

Letter to Editor

Dear Dr. Grégoire,

Thank you very much for handling our manuscript and having provided several important suggestions. In an effort of thoroughly addressing your and the two reviewers' concerns we also attach our response to your comments as follows.

1) The statement mentioned in the abstract (line 21-23) that “High nucleic acid (HNA) bacteria show relatively high cell-SPECIFIC productivity, respiration, and utilisation of the semi-labile dissolved organic carbon pool compared to their low nucleic acid (LNA) bacteria counterparts.” is highlighted as an important result of model simulations performed in this paper although it results from model parameterization (and hence assumptions). Reviewer #2 pointed out that this conclusion can be considered as an output from the model if it applies to the non-normalized quantity. However, as written, this refers to the biomass-normalized quantities. Please carefully check the detailed comment of reviewer #2 and provide a clear answer.

We agree with reviewer #2. The fact that cell-specific BP, respiration, SDOC uptake rates were significantly higher for HNA bacteria compared to those for their LNA counterparts (Section 3.2) is mainly because of the way the parameter optimization was conducted (now detailed in Text S3). The higher initial parameter values assigned for HNA bacterial growth, RDOC excretion, mortality, and respiration rates (Table 1) might drive not only their faster cell-specific growth rates but also their higher DOC uptake rates to coexist with LNA bacteria when the loss rates were relatively large for HNA bacteria. Though driven by the model assumptions, the important aspect of these results lies in the fact that the model can leverage such assumptions to examine the implications for the WAP food-web dynamics and biogeochemistry (line 356-362).

2) The model description needs to be revised and the number of state variables to be checked. At the time of the first review, the GMD paper was not available. Since it is now published, I suggest that you keep in the main text a summary of the model main characteristics and put in the appendix details on the formulation. These details have been requested by Reviewer #2 and this is important that it remains accessible in the manuscript but to improve the clarity of the paper and make it more accessible, transfer some parts in the supplementary section as suggested by reviewer#3 in his/her comment.

As suggested by both reviewers, we have greatly reduced and simplified the Methods section on the model processes, and much of the relevant sections are now redirected to Kim et al. (2021) or Supplementary Material of our study. As suggested, we also have removed the model equations previously in Appendix and only focused on the formulation of the newly added processes and equations for HNA and LNA bacteria in main text (line 78-107).

3) Both reviewers do not understand the concept of bacterial models and fmodes. This has to be absolutely clarified. To refer to Bowman et al., (2017) is not enough. Please clearly explain, in this paper, to what functions the fmodes refer. This comment has been mentioned many times by reviewer #2 and the third reviewer has exactly the same concern. Please carefully check and answer the detailed comments of Reviewers #2 and #3.

We completely agree with your assessment. Thus, after significant thought, we have decided that it would be best to completely remove the Methods section on the functional modes given that the functional modes did not have significant relationships with the modelled bacterial C stocks and rates and other ecosystem functions and were not presented in Results and Discussion to begin with. Thus, Removing this part does not change the findings and conclusions of our study. We also have made the Methods section on the taxonomic modes more accessible and in line with our mechanistic modeling perspectives, as well as moved the taxonomic mode figure from main text to Supplementary Material (line 108-121, Fig. S10).

4) Please explain why the variability in model simulations is lower compared to that from observation.

The lower variability in the simulations should be considered as the limitation of our model and we have included the relevant information in the revised manuscript. The model results have less variability than the observations likely because we do not have sufficient information in the model forcing to capture all of the small-scale and high-frequency sources of variability, such as local circulation and tidal flow near Palmer station in the 4 study years modelled in this study. By contrast, our model adequately captures seasonal variations in modelled ecosystem dynamics. This is likely because such high frequency processes do not strongly rectify into the seasonal cycles in the system (line 382-386).

Response to Referee #1 (Reviewer #2)

Third review of "Modelling polar marine ecosystem functions guided by bacterial physiological and taxonomic traits" by H. H. Kim et al.

I think this is the first time I am writing a third review for a manuscript. In their second revision, the authors have now clarified/corrected most of the technical and language problems of the previous versions. Nevertheless, substantial problems remain, as detailed below, so that I still can not recommend publication of the ms as is. While the amount of required changes is quite small, the confusion of assumptions and results is sufficiently severe, so that it should be considered a major revision.

