

Zooplankton grazing is the largest source of uncertainty for marine carbon cycling in CMIP6 models



Open Access This file is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. In the cases where the authors are anonymous, such as is the case for the reports of anonymous peer reviewers, author attribution should be to 'Anonymous Referee' followed by a clear attribution to the source work. The images or other third party material in this file are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

Web links to the author's journal account have been redacted from the decision letters as indicated to maintain confidentiality

20th Sep 22

Dear Dr Rohr,

Please accept my apologies for the delay in obtaining reports on your manuscript titled "Marine carbon cycling and sequestration is extremely sensitive to zooplankton grazing in biogeochemical models". Your manuscript has now been seen by 3 reviewers, whose comments are appended below. You will see that they find your work of some potential interest. However, they have raised quite substantial concerns that must be addressed. In light of these comments, we cannot accept the manuscript for publication, but would be interested in considering a revised version that fully addresses these serious concerns.

We hope you will find the reviewers' comments useful as you decide how to proceed. We would be willing to consider a substantially revised manuscript, should additional work allow you to address these criticisms and meet the following editorial thresholds:

- Provide compelling quantitative new insights into the differences in the sensitivity of carbon export to zooplankton grazing pressure between different biogeochemical models;
- Provide comprehensive method details, including discussion of assumptions and associated limitations and uncertainties;
- Consider using a more complex reference model;
- Provide a balanced in-depth discussion of the wider related literature and clearly frame the advance your findings offer in this context.

If you choose to take up this option, please either highlight all changes in the manuscript text file, or provide a list of the changes to the manuscript with your responses to the reviewers.

Please bear in mind that we will be reluctant to approach the reviewers again in the absence of substantial revisions.

If the revision process takes significantly longer than three months, we will be happy to reconsider your paper at a later date, as long as nothing similar has been accepted for publication at Communications Earth & Environment or published elsewhere in the meantime.

We understand that due to the current global situation, the time required for revision may be longer than usual. We would appreciate it if you could keep us informed about an estimated timescale for

resubmission, to facilitate our planning. Of course, if you are unable to estimate, we are happy to accommodate necessary extensions nevertheless.

We are committed to providing a fair and constructive peer-review process. Please do not hesitate to contact us if you wish to discuss the revision in more detail.

Please use the following link to submit your revised manuscript, point-by-point response to the reviewers' comments with a list of your changes to the manuscript text (which should be in a separate document to any cover letter) and any completed checklist:

[link redacted]

** This url links to your confidential home page and associated information about manuscripts you may have submitted or be reviewing for us. If you wish to forward this email to co-authors, please delete the link to your homepage first **

Please do not hesitate to contact me if you have any questions or would like to discuss the required revisions further. Thank you for the opportunity to review your work.

Best regards,

Clare

Clare Davis, PhD

Senior Editor

Communications Earth & Environment

www.nature.com/commsenv/

@CommsEarth

EDITORIAL POLICIES AND FORMAT

If you decide to resubmit your paper, please ensure that your manuscript complies with our editorial policies and complete and upload the checklist below as a Related Manuscript file type with the revised article:

Editorial Policy [Policy requirements](https://www.nature.com/documents/nr-editorial-policy-checklist.pdf) (Download the link to your computer as a PDF.)

For your information, you can find some guidance regarding format requirements summarized on the following checklist: (<https://www.nature.com/documents/commsj-phys-style-formatting-checklist-article.pdf>) and formatting guide (<https://www.nature.com/documents/commsj-phys-style-formatting-guide-accept.pdf>).

REVIEWER COMMENTS:

Reviewer #1 (Remarks to the Author):

General comments

In this interesting paper, Rohr et al. describe the incredible sensitivity of global primary production, secondary production and export production to a derived, theoretical variable they call the 'Grazing Pressure' in a range of CMIP6 models. Having carefully read this work, I am torn with regard to the recommendation I am expected to give. On the one hand, I think the message with regard to the completely underestimated importance of the higher trophic levels and top-down controls on carbon fluxes in these models is very important, and also supported by recent observational evidence (e.g. the recent metagenomic information; Guidi et al., Nature, 2016). The traditional bottom-up view of marine ecosystems in the classical biogeochemistry modeling community is no longer tenable, and must be revisited. On the other hand, the marine ecosystem modeling community has known this for years, if not decades, and every modeler who has ever tuned a model has made the experience that tinkering around with the grazing pressure changes everything in these PFT-type models. Hence, the findings presented in this work are not entirely novel, nor very surprising, and could equally well be published in a more topical journal. Regardless of this strategic decision, the paper needs major revisions before it can be accepted.

In my view, several major flaws need to be resolved before this work can be published here or elsewhere. These mostly affect the presentation and justification of the methodology and results, and not the actual conclusions, which I think are sound and in line with what we know about these models.

First and foremost, the paper seems oblivious to the current state-of-the-art in the current marine ecosystem modeling community with regard to the role of zooplankton for marine ecosystem dynamics and the regulation of derived carbon cycle variables such as NPP, phytoplankton biomass dynamics, or export. It seems to neglect a whole decade of literature on this topic, starting from the work of Sailley et al., *Ecological Modelling*, 2013, who already examined the differences in top-down pressures and food-web structure between CMIP-type models and came to the absolute same conclusions that the current authors do (that food-webs differ (even intended ones from those effectively realized) and that future projections of the carbon cycle variables are shaky at best) to the work by Vallina et al., *Prog. Oceanog.*, 2014 or Prowe et al., *Prog. Oceanog.*, 2012 on the difference of different grazing formulations on phytoplankton biomass and diversity (which both are closely linked to NPP), nor is it discussing the work by Le Quéré et al., *Biogeosciences*, 2016 on the importance of zooplankton community structure on Southern Ocean NPP (which claims we need multiple higher trophic levels and trophic cascades to correctly represent NPP and biomass dynamics). Further important recent work discussing the importance of grazing in structuring simulated marine ecosystems isn't cited either (e.g. Dutkiewicz et al., *Biogeosciences*, 2020; Chenillat et al., *PLOS One*, 2021), and the authors also haven't really fully exploited the potential of the work of Laufkötter et al., *Biogeosciences*, 2015, 2016 to strengthen their case, with both of the latter papers having made a relatively refined attempt at tracing carbon across the food-web and quantifying export efficiency, with important consequences for the work presented here. It also does not compare its findings to important conceptual work on differences in food-web structure between different oceanic environments (Everett et al., 2022), nor has it taken full advantage of another paper which I think it cites (citation truncated; Everett et al., *Front. Mar. Scie.*, 2017). It also neglects an absolute wealth of zooplankton observations to back up its claims, starting from novel approaches to detect zooplankton biomass from space (Basedow et al., *Scientific Reports*, 2019; Druon et al., *Scientific Reports*, 2019; Strömberg et al., *Journal of Marine Systems*, 2009), to extensive zooplankton biomass and trait observations that modelers use to validate their models (Buitenhuis et al., *ESSD*, 2013 on microzooplankton, O'Brien and Moriarty, *ESSD*, 2013 on mesozooplankton; Moriarty et al., *ESSD* 2013 on macrozooplankton and plenty of literature on the functional biogeography of zooplankton (multiple papers by P. Brun and F. Benedetti, as well as Barton et al., *Ecology Letters*, 2013)). This means that the paper currently fails to identify and highlight the elements of novelty in their work to a sufficient degree.

Second, the methods section is poorly composed and many methodological choices remain ill justified and explained, so that it is difficult to judge whether the paper's new 'Grazing Pressure' metric is helpful. In my view, it would be fully possible to calculate the predator- and prey-specific and total grazing rates of the CMIP models offline on a pixel-by-pixel basis, based on the regular CMIP6 output, the model equations and some assumptions with regard to missing variables that weren't submitted to the CMIP6 archive (e.g. replacing missing submissions by observational fields, or ensemble means across all other models). On a monthly scale, you will usually get to between 10-20% within the actual value

calculate by the model online with an offline calculation, which is quite good, and could be used to see how the true grazing pressure changes in space and time, and as a function of prey biomass. The authors were unable to convince me that their theoretical 'Grazing Pressure' variable, which uses very coarse averaging in space and time to produce variables that eventually do not reflect community structure and dynamics anywhere in the model domain, was better than the actual offline computation of zooplankton biomass, and the de facto grazing pressure in these models. Having worked a lot with grazing formulations of different models, I could have guessed some of the responses shown in their figures without any knowledge of 'Grazing Pressure', just by simply looking at model parameters and structure. Hence, the authors (a) must give equations to show how 'Grazing Pressure' is calculated for each model (are you working with annual or monthly means, the surface ocean or the euphotic zone, how is the seasonal dynamics of grazing pressure reflected in your metric? why is one global number more informative than zonal means, what does an annual mean represent in seasonally highly varying systems? etc.) (b) justify why looking at 'Grazing Pressure' would be better than looking at the actual equations themselves (and quantitatively, please, using more than just rather hand-waving claims; you could do sensitivity tests here as to the comparability of the actual grazing pressure in WOMBAT as computed pixel-wise), and (c) discuss in detail how their findings fit into the context of previous work. Furthermore, we need to see how you take into account the substantial differences in multi-prey formulations (which you now just state is an additional uncertainty, but in our experience changes grazing pressure substantially, see respective literature, for instance by Vallina et al., *Ecological Modelling*, 2011; Vallina et al., *Prog. Oceanography*, 2014), and address this theoretically. Last but not least, I'd like to see an uncertainty range in form of shading for all the theoretical curves that takes into account the spatio-temporal variability of 'Grazing pressure' (Figs 1,2,3 and so on), as it varies across environments, seasons and model set-ups. A single dot for a hypothetical annual mean is unlikely to reflect the dynamical variation each model is capable of in space and time.

Third, the paper should reflect more on the relative sources of uncertainty and variability in their 'grazing pressure' and carbon cycle metrics stemming from (a) the grazing formulation itself, (b) the way multi-prey interactions are parameterized (see e.g. Vallina et al., *Ecological Modelling*, 2011 and Vallina et al., *Prog. Oceanog.*, 2014 as to how to remedy suboptimal grazing), (c) the effect of higher trophic level complexity (see Le Quéré et al., *Biogeosciences* 2016), and (d) the way carbon is channeled to multiple higher trophic levels with different storage capacity (see e.g. Laufkötter et al., *Biogeosciences*, 2016). I won't even mention behavior & life stages (Clerc et al., *Theoretical Ecology*, 2021), DVM (e.g. Kelly et al. *Front Mar. Sci.*, 2019), trophic amplification (Kwiatkowski et al., *Global Change Biology*, 2018) and so on. The former four issues act in concert to generate uncertainty in carbon fluxes within the marine ecosystem, and the grazing formulation is just one part of the uncertainty. And since we know a lot about that particular aspect already, setting this one source of uncertainty into the context of the uncertainty associated with the other three issues seems warranted. Furthermore, spatio-temporal information on the zonal and seasonal patterns of grazing are essential dimensions of variability. NPP and EP aren't generated in the annual mean, and everywhere – they are generated during specific high productivity months, and at certain specific latitudes. And we know that the models poorly represent the spatio-temporal dynamics of bloom formation and transfer across the ecosystem (Hauck et al., *Global Biogeochemical Cycles*, 2015; Hashioka et al., *Biogeosciences*, 2013). The author has the entire range of CMIP6 models at his hands to delve deeper into some of these issues. Broadening the work to a

more general discussion of the aspects above and including spatio-temporal variability as an essential source of information on how well these models really do would increase the novelty and relevance of the current work.

Fourth, I question the choice that the emphasis of the paper being put on WOMBAT results, whereas one could have learnt a lot from a more detailed analysis of the CMIP6 model suite. WOMBAT's ecosystem structure is far too simplistic to correctly capture the huge uncertainties in model ecosystem dynamics due to the above-mentioned issues. Hence, I personally would not use the standard version of WOMBAT as a reference model (I'd add further trophic levels; see e.g. Ward et al., *Prog. Oceanog.*, 2013), but since that is the model the authors are familiar with, that probably can't be helped. Alas, based on the current text, I really cannot understand how simulations conducted with WOMBAT are reflective of the carbon dynamics in the CMIP5 models, and this link needs to be strongly emphasized in a revised version of the paper. All of the CMIP models are heavily tuned not only to match NPP estimates, but also export at 100 m. Hence, as you are varying WOMBAT through the range of realized values of 'Grazing Pressure' across models, you are looking at variations in NPP, GSP, EP that could lead to the exclusion of the respective parameterisation in the original CMIP models, because NPP and EP could potentially not be tuned to fit the observed range. So what do you prove with WOMBAT? In essence, this model is somewhat too simple to assess the role of zooplankton in these models, since it doesn't incorporate any of the known sources of variations in these models (multi-trophic level interactions (Le Quéré et al. 2016), prey-switching (Vallina et al. 2014), food-web links (Sailley et al. 2013) and so on, except for one single group with one single grazing equation. Why would you use a model that does not resolve higher trophic levels to prove that higher trophic levels matter? Please express the novelty and relevance of your findings far more clearly, and work more with the CMIP models more to support your main take home messages.

