mentioned, which hints to bacteria contaminated cultures. In these cultures, regeneration of e.g. ammonium and phosphate takes part. If half saturation constants, maximum uptake rates e.g. are based on non-axenic cultures with the assumption of the absence of regeneration, the values are likely too high, since the experiment will have a nutrient source not accounted for, which leads to underestimations of nutrient uptake, or in the worst case overestimation of growth efficiency if ammonium is not measured at all.

Changes:

“These latter models have been often developed and tuned based on cultivation experiments in which microbial remineralization reactions were assumed to be absent (e.g. Geider et al., 1998; Flynn, 2001) despite the fact that most algae cultures, likely including Geider et al., (1998) and Flynn (2001) are not axenic. Parameters estimated by fitting axenic models on non-axenic experiment may be misleading, mostly by an inflated efficiency of DIN uptake.”

Pg. 3, ln. 72: this process was well-known long before this citation (Flynn, 1997); dig a bit deeper

We chose the citation by Flynn (1997) due to the modelling component in the paper, but agree that we should cite earlier literature. Thus we included the review by Morris (1974) and the review by Dortch (1990).

Dortch, Q.: The interaction between ammonium and nitrate uptake in phytoplankton, Marine ecology progress series, 61(1), 183-201, 1990.

Morris, I.: Nitrogen assimilation and protein synthesis, Algal physiology and biochemistry, 10, 1974.

Pg. 3, ln. 72: “iron has a strong control on silicate uptake” - I’m not sure that this is quite right; Si:C ratios are affected by Fe availability, but this is through continuing Si uptake but reduced C/N uptake and no cell division; my understanding is that Si *uptake* (within a certain range of Si:C) is not immediately affected by Fe; also the recent source given for this statement, Hohn et al., 2009, is a modelling PhD thesis

We agree that the control of Fe on Si is controversial and not well documente din earlier literature and that a modelling PhD thesis is not the best support for this hypothesis. Thus, we changed the statement in the following way:

Change:

“…C and N uptake is reduced under Fe limitation, while Si uptake continues, leading to increasing Si:C/N ratios (Werner, 1977; Firme et al., 2003),…”

Pg. 3, ln. 75: “ultimately too complex” - they add computational expense to large-scale ecosystem models; it’s not clear that they are "too complex" (or even what is meant by this)

We changed it to “ultimately too computationally expensive when implemented in a global biogeochemical model”

Pg. 3, ln. 79: is phosphate limiting in the Southern Ocean?; in parts of its northern extent, yes, but in the south its concentrations are high, no?
Pg. 3, ln. 79: "coastal"?; is there a distinction to be drawn with deep Arctic locations?
We clarified the sentence by: 1) removing the statement about phosphate, which is indeed only limiting in the northern parts of the Southern Ocean, 2) by removing “coastal”, since Fe is not limiting and DIN is the primary limiting nutrient in most Arctic marine systems and by 3) adding supportive citations.

Change:

“In contrast to Antarctica, DIN is the primary limiting nutrient and iron is not limiting in most Arctic systems (Tremblay and Gagnon, 2009; Moore et al., 2013)”

Pg. 3, ln. 81: yes and no; if simple lab experiments exclude factors such as zooplankton excretion which might help fuel phytoplankton growth in parallel with bacterial remineralisation, then it is questionable that they are demonstrating something that’s important in the real ocean
We agree that zooplankton N excretion may have an additional role, but as mentioned above, argue that bacterial N regeneration is quantitatively more important. However, we relativized the sentence in the following way:

Change

“While simple lab experiments cannot represent all nutrient dynamics found in the environment (e.g. N excretion by zooplankton), they can focus on the quantitatively most important dynamics, to facilitate the development of simple, but accurate multinutrient models scalable to larger ecosystem models.

”

Pg. 3, ln. 86: how "associated" is this?; is it something that lives in direct physical contact or shares the same waters?

The bacteria cultures were obtained from the diatom culture directly plated onto LB agar plates. This means they grew together with the diatom outcompeting other bacteria in an environment heavily influence by the algae, which was the only carbon source to the system. Microscopy showed bacteria attached to the diatoms (mostly in the stationary phase), but mostly free-living. However, since the bacteria are heterotroph and there was no other carbon soruce than the DOM coming from the diatom culture, we see them as associated.

Change:

“…inoculation with bacteria cultures, isolated beforehand from the non-axenic culture.”

Pg. 3, ln. 87: again, what specifically is the issue with complexity?; is it model cost, or is there some other aspect of complexity that disfavours inclusion in large-scale models?
Yes, it is model cost. We changed it to: “... aiming to keep the number of parameters, and computational costs low to allow its use in larger ecosystem models.”

Besides the computational costs a large number of parameters, as used in detailed physiological models, is also more difficult to tune or verify with experimental/environmental data, which leads to the issue of overfitting.

Pg. 3, ln. 93: good hypotheses!; however, you do not clearly return to them (e.g. "Regarding the hypotheses framed for this study . . .")
As mentioned above, we changed the hypotheses in the following way:

We hypothesize that: I) Bacterial regeneration extends a phytoplankton growth period and gross carbon fixation; II) Diatoms continue photosynthesis under silicate limitation at a reduced rate if DIN is available; III) Cultivation experiments are powerful for understanding the major spring bloom dynamics.

Pg. 4: this all sounds good, but my expertise in laboratory work is very limited
Thank you

Pg. 4, ln. 111: just for simplicity in the labelling, you might want to come up with nice short names for these experiments; e.g. BACT- (for the axenic) and BACT+ (for the non-axenic), or similar; this will make it easier to refer to the experiments in clear, non-wordy ways later on
We appreciate the suggestion and used the abbreviations BAC- and BAC+ throughout the corrected manuscript and figures.

Pg. 5, ln. 143: the origin of the f-ratio should be cited so that less familiar readers can understand what it is
We added following citation:
Eppley, R. W.: Autotrophic production of particulate matter, Analysis of marine ecosystems/AR Longhurst, 1981.