Thank you for reviewing our manuscript again. We have addressed your concerns as follows.

My main problem now lies in the still unresolved problem that the authors present an immediate consequence of their model assumptions as a result of their study, namely that the "High nucleic acid (HNA) bacteria show relatively high cell-specific productivity, respiration, and utilisation of the semi-labile dissolved organic carbon pool compared to their low nucleic acid (LNA) bacteria counterparts." (Abstract, lines 21–23). The authors' response does not really address this problem as it only applies to the total, rather than the biomass-normalised, rates. Please note that I have no problem accepting as a result the finding that the total (not biomass-normalised) rates are higher for HNA than for LNA, and I also think that this would be actually much more relevant in terms of both ecology and biogeochemistry. I also do not question that the growth rate of the HNA may potentially be lower at low labile DOC concentration because the HNA have a higher half-saturation concentration. But this does not apply (at least not for the parameters shown in Tables S2–S6) for the labile DOC concentrations in this study (Fig. 8, LDOC). In addition, several (not optimised or constrained) loss-rate parameters (RDOC production, mortality, respiration) are higher for HNA, and these must be compensated by faster DOC uptake in order to allow coexistence of HNA and LNA. Clearly, therefore, the higher biomass-specific rates of HNA are imposed by the model assumptions and must not be presented as a result.

We agree with your comments. The fact that cell-specific BP, respiration, SDOC uptake rates significantly higher for HNA bacteria compared to those for their LNA counterparts (Section 3.2) is mainly because of the way the parameter optimization was conducted (now detailed in Text S3). The higher initial parameter values assigned for HNA bacteria's growth, RDOC excretion, mortality, and respiration rates (Table 1) might drive not only their faster cell-specific growth rates but also their higher DOC uptake rates to coexist with LNA bacteria when the loss rates were relatively large for HNA bacteria. Though driven by the model assumptions, the important aspect of these results lies in the fact that the model can leverage such assumptions to examine the implications for the WAP food-web dynamics and biogeochemistry (line 356-362).

The description of the state variables on lines 80–102 is still wrong. The statement that the model has 12 state variables (line 81) is simply not true. I do not understand the hesitation of the authors to correct this obvious mistake.

We have 12 model state variables as stated in the previous version of the manuscript. The confusion may have come from the fact that there are 14 boxes in the model schematic (Fig. 1). However, high level and RDOM are not state variables because their time derivatives are not calculated in the model, even if there are C flows going into them as the sink terms of other model state variables. If we counted N and P components of the model compartments the number of state variables increases further. As the other reviewer suggested as well as we are not sure where this confusion on the number of state variables comes from, we have decided to omit this information in the revised version, given that the actual total number of state variables is not important.

The problems with the bacterial modes and fmodes largely remain. For example, I had asked what functions the functional modes refer to. The authors refer (also in their response letter) to Bowman et al. (2017) for details but all that Bowman et al. (2017) write about the fmodes is this: "Based on inspection of the within-cluster sum of squares plot, we identified [...] eight modes based on inferred metabolic pathways (not shown)." I think it is impossible to judge the validity of the statements regarding the fmodes without concrete information about the associated actual functions (metabolic pathways). How can one know whether these functions are selected in a meaningful or useful manner if no information about them is provided? Regarding the bacterial modes, in their response letter, the authors write that they changed the example in Fig. 6 to Dokdonia. This does not address the problem I described. I never had any problem understanding the Dokdonia case. My problem is understanding how the assignment of the modes to species works and I explained the (apparent?) contradiction between the authors' definition and what is shown in Fig. 6 with the example of *Candidatus Pelagibacter ubique* and *C. Thioglobus singularis*. Using a set of species where this problem does not show up obviously does not help here. The unclear presentation of the mode concept and the above problem of the confusion of assumptions and result are the reasons why I grade the scientific significance and quality as poor. The presentation quality gets a fair grade mainly because of the extremely poor language quality. While I do appreciate the additional information about the model and the corrections in this latest revision, I do in fact expect that authors supply this kind of information already with the initial submission.