Fifth, the paper must work on its selection of figures. While some of those presented are generally informative, others are mostly summarizing model sensitivity studies, either conducted with WOMBAT, or based on the CMIP6 model suite (e.g. Figs 1,3,4). Hence, some figures presented in the main text feel like supplementary information, based on theoretical calculations with little to no relevance for the actual global ocean (see my comments above). What we would like to know, or at least what I would like to know, is what lessons we should learn, i.e. a conceptual diagram or schematic how (different versions of) zooplankton community structure and grazing influences NPP, GSP and EP, and how this is likely to affect future projections of the carbon cycle variables in the CMIP6 models (as opposed to 'in WOMBAT').

Specific comments:

Introduction

One of the largest gaps in our understanding of marine carbon cycling is the role of zooplankton. – Add relevant references. And is this true? Justify.

Yet grazing dynamics are poorly constrained (7) and difficult to model (8), hampering our understanding of the ocean's – Referencing utterly insufficient, look at works by Vallina (multiple), Sailley et al., Ecological Modelling, 2013, Chenillat et al., PLOS One, 2021 (and references therein); Dutkiewicz et al., Biogeosciences, 2020 and so on. Also have a look at Séférian et al., Current Climate Change Reports, 2020 on improvements in model complexity between CMIP5 and CMIP6 models.

Zooplankton species exhibit a great deal of functional diversity and their relative distribution varies widely (9). – Please cite all the relevant work by Benedetti et al., Brun et al. and Kiorbøe et al.. See also Barton et al., Ecology Letters, 2013 on trait biogeography.

Yet marine biogeochemical (BGC) models represent the integrated behavior of all zooplankton using a strictly limited number of zooplankton groups, especially in computationally expensive IPCC coupled-climate models. – Cite Le Quéré et al., Global Change Biology, 2005 on PFTs, Laufkötter et al., Biogeosciences, 2015, 2016, Sailley et al., Ecological Modelling, 2013 and so on.

Further, while chlorophyll, NPP, and phytoplankton biomass can be estimated from space, global in-situ zooplankton observations cannot – Not true, see recent papers cited in general comments above on the detectability of zooplankton from space. Rephrase.

making simulated zooplankton biomass much more difficult to validate against spatially-resolved global data sets (22). – Not a discussion paper anymore, and see also Sailley et al. 2013 and think about the availability of further global data sets now available for different size classes and taxonomic groups, e.g. used in Buitenhuis et al., Global Biogeochemical Cycles, 2019

rooted primarily in theoretical predator-prey ecology and small-scale, in-vitro, laboratory experiments (23) – Referencing not up to date with literature on the multiple trait collections now used to validate models.

While these values can be estimated empirically (26, 27), measured parameter values span over three orders of magnitude (23), vary with zooplankton size, species, and age (26), and are not likely representative of grazing dynamics in the open ocean (28). – (28) seems wrong reference here.

Unsurprisingly, Rohr et al. (23) surveyed 40 BGC models, dating back to foundational NPZD box-models (29, 30) and found that the values used for K and G both span over two orders of magnitude. – You are not the first to survey different models, see e.g. early attempts in Sailley et al., *Ecological Modelling* 2013, or the work by Hashioka et al., *Biogeosciences*, 2013 on the quantification on top-down versus bottom-up pressures to explain bloom dynamics.

how the distribution of prey modifies grazing behavior and influences the integrated grazing pressure exerted on the full phytoplankton community – references needed, see e.g. the work by Vallina et al., and Vallina and Le Quéré.

Despite the uncertainty associated with the representation of grazing, there has been no global study to quantify the sensitivity of marine carbon cycling to the degree of grazing pressure prescribed in coupled ocean-BGC models. – Not true, see Sailley et al., *Ecological Modelling*, 2013 and Laufkötter et al., *Biogeosciences*, 2015, 2016. And Vallina et al., *Prog. Oceanog.* 2014 on the effect of the grazing formulation on diversity and biomass distribution. Just because your metric was not used, this doesn't mean that the community hasn't looked at the role of zooplankton or grazing, wouldn't you agree?

Together, this could have contributed to the persistent uncertainty surrounding NPP (32) and zooplankton biomass (8) in future climate projections – see also Laufkötter et al., *Biogeosciences*, 2015.

the lack of adequate global, in-situ grazing observations to constrain them (35) - Really? What about all the important and extensive work on grazing rates that there is? Multiple papers by

Calbet? Landry? Ohman? Compilations made by marine ecosystem modellers for each new zooplankton FT included? See also references in Barton et al., *Ecology Letters*, 2013.

A better understanding of zooplankton grazing and its influence on marine carbon cycling could substantially improve estimates of past, present, and future oceanic carbon sequestration and transfer to higher trophic levels – please cite relevant work, from global to regional modeling and from all over the marine ecosystem modeling world – Dutkiewicz, Le Quéré, Hashioka, Bopp, Aumont, and many more. Interesting in this context may also be the discussion in Nissen et al., *Biogeosciences*, 2021.

Results

2.1 Models - Here, I lack a discussion of sub-optimal grazing and the KTW formulation by Vallina et al., Prog. Oceanog. 2014 in the context of advantages and disadvantages of each formulation. None of the models here account for the fact, that overall grazing decreases as the number of prey increases, which is unlikely to reflect the behavior of zooplankton in nature.

2.2 Grazing pressure – Metric used is ill-defined and ill-motivated. Why not simply calculate the actual grazing rates offline, since you know the grazing functions, and have prey fields and temperature as outputs in CMIP6? What is gained by the very coarse global averaging? And what are you actually even averaging? Rather than saying you use some unspecified ‘mean’, let us know if you used monthly or annual fields to average, if you are using surface or depth integrals, etc. This very important information is partly (!) given in the Methods, but must be stated in the main text, since else it is not possible to evaluate the suitability of the metric used. And I do not believe I learn so much from that metric, actually. It seems to me that it is only useful to be able to determine the range within which you fit WOMBAT. So for me it is collapsing complexity and variance into one single variable that is only useful for comparison with an NPZD model, which, in my view, doesn’t account for the necessary complexity in zooplankton to even assess the role of higher trophic levels.

We define Grazing Pressure, $g(P)$ ($d-1$), as the mean rate at which phytoplankton biomass would be lost to grazing per unit zooplankton in a standardized prey field. – how does the ‘standardized prey field’ reflect nature? Phytoplankton are highly diverse, coexistence of multiple types of plankton is standard, and distributions are patchy and highly variable. How do you reconcile this with your choice of ‘one value fits all’ test fields?

Grazing Pressure can be used to compare the relative strength of grazing assumed in models with disparate formulations to that implied by empirically estimated functional response curves. – Yes, and so could the actual grazing equations. Why is your approach better?

Here, we compute Grazing Pressure as the mean of two end-member scenarios, both with the same standardized, mean, global phytoplankton concentration as in the simple scenario described above, but now with two different distributions. In one ‘segregated’ scenario plankton are assumed to be completely heterogeneously distributed into unique niches with no competition, while in one ‘mixed’ scenario plankton are assumed to be completely homogeneously distributed throughout the ocean. – I don’t see why these two cases are representative of global ocean marine ecosystems. And your explanation below does not convince me. Communities with a very high dominance of one PFT all others occur, perhaps, in bloom-situations at mid latitudes. Why is this interesting? For the majority of the globe and during most times, phytoplankton coexist and/or closely succeed one another, with huge differences in the communities between seasons. You bring up theoretical scenarios, but then you still want to make inferences for the carbon cycle variables as projected by models that aim to represent real plankton systems. You lose me here as to how these two scenarios are relevant beyond a merely theoretical (and thus not easily applicable) knowledge-gain.

Experimental estimates of Grazing Pressure can be determined from the functional response curves fit empirically to > 60 laboratory dilution experiments reviewed in the literature (26, 27). – Experiment absolutely insufficiently explained (here and in Methods). Which organisms are being studied, under which conditions? Here, the prey species matter just as much as the zooplankton predators. List needed.

These empirical estimates of Grazing Pressure differ by over three orders of magnitude, ranging from < 0.003 d⁻¹ to > 3 d⁻¹ and vary widely with zooplankton species and functional group. – yes, as I said above. We need a supplementary table where you show what these 60 experiments really studied.

When zooplankton are grouped into crustacean mesozooplankton (Fig.

1A; orange) and microzooplankton (Fig. - Why? Rationale behind the grouping? Here the issue is that many organisms previously considered microzooplankton (e.g. heterotrophic dinoflagellates) are mixotrophs, that mesozooplankton previously only contained copepods, which are abundant, but new data shows other zooplankton functional groups are closely associated with export (Rhizaria for example, Guidi et al., Nature, 2016), and that macrozooplankton has been shown to be very important in high latitude systems (e.g. Le Quéré et al., Biogeosciences, 2016; Hashioka et al. 201X).

the range of Grazing Pressure assumed in surveyed CMIP6 models spans 84x – 84x what? Revise.

substantially less Grazing Pressure than would be expected from an ocean populated exclusively with slow-grazing crustacean mesozooplankton. – why is this your reference here? Role of macrozooplankton? Also, reference needed for mesozooplankton growth rates.

all of which assume Grazing Pressure similar or greater than would be expected from an ocean populated exclusively with fast-grazing microzooplankton. – As above. References needed.

models, there does not appear to be a relationship between Grazing Pressure and model complexity. – Claim not backed up by evidence. And even if this were true for different models, what about the same model run with different trophic complexity? Complexity has been shown to affect food web structure substantially, see Le Quéré et al., 2016, Prowse et al. 2012, Vallina et al. 2014. Can it be that you don't see this in your metric, because models are highly tuned to get NPP and export right, regardless of their structure? And if so, what would that imply for your statement? In my experience, the addition of prey species decreases grazing pressure in these models (unless KTW is used, see Vallina and Le Quéré, 20XX), and the addition of higher trophic levels strongly affects biomass partitioning in the lower trophic levels (Prowse, Vallina), with tri-trophic interactions strongly altering zooplankton grazing pressure across all zooplankton functional groups represented (Buitenhuis et al., 2019). While I do not argue that the

total grazing pressure, as you calculate it, here may not depend on complexity, as far as you can diagnose it from model with a generally very low PFT complexity, additional complexity has been shown to affect, e.g., surface chlorophyll-a and NPP (Le Quéré et al. 2016).

everything from slow-grazing krill to fast-grazing ciliates – krill is usually considered macrozooplankton

satellites, it is unclear what the isolated effect of such dramatic differences in the assumed grazing behavior is on marine carbon cycling. – No. See Laufkötter et al. 2015, 2016. And these models aren't just tuned to fit NPP, they are also tuned to fit EP.

global, mean annual NPP (NPP; Fig 1B - green) declines rapidly across a narrow range of plausible Grazing Pressure - Well, what is plausible here? Fig. 1A shows a very wide range?

Increasing Grazing Pressure monotonically decreases NPP from a state of complete nutrient utilization to nearly zero. - Absolutely unsurprising. If the model didn't do this, there would be a mistake in your code...

maximum that is determined by the 'bottom-up' parameters that describe phytoplankton growth and is largely invariant to the grazing formulation. – As above. This is exactly as it should be!

S4K), leading to high depth-integrated phytoplankton biomass – Obvious.

Pressure, nutrient stress is entirely relieved but NPP approaches zero. – Completely obvious from the equations. What is your target audience? Someone who has never worked with such a model before?

The decline in export efficiency is driven primarily by a non-linear reduction in particle aggregation, which scales quadratically with each biomass pool (Fig. S3), such that less biomass sinks less efficiently. – In. Your. Model. Please have a look at Laufkötter et al., 2016 as to how other models funnel carbon through the food web. This is not the case for all models, and does not depend on the grazing pressure solely. What about the ballast effect, btw? Has been shown to strongly modify particle export patterns, see e.g. Nissen et al. 202X.

General comment here: Why focus on your own model here only, when we know how export works for a range of the other models already?

If Grazing Pressure is too low, zooplankton starve because grazing rates are too slow to sustain their biomass. By contrast, if Grazing Pressure is too high, zooplankton starve because overgrazing depletes the food supply before zooplankton biomass

has an opportunity to accumulate. Similarly, - Again it feels like you are just explaining the model equations to us, not your findings.

contour), regionally integrated NPP, EP, and GSP all behave similarly to the global mean (Fig. 2D, H, - strange, since it is usually the largest surface areas that determine the global mean? And what we usually see in these models in bioregionalisation attempts is a very clear separation between tropical and extra-tropical regions, that appears in marine ecosystem structure, but also global biogeochemical cycling (consult the many papers on bioregionalisation).

to match observed NPP, which has been estimated from remote

sensing at roughly 44–67 PgC/yr (42). – see newer estimates in Buitenhuis et al., Global Biogeochemical Cycles, 2013.

to match observed NPP, which has been estimated from remote sensing at roughly 44–67 PgC/yr (42). – As above. 84x what? And redundant with above.