Pg. 5, ln. 144-146: this is a little confusing; perhaps spell it out with equations?
We added following equation:

"Equation C1. F-ratio estimation in the cultivation experiments with the average PON concentrations at day 13 to 15 ($PON^{d13\text{-}15}$) for the BAC- and BAC+ treatments."

$$f-ratio = \frac{PON_{BAC-\ \ d13-15}}{PON_{BAC-\ \ d13-15} + PON_{BAC+\ \ d13-15}}$$

Pg. 5, section 2.2: I don't think it ever hurts to have a schematic of a model's dynamics to supplement equations and (especially) verbal description
We added a schematic of the model dynamics of G98 and the extended model and briefly described the main dynamics focusing on the controls of photosynthesis, nitrogen assimilation and chlorophyll synthesis by C:N and Ch:N ratios, DIN concentrations, and light.

The schematic figure is shown above (Fig. 1). The following details, were added to the text:

Change:
"The Geider et al. (1998) model (G98) describes the response of phytoplankton to different nitrogen and light conditions and is based on both intracellular quotas and extracellular dissolved inorganic nitrogen (DIN) concentrations, allowing decoupled C and N growth (Fig. 1). Within this model, light is a control of photosynthesis and chlorophyll synthesis. C:N ratios and DIN concentrations control nitrogen assimilation, which is coupled to chlorophyll synthesis and photosynthesis. Chl:N ratios are controlling photosynthesis and chlorophyll synthesis."…"The EXT model keeps all formulations of the G98 and adds dynamics and interactions of silicate, nitrate and ammonium uptake, carbon and nitrogen excretion and bacterial remineralisation (Fig. 1)."…

Pg. 5, section 2.2: similarly, this section would be a lot clearer if you spelled out which models you were using, and ensured that the later plots use the same nomenclature; I initially misread the work ; I reckon it's: 1. G98; 2. Extended; 3. G98 – excretion; 4. Extended – excretion
We clarified the model runs used in the manuscript and used consistent nomenclature: 1. G98, 2. EXT (by default with excretion) 3. $EXT_{\text{-excr}}$. G98 does not have an excretion compartment.

Change:

“We used a dynamic cell quota model by Geider et al. (1998) to describe the BAC- experiment (G98). We then extended the G98 model to represent the role of silicate limitation, bacterial regeneration of ammonium, and different kinetics for ammonium and nitrate uptake (EXT) and fitted it to the BAC+ experiment while retaining the parameter values estimated for G98.”…” For testing the importance of DON excretion we also ran the EXT model without DON excretion (EXT–excr).“

Pg. 5, section 2.2: stating up front an outline about the modelling strategy might help (i.e. two models, tuned to the lab work, DOM addition, etc.)
We added a summary of which models were used for which experiment in the beginning in the now estensive chapter describing the fitting routines.

Change:

“Parameter of the G98 model were fitted to the BAC- experiment data and the EXT model was fitted to the BAC+ experiment data. The G98 parameter values were fitted first and retained without changes for the EXT model fitting.”

Pg. 5, ln. 149: some model equations by the looks of things; the model description appears incomplete
We added the missing equation and double-cheked for any other incomplete model descriptions.

Changes in Table:

7a)	Silicate uptake *(Monod kinetics after Spilling et al., 2010)*	$\frac{dSi_d}{dt} = V_S^C = \left(V_{max} Si_d \frac{Si_d - S_{min}}{K_{si} S_{min}}\right) C$
7b)		$\frac{dSi_p}{dt} = \frac{-dSi_d}{dt} 14$

Pg. 6, ln. 164: equation for dSi seems missing in appendix
We added the missing equation in Table 7

$$\frac{dSi_d}{dt} = V_S^C = \left(V_{max} Si_d \frac{Si_d - S_{min}}{K_{si} S_{min}}\right) C$$

Pg. 6, ln. 166: “80% reduction” - is this where the 0.2 in the equations comes from (i.e. 1 - 0.8 = 0.2)?
Yes, we clarified it in the following way: “Werner (1978) found that silicate limitation can lead to a 80% reduction in photosynthesis and a stop of chlorophyll synthesis in diatoms within a few hours. Hence, we added a parameter for the reduction of photosynthesis under silicate limitation and formulated a stop of chlorophyll synthesis under silicate limitations.”

However, we changed the fixed 80% value to a tunable parameter and rerun the fitting routine and sensitivity analysis as described above.

Pg. 6, ln. 167: some syntheses would suggest that N dynamics *are* coupled to Si dynamics: e.g. Martin-Jezequel, V., M. Hildebrand, and M. A. Brzezinski, Silicon metabolism in diatoms: Implications for growth, J. Phycol., 36, 821 – 840, 2000.

We argue that N dynamics are not directly coupled to Silicate limitation, but indirectly via reduced photosynthesis and inhibited chlorophyll production. The reference by Martin-Jezequel shows no direct coupling of N and Si, but overall different controls for Si and N/P, where Si is tighly linked to the cell cycle, fueled by heterotrophic respiration, while N/P are controlled by photosynthesis. Overall, Martin-Jezequel et al. supports our assumption of decoupled Si and N metabolism and is included in the manuscript as additional support:

Change:

“N and Si metabolism have different controls and intracellular dynamics, with N uptake fuelled by photosynthesis (as PCref in G98) and Si mainly fuelled by heterotrophic respiration (Martin-Jezequel et al., 2000). In general, we assume that nitrogen metabolism is not directly affected by silicate limitation (Hildebrand 2002, Claquin et al., 2002), but we expect cellular ratios to be affected by reduced photosynthesis and chlorophyll synthesis under Si limitation (Hildebrand, 2002; Gilpin, 2004).”

However, we acknowledge that there is 1 study by Gilpin et al., 2004, discussing a coupling of N:Si. Hence, we added it in the discussion:

Change:

“Studies on the coupling of silicate limitation on C, N, and Chl show inconclusive patterns, including a complete decoupling (Claquin et al., 2002), a relation of N to Si (Gilpin et al., 2004) and reduction of photosynthesis (Werner, 1978; Gilpin et al., 2004) while no new chlorophyll is produced (Werner, 1978; Gilpin et al., 2004).”…” Our cultivation study shows”…” ii) that coupling of Si:N:C:Chl is present. We do not expect a direct Si:N coupling, due to different controls of Si and N metabolism (Martin-Jézéquel et al., 2000.), but suggest indirect coupling via reduced photosynthesis.”