Thank you for your comments. we have decided that it would be best to completely remove the Methods section on the functional modes, given that the functional modes did not have significant relationships with the modelled bacterial C flows and other ecosystem functions and were not presented in Results and Discussion to begin with. Thus, please note that removing this part does not change any of the findings and conclusions in our study. The revised manuscript now only focuses on the taxonomic modes. We also have moved the taxonomic mode figure (previously Fig.

6) to Supplementary Material (now Fig. S10) to simplify the relevant Methods section and make it more accessible (line 108-121). The revised manuscript now briefly discusses what each mode is associated with regard to bacterial taxonomy in the Results section (line 264-267). We hope these efforts address your concerns.

The sentence on lines 447–448 is still unclear to me. "... the values above ..." (above what?) This is followed by "... detect limits ..." It appears that two sentences were merged and something was lost, e.g., "... the values above those required by ..." "... this was used to detect the limits or indicate the lack of macronutrient limitation ..."

Thank you – we fixed that sentence to: The modelled nutrient stocks were above the detection limits, indicating no evidence of macronutrient limitations at the study site (line 371-372).

Response to Referee #3 (Reviewer #3)

This paper uses data from a 1D ecosystem model alongside bacterial genomic data to give insight into the role of bacteria in ecosystem functioning at a site in the West Antarctic Peninsula. The ecosystem model, recently accepted for publication in a separate paper, has been modified to include two functional types of bacteria, HNA and LNA, and it is able to assimilate biomass of these types using flow cytometry data. Outputs from the model are examined in association with bacterial groupings derived from previously published taxonomic analysis of genome data from the same location. This is interesting work, and I was pleased to see a study that combine ecosystem model and genome data. The paper has been much improved by the previous cycles of review. However, I found it difficult to read and I recommend a little further work before it is published, to make it accessible to a wider readership.

[Thank you for taking time to review our manuscript and also for your positive feedback. Please see below how we have addressed your comments on the previous version of this manuscript.](#)

First, now that the paper describing the ecosystem model has been accepted for publication (Kim et al., 2021) the methods section can be reduced and simplified. Section 2.1 just needs a summary of the model previously published: the key points are the main functional types, which elements are tracked (C,N,P) and whether there is flexible stoichiometry. Of course the differences from Kim et al. (2021) need to be covered, but this can be relatively brief. I would remove the equations and the references to them – I know that a previous reviewer requested them, but they are identical to the ones in Kim et al. (2021) and can now be read there; lines 80-117 will be much easier to read without the references. If the equations are kept the symbols need to be explained. The description of the data assimilation scheme in sections 2.3-2.5 is very similar to that in Kim et al. (2021) and does not need to be repeated in such detail. Only a summary is needed in this paper, with the emphasis on the differences compared to the first paper, e.g. the difference in calculation of the target error. Reduced in this way, the methods section would be easier to read and give more weight to the work on bacteria, which is the novel part of this paper.

[Thank you for your suggestion. As you suggested, we have greatly reduced and simplified the Methods section by moving many of the details on parameter optimization processes and target error adjustment to Supplementary Material \(now in Text S3, S5\). However, we still provide a summary of the key model aspects and in particular what is new to this modified model compared to the original WAP-1D-VAR v1.0 \(line 78-107, 129-132, 181-183\). Throughout the manuscript, we now redirect the reader to Kim et al. \(2021\) for more thorough presentations of the details of the base model. As suggested, we also have removed all the model differential equations of other ecosystem variables previously presented in main text as well as from Appendix.](#)

Second, I suggest giving a slightly fuller explanation of the genomic data, for readers like myself

who are less familiar with this than with the modelling. The terms mode, functional mode, closest estimated genomes and closest completed genomes all seem to be specific to Bowman et al. (2017) and as such need to be explained more fully (or omitted if they are not needed – I don't see that referring to closest estimated genomes and closest completed genomes is required for the discussion here). I understand that taxonomic modes are groupings based on 16S rRNA gene sequence, but I don't understand where the fmodes come from. Line 134 says that “functional modes (fmodes hereafter) were derived from predicted community metabolic structure” – what data was this based on? From reading Bowman et al., 2017, I think it is also the 16S data, but the phrasing here seems to imply that it comes from a different source. I'm also not clear how the distinction between modes and f-modes can be interpreted: in section 3.3 it is stated that “fmodes did not have a significant relationship with any of the modelled ecosystem functions examined” – how can we interpret that in terms of the bacterial contribution to ecosystem functioning?