This is less a reflection of any particular model's skill, but rather a testament to the extreme sensitivity of carbon cycling to grazing, which remains poorly constrained by observations. - Too easy – please cite relevant work. Have a look at Barton et al., Ecology Letters, 2013 and references therein, e.g. Hansen et al., 1997. And look at the seminal works of Calbet, L&O, 2001, Calbet and Landry, L&O, 2004, as well as Steinberg and Landry, Ann Rev. Mar Sci., 2017 and Landry & Calbet, ICES Journal, 2004.

pathways, the internal dynamics can yield substantial differences in the efficiency of carbon subsequently transported to depth and higher trophic levels. – Yes, but export is usually also tuned.

Methods”) and found the sensitivity of all variables to Grazing Pressure increases dramatically when a type II response is employed – see here also Nissen et al., Biogeosciences, 2018.

lines). For example, the collapse in NPP from 40 to 10 PgC/yr requires a roughly 32x change in Grazing Pressure when using a type III response, but only a 6x change when using a type II. – 32x and 6x what? Reference?

simulations (45), suggesting that predictions of future ocean carbon sequestration, artificially enhanced or not, are extremely sensitive to how fast we assume zooplankton are grazing. – Yes, but we also tune export. You did not.

We did not explicitly simulate any change in ocean temperature; however, another way to interpret these results is as the isolated effect of temperature-modified grazing rates on carbon cycling in a warming ocean. – What? Rephrase.

by roughly two-fold for every 10°C of warming (i.e. $Q_{10} = 2$, see (8)) – Yes, but models now use size-specific Q_{10} functions, see Le Quéré et al., Biogeosciences, 2016.

by roughly two-fold for every 10°C of warming (i.e. $Q_{10} = 2$, see (8)). Moving across an intermediate range of values, a 2x increase in Grazing Pressure from $\log_2(g(P)) = -4$ to -3 leads to a 7.6, 1.6, and 1.83 PgC/yr drop in NPP, EP, and GSP, respectively. Not sure I understand this correctly. Why would you assume a ten (!) degree warming here for this estimate? What is your ‘intermediate range’? Be specific.

translates to a 2.7%, 2.9%, and 1.5% decline in productivity per degree of warming, - Not really valid, as many other metabolic rates also scale with the same Q_{10} , and among these, phytoplankton growth is one of them. Caveat this finding.

perspective, that is nearly another PgC/yr that would need to be removed to counteract the feedback of warming on zooplankton grazing rates by – I don’t buy this, since phytoplankton growth scales with similar temperature dependencies.

From a fisheries perspective, GSP declines by 1.5%/°C despite a 1.45%/°C increase in ecosystem efficiency. – Same here – what are the assumptions here? That only grazing scales with temperature? We know that this isn’t true. And lack of precision in how calculations are described.

The whole climate change topic also has data-based hypotheses to be considered: Brun et al., Nature Ecol. Evol., 2019, for example. Please include in your discussion.

Crustacean mesozooplankton dominant current estimates of global zooplankton biomass (22, 56), but they are not the only important functional group – dominate instead of dominant, important group for what? And please cite relevant work such as Buitenhuis et al., ESSD, 2013 and Le Quéré et al., Global Change Biology, 2005.

Therefore, it is reasonable to consider that the best first order estimates of global mean Grazing Pressure in models are those that fall between what is predicted empirically for an ocean of relatively slow-grazing mesozooplankton and faster-growing microzooplankton, i.e. a mix of meso- and microzooplankton. – This seems obvious. But can you cite the relevant sources? Evidence on relative biomasses of different zooplankton functional groups (Bar-On et al., Science, 2018; Buitenhuis et al., ESSD, 2013)?

may be adequately resolving the variability and the associated functional diversity therein. – Cite appropriate references, e.g. Benedetti et al., Ecography, 2022, Brun et al. 2016 (data base), Barton et al. Ecology Letters, 2013 and so on.

Mesozooplankton, which generally graze much slower than microzooplankton (Fig. 1A), are able to consume much larger prey, such as dinoflagellates and diatom assemblages - What? Many copepod species are omni- or carnivores, see Bendetti et al. 2015, Benedetti et al., Ecography, 2022.

In turn, mesozooplankton and microzooplankton tend to dominate different niches, thriving in more and less productive biomes, respectively (58, 59). – Bad referencing. And is this true? Check the (biomass) data to support this claim.

This plays an critical role in regulating the biological pump but is rarely resolved (55). – NEMURO had DVM a decade ago. See also Aumont et al., Global Biogeochemical Cycles, 2018, Bianchi et al., Global Biogeochemical Cycles 2013 and multiple others.

Other groups could be equally important to resolve. – Cite community recommendations.

For example, larvaceans and salps are extremely abundant mesozooplankton that can filter bacteria and picophytoplankton, unavailable to crustaceans, and generally grow an order of magnitude faster than crustacean zooplankton – But then look at de Vargas et al., Science, 2016 and Guidi et al., Nature, 2016. Lots of other and unexpected groups associated with carbon cycle dynamics, e.g. rhizaria.

In turn, improving the representation of zooplankton grazing could result in dramatic improvements in marine BGC models and our predictions of future ocean states. – Agree, but you are not the first to diagnose this. Cite relevant references.

As stratification under climate change likely leads to greater light availability but fewer nutrients, the phytoplankton community may shift toward smaller phytoplankton with faster nutrient uptake but higher light requirements (53). – Already found in Bopp et al., GRL, 2005 (!).

but we believe further attention to grazing could substantively improve simulated carbon cycling and projections of both future climate scenarios and global food production. – Cite relevant references (see above). And redundant with previous.

With increasing computational capacity, recent BGC model development has largely focused on expanding the complexity of the phytoplankton community (65). – I disagree, see Buitenhuis et al., Global Biogeochemical Cycles, 2019, Henson et al., Science Advances, 2021 (see model structure in supplement), Clerc et al. 2021 and many more.

However, the functional traits and population dynamics of the mean phytoplankton community may actually be more sensitivity to variability in how they are grazed, rather than how they grow (66) – cite Hashioka, Nissen, multiple papers by Vallina, Prowe and many more! This has been shown!

Methods

S3. We did not use more a complex model, with multiple plankton pools or preferential feeding, as we believe that we must first better understand the impact of grazing in a relatively simple global BGC model. – This, in my view is a flawed decision, probably based on a lack of overview over the literature, as highlighted in my general comment above. We know a lot more about the effect of zooplankton than what you suggest in these models, and you are not taking into account what we know to a sufficient degree. A more thorough literature search will allow you to revise the focus of your work, to better place your paper into the context of the current literature, and to identify the elements of novelty that your study brings in a more convincing way.

Finally, note that WOMBAT quantifies biomass in nitrogen units and uses disk parameters to describe the functional response – I do not know what disk parameters are? Reference needed.

and then from Michaelis-Menten parameters into disk parameters (e – Lack of equations and references.

by a model, we define Grazing Pressure, $g(P)$ (d^{-1}), as the mean rate at which phytoplankton biomass would be lost to grazing per unit zooplankton in a standardized prey field. - this is an extremely poor description of the main variable of your entire work. What are you averaging over? Surface only or depth-integrated? Annual or monthly? Which years and time period, or single year? Resolution?

The key is describing the standardized prey field such that it remains as standardized as possible across different model configurations, with different numbers of phytoplankton, zooplankton and prey groups (see Table 1). To start, we assume a constant, standardized, mean, global, phytoplankton concentration, P $mmolC/m^3$, which is inclusive of all integrated phytoplankton groups in a given model. – as above. – First sentence is an tautology. And again, what are you calculating? Estimates made per model grid point? Annual? Monthly or daily average?

As such, P must be chosen to be generally representative of the global mean phytoplankton concentration that might be expected to occur across all models being compared. - So what was chosen? And I disagree – the global mean isn't representative. You have oligotrophic and eutrophic regions, with each covering different fractions of the global surface ocean. And in the mid-latitudes, you have huge seasonal differences. So how can this one number be representative of anything. Uncertainty analysis/range bars for real open ocean conditions would be helpful, and may convince me that your metric is actually useful to describe CMIP models that attempt to describe real marine ecosystems.

To take into account multiple zooplankton, Grazing Pressure is averaged across all zooplankton groups, which are assumed to exist in equal proportions. Why? Differences between the zooplankton groups are vital for global biogeochemical cycling (Si, N, $CaCO_3$, DMS,...). And these groups do not exist in similar proportions. Why don't you use the actual equations and simulated biomass ratios at each grid point to study the grazing pressure across models?

as the mean of two end-member scenarios, both with the same standardized, mean, global, phytoplankton concentration as in the simple scenario described above, but now with two different distributions. – The mean of two very different scenarios is statistically not very telling, and the mean scenario across ensemble members with totally different assumptions is likely to be neither representative of one, nor of the other scenario, and, more importantly, not representative for the real ocean either. This seems nonsensical to be. Need a far more exhaustive supplementary to back up your methodological decisions.

In models where non-plankton prey (i.e. detritus, bacteria) is available, all other prey types are assumed to exist in equal concentrations to phytoplankton, that is P . – Arbitrary and utterly unmotivated assumption. You could have chosen these to be Z . And which models and which non-plankton prey species did you encounter in your work?

Neither distribution is realistic, but together they offer useful bounds on how fast zooplankton could be expected to graze down phytoplankton in given model, without having to integrate it. – If they offer bounds, and both aren't realistic, why (a) were they chosen, and (b) why would you average them? And what don't you want to 'integrate'? Need to motivate, and justify and test and evaluate your methodological choices.

That is, at a given location we assume one phytoplankton and one zooplankton have out competed all other plankton groups, driving them to extinction. – Essentially, it is exactly this behaviour that any such model I have seen tries to avoid at all costs... I am at a loss why this would be your underlying assumption? This is a complete simplification of the food-web, with known impacts on biomass dynamics, see references above..

That is, the concentrations of the dominant phytoplankton, as well as any non-plankton prey options, are equal to P , while all other phytoplankton or zooplankton concentrations are assumed equal to 0. – Why? Then, it could be that grazing pressure is assessed based on grazing on detritus and one prey species alone? How is that representative of real ocean conditions?

In both scenarios, to take into account model differences in temperature limitation Grazing Pressure is computed at 16°C, roughly the global, pre-industrial mean sea surface temperature (79). - Surface ocean mean temperature is 16degC? Which depth range was taken? Why? Zooplankton live in upper few 100 m of the ocean, phytoplankton mostly in upper 100 m... so which temperature is relevant here? Justify your assumptions.

Here, we set P equal to the mean phytoplankton concentration simulated across all simulations conducted in our study, $P = 0.56 \text{ mmolC/m}^3$. – What environment is this representative for?

This is appropriate for comparing the influence of Grazing Pressure between our runs; however, we note that this value is not necessarily representative of the mean phytoplankton concentration in other simulations or the values observed from space. – So. Why. Was. A. Nonsensical. Value. Chosen???

These values were calculated using roughly 18 years (July 2002 - April 2021) of data from the Carbon based Productivity Model (42). - Why this model? Justify your choices.

S2). It is important to reiterate that the absolute value Grazing Pressure for a given model is of little consequence and necessarily varies with the value of P. – The repeated statements that you methodological choices ‘do not matter’ and ‘have little consequence’ and ‘aren’t supposed to represent the real world’ (see throughout manuscript) do not increase my confidence in your analysis. To me, this means that perhaps you haven’t sufficiently tested this and you lack the ambition to understand how real system works, since very little robustness analysis is presented in your supplement, and you also state repeatedly that your findings aren’t relevant for real-world problems. Yet, you claim we learn more about CMIP models and their capacity to simulate future carbon cycle characteristics.

Fig 1: Figure must be stand-alone. Revise figure caption and color schemes. We do not know how the empirical estimates have been calculated (need to describe this in your Methods). We do not know what was measured, not what was simulated, and neither by which model. B must tell us what we simulated by which model(s). Colors unsuitable for the many people affected by red-green blindness. And I would like to see variability envelopes around each of your dots.

Fig. 2: Figure must be stand-alone. Revise figure caption and color schemes. Which models are you using for this figure? Same issue of red-green blindness unfriendly color scheme.

Fig. 3: Why are some axis titles written in red color? This looks like a supplementary figure to me. Color scheme is better.

Fig. 4: Again, state which model this figure is based on. Revise color scheme.

Table 1: Very nice. Revise color scheme. See colorbrewer.org.