Pg. 6, ln. 171: make it clear here that your model has labile and refractory DON

We agree that this needs to be clarified.

Change:

“It was assumed that this process is faster for freshly excreted DON compared to DON already present in the medium. Thus, we implemented a labile (DONl) and refractory (DONr) DON pool with different remineralization rates (rem, remd).“

Pg. 6, ln. 173: it seems unlikely that the bacteria would simply "give up" on remineralisation if the C:N ratio is too high; perhaps expand on why Tezuka suggests this is happening

We do not suggest a complete stop of remineralisation, but a net release of nitrogen of 0, since bacteria need more DIN on their own, rather than having the luxury of releasing it to the environment. We mention now two papers as support. Both papers base their fining on net changes in DIN. We tried to clarify it by following change:

Change:

“After Tezuka (1989), net bacterial regeneration of ammonium occurs at DOM C/N mass ratio below 10 and is proportional to bacterial abundances. Higher thresholds up to 29 have been found (e.g. Kirchmann, 2000), but we selected a lower number to stay conservative.”

Pg. 6, ln. 175: this is unclear; when you say "substrate" what do you mean?; typically substrate is used to indicate a resource consumed by an organism; here you’re talking about phytoplankton, so DIN and DIC would appear to be meant - but DIC will likely be much higher than 10x DIN

We refer to DOM as substrate for bacteria and clarified it: “DOM C/N ratios…. “

Pg. 6, ln. 177: does this mean that bacteria won't remineralise material with a C:N > 10?; that seems a little unlikely

As for line 173 we changed "bacterial remineralization" to "net bacterial ammonium regeneration".

Pg. 6, ln. 178: as the paper makes a fuss earlier about other models glossing over bacterial remineralisation, this simplified form is surprising

The main improvement of the model is to include a remineralisation rate controlled by: i) bacteria biomass, ii) substrate (DOM) C:N ratios, and iii) substrate origin (autochthonous, allochthonous). Other models typically have a fixed remineralisation rate either only dependent on the DOM/POM, or not controlled by any environmental variable. Thus, we still see our extension as a considerable improvement and consider a simple logistic growth estimate sufficient.

We could of course model bacteria growth via Michaelis-Menten kinetics based on 2 DOM pools, but this would not have any effect on the parameterization or modelling of algae physiology, which is the main goal of the paper, while increasing the number of parameters and computational costs, which we tried to keep low. Since, the aim of the model is not to model bacteria growth, but algae growth and intracellular C:N:Chl ratios we do not see that a more accurate and more complex model of bacteria growth would improve the manuscript.

Pg. 6, ln. 185: Table A6 - it looks to me like some equations are missing

We added the missing equations as mentioned above.

Pg. 6, ln. 186: which order of RK?; e.g. 3 or 4 (or higher)

We used the 2nd-3rd order Runge-Kutta method with automated stepsize control and added this information in the manuscript.

"The differential equations were solved using the ode function of the deSolve package (Soetaert et al., 2010) with the 2nd-3rd order Runge-Kutta method with automated stepsize control."

Pg. 6, ln. 185-191: this description of model tuning is far too brief; I'm not sure what's going on here; readers unfamiliar with R (I am one) will not understand what these different packages are doing or what their underlying assumptions are; this aspect of the modelling is too important to be glossed over so quickly; in general, to avoid the appearance of having just used the first package that occurred to you, expanding on the detail of the tuning (tools, approach, goal) would greatly benefit this description (hence and on Pg. 7)

We added a more detailed and extensive description of what the R packages are doing. deSolve is the most widely used solver for differential equations in R, and FME is a package for model fitting and sensitivity analysis developed as add on to deSolve. The tuning approaches via

1st manual fitting (based on RMSE error and graphical comparisons),

2nd automated fitting of selected parameters (avoiding collinearity/ linear dependence of sensitivity of 2 parameters = unidentifiable parameters), and choosing the more sensitive parameter in case of conflicts) via the Pseudorandom algorithm (searching for a global optimum),

3rd fine tuning for a local optimum using the Nelder Mead algorithm. 4th check if the new parameters give a better fit regarding graphical comparisons and RMSE.

See the added chapter above in this response.

Pg. 6, ln. 192: the text should be clear on which observed variables were used to fit the model, why these were favoured, and whether any weighting was made to account for those judged better observed or more important; I would naively expect nutrient concentrations to be of prime importance but it's unclear what criteria the authors are using here (see my later remark about ammonium)

We also added more details about the observed variables used for tuning. We used POC, PON, Chl, and DIN (NOx + NH4) with standardized values (POC, 10xPON, 10xChl, DIN/10) and no further weighting. Due to rather poor quality of the NH4 data, we did not fit the model to NOx and NH4 separately.

Details about the parameters tuned and the constraints are also given in Table A3. Parameters were partly given by measured data, or tuned after constraining with measured or published constraints. In case of strong collinearity, only the most sensitive of the collinear parameters was tuned.

See the added chapter above in this response.

Pg. 7, ln. 197: this seems rather unsatisfactory; I would expect parallel runs with the same parameter values to be performed for axenic and not-axenic simulations, with an automated process (e.g. a genetic algorithm) to evaluate cost (i.e. misfit) before somehow generating new parameter values and iterating; having a manual component seems odd

We did do parallel runs with the same parameter values. The G98 model was fitted to BACT- data, but the resulting parameters were used for a G98 model run of both BACT- and BACT+ and kept without changes or further fitting as part of the EXT model. For the EXT model only the extended parameters relevant for describing our key observed variables (POC, PON, Chl, DIN) were fitted with the same rigorous fitting approach used for G98. The resulting parameters were used for the model of both BACT+ and BACT-.
We corrected the text in the manuscript to clarify what we did

We argue for an initial manual tuning approach in order to account have control of the model dynamics and to obtain good start parameters for the automated tuning approach and sensitivity/ collinearity analyses.

See the added chapter above in this response.