Thank you for your comments. We agree that our bacterial mode description was not as clear as it should have been and it caused lots of confusion to both you and the other reviewer. We have realized that presenting the details on genomic and statistical analyses for mode extraction would make our manuscript even more difficult to understand, so we have decided to greatly simplify the Methods section on bacterial modes. First, we have completely removed the Methods section on functional modes as they did not show any significant relationships with the modelled bacterial stock and flows or other key ecosystem functions. Second, we have shortened the Methods section on taxonomic modes and changed the wording to be more in line with our numerical modeling aspects (line 108-121). Third, we have moved the taxonomic mode figure to Supplementary Material (Fig. S10) and only briefly discuss how each mode represents its specific taxonomic attributes in the Results section (line 264-267).

Third, section 2.2 and Figure 2 led me to expect more integration of the genomic data into the model than I actually found. Figure 2 shows the data-driven part of the model, where the bacterial modes are used, but I could not find any description of how this is done in the methods. Line 133 refers to “the data-driven part representing how bacterial modes (Bowman et al. 2017) are compared to final model outputs based on optimized model parameters from the dynamic part”. So is this just a comparison between model outputs and bacterial modes, as presented in sections 3.3 and 4.3? In what sense is the modelling framework data driven, i.e. what does the arrow in Figure 2 pointing from the bacterial modes to the model field represent?

Thank you for your comments. Your assessment is correct that we compare the optimized model outputs from the ecosystem model (Fig. 1) with the bacterial modes from Bowman et al. (2017). We refer to the bacterial mode part as “data-driven” because Bowman et al. (2017) uses a data-driven approach to derive the taxonomic modes; in particular, they used an unsupervised machine learning algorithm called Kohonen's self-organizing maps (Kohonen, 2001) to reduce the high dimensionality of 16S rRNA gene sequence abundance data to a single taxonomic mode. We have

elaborated the Methods section to make this aspect clear (line 108-121). We also removed the arrow in Fig. 2 from the “bacterial modes” in the data-driven part to the “model outputs” in the mechanistic part as the bacterial modes themselves are not directly incorporated to the ecosystem model in contrast to HNA and LNA C biomass data, which are assimilated into the simulation.

My main scientific concern is that in many cases the model values have a much lower standard deviation than the observations (Figure 3, Figure S2), and more so than in the previous model (Kim et al, 2021). So the model appears to be missing much of the variability observed in the field measurements. Do the authors have any comment on this?

Thank you for making this point. The model results have less variability than the observations likely because we do not have sufficient information in the model forcing to capture all of the small-scale and high-frequency sources of variability, such as local circulation and tidal flow near Palmer station in the 4 study years modelled in our study. By contrast, our model adequately captures seasonal variations in modelled ecosystem dynamics. This is likely because such high frequency processes do not strongly rectify into the seasonal cycles in the system (line 382-386).

With these modifications, I think the paper will be a useful contribution to the literature, with relevance beyond the particular study area.

A few specific points:

Abstract: the abbreviation WAP needs to be explained.

Line 61: r missing from Palmer

Line 81: I agree with the previous review comment that there are more than 12 state variables, and it is still not clear that only the carbon stocks are being considered. But why give a number? I don't see that this is important for the rest of the text.

Line 83: LDOC in the text is given as LDOM in Figure 1. Is it the same thing? Similarly for SDOC, RDOC.

Line 108: rpresent instead of present

Line 168: avialable instead of available

Line 291 and Figure 3: it is not specified how the standard deviation is normalized.

Tables S2-S6: it would be helpful to explain the abbreviations OP and CS in the legends.

Text S1 to S2: I think much of this is now part of Kim et al. (2021) and can be removed.

Thank you – all of these technical comments have now been fixed.