Additional figs: I’d like to see a schematic that summarises the relevance of your findings with WOMBAT for CMIP models and future projections.

Reviewer #2 (Remarks to the Author):

The authors present an interesting manuscript on an important topic. I find their discussion of the differences between grazing parameterizations in IPCC-class models to be enlightening. However, I think

there are substantial issues with the authors' attempt to simplify the different models into a single metric. As I describe below, I do not find this metric useful. Instead, I think it is far more important to focus on the functional responses of zooplankton within each model. For instance, extensive research has shown how the fundamental behavior of models can change with switches between Type II and Type III grazing, switches between Holling's Disk and Ivlev grazing formulations, inclusions of additional trophic levels, and switches between zooplankton losses mediated primarily by linear or quadratic closure terms. These topics are all likely to fundamentally alter how zooplankton respond to changing bottom-up forcing, and hence will alter the predicted responses of the marine ecosystem to climate change. Differences of $g(P)$, however, mostly reflect differences in zooplankton biomass, since most of the models have tuned bulk grazing to approximately match NPP. I elaborate on this issue below. However, I do think there is a lot of merit in the authors' broader analysis of differences between models.

Fundamentally, what matters is the functional response of zooplankton to changing phytoplankton concentrations (and in this case "functional response" depends on the interests of the study and likely will include either bulk grazing pressure, zooplankton secondary production, or carbon export efficiency). Unfortunately, the authors' $g(P)$ metric gives us no information about this response. To a certain extent, this issue is really highlighted by the authors' analysis with the WOMBAT model. Within the WOMBAT model, increasing $g(P)$ to values found in other model's causes NPP to crater. This is, of course, because WOMBAT is tuned to behave as if the single zooplankton is a metazoan zooplankton (krill or copepod). As such, it likely obtains a large biomass in the model, with slow turnover rates (turnover rates that are likely modulated mostly by the zooplankton mortality term and/or the zooplankton respiration term). These closure terms are low and hence WOMBAT has high zooplankton biomass leading to low $g(P)$. In another model in which the single zooplankton state variable is tuned to act like a protist, a much higher mortality rate would be prescribed, leading to more rapid turnover, lower biomass and hence a higher $g(P)$ necessary to balance NPP. This is a very simple way of looking at things, but it helps to clarify two issues: first, it is bulk grazing pressure (an emergent property of both $g(P)$ and zooplankton biomass) that matters for NPP and second the functional response of zooplankton is controlled equally by its grazing rate AND its mortality rate – considering one without considering the other is not particularly helpful. Issues with using $g(P)$ are further exacerbated when there are multiple zooplankton. The term $g(P)$ really loses meaning when multiple zooplankton trophic levels are present. Consider a model with a single (protistan-like) zooplankton. It will have a high $g(P)$. Now consider a second model that builds on the first by adding an herbivorous mesozooplankton (e.g., a small copepod). Most of the grazing in the model should still be mediated by the protist (70% of NPP is the canonical global value), while the copepod feeds on roughly 15% of NPP. However, the copepod, with its slow turnover rates, will have much higher biomass than the protist. Hence the global $g(P)$ will be much lower, because bulk grazing will remain similar but total zooplankton biomass will be much higher. Nevertheless, the model will probably still exhibit fairly similar grazing behavior, because most of the grazing is still mediated by that rapid protist. Next consider a third model that adds an omnivorous macrozooplankton (e.g., a large copepod or krill). The protist still does most of the grazing, the small copepod consumes ~10% of NPP, the large copepod consumed ~5%, but the long-lived large copepod adds substantially more biomass and $g(P)$ drops again. These models would all behave in functionally similar ways with respect to top-down impacts on NPP, although $g(P)$ would be very different in the

three models. Furthermore, the fact that the authors do not actually calculate $g(P)$ based on emergent biomasses and grazing rates, but instead calculate it based on assuming equal biomasses of all grazers is not at all realistic.

Other comments:

The manuscript had many supplemental figures that were frequently referenced. I found this distracting, because I frequently had to look to the supplement. I would recommend including more of these figures in the manuscript.

Section 2.5: Description of WOMBAT: Authors state that WOMBAT grazing pressure can be modified by three variables, but then only tell us what two of them are.

Within the text and Supp. Fig. 2 the authors use different names for the same variables, which is confusing.

While I like Fig. S2, the equations should be simplified. I spent a lot of time going through these equations (and substituting in multiple variables for each other) only to find in the end that they are mostly very standard equations (e.g., GSP is a very typical formulation, but that is not clear without working through the math). There are other issues as well, such as the fact that variables are sometimes written out in strange ways (for instance, it took me a little while to realize that 'gambio' was just a variable that is probably normally written as Greek letter gamma subscript bio). Also, in the zooplankton equation, mortality is written as mZ so I assumed it was a linear mortality term with parameter m as the daily mortality rate. About an hour later, I realized that apparently the authors were using the term mZ as a term meaning zooplankton mortality that was defined in the export production section as $m_{\text{uezbio}}[Z]^2$ (i.e., quadratic mortality). I think I now understand this, but why didn't they simply write in both the zooplankton and detritus equations: $m[Z]^2$ so that it would be clear and easily readable? These issues are present throughout the equations and add substantial confusion to what is otherwise a nice figure.

While I like most of the figures, there are two things that can be improved. First the logarithmic axes would be much more useful if instead of giving the log-transformed values they gave the actual values. For instance, in Figure 1, the x-axes would be far more readable if instead of -9, -8, -7, -6, etc. it was 0.002, 0.004, 0.008, 0.0156, etc. Also, it would be helpful if they had actual legends. Having legends that are kind of hidden in the y-axes was unnecessarily confusing for me, since there is room on the figures for a proper legend.

Reviewer #3 (Remarks to the Author):

Dear authors,

I have now reviewed the submitted manuscript.

Overall the manuscript is well written, interesting, timely but may lack essential aspects dealing with zooplankton dynamics and its representation in models.

The strength and originality of the work is to propose a unified methodology to inter-compare the impact of modifying zooplankton grazing on BGC models and to compare different formulations and parametrization on export flux and trophic transfer by running simulation in one of the simplest models.

BGC models are already further than CMIP model and they should be integrated in this article.

Communications Earth & Environment is indeed a journal where the most advanced studies should be addressed.

I would suggest major revision and hope for the authors to address the main comments.

The information that zooplankton is systematically not well represented in BGC models is known for decades and so is not a surprise. Solving this issue is an important aspect in many modelling frameworks.

Many articles have been describing the gaps and propose solutions. They all stress the possible impact on food web transfer and export efficiency. Some new developments are existing but the authors do not report about them in the same way they do for the CMIP.

The different CMIP models have been intercompared for different biological variables including zooplankton. For example in Petrick et al., 2022 (cited by the authors), the authors demonstrate the sensitivity of zooplankton in different models by using a Taylor representation.

Using the same framework in addition to their analysis would allow to intercompare based on similar methods. This would be possible in the second section of the articles where the authors run simulation.

The weak part is that the authors do not include in their quantitative analysis and discuss enough the already existing models that have complex zooplankton representation (for example in size and taxa) and are already improvement relative to CMIP models.

The prey field of zooplankton is not only phytoplankton but also POC and zooplankton itself. Filter feeding is not the only mechanisms and there is prey selection. Some models have included it but are not used in the quantitative approach.

for example:

Ward et al., J. Plankton Res. (2013) 0(0): 1–17. doi:10.1093/plankt/fbt097

Heneghan et al., 2020 Ecological modeling <https://doi.org/10.1016/j.ecolmodel.2020.109265>

Gehlen et al., 2006: Biogeosciences, 3, 521–537,

Including these models in their quantitative analysis of sensitivity would really make this paper up to date and would inform about existing improvement. It would also quantify the improvement (or not?)

The current version of the manuscript do not really propose and discuss future evolution. They do not really address the possibility of joint co development in observation and modeling. The paper by J. D. Everett, et al., Frontiers in Marine Science 4, 77 (2017) is cited but only in the beginning.

Finally, recent observations have shown that Rhizarians (Biard et al., 2016, Nature Letter, doi:10.1038/nature17652) are important biomass in the ocean and their inclusion in future BGC models should be taken into consideration.

I have also minor comments:

- I noticed minor typos. Should be checked as I'm not a native english speaking person.

- I noticed an inconsistency in bibliographic formatting:

for example reference 58

REVIEWER COMMENTS:

Reviewer #1 (Remarks to the Author):

First and foremost, we would like to sincerely thank you for your helpful and stimulating comments. In particular, your suggestion to compute and investigate offline grazing dynamics has vastly improved the manuscript. See below for our point-by-point response to your specific comments and note we have numbered your comments for reference in our response.

General comments

1. In this interesting paper, Rohr et al. describe the incredible sensitivity of global primary production, secondary production and export production to a derived, theoretical variable they call the ‘Grazing Pressure’ in a range of CMIP6 models. Having carefully read this work, I am torn with regard to the recommendation I am expected to give. On the one hand, I think the message with regard to the completely underestimated importance of the higher trophic levels and top-down controls on carbon fluxes in these models is very important, and also supported by recent observational evidence (e.g. the recent metagenomic information; Guidi et al., Nature, 2016). The traditional bottom-up view of marine ecosystems in the classical biogeochemistry modeling community is no longer tenable, and must be revisited. On the other hand, the marine ecosystem modeling community has known this for years, if not decades, and every modeler who has ever tuned a model has made the experience that tinkering around with the grazing pressure changes everything in these PFT-type models. Hence, the findings presented in this work are not entirely novel, nor very surprising, and could equally well be published in a more topical journal. Regardless of this strategic decision, the paper needs major revisions before it can be accepted.

We have revised the manuscript to include a detailed analysis of emergent grazing dynamics in CMIP6 models and their relationship with the Prescribed Grazing Index (our revised ‘theoretical metric’ -- previously called ‘Grazing Pressure’). In doing so we believe we have presented more novel results, defended them more meticulously, and incorporated them into the existing literature more thoroughly. Briefly, we have:

1. Incorporated an extensive analysis of the emergent grazing dynamics (Table 1) and emergent zooplankton mortality rates (Table S3) calculated offline, pixel by pixel, using the monthly mean model output from historical simulations of 11 CMIP6 models (see **Section 2.3, 2.4, & 4.5**).
2. Shown quantitatively that the largest source of uncertainty in marine carbon cycling across CMIP6 models appears to be emergent grazing pressure, which now refers to the emergent phytoplankton specific loss rate to grazing (i.e. $g(P)Z/P$), not our diagnostic metric (see **Section 2.3**). This key result is now reflected in the revised title.
3. Shown that large inter-model variability in emergent grazing pressure is driven by large variability in emergent specific grazing rates, which is not tuned out via compensating variability in each model’s closure scheme (see **Section 2.3**)
4. Refined the definition and description of our metric, now called the ‘Prescribed Grazing Index’ (*PGI*). Notably, this innate property of a model’s grazing formulation is now evaluated

using the global, median observed plankton, bacteria and POC distributions rather than the unrealistic end-member scenario previously employed (see **Section 2.2, 2.4, 4.2**).

5. **Importantly**, shown that despite reducing many dimensions of variability and important qualitative differences in food web structure, prey refuge, and active/passive switching, into a single metric, the Prescribed Grazing Index is a strong predictor of emergent grazing dynamics. We show that the Prescribed Grazing Index is well correlated with not only inter-model variability in emergent specific grazing rates, but also grazing pressure at global, regional, and seasonal scales (see **Section 2.4.1**).
6. Motivated the WOMBAT sensitivity study (which now is a smaller part of the manuscript) with the established relationship between in the Prescribed Grazing Index and emergent grazing dynamics (see **Section 2.4.2**).
7. Throughout the manuscript, incorporated more descriptions of spatio/temporal variability, now including some form of either zonal, regional or seasonal variability in **Figures 2-7, S3-S6, and S9**. In particular, **Figures 4 and S3-6** highlight the emergent zonal and seasonal variability in marine carbon cycling across models (see **Section 2.3.2**).
8. Throughout the manuscript leveraged more of the observational record to include derived global and zonal observational estimates, with uncertainty bounds, for comparison with of all model variables (see **Section 4.4**).
9. Throughout the manuscript, weaved a much richer conversation with the existing modelling and observational literature, and incorporated all suggested references.

2. In my view, several major flaws need to be resolved before this work can be published here or elsewhere. These mostly affect the presentation and justification of the methodology and results, and not the actual conclusions, which I think are sound and in line with what we know about these models.