Pg. 7, ln. 198: what are these "known limitations"?; also, it's noticeable in the plots that the model solutions inflect strongly around the lag/stationary phase time point - is the model somehow different either side of this division?

The known limitations is that parameter tuning of the G98 in earlier attempts did not allow modelling the lag phase (Pahlow, 2005); later, howerver, Smith and Yamanaka (2007) showed that the Geider model can be brought to reproduce an initial lag phase.

The strong change around the beginning of the stationary phase is based on the threshold based approach to responses of Photosynthesis and Chl synthesis after Silicate limitation. Once silicate falls below a threshold, the physiology changes, which can be seen as a sudden change. Threshold based approaches are common in other dynamic models (e.g. threshold for cell division in Ross & Geider, 2009, threshold deciding which limiting nutrient decides the growth in Vichi et al., 2007).

Pahlow, M. (2005). Linking chlorophyll-nutrient dynamics to the Redfield N: C ratio with a model of optimal phytoplankton growth. Marine Ecology Progress Series, 287, 33-43.

Smith, S. L., & Yamanaka, Y. (2007). Quantitative comparison of photoacclimation models for marine phytoplankton. Ecological Modelling, 201(3–4), 547–552. https://doi.org/10.1016/j.ecolmodel.2006.09.016

See also the added chapter above in this response.

Pg. 7, ln. 202: "Colinearlity" - do you mean that you're looking for linkages between parameters here?

Collinearity is a measure for the parameter identifiability in complex simulation models (Brun et al., 2001) and allow identifying which parameter(s) (sets) can be uniquely constrained from the data.
If the perturbation of two different parameters can lead to the same change in the output variables, they are collinear, which makes them unidentifiable. Parameters were considered collinear and not identifiable in combination with a collinearity index higher than 20 as described in (Brun et al., 2001). In this case, only the more sensitive parameter was fitted.

Brun, R., Reichert, P. and Kunsch, H. R., 2001. Practical Identifiability Analysis of Large Environmental Simulation Models. Water Resour. Res. 37(4): 1015–1030.

See the added chapter above in this response.

Pg. 7, ln. 205: ammonium was "constrained loosely" - perhaps given later results this was a mistake?

Due to potential uncertainties associated with the ammonium data (e.g. immobilization in the biofilm by adsorption and micro-pockets, leakage of intracellular NH4 during filtration, freeze-thaw cycle), and high variability in published parameters (e.g. Eppley et al., 1969), we used wider constraints for ammonium related parameters. We do not agree that narrower constraints would lead to a better model fit to ammonium, since the new values would be within the same parameter space/ constraints. However, for consistency and usability of the model in other settings we narrowed down the constraints of published half saturation constants by Eppley et al., 1969. The reason for the poor fit is partly the lower weighting of the ammonium output during model fitting (We only fitted to DIN (NH4+NO3) and not separately to NH4 and NO3, but also the uncertainty of the ammonium values which likely include immobilized ammonium from algae cells, and the biofilm. We did a test run where we fitted the EXT model to POC, PON, Chl, NH4 and NO3, but the overall fit was substantially worse (RMSE = 9, instead of 2 with the DIN fit) with parameter values reaching the limits of the constraints.

Eppley, R. W., Rogers, J. N., & McCarthy, J. J. (1969). HALF-SATURATION CONSTANTS FOR UPTAKE OF NITRATE AND AMMONIUM BY MARINE PHYTOPLANKTON 1. Limnology and oceanography, 14(6), 912-920.

See the added chapter above in this response.

Pg. 7: ecosystem models have notoriously non-linear misfits in their parameter space; when this is highly multidimensional (as here) it can be difficult for optimisation to find the global minimum misfit; how has this been achieved here?

We agree that his is a potential problem and therefore we approached the problem from different angles.
First, we started with extensive manual tuning, as this gives a lot of insight for the modeler on how an optimal fit can be achieved and which parameters influence the results the most.
Secondly, we applied an automated parameter fitting procedure, which started with a collinearity analysis to make sure we are working with a parameter set that can actually be identified from the data. This reduces the risk of getting stuck in a local minimum. Subsequently, we ran a pseudorandom optimization routine to ensure a better coverage of the (identifiable) parameter space to increase the chance of approaching the global minimum randomly. The automated optimization routine ended with a directed descent algorithm, i.e. the Nelder Mead algorithm, that ensures quick convergence to the minimum.

See the added chapter above in this response.

Pg. 7, ln. 220: “stationary phase” - how exactly defined here?; particularly in the context of Figure B3c, which shows chlorophyll concentrations peaking _2 days later in the bacterial incubations

We defined the stationary phase by the sudden increase in phosphate and ammonium, silicate and DIN (for axenic cultures) values falling below minimum values in the model, and the Quantum yield dropping below 0.63. Since the explanation of all of these evidence is spread over the page, we changed the term “stationary phase” to “day 8”, which is less objective.

Pg. 8, ln. 234: can you explain if these values are meaningful, or is it just the relative values between phases that’s important?

The Quantum yield is a ratio based on variable fluorescence of chlorophyll. The number ranges between 0 and 1 and show how efficiently energy is transported after adsorption. Generally, high numbers indicate fit and active cells, while low numbers indicate stressed algae cells. Low N:C ratios are one stressor described to lead to inefficient energy transfer (low QY, Cleveland and Perry, 1987).

Cleveland, J. S., & Perry, M. J.: Quantum yield, relative specific absorption and fluorescence in nitrogen-limited Chaetoceros gracilis. Marine Biology, 94(4), 489-497, 1987.

Change:

"The maximum photosynthetic quantum yield (Fv/Fm) is commonly used as a proxy of photosynthetic fitness (high QY), indicating the efficiency of energy transfer after adsorption in photosystem II. Low values are typically related to stress, including for example nitrogen limitation (Cleveland and Perry, 1987). We found an increase in QY from approx. 0.62 to 0.67 d-1 in the exponential phase and a decrease to approx. 0.62 in the BAC+ treatment after 8 days and to approx. 0.58 in the BAC- treatment (Table A8)."

Pg. 8, ln. 279: This seems a pretty serious deficiency given the focus of this paper; I would interpret this as potentially a problem at the tuning stage; did you consider weighting fitting ammonium more heavily?

We suggest that the poor fit to ammonium is mainly related to the measurements rather than the model (immobilized ammonium in the biofilm). Hence, we did not put a strong weighing on ammonium for the fitting routines but fitted the parameter to DIN (NO3 + NH4) instead. When we did try to fit the model with heavy weighting on ammonium, we could still not reproduce the high ammonium concentrations in the stationary phase, while having a substantially worse fit for the other measured variables (RMSE=8.8). We discuss this limitation in the manuscript as follows:

Changes:

"The model was only fitted to total DIN, due to the potential uncertainties related to immobilized ammonium in the biofilm. In fact, a test run, fitting the EXT model to NO3 and NH4 separately lead to a substantially worse overall fit (RMSE=8.79)."

"While not all ammonium measured is also available for algae growth, discussion of the dynamics (decrease in the start, increase with the onset of the stationary phase), especially if also shown in the EXT model, are still useful to understand multinutrient dynamics (e.g. regeneration). Considering the overall higher concentrations of NO3, compared to NH4, discussions of total DIN dynamics, DIN.DIP ratios, and limitations are also meaningful."

"Fine scale DIN dynamics caused by ammonium – nitrate interactions were represented well (Fig. 6a). However, at the onset of the stationary phase, ammonium concentrations of the model were one order of magnitude lower than in the experiment, showing a major weakness (Fig. 6c). Increased weighting of ammonium during the model fitting led to a slightly better fit to ammonium, but a substantially worse fit of the model to POC, PON, and Chl ($RMSE_{EXT}$=8.79), indicating that the problem lies with the ammonium data (immobilized ammonium)."

Pg. 9, ln. 283: "complexity" is an unusual way to describe a lack of sensitivity (which is what you seem to be suggesting); also, given the extended model performs no better (worse?) than the G98 model is this not to be expected?; i.e. you've added a means for the model to be different, but this means is far less powerful than what the model already has

We agree that complexity is not the best fitting term and changed it to "sensitivity" or "added parameters".

Changed to: "…was more sensitive than any of the original model parameters. Hence, the added parameters of the extended…"

Change:

"The sensitivity analysis (Fig. B1, Table A1) revealed that the sensitivity of the added parameters in EXT is overall comparable to the sensitivity of the original parameters in G98. The model outputs were most sensitive to P_C^{Ref} (L1=0.8, L2=1.5), which is a parameter in both G98 and EXT. The most sensitive added parameters in EXT were the remineralisation rate of refractory DON (rem_d, L1=0.24), the half saturation constant for ammonium (K_{nh4}, L1=0.08) and the inhibition of photosynthesis under Si limitation (Si_{PS}, L1=0.08), which was comparable to other sensitive parameters of the G98 model (Q_{max}, R_C, α_{Chl}, ζ, n, I, Θ_N^{max}, Table A1). Small perturbations of the parameters only indirectly related to the fitted output variables did not lead to changes in POC, PON, Chl, or DIN."

Pg. 9, section 4: this discussion seems far too long for what's quite a simple set of Experiments

We agree that we originally thought this was simple set of experiments. However, the additional model interpretation, though very valuable we believe, does warrant a lengthier discussion. Also the detailed and thorough reviews for this manuscript made it impossible for us to substantially shorten the discussion and we believe that shortening the discussion would not be possible while addressing all comments of the three reviewers. We still tried to keep it as short as possible with the suggested changes.

Pg. 10, section 4.1: there's nothing in here about the (hard-wired!) 80% adjustment to growth rates caused by low silicate; this appear to be an unchangeable assumption

Firstly, we changed the 80% formulation into a tuneable parameter. Secondly, we added a sentence in the discussion. "Photosynthesis was reduced by 80% after silicate became limiting, which is in accordance with earlier experimental studies (Tezuka..)."

Pg. 10, ln. 310: do values of the f-ratio from bottle experiments relate well to those measured from the open ocean?; I can't think of any reason to suspect that they will, not least because there are no nitrifying bacteria including in the cultures here

We do not expect that the f-ratio in our bottle experiment is representative for the open ocean, but compare the values as starting point for discussing why. We argue that a discussion of the differences between the bottle experiment and open ocean values (e.g. grazing, nitrification) can show the limitations of the experiment and thereby the limitations of our model. We also add a reference to nitrification to the lacking processes.

Change:

"While we do not expect the f-ratio in our bottle experiment to be directly comparable to open ocean system, which does include a variety of algal taxa beyond C. socialis, a comparison can aid to identify limitations in our experiment and model. Regenerated production is significant in polar systems and our estimated experimental f-value of 0.31 is slightly below the average for polar systems (Harrison and Cota, 1990, mean f-ratio=0.54). Nitrification is a process supplying about 50% of the NO3 used for primary production in the oceans, which may lead to a substantial underestimation of regenerated production (Yool et al., 2007), inflating the f-ratio interpreted as estimate for new production, potentially also in the study by Harrison and Cota (1990). The absence of vertical PON export in our experiment may be another explanation…"

Pg. 11, ln. 357: ah-ha, computational cost is finally mentioned

We also added this information in previous formulations of "cost" and "complexity" in the corrected version of the MS.

Pg. 11, ln. 351: I don't think it's ever made clear why there may be a preference for NH4 over NO3; it would be good to include mention of this so that readers understand why this aspect may be important in the work here

The conversion of NH4 to biomass NH3 is energetically much cheaper, making it the preferred source. We added following information:

"Due to the metabolic costs related to nitrate reduction to ammonium, ammonium uptake is preferred over nitrate, potentially leaving more energy for other processes (Lachmann et al., 2019). Ammonium can even inhibit or reduce nitrate uptake over certain concentrations (Morris, 1974). The dynamics are mostly controlled by intracellular processes, such as glutamate feedbacks on nitrogen assimilation, cost for nitrate conversion to ammonium, or lower half saturation constants of ammonium transporters (Flynn et al., 1997)."

Lachmann, S. C., Mettler-Altmann, T., Wacker, A., & Spijkerman, E.: Nitrate or ammonium: Influences of nitrogen source on the physiology of a green alga, Ecology and evolution, 9(3), 1070-1082, 2019.