First and foremost, the paper seems oblivious to the current state-of-the-art in the current marine ecosystem modeling community with regard to the role of zooplankton for marine ecosystem dynamics and the regulation of derived carbon cycle variables such as NPP, phytoplankton biomass dynamics, or export. It seems to neglect a whole decade of literature on this topic, starting from the work of Sailley et al., Ecological Modelling, 2013, who already examined the differences in top-down pressures and food-web structure between CMIP-type models and came to the absolute same conclusions that the current authors do (that food-webs differ (even intended ones from those effectively realized) and that future projections of the carbon cycle variables are shaky at best) to the work by Vallina et al., Prog. Oceanog., 2014 or Prowe et al., Prog. Oceanog., 2012 on the difference of different grazing formulations on phytoplankton biomass and diversity (which both are closely linked to NPP), nor is it discussing the work by Le Quéré et al., Biogeosciences, 2016 on the importance of zooplankton community structure on Southern Ocean NPP (which claims we need multiple higher trophic levels and trophic cascades to correctly represent NPP and biomass dynamics). Further important recent work discussing the importance of grazing in structuring simulated marine ecosystems isn't cited either (e.g. Dutkiewicz et al., Biogeosciences, 2020; Chenillat et al., PLOS One, 2021), and the authors also haven't really fully exploited the potential of the work of Laufkötter et al., Biogeosciences, 2015, 2016 to strengthen their case, with both of the latter papers having made a relatively refined attempt at tracing carbon across the food-web and quantifying export efficiency, with important consequences for the work presented here. It also does not compare its findings to important conceptual work on differences in food-web structure between different oceanic environments (Everett et al., 2022), nor has it taken full advantage of another paper which I think it cites (citation truncated; Everett et al., Front. Mar. Scie., 2017).

Changed. We have dramatically overhauled the manuscript to discuss these (and many other) pertinent studies. Please see our response to specific comments below for many examples of their inclusion.

3. It also neglects an absolute wealth of zooplankton observations to back up its claims, starting from novel approaches to detect zooplankton biomass from space (Basedow et al., Scientific Reports, 2019; Druon et al., Scientific Reports, 2019; Strömberg et al., Journal of Marine Systems, 2009), to extensive zooplankton biomass and trait observations that modelers use to validate their models (Buitenhuis et al., ESSD, 2013 on microzooplankton, O'Brien and Moriarty, ESSD, 2013 on mesozooplankton; Moriarty et al., ESSD 2013 on macrozooplankton and plenty of literature on the functional biogeography of zooplankton (multiple papers by P. Brun and F. Benedetti, as well as Barton et al., Ecology Letters, 2013)). This means that the paper currently fails to identify and highlight the elements of novelty in their work to a sufficient degree.

Changed. Thank you for this comment. We have incorporated each of these (and other) observational works into the Results and Discussion of the revised manuscript. Further, we have leveraged them for explicit analysis (see Section 4.4). For example, we now use the MAREDAT database (e.g. Buitenhuis et al., 2013; O'Brien and Moriarty, 2013; Moriarty et al., 2013) to not only refine a more realistic standardized prey field with which to evaluate the *PGI* (see response to **Comment 4** below), but also derive global and zonal estimates of biomass with which to compare the newly included emergent model output (see immediately below).

Line 729: “Mean annual, globally-integrated biomass estimates (Figure 5A-D) were computed entirely from the MAREDAT dataset (Buitenhuis et al., 2013). First, for each phytoplankton and zooplankton PFT considered in the standardized prey field described above (Buitenhuis et al., 2012a; 2012b; O'Brien et al., 2012; Moriarty et al., 2012; Leblanc et al., 2012; Vogt et al., 2012) we computed the mean annual carbon biomass inventory by constructing global, monthly biomass profiles (0-200 m), then integrating in depth, averaging in time, and multiplying by the global ocean area between 60S-60N. Note, very high latitudes were excluded here and for model output to avoid regions where models tend to perform poorly and observations are sparse. Monthly biomass profiles were created by computing the median biomass concentration at each depth horizon, with uncertainty bounds approximated using the 25th and 75th percentiles. We use the median +/-25% rather than the mean +/- 1 standard deviation to estimate the global carbon inventory because field experiments, particularly those specifically targeting blooms, are typically designed to sample where plankton are more abundant, likely creating a larger positive bias in the mean. Uncertainty bounds reflect the middle 50% of the MAREDAT data to be consistent with uncertainty estimation of grazing pressure by \cite{schmokerMicrozooplanktonGrazingOceans2013} (see below). Mean annual phytoplankton biomass (0.59 *PgC* (0.37, 1.25); Figure 5E), zooplankton biomass (0.33 *PgC* (0.20, 0.78); Figure 5F), total phytoplankton and zooplankton biomass (0.92 *PgC* (0.66, 1.7); Figure 5G), and the ratio of zooplankton to phytoplankton biomass (0.57 (0.25, 1.6); Figure 5H) are then computed by operating on the mean annual global PFT inventories and propagating uncertainty bounds accordingly.”

Line 729: “Zonally resolved observational estimates of grazing pressure (Figure 4B), zooplankton specific grazing rates (Figure 4E), zooplankton biomass (Figure 4H), phytoplankton specific division rates (Figure S4E), phytoplankton biomass (Figure S4H), bulk phytoplankton grazing mortality (Figure S4B), total plankton biomass (Figure S6B) and the ratio of zooplankton to phytoplankton (Figure S6E) were computed in a similar manner to the global estimates described above. Following Schmoker et al. (2013), zonal bins were partitioned into the northern high latitudes (>60°N), northern westerlies (30-60°N), tropics (30°S - 30°N), southern westerlies and southern high latitudes (<60°S). No explicit seasonal delineations were made; however, the vast majority of field studies included in Schmoker et al. (2013) and Buitenhuis et al. (2013) were conducted outside of the winter months, so observational estimates are thus considered more representative of “summer”, and plotted accordingly.

First, zonal distributions of biomass for each PFT (i.e Figure 3 in Buitenhuis et al. (2013)) were downloaded from the MAREDAT Pangea data base (Buitenhuis et al., 2012a; 2012b; O'Brien et al., 2012; Moriarty et al., 2012; Leblanc et al., 2012; Vogt et al., 2012) and binned as described above. Similar to the global estimates, we found the median biomass for each PFT in each depth layer and zonal band, then integrated these profiles across depth and functional group to yield the median phytoplankton, zooplankton, and total plankton inventory. Again, uncertainty bounds were computed using the same method but for the 25th and 75th percentiles. The ratio of zooplankton to phytoplankton biomass is then computed for each zonal band with uncertainty propagated accordingly.”

4. Second, the methods section is poorly composed and many methodological choices remain ill justified and explained, so that it is difficult to judge whether the paper's new 'Grazing Pressure' metric is helpful. In my view, it would be fully possible to calculate the predator- and prey-specific and total grazing rates of the CMIP models offline on a pixel-by-pixel basis, based on the regular CMIP6 output, the model equations and some assumptions with regard to missing variables that weren't submitted to the CMIP6 archive (e.g. replacing missing submissions by observational fields, or ensemble means across all other models). On a monthly scale, you will usually get to between 10-20% within the actual value calculate by the model online with an offline calculation, which is quite good, and could be used to see how the true grazing pressure changes in space and time, and as a function of prey biomass. The authors were unable to convince me that their theoretical 'Grazing Pressure' variable, which uses very coarse averaging in space and time to produce variables that eventually do not reflect community structure and dynamics anywhere in the model domain, was better than the actual offline computation of zooplankton biomass, and the de facto grazing pressure in these models.

Changed. Thank you – this was an excellent suggestion and we have used it. We have heavily revised the manuscript to better refine, support and incorporate the 'Prescribed Grazing Index' (previously 'Grazing Pressure'). In doing so we have:

1. Renamed 'Grazing Pressure' to the 'Prescribed Grazing Index'.
2. Reformulated the definition and derivation of the Prescribed Grazing Index to better reflect realistic community structure by using observed global mean fields, rather than unrealistic endmember scenarios.
3. Rewritten the methods (and relevant results) to more clearly explain the value of the Prescribed Grazing Index as a metric to compare innate properties of the prescribed grazing formulation across models, rather than estimate emergent properties of the dynamical simulation.
4. Included a section on the 'Limitations and Sensitivity of the Prescribed Grazing Index'
5. **Importantly**, substantially revised the manuscript to focus on an extensive analysis of the emergent simulated zooplankton specific grazing rates (and many other variables), calculated offline, pixel by pixel, using the monthly mean model output, for all considered CMIP6 models, opposed to just WOMBAT.
6. Incorporated into our analysis the relationship between the Prescribed Grazing Index (an innate property) and the emergent properties of each simulation. We now show that the Prescribed Grazing Index is well correlated with emergent properties of the models grazing dynamics at most latitudes during summer and winter, confirming that it is a useful metric.
7. Examined how other emergent aspects of simulated marine carbon cycling vary (or do not vary) with differences in the Prescribed Grazing Index, both in CMIP6 models, and the sensitivity study in WOMBAT which was previously presented.
8. Finally, include in the SI, an analysis of how emergent grazing rates vary with total phytoplankton abundance (i.e. the Apparent Functional Response).

Points 1-5 are discussed in further detail below. Points 6-8 are discussed in further detail in response to other specific comments below.

1. 'Grazing Pressure' was a poorly selected name, as it does not refer to emergent 'grazing pressure' in the sense of Prowe et al., (2012) – i.e. the phytoplankton specific loss rate to grazing ($g(P)Z/P$) – and it is intentionally **not** meant to directly estimate any emergent

property of a given model run. Instead, it is meant to quantify a reduced approximation of the magnitude of grazing rates prescribed by the innate grazing formulation in each model. We have renamed the metric ‘Prescribed Grazing Index’ to better reflect this distinction.

2. To standardize the ‘Prescribed Grazing Index’ across models, and thus obviate other inter-model differences that drive phytoplankton dynamics, it is necessary to evaluate it in a standardized prey field. Instead of using the two, admittedly confusing, hypothetical end-member scenarios, we now evaluate the ‘Prescribed Grazing Index’ at the observed global median plankton distribution (see Table S1, and Section 4.2 and 4.4 below).

Table S1:

Field	MAREDAT Group	Observed Global Median Prey Field	Mean Modelled Global Median Prey Field	Observational Reference
Phytoplankton ($mmolC\ m^{-3}$)				
Small	Picophytoplankton	0.46	0.37	(28,154)
Large	Diatoms + Phaeocystis	0.33	0.24	(28,155,158)
Diazatrophs	Diazatrophs	2.5e-3	0.02	(28,157)
Total	Pico+Diat +Phae+Diat	0.79	0.61	(28,154-158)
Zooplankton ($mmolC\ m^{-3}$)				
Small	Microzooplankton	0.26	0.17	(28,93)
Medium	Mesozooplankton	0.23	0.16	(28,94)
Large	Macrozooplankton	0.02	0.04	(28,154)
Total	Micro+Meso+Macro	0.51	0.34	(28,93,94,154)
Other Carbon Reservoirs ($mmolC\ m^{-3}$)				
Bacteria	Picoheterotrophs	0.55	0.57	(28,151)
Detritus	-	2.65	1.12	(28,150)
Temperature ($^{\circ}C$)				
SST	-	15.30	15.61	(149)

3. We have rewritten the Methods to better explain the metric:

Section 4.1 Line 541: “The Prescribed Grazing Index (*PGI*) is computed as the zooplankton specific grazing rate, integrated across all phytoplankton prey options, prescribed by a given grazing formulation (Table 1) when evaluated at the observed global median phytoplankton and zooplankton plankton distribution and long-term mean sea surface temperature. In models with multiple zPFTs, zooplankton are assumed to be partitioned based on the observed global median distribution (Table S1) and the *PGI* is their biomass-weighted mean specific grazing rate on all phytoplankton prey options. When non-phytoplankton prey is included in the food-web (i.e. bacteria, detritus, or other zooplankton), it is accounted for in so much as it offers refuge for phytoplankton by diverting grazing effort. Mathematically, this is equivalent to evaluating Equation 3, which is used to compute the offline emergent dynamics (see Section 4.5), except here a standardized prey field is used instead of the emergent monthly mean biomass fields. By using a standardized prey field, we can obviate myriad other inter-model differences that drive the emergent distribution and abundance of prey and thus quantify a reduced approximation of the magnitude of grazing rates prescribed by the innate grazing formulation in each model. Accordingly, the *PGI* does not necessarily reflect the emergent grazing dynamics of a given simulation, which further depend on the formulation of bottom-up controls, physics, zooplankton mortality rates, and other aspects of the grazing formulation that cannot be captured with a single value.