Pg. 12, ln. 360: the authors note different conceptual models for the Si:N relationship in this section, but stick instead with a highly simplified approach from a review almost 40 years old; and also remove this relationship from the tuning exercise undertaken; I would expect to see more justification for this - or potentially some form of model sensitivity analysis to evaluate how important it is

We argue that N dynamics are not directly coupled to Silicate limitation, but indirectly via reduced photosynthesis and inhibited chlorophyll production. The reference by Martin-Jezequel shows no direct coupling of N and Si, but overall different controls for Si and N/P, where Si is tighly linked to the cell cycle, fueled by heterotrophic respiration, while N/P are controlled by photosynthesis. Overall, Martin-Jezequel et al. supports our assumption of decoupled Si and N metabolism and is included in the manuscript as additional support.
However, we acknowledge that there is 1 study by Gilpin et al., 2004, discussing a coupling of N:Si. Hence, we added it in the discussion:

Change:

“Studies on the coupling of silicate limitation on C, N, and Chl show inconclusive patterns, including a complete decoupling (Claquin et al., 2002), a relation of N to Si (Gilpin et al., 2004) and reduction of photosynthesis (Werner, 1978; Gilpin et al., 2004) while no new chlorophyll is produced (Werner, 1978; Gilpin et al., 2004).”…” Our cultivation study shows”…” ii) that coupling of Si:N:C:Chl is present. We do not expect a direct Si:N coupling, due to different controls of Si and N metabolism (Martin-Jézéquel et al., 2000.), but suggest indirect coupling via reduced photosynthesis.”

We agree that the 80% reduction should not be a fixed parameter, but tuneable. We adjusted the model accordingly. We also included the parameter to the sensitivity analysis and repeated the fitting routine.

Pg. 12, ln. 375: is a biofilm something one might expect in the natural system?; it doesn’t seem to be the sort of thing that would form in free water; also, it’s unclear from the methods whether there’s any agitation of the cultures to mimic ocean mixing

We would not expect biofilm formation in open oceans, but aggregation, which is commonly found in the end of spring blooms increasing the vertical export (e.g. Thornton, 2002). Both processes are similar in the way that algae aggregate via EPS facilitating a specific and active microbiome. We added a sentence about biofilm as proxy for marine snow in the discussion. Ocean mixing was mimicked by inverting all bottles 2-3 times a day (added to the methods).

Change in the discussion:

“While we would not expect biofilms in the open ocean, aggregation of algae cells, facilitated by EPS is common towards the end of spring blooms, increasing vertical export fluxes (e.g. Thornton, 2002). Chaetoceros socialis is in fact a colony forming diatom building EPS-rich aggregates in nature (Booth et al., 2002).”

Change in the methods:

“The cultures were incubated at 4°C and 100 µE m-2 s-1 continuous light and mixed 2-3 times a day to keep the algae and bacteria in suspension.”

Pg. 13, ln. 392: the value of the f-ratio has been questioned as the wider role of nitrifying bacteria has been recognised; perhaps rephrase talking instead about the balancing roles of export and remineralisation?

We replaced the term f-ratio by “regenerated production” and added that the higher regenerated production is due to increased remineralization compared to export.

Pg. 13, ln. 406: consider: Kamatani, A., Dissolution rates of silica from diatoms decomposing at various temperatures, Mar. Biol., 68, 91– 96, 1982

We included the reference

Pg. 14, ln. 426: model availability?; might be good to include the code too - it's simple Enough

The R code is now available at github.

Pg. 21, Figure 1: presumably the gap between (NOx + NH4) in the two experiments is due to N getting stuck in (dead) organic matter?; bar PON / POC, was anything about this recorded in the experiments?

We did not differentiate between life and dead organic matter, but assume mostly life organic matter until the stationary phase where biofilm formation played a role indicating EPS production, which can contribute to the measured PON and POC.

We agree that NH4 adsorption to organic matter (EPS) can play an important role and is likely one of the main explanations for the poor (lower) model fit of ammonium to the measured data. In addition, NH4 may be immobilized in micro-pockets of the biofilm unavailable for algae uptake.

However, we attribute the gap of DIN between the experiments mainly to a) increased NH4 regeneration in BACT+, with some ammonium likely immobilized in the biofilm (= higher NH4 concentrations), and b) preferred NH4 uptake over NO3 and NO3 uptake inhibition by NH4 leading to higher NO3 concentrations in the BACT+ treatment due to slower uptake. The PON/POC ratios change due to carbon overconsumption (Schartau et al. 2007), which is most relevant under N limitation, whule Si limitation has a more direct effect on photosynthesis (Lippemeier et al., 1999, Thangaraj et al., 2019; Liu et al., 2020). All these 3 dynamics are part of the extended model taking bacterial processes and NH4-NO3 interactions into account.

Pg. 21, Figure 1: the span of PO4 at day 14 (5-55) seems implausible given its narrow span at day 11 (30-35); especially as it narrows again at day 15 (5-18)

We argue that the large range is plausible since it is i) based on 1 data point, which may be an outlier and ii) it corresponds with high variation in bacteria abundances, which are ultimately responsible for the high PO4 value presumably originating from remineralization. Especially towards the end of the experiment it is not implausible that the different bottles behave differently.

Change in figure legend:

"c) PO42- with a potential outlier at day 14 leading to a negative peak"

Pg. 22, Figure 2: not so axenic, eh?; is this contamination in the axenic incubations from repeatedly opening the vessels?

As mentioned in the results, the bacteria growing towards the end are still in so low abundances compared to the bacteria enriched experiment, that it is effectively axenic. Obtaining and especially maintaining axenic diatom cultures is challenging and does typically not last very long. Since we used independent bottles during the experiment, contaminations during the course of the experiment are not possible (bottles were not opened before the sampling day). However, antibiotic treatments attack mostly active bacteria cells susceptible to the antiobiotics, while endospores and antibiotic resistant bacteria can survive. We believe that the bacteria starting to grow at day 14 originate form endospores activated by the high concentrations of DOM excreted by the stressed algae.