The standardized prey field used to evaluate the *PGI* is based on the global median picophytoplankton, diatom, *Phaeocystis*, diazotroph, microzooplankton, mesozooplankton, macrozooplankton concentrations (Buitenhuis et al., 2013) and the global mean long-term average sea surface temperature from the NOAA ERSSTv5 product (Huang et al., 2017). The observed global median particulate organic carbon concentration (Stramski et al., 2022) and bacteria concentration from the upper 200 m (Buitenhuis et al., 2012) are included when available as prey. See ‘Description of Observations’ below. To account for differences in the structure of the food webs, some assumptions must be made about how biomass is partitioned in models with different numbers of PFTs. However, total phytoplankton and

zooplankton biomass is kept constant. In models with just one phytoplankton or zooplankton group, total phytoplankton and zooplankton biomass is used. In models without diazotrophs or macrozooplankton, those pools are added to the 'large phytoplankton' and 'medium zooplankton' pools, respectively. *Phaeocystis*, which no surveyed model includes but compete with diatoms for nutrients, particularly in the Southern Ocean (Nissen et al., 2021), are included in the 'large phytoplankton' pool.

Note that while these values are deliberately agnostic to the emergent dynamics of any specific model run, thereby providing a standardized field to obviate other inter-model differences and isolate innate properties of the grazing formulation, they compare well to the inter-model mean of the median simulated biomass fields (Table S1). In both models and observations, the concentration of biomass increases from total zooplankton to bacteria to total phytoplankton to detritus, and the relative distributions are generally consistent. Finally, the zooplankton biomass-weighted temperature from the models (i.e. the temperature the median simulated zooplankton experiences) is close to the standardized observed surface temperature used (15.30 and 15.61°C, respectively). Together, this confirms that the standardized observed values are representative of the median values simulated across models.

Section 4.4 Line 695: “The global median standardized prey field used in Prescribed Grazing Index is primarily based on field observations collected by from the MARine Ecosystem biomass DATa (MAREDAT) initiative. MAREDAT is a compilation of global, gridded biomass datasets for a range of PFTs and is described in detail in (Buitenhuis et al., 2013). Here, we use the reported global median epipelagic (0-200m) concentration of microzooplankton (Buitenhuis et al., 2012a) mesozooplankton (O’Brien et al., 2012), macrozooplankton (Moriarty et al., 2012), picophytoplankton (Buitenhuis et al., 2012b), diatoms (leblanc et al., 2012), diazotroph (Luo et al., 2013), *Phaeocystis* (Vogt et al., 2012) to describe the phytoplankton and zooplankton distribution and in models in which bacteria is available as prey, use global median picoheterotroph concentration (which also include Archaea) (Buitenhuis et al., 2012). Briefly, all MAREDAT data is gridded onto a 1° by 1° grid with 33 vertical layers and a climatologic monthly time step. Throughout, data has been quality controlled by contributing researchers and Chauvenet's criteria was used to omit anomalously high values. Note, calcifying PFTs (coccolithophores, foraminifers, and pteropods) are reported but were not included as no model explicitly represented them and their median global values are small relative to other PFTs.

Uncertainty in these global values is fairly large due to patchiness and sampling bias as discussed in detail in (Buitenhuis et al., 2013). However, estimates of total phytoplankton ($0.79 \text{ mmolC m}^{-3}$) are generally consistent with independent of estimates of the median phytoplankton biomass ($0.97 \text{ mmolC m}^{-3}$) estimated from 18 years (July 2002 - April 2021) of remote sensing data from the Carbon-based Productivity Model (Westberry et al., 2008). Note, in high latitude regions with summer remote sensing coverage but without winter coverage, winter time concentrations were set to 0, assuming that any overwintering biomass persisting through the darkness was trivial to the global median.

“Temperature and detritus are also needed to compute the *PGI* for some models, but are not included in the MAREDAT database. In models in which grazing is temperature-limited, rates are computed the global (-60S-60N) mean long-term average (1854-2022) sea surface temperature from the NOAA ERSSTv5 product (Huang et al., 2017). In models in which detritus is available as prey, detritus is estimated as the global median area-weighted particulate organic carbon (1-200 m) field observations reported in Table 1 of (Stramski et al., 2022) less the global median living biomass concentrations from Table S1.”

4. We have included, in the Methods, a refined section on the ‘Limitations and Sensitivity of the Prescribed Grazing Index’.

Line 583: “One limitation of the *PGI* is that it is not well suited to comparing models with much more trophic complexity than those studied here. That is because evaluating the *PGI* for a model with many more functional groups or an allometrically-scaled size-spectrum within functional groups would require many further assumptions about how the standardized prey field is partitioned across plankton groups. While some of these assumptions could potentially be made using global estimates of size-spectra (Everett et al., 2021), the *PGI* becomes a less interesting metric for models that explicitly resolve greater trophic complexity, as it is designed to reflect difference in how simpler models assume that complexity is averaged over and parameterized out.

A second, and larger limitation of the *PGI* is that it collapses many modes of variability into a single dimension. While this is intentional, and useful to quantitatively compare the integrated influence of multiple models' disparate grazing formulations along a single axis, it means that two grazing formulations could yield a different hierarchy of *PGI* if evaluated using a different standardized prey field. For example, even in the simplest case - a single zooplankton group grazing a single phytoplankton group - using a type II versus type III response function with the

same parameter values would yield a higher *PGI* if evaluated below $K_{1/2}$ and but a lower *PGI* if evaluated above $K_{1/2}$ (Figure S1).

This underscores the importance of selecting an appropriate standardized prey field. Using global median values rather than high phytoplankton concentrations that characterize large blooms (during which global NPP and EP disproportionately occur) is important for two reasons. First, almost all modern grazing formulations, and all of those surveyed here, saturate at high phytoplankton concentrations to simulate the satiation of predators once they become limited by prey consumption rates, rather than capture rates (Gentleman et al., 2008). In turn, if the goal is to capture as much variability as possible across different formulations, then they should be evaluated before they have begun to significantly saturate. Second, and more importantly, the dynamical system is more sensitive to the values of grazing at low phytoplankton concentrations than high ones (Rohr et al., 2022). Intuitively this makes sense as all high phytoplankton concentrations must first pass-through lower ones and thus cannot come to be if grazing rates are fast enough at low phytoplankton concentrations to stifle a bloom before it happens.

However, to test the sensitivity of the *PGI* to the selection of the standardized prey field we evaluated it using a total phytoplankton concentration ranging from 0.1-5.0 mmolC/m³, compared to the MAREDAT global median value of 0.79 mmolC/m³ (Figure S2A). In all estimations the relative distribution of biomass across groups was kept the same. Generally, models with a larger *PGI* also have more uncertainty when evaluated across the range of prey fields. This is because they typically have lower $K_{1/2}$ or higher $g_{\max}$ values, leading to larger direct interaction strengths at low to moderate prey concentrations, and thus larger changes to grazing with respect to changes in prey biomass (Sailley et al., 2013). In our result we express this uncertainty by including error bars representing the *PGI* evaluated at $\pm 25\%$ of the global median phytoplankton concentration, illustrating how *PGI* generally becomes less precise as it increases (Figures 2, 5, 6).

Still, when looking across the full distribution of *PGI* values (Figure S2A), the hierarchy of median values (horizontal black lines) is very similar to that obtained when solely using the standardized MAREDAT field (red stars). Moreover, while the difference in *PGI* between models decreases when evaluated at larger prey concentrations, there is still a 680% difference between the model with the fastest and slowest *PGI* when evaluated at a much larger total phytoplankton concentration of 5.0 mmolC/m³, and over 10-fold when evaluated below 0.5 mmolC/m³ (Figure S2a, black trace). By comparison, this ratio ranges from only 9-10 fold when calculated using the empirically measured functional response curves for micro- and meso-zooplankton (see "Description of Observations" below).

Combined with the fact that the *PGI* is well correlated with simulated specific grazing rates and grazing pressure globally (Figure 5) and at most latitudes, during summer and winter (Figure 4), it appears that it is a useful tool to compare large-scale differences in the rate zooplankton are assumed to graze between models. That is, while small differences in *PGI* are not particularly interesting, and sensitive to the definition of the standardized prey field, large differences are. A 70-fold difference in the *PGI* means zooplankton in CESM2 would graze 7,000% faster than in MEDUSA v2 if subject to the same temperature and prey. This implies something fundamentally different about the model dynamics and likely requires a substantially different parameterization of bottom-up controls to prevent complete nutrient limitation and or ecosystem collapse."

5. Having worked a lot with grazing formulations of different models, I could have guessed some of the responses shown in their figures without any knowledge of 'Grazing Pressure', just by simply looking at model parameters and structure. Hence, the authors (a) must give equations to show how 'Grazing Pressure' is calculated for each model (are you working with annual or monthly means, the surface ocean or the euphotic zone, how is the seasonal dynamics of grazing pressure reflected in your metric? why is one global number more **informative than zonal means**, what does an annual mean represent in seasonally highly varying systems? etc.)

Changed. We have:

- Provided in **Table 1** the grazing formulation for each model from which the 'Prescribed Grazing Index (PGI)' is computed.
- Provided in **Section 4.5** the general diagnostic equations through which model specific equations were used to compute both emergent properties offline and the *PGI*.

- Described in **Section 4.2** (see response to **Comment #4** above), how each set of model-specific evaluated using the global median (0-200 m) observed biomass (and POC) distributions described in Table S1 (see response to **Comment #4** above).
- Seasonal and zonal dynamics are not explicitly reflected in the *PGI* but **are** explicitly computed for the emergent properties and shown in **Figure 4** (see immediately below) and **Figures S3-6**. Here, all depth averaged rate terms are biomass weighted across the top 200 m to reflect the actual distribution of simulated biomass. The general equations used to compute offline variables are provided in detail in **Section 4.5** and the model-specific equation for the grazing and mortality framework are provided in **Table 1** and **Table S3**.
- In **Figures 4** and **Figures S3-S6** we show the correlation between the *PGI* (an innate property) and various model output (emergent properties) across latitudes during winter and summer. In particular, emergent specific grazing rates are well correlated with *PGI* at essentially all latitudes, during winter and summer (see **Figure 4F**). Thus, differences in each model grazing formulation evaluated at the global, annual median observed biomass concentration (i.e. the *PGI*) do appear to reflect differences in the emergent dynamics across the full range of seasonal and zonal variability.

Modified from **Figure 4**:

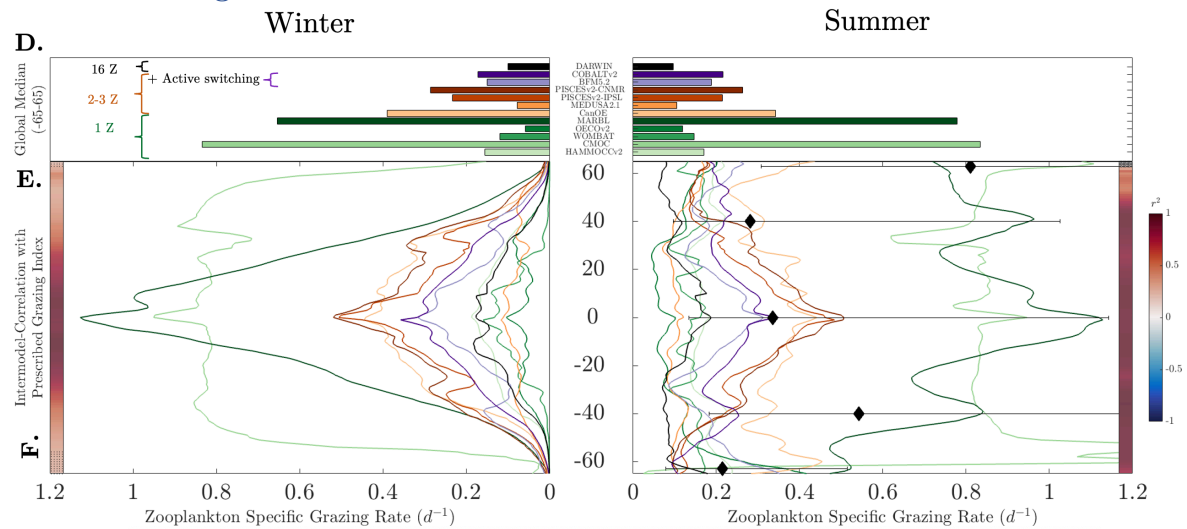


Figure 4: The Distribution of median **A-C**) grazing pressure, **D-F**) zooplankton specific grazing rates, and **G-I**) zooplankton specific mortality rates. **A, D, G**) The global median of each variable is plotted during the climatologic winter (December/June above/below the equator) and summer (vice versa). Medians are computed over the top 200m of the water column and weighted by **A**) phytoplankton or **D, G**) zooplankton biomass. Model colors correspond to the traces below. **B, E, H**) The zonal distributions are plotted for each model using a sliding 3° latitudinal band. **B, E**) Diamonds correspond the median observed estimate with error bars for the 25th and 75th percentiles (see “Methods”). **C, F, I**) On the boarder, the inter-model correlation with the *PGI* is plotted for each zonal band and season, with statistically insignificant relationships x'd out. Lighter x's indicate a casual, but statistically insignificant, relationship with $.2 > p > .05\%$.