Change in the methods:

"We ensured sterile conditions during the experiment by keeping the cultivation bottles closed until sampling. However, endospores may survive the antibiotic treatment in low numbers and start growing especially towards the end of the experiment."

Pg. 23, Figure 3: so as well as having less NOx and NH4, the axenic experiments have less PON; where is the N going?

We suggest that the N is contributing to a higher DON pool (not measured) in the axenic experiments, which is not shown in Figure 1 and 3. The DON could be remineralized in the experiments with bacteria yielding higher NOx, NH4+ and PON.

We hope that our schematic representation of the model added to the methods helps to clarify it (See Fig. 1 above).

Pg. 24, Figure 4: it's idle curiosity, but what happens if you extend your model runs past the time point that the laboratory cultures ran?; the model should permit this
we added the output of a prolongated model run in the supplement (See above).

Pg. 24, Figure 4: the inflections on some of the model plots here look rather artificial; can you explain why there are such sharp transitions around the 8-day mark?
The sharp transition is due to the threshhold based formulation of reduced photosynthesis and inhibited Chl synthesis under Si limitation. At day 8, the silicate limitation threshold is reached and Photosynthesis is reduced and Chl synthesis inhibited. As describe dabove, threshold based modelling approaches are not uncommon.

Pg. 24, Figure 4: the spikes in chlorophyll in the cultures seem difficult to believe; do you think they are perhaps artifacts / measurement error?
Yes, as mentioned above these spikes represent a single data point that can be measurement artifacts and added this information to the figure legend.

Pg. 25, Figure 5: given that the key is the same in all of the plots, it would be better to not use it in plots where it interferes with the data (e.g. 5c)
We now only put the legend in plot a and mention that the legend is valid for all subplots.

Pg. 25, Figure 5: why are the fits without the excretion term all flat?; that's not what I'd expect at all; actually, I now realise that you're using two sets of dotted lines on this plot; one for the model output, one for the limiting concentration of the nutrients; this should be changed as it's a very confusing presentational choice
We changed the style of the lines.The fit without excretion is not the flat line, but the dotted line close to the +excr model. The –excr model simply modelled the excretion fraction of the +excr model into the maintenance respiration term (general loss without being available for remineralization). Since our system was highly affected by ambient DOM (likely terrestrial), the difference is little, showing that the regenerated production in our experiment is mostly caused by terrestrial DOM regeneration rather than freshly produced DOM regeneration.

Pg. 26, Table 1: the text reads as if these crosses denote both (a) remineralisation, and (b) variable stoichiometry?; that seems a lot for one cross to bear!; however, in the table, it looks like you separate out the stoichiometry - I think this sentence needs rewording
Pg. 26, Table 1: as a stylistic aside, a cross is not necessarily the best way to denote that a model includes something; conventionally, ticks are used, with ticks and crosses meaning opposite things
Pg. 26, Table 1: where other models are presented, these are often older versions of these models; might it be better to report their current versions?
We clarified the table caption and used ticks and crosses instead.
While the full ecosystem scale models may have more recent versions with updated formulations, we give the original reference to the biogeochemical compartment of the ecosystem scale models, which are still quite old. We will however, added references to the most recent full-scale models used in addition to the reference only describing the algae growth formulations. We added following references to more recent ecosystem scale model formulations:

BFM model:_ Smith, K. M., Kern, S., Hamlington, P. E., Zavatarelli, M., Pinardi, N., Klee, E. F., & Niemeyer, K. E. (2020). BFM17 v1. 0: Reduced-Order Biogeochemical Flux Model for Upper Ocean Biophysical Simulations. *Geoscientific Model Development Discussions*, 1-35.

ReCom-2 model: Schourup-Kristensen, V., Wekerle, C., Wolf-Gladrow, D., Völker, C. (2018): Arctic Ocean biogeochemistry in the high resolution FESOM 1.4-REcoM2 model, Progress in Oceanography, 168, 65-81,doi:10.1016/j.pocean.2018.09.006.

MEDUSA model: Henson, S. A., Cole, H. S., Hopkins, J., Martin, A. P., & Yool, A. (2018). Detection of climate change-driven trends in phytoplankton phenology. *Global Change Biology*, *24*(1), e101-e111.

NEMURO model: Anju, M., Sreeush, M. G., Valsala, V., Smitha, B. R., Hamza, F., Bharathi, G., & Naidu, C. V. (2020). Understanding the Role of Nutrient Limitation on Plankton Biomass Over Arabian

Sea Via 1-D Coupled Biogeochemical Model and Bio-Argo Observations. *Journal of Geophysical Research: Oceans*, *125*(6), e2019JC015502.

SINMOD model: Alver, M. O., Broch, O. J., Melle, W., Bagøien, E., & Slagstad, D. (2016). Validation of an Eulerian population model for the marine copepod Calanus finmarchicus in the Norwegian Sea. *Journal of Marine Systems*, *160*, 81-93.

NPZD model: Gruber, N., Frenzel, H., Doney, S. C., Marchesiello, P., McWilliams, J. C., Moisan, J. R., Oram, J. J., Plattner, G., and Stolzenbach, K. D.: Eddy-resolving simulation of plankton ecosystem dynamics in the California Current System, Deep Sea Research Part I: Oceanographic Research Papers, 53(9), 1483-1516, 2006.

And we added and discussed following culture-scale model suggested by reviewer 4:

Flynn, K. J., Skibinski, D. O., & Lindemann, C. (2018). Effects of growth rate, cell size, motion, and elemental stoichiometry on nutrient transport kinetics. PLoS computational biology, 14(4), e1006118.

Pg. 27, Table A1: this is confusing; why not have separate columns for G98 and the extended model?; also, this table implies that some properties are not in the model, but you seem to have equations for them; meanwhile, there are other properties, e.g. dSi, for which no equation is presented

We had only a G98 column because all state variables are part of the extended model (The EXT model is the G98 model with added variables). We added a column for EXT besides G98 with ticks for every state variable for clarification. We also mentioned the equation for each state variable in table A6/A7 and added the missing equations.

Pg. 27, Table A1: you appear to be using underscores rather than minus signs in units at the base of this table

We changed it.

Pg. 30, Table A4: what do all of the columns mean here?; some explanation would be Useful

We shortened the table slightly and explained all columns in detail in the corrected version.

Change:

Table A4. Output of the sensitivity analysis (senFun of the FME package in R) with the value for each parameter and different sensitivity indices obtained after quantifying the effects of small perturbations of the parameters.on the output variables (POC, PON, Chl, DIN). The L1 and L2 norms are normalized sensitivity indices defined as $L1 = \sum \frac{|S_{i,j}|}{n}$ and $L2 = \sqrt{\frac{S_{i,j}^2}{n}}$ with $S_{i,j}$ being the the sensitivity of parameter i for model output j.

par	value	L1	L2	Mean	Min	Max
			G98			
ζ	1.00	0.10	0.19	-0.02	-0.15	0.98
R^C	0.07	0.04	0.05	-0.03	-0.08	0.14

Pg. 31, Table A5: please choose a table size that doesn't line-break your units

We adjusted the table size

Pg. 32, Table A6: maybe pull the ODEs together in one place then follow-up with the separate terms afterwards?; it's a little difficult to parse the equations otherwise

we changed the order accordingly.

Pg. 33, Equation 1: if there's a conditionality on a single term in an equation (as here) better to have a single ODE and put the conditionality inside this term (i.e. it's this value if X, zero if Y); this is easier to follow and makes it much easier to see where the important parts of the model's behaviour lie; duplicating the equations for the sake of a single term in them does not make things clear

We changed the equations accordingly.

Pg. 33, Equation 1: you should note somewhere that organic C is removed from an unmodelled reservoir of DIC; unmodelled because it's always in excess of the ecosystem model's requirements

We added the information to the schematic figure in the methods and mentioned it next to the equation.

Pg. 35, Equation 15: the presentation of equations 14 and 15 around the 14e3 divisor is different; this is an unnecessary confounding factor that makes the equations less readable

We changed the form of eq 14 to the same format as in eq 15.

Pg. 35, Equation 16: why is this a hard-wired number (0.2) and not a parameter?; even if it's not something you change in your study (which seems a little strange given what you do change), having this as a clearly parameter rather than an undescribed constant is important

We changed this parameter to a tuneable parameter and included it into the sensitivity analyses and parameter fitting exercise.

Pg. 36, Figure B1: I don't understand what this plot is showing; please explain what it means for a line to deviate from zero here; also, why is sensitivity time-variable in any case?; and why is it not monotonically variable in time?; I also note that it looks like DIN is super-sensitive compared to the other properties - is that a correct interpretation of this plot?

The sensitivity analyses in the FME package tests the sensitivity of the model output (here DIN, POC, Chl, DIN) with changing parameter values within the predefined constraints. The plot shows the deviation from the model output towards the measured data over time. We realized that this figure is too complex while adding little information to the manuscript and removed it.

Pg. 37, Figure B2: a full explanation for what this plot is showing is critical; it is very difficult to understand what's being shown; also does the frequent occurrence of "NA" imply that some parameters should be excluded from this analysis?

The plot shows pairwise comparisons of parameter sensitivity/ sensitivity functions. On the upper right the pairwise data are shown for each tuneable parameter with the boundaries/constraints given in table A3. The sensitivity is given for POC (blue), PON (red) and Chl (green). The correlation coefficients are given in the lower left corner. NAs indicate no correlation because of low sensitivity. We realized that this figure is too complex while adding little information to the manuscript and replaced it with following table.

Change:

Table A8. Output of the collinearity or parameter identifiability analysis using the collin function of the FME R package (Soetaert et al., 2010b). A subset of any combinations of two parameter with a collinearity above 20, indicating non-identifiable parameter combinations is given (Brun et al., 2001).

ζ	R^C	θ^N_{max}	Q_{min}	Q_{max}	α^{Chl}	I	n	K_{no3}	P^C_{ref}	collinearity
1	0	1	0	0	0	0	0	0	0	31
1	0	0	0	1	0	0	0	0	0	59
1	0	0	0	0	1	0	0	0	0	42
1	0	0	0	0	0	1	0	0	0	42
1	0	0	0	0	0	0	1	0	0	74
0	1	0	0	0	0	0	0	0	1	22
0	0	1	0	1	0	0	0	0	0	32
0	0	1	0	0	1	0	0	0	0	26

Pg. 38, Figure B3: the key seems to omit reference to the bacterial model
We added the information in the legend.

Pg. 38, Figure B3: the failure of the model to capture the observed behaviour of the PON seems quite significant, but is not well-described in the text; it is also noticeably different from that of POC, which suggests interesting POM dynamics in the model that I would not expect; do the authors know what is going on here?
After G98 carbon is continuously fixed, even under nitrogen limitation (Carbon overconsumption, Schartau et al., 2007), while nitrogen is slowly used up for maintenance (maintenance respiration term), leading to a decoupling of POC and PON. The main reason for the failure of the G98 model is the neglection of bacterial DIN regeneration. Thus, the PON dynamics are quite well modelled for the BACT- experiment, while the BACT+ experiment shows severe limitations. In fact, this is one of the main arguments showing the need to include bacterial regeneration. The model may be tuned to an artificially better fit the BACT+ treatment by increased DIN uptake efficiencies, but this would lead to a substantially poorer fit to the BACT- experiment. As discussed on p2 l.63 this fitting of the G98 model without a bacterial regeneration component on non-axenic culture experiment can lead to misleading interpretations and kinetic parameters (e.g. half saturation constants). We added this information to the discussion.

Schartau, M., Engel, A., Schröter, J., Thoms, S., Völker, C., & Wolf-Gladrow, D.: Modelling carbon overconsumption and the formation of extracellular particulate organic carbon, 2007.

Pg. 38, Figure B3: would quartile or decile range be better here?; this may make your experiments look more messy than they actually are (i.e. it looks like you may have an outlier experiment); this may not be possible given the number of replicates

With three measured values per day and treatment we prefer to show all values separately instead of artificially calculating error estimates (e.g. quartiles, deciles, standard deviations).