(b) justify why looking at ‘Grazing Pressure’ would be better than looking at the actual equations themselves (and quantitatively, please, using more than just rather hand-waving claims; you could do sensitivity tests here as to the comparability of the actual grazing pressure in WOMBAT as computed pixel-wise),

Changed. We have now used the actual equations to compute both the *PGI* **and** emergent grazing terms. Moreover, we gone further and compared the *PGI* to the emergent grazing pressure (and many other properties) in not just WOMBAT but all 11 CMIP6 models and show explicitly that variability in the *PGI* is a strong predictor of emergent grazing dynamics (see **Figure 5A, B; Section 2.4** quoted below).

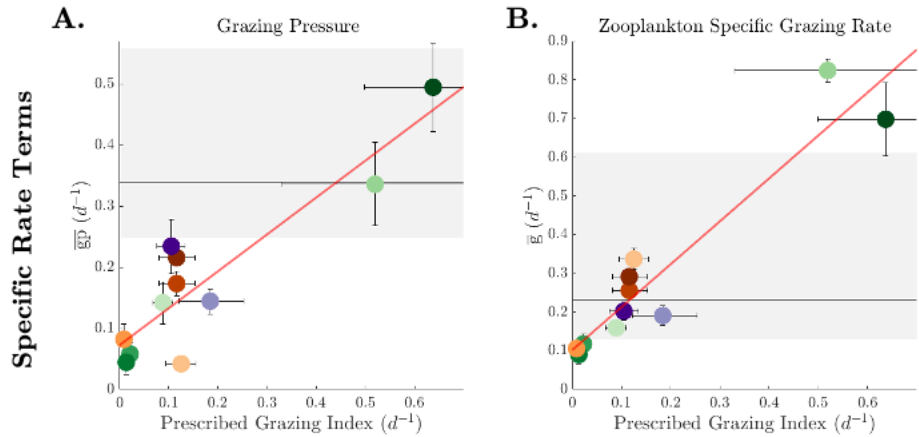
Line 236: “Large differences in emergent grazing dynamics at the global, regional and seasonal scales appear to be driven by large differences in the magnitude of the grazing formulation, as quantified by the *PGI*. Here we investigate the relationship between the *PGI* and emergent grazing dynamics and broader marine carbon cycling in CMIP6 models (Section 2.4.1) and in a controlled sensitivity study (Section 2.4.2). Qualitative differences in the functional response are explored “Supplemental Discussion 1”.

Despite reducing many modes of variability to a single value, the *PGI* is a strong predictor of not only global annually averaged specific grazing rates (Figure 5A; $r=.82$), but also grazing pressure (Figure 5B; $r=.78$). In fact, emergent grazing pressure is better correlated with the *PGI* at the global scale than emergent zooplankton mortality (m), biomass (Z) or even specific grazing rates (g) (Figures 2C-E). Moreover, despite substantial spatial-temporal variability in zooplankton biomass (Fig 4H), the *PGI* remains significantly, positively correlated with specific grazing rates (Figure 4F) and grazing pressure (Figure 4C) at most latitudes, during both winter and summer. In turn, even though most NPP and EP disproportionately occur during blooms, when the biomass is much higher than the global median, it appears that understanding how fast that models assume their zooplankton will graze at lower concentrations (as quantified per the *PGI*) is an excellent predictor of the emergent grazing pressure and zooplankton specific grazing rates across the entire dynamical system. That is, the rate of grazing prescribed for low to intermediate biomass stocks appears to be an important control on grazing dynamics across all prey levels via its feedback on the size of the prey population.

Given the close link between the *PGI* and emergent grazing dynamics, we examine if large differences in the *PGI* can explain other modes of uncertainty in the marine carbon cycle (Figure 5). Notably, there is no significant relationship between the PGI and NPP, EP or export efficiency at global (Figures 5I, J, M), regional or seasonal scales (Figure S3). Despite the importance of top-down controls (Laufkötter et al., 2015; Prowe et al., 2012; Dupont et al., 2023; Laufkötter et al., 2016; Karakus et al., 2021; Negrete-garcia et al., 2022), this is unsurprising because CMIP6 models are highly tuned towards observed estimates of NPP and EP. However, because BGC models are otherwise largely undetermined (WardParameter et al., 2010), this means that large uncertainty in grazing pressure must be compensated for elsewhere. While the *PGI* is correlated with faster zooplankton specific mortality rates (Figures 5D; S4C), these differences are already accounted for within grazing pressure and are insufficient to constrain its variance. In turn, models with a larger *PGI* transfer more carbon more efficiently from primary producers to higher trophic levels (Figures 5K, L, N; S5) and must balance larger grazing losses and lower phytoplankton biomass (Figures 5E, S4I) with faster phytoplankton specific division rates (Figures 5C, S4F). Thus, NPP appears to be tuned via faster turnover of a smaller phytoplankton population. However, because the large variance in specific grazing rates is not adequately balanced by zooplankton mortality, there is not necessarily higher turnover in the zooplankton population. In turn, zooplankton biomass exhibits nearly twice as much inter-model uncertainty as phytoplankton biomass (Figure 5A, E, J) and a weaker, statistically not significant relationship with *PGI* (Figure 5F).

In short, large differences in the *PGI*, and thus grazing pressure, appear to be tuned out in terms of historical NPP via faster nutrient uptake, but models with similar NPP can have very different standing stocks of phytoplankton (Figure 5E), zooplankton (Figure 5F), total biomass (Figure 5G) and zooplankton to phytoplankton ratios (Figure 5H).”

Modified from **Figure 5:**



Variable	Observed Values		CMIP6 Statistics			Sensitivity to <i>PGI</i>	
	Best Estimate	Uncertainty	$\bar{x}$	σ	$\sigma \bar{x}^{-1}$	r^2	p
Specific Rate Terms (d^{-1})							
Grazing Pressure	.35	.25 – .56	.18	.14	.78	.8	<.001
Z. Specific Grazing Rate	.23	.12 – .61	.30	.24	.82	.88	<.001

Modified from Figure 5. Inter-model variability in (A) emergent grazing pressure and (B) specific grazing rates is well correlated ($r^2=0.8, 0.88$, respectively) with the Prescribed Grazing Index at the global scale

and (C) discuss in detail how their findings fit into the context of previous work.

Changed. We have extensively revised the manuscript to put our findings in the context of previous work. Please see our response to specific comments below for many (but not all) aspects of the improved discussion.

6. Furthermore, we need to see how you take into account the substantial differences in multi-prey formulations (which you now just state is an additional uncertainty, but in our experience changes grazing pressure substantially, see respective literature, for instance by Vallina et al., Ecological Modelling, 2011; Vallina et al., Prog. Oceanography, 2014), and address this theoretically.

Changed. This is discussed throughout, particularly in **Section 4.1** (see response to specific comments below). Moreover, these differences are fully captured in the newly computed, offline emergent dynamics (see **Section 4.5**) and captured in a reduced manner in the *PGI* (see **Section 4.2**, Quoted in response to **Comment #4** above).

See also our **Response to Reviewer #2** concerning their hypothesis that models with different food webs are tuned via differential turnover rates in different zooplankton groups to exert a relatively similar community-integrated top-down control on NPP. We argue that evidenced by the very large uncertainty in emergent grazing pressure (see **Figure 3**), which integrates the entire multi-trophic community response, this is not the case.

7. Last but not least, I'd like to see an uncertainty range in form of shading for all the theoretical curves that takes into account the spatio-temporal variability of 'Grazing pressure' (Figs 1,2,3 and so on), as it varies across environments, seasons and model set-ups. A single dot for a hypothetical annual mean is unlikely to reflect the dynamical variation each model is capable of in space and time.

Changed. Uncertainty in the Prescribed Grazing Index has been quantified by evaluating it at +/- 25% of the total phytoplankton concentration and associated horizontal error bars have been added in

Figures 2, 5, and 6 (see expert from Figure 5 in our response to **Comment 5** above). While the *PGI* is intentionally designed to reduce many modes of variability into a single value, we now show how this value relates variability in emergent dynamics stemming different model set-ups across a range of environments and spatio-temporal variability (see **Section 2.3.2, 2.4, Figure 4**). We now show explicitly, that the *PGI* does appear to largely ‘reflect the dynamical variation each model is capable of in space and time’ (see above). Thank you for pushing us to investigate this.

8. Third, the paper should reflect more on the relative sources of uncertainty and variability in their ‘grazing pressure’ and carbon cycle metrics stemming from (a) the grazing formulation itself, (b) the way multi-prey interactions are parameterized (see e.g. Vallina et al., *Ecological Modelling*, 2011 and Vallina et al., *Prog. Oceanog.*, 2014 as to how to remedy suboptimal grazing), (c) the effect of higher trophic level complexity (see Le Quéré et al., *Biogeosciences* 2016), and (d) the way carbon is channeled to multiple higher trophic levels with different storage capacity (see e.g. Laufkötter et al., *Biogeosciences*, 2016). I won’t even mention behavior & life stages (Clerc et al., *Theoretical Ecology*, 2021), DVM (e.g. Kelly et al. *Front Mar. Sci.*, 2019), trophic amplification (Kwiatkowski et al., *Global Change Biology*, 2018) and so on. The former four issues act in concert to generate uncertainty in carbon fluxes within the marine ecosystem, and the grazing formulation is just one part of the uncertainty. And since we know a lot about that particular aspect already, setting this one source of uncertainty into the context of the uncertainty associated with the other three issues seems warranted.

Changed. We have explicitly included and discussed other important sources of uncertainty. We explicitly compute the inter-model uncertainty (defined with the coefficient of variation across median model values, computed at various scales) in 14 aspects of the marine carbon cycle (see **Figure 3** below). We find the two largest sources of uncertainty are specific grazing rates and grazing pressure. Surprisingly, we find that uncertainty in specific grazing rates is not adequately compensated for by tuning of the turnover time of higher trophic levels, leading to even larger uncertainty in grazing pressure. Note the emergent grazing pressure explicitly captures differences in the grazing formulation, multi-prey preferences, food web structure, and the mortality/storage capacity of higher trophic levels.

Additionally, in the SI we explore the way in which differences in the Apparent Functional Response (an emergent properties of a-d) appears to drive differences in carbon cycling and community composition, creating ecological niches for higher export and ecosystem efficiency (see **Supplemental Discussion 1; Figure S11**).

Figure 3:

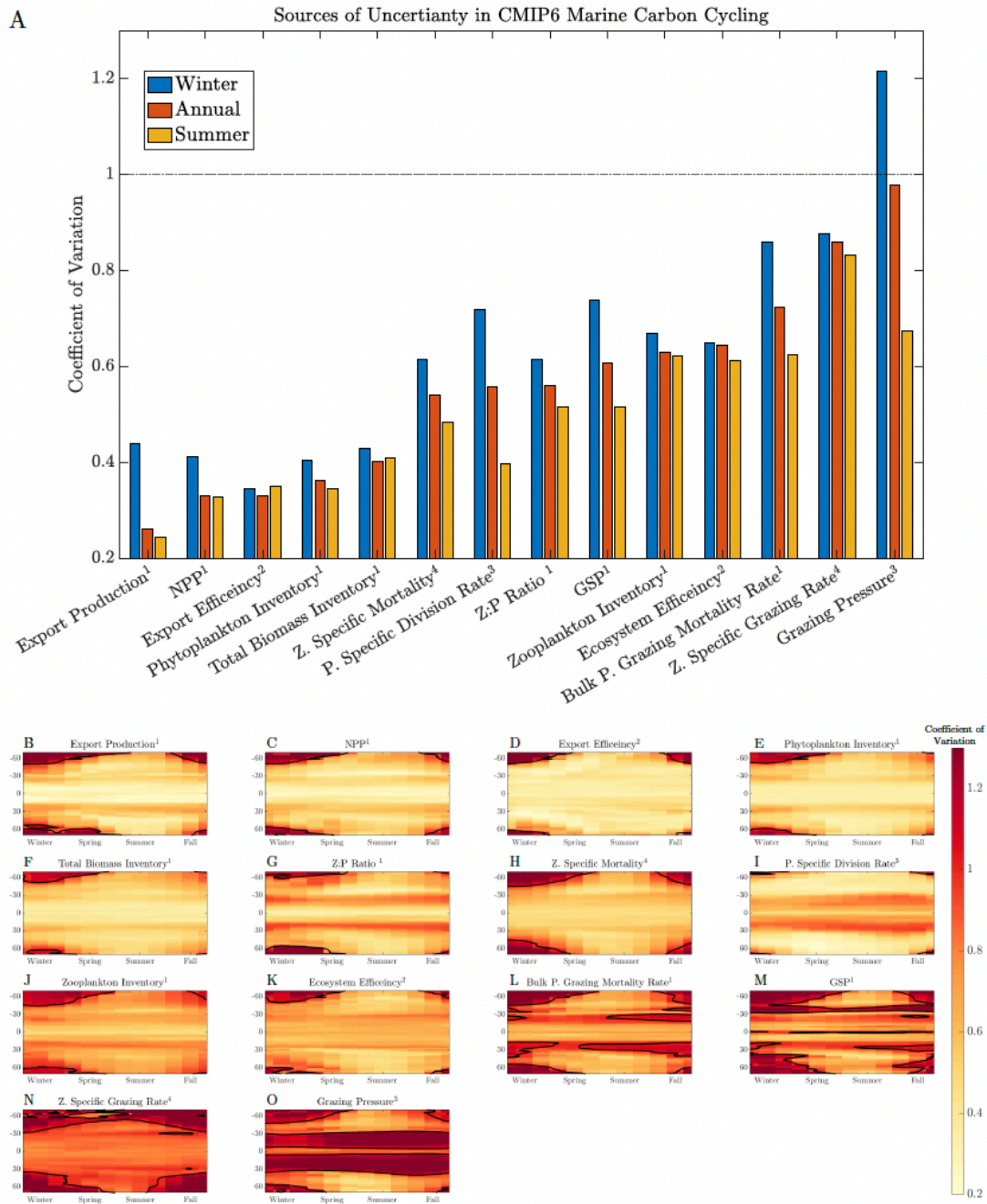


Figure 3: Inter-model uncertainty in marine carbon cycling. **A)** The coefficient of variation (standard deviation divided by the mean) is shown for the annual (orange), winter (blue), and summer (yellow) median of major biological marine carbon fluxes, stocks, and rates. Winter values are computed using the December and June for the northern and southern hemispheres, respectively. Vice versa for summer. Medians are computed over the top 200m of the water column and weighted by either (1) volume, (2) NPP, (3) phytoplankton biomass or (4) zooplankton biomass to account for the distribution of where/when the relevant carbon cycling occurs. Note, qualitatively similar values are computed using weighted-means, instead of medians, and are reported in Figure 5. **B-O)** Hovmöller diagrams (month vs. latitude) of the seasonal and zonal distribution of inter-model uncertainty are provided for each variable. Areas in which the standard deviation is greater than the mean are contoured in black.

Excerpt from **Figure S11**:

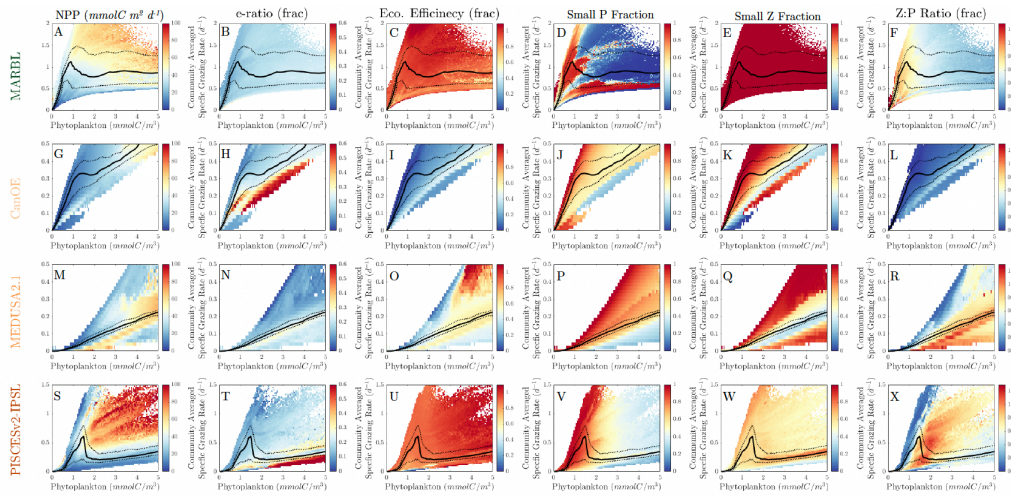


Figure S11: Sensitivity of marine carbon cycling and community composition to the Apparent Functional Response. **A-OO**. The Apparent Functional Response is plotted as per Figure S10. However, now instead of the grid-cell frequency on the color axis, the mean NPP, e-ratio, export efficiency, fraction of small phytoplankton, fraction of small zooplankton, and total zooplankton to phytoplankton ratio of all qualifying grid-cells is plotted. Note, in COBALTv2, medium sized PFTs are included as small.

9. Furthermore, spatio-temporal information on the zonal and seasonal patterns of grazing are essential dimensions of variability. NPP and EP aren't generated in the annual mean, and everywhere – they are generated during specific high productivity months, and at certain specific latitudes. And we know that the models poorly represent the spatio-temporal dynamics of bloom formation and transfer across the ecosystem (Hauck et al., Global Biogeochemical Cycles, 2015; Hashioka et al., Biogeosciences, 2013). The author has the entire range of CMIP6 models at his hands to delve deeper into some of these issues. Broadening the work to a more general discussion of the aspects above and including spatio-temporal variability as an essential source of information on how well these models really do would increase the novelty and relevance of the current work.

Changed. Thank you for this suggestion. We have now explicitly included spatio-temporal information on the zonal and seasonal patterns of emergent model properties in **Figures 2-7, S3-S6 and S9**. In **Section 2.3.2** we now analyse zonal and seasonal variability in emergent grazing dynamics across CMIP6 models and evaluate them against not only observations but also a more complex BGC model: a ~50 PFT version of DARWIN, generously provide by Stephanie Dutkiewicz.

Line 194: “Uncertainty in grazing pressure (Figures 3O, 4A-C) and specific grazing rates (Figures 3N, 4D-F) is even larger at smaller spatial and temporal scales. During the winter, when relieved grazing pressure regulates the phytoplankton seed population required to initiate a spring bloom (Freilich et al., 2021; Behrenfeld et al., 2013), the median global grazing pressure is 43-times larger in MARBL than OECOV2 (Figure 4A; left panel) and the inter-model standard deviation is 20% larger than the inter-model mean (Figures 3A, O). During summer, when zooplankton biomass is higher (Figures 4G-I) and slow-grazing meso- or macrozooplankton may regulate standing stocks of phytoplankton, especially in the Southern Ocean (Le Quere et al., 2016), the largest source of uncertainty in models is specific grazing rates (Figures 3A, N).

Zonally, the inter-model uncertainty in grazing pressure is largest in the tropics, regardless of season, with a roughly 20-times difference between MARBL and the three models with the lowest median grazing pressure (Figure 4B). This is partly driven by substantial enhancement in the specific grazing rates of models with temperature-limited grazing (MARBL, PISCESv2, BFM5.2 and COBALTv2) in the tropics (Figure 4E). Variance in zooplankton biomass, however, is largest at the poles, where models vary from almost 0 (CMOC, CanOE) to over 400 (COBALTv2, MEDUSA2.1) mmolC m^{-2} in the Southern Ocean for example (Figure 4H).

Variability in grazing pressure across models is not well constrained by observations (see ``Methods"). Qualitatively, observations of grazing pressure agree with the zonal distribution simulated in models, peaking at the equator, then tapering off towards the poles (Calbet et al., 2004; Schmoker et al., 2013; Calbet et al., 2001), but the models simulate much lower values and exhibit a much larger spread. In no region do more than 65% of models fall with the middle 50% of zonally binned, basin-scale observational estimates (Fig 4B; black diamonds). While this underscores how poorly constrained grazing pressure is in models, it is important to remember that field measurements are sparse and subject to substantial sampling biases (Calbet et al., 2001; Schmoker et al., 2013). Thus, a small uncertainty window is probably a consequence of a small sample size, rather than low natural variability (see ``Methods"), suggesting that the solution is not simply nudging models toward current observations but also further constraining the observations.

Perhaps a more useful comparison is a more complex BGC model. While no CMIP6 model included more than three zooplankton or phytoplankton groups, we have included in Figure 4 (black trace) the DARWIN model of (Follett et al., 2022), which includes 31 pPFTs and 16 zPFTs. Only BFM5.2, COBALTv2 and MARBL come close ($r^2 > .6$) to matching the zonal variability in grazing pressure simulated in DARWIN during both winter and summer. MARBL, however, exhibits by far the largest magnitude bias, while BFM5.2 has the second smallest. This suggests that active prey switching, only included in COBALTv2 and BFM5.2, is an essential ingredient in implicitly simulating the more diverse ecosystem explicitly resolved in DARWIN (Adjou et al., 2012; Prowe et al., 2012; Vallina et al., 2011; Chenillat et al., 2021). Interestingly though, grazing pressure in DARWIN is achieved via much lower specific grazing rates (Fig 4E) and higher zooplankton biomass (Fig 4H) than in BFM5.3 or COBALTv2. This is likely due to a decline in the direct interaction strength between predator-prey in a more complex food web (Sailley et al., 2013) combined with a reduction in non-linear mortality to explicit and implicit predation, which will be smaller in aggregate when biomass is distributed across more discrete biomass pools."

10. Fourth, I question the choice that the emphasis of the paper being put on WOMBAT results, whereas one could have learnt a lot from a more detailed analysis of the CMIP6 model suite. WOMBAT's ecosystem structure is far too simplistic to correctly capture the huge uncertainties in model ecosystem dynamics due to the above-mentioned issues. Hence, I personally would not use the standard version of WOMBAT as a reference model (I'd add further trophic levels; see e.g. Ward et al., Prog. Oceanog., 2013), but since that is the model the authors are familiar with, that probably can't be helped. Alas, based on the current text, I really cannot understand how simulations conducted with WOMBAT are reflective of the carbon dynamics in the CMIP5 models, and this link needs to be strongly emphasized in a revised version of the paper. All of the CMIP models are heavily tuned not only to match NPP estimates, but also export at 100 m. Hence, as you are varying WOMBAT through the range of realized values of 'Grazing Pressure' across models, you are looking at variations in NPP, GSP, EP that could lead to the exclusion of the respective parameterisation in the original CMIP models, because NPP and EP could potentially not be tuned to fit the observed range. So what do you prove with WOMBAT? In essence, this model is somewhat too simple to assess the role of zooplankton in these models, since it doesn't incorporate any of the known sources of variations in these models (multi-trophic level interactions (Le Quéré et al. 2016), prey-switching (Vallina et al. 2014), food-web links (Sailley et al. 2013) and so on, except for one single group with one single grazing equation. Why would you use a model that does not resolve higher trophic levels to prove that higher trophic levels matter? Please express the novelty and relevance of your findings far more clearly, and work more with the CMIP models more to support your main take home messages.

Changed. The focus of the study is now on a detailed analysis of the CMIP6 model suite, although we still have included some WOMABT sensitivity analysis. However, we believe we have better established the link between the *PGI* in WOMBAT and more complex food webs in other CMIP6 models by showing first that the *PGI* in those models is well correlated with their emergent dynamics, suggesting that the *PGI* can either capture (or dominate) the many other modes of variability in the construction of the food web and grazing formulation therein. See response to **Comment 5** above and further comments below.

11. Fifth, the paper must work on its selection of figures. While some of those presented are generally informative, others are mostly summarizing model sensitivity studies, either conducted with

WOMBAT, or based on the CMIP6 model suite (e.g. Figs 1,3,4). Hence, some figures presented in the main text feel like supplementary information, based on theoretical calculations with little to no relevance for the actual global ocean (see my comments above). What we would like to know, or at least what I would like to know, is what lessons we should learn, i.e. a conceptual diagram or schematic how (different versions of) zooplankton community structure and grazing influences NPP, GSP and EP, and how this is likely to affect future projections of the carbon cycle variables in the CMIP6 models (as opposed to ‘in WOMBAT’).

Changed. We have included many new Figures, that illustrate variability in the innate (**Figure 1, Table 1**) and emergent (**Figures 2-5**) dynamics across all CMIP6 models. Note that these new figures also reflect the reframed focus of the manuscript, which now primarily highlights the uncertainty in emergent grazing pressure and how it is driven by large first-order difference in the grazing formulation that can be captured by the *PGI* metric. Only then does it look at how differences in the *PGI* can influence NPP, GSP, EP. Now, however, these avenues are explored through both the CMIP6 suite (**Figure 5**) and the controlled WOMBAT sensitivity study (**Figure 6-8**).

We have also designed a schematic (**Figure 1**) to broadly highlight differences in the grazing structure/food web of CMIP6 models and illustrate the large differences in the ‘magnitude’ of grazing, as described by the *PGI*, they assume. We then use the rest of the manuscript to show how these differences are (or are not) associated with various emergent properties of the marine carbon cycling across both the entire CMIP6 model suite and WOMBAT specifically.

Specific comments:

Introduction

12. One of the largest gaps in our understanding of marine carbon cycling is the role of zooplankton. – Add relevant references. And is this true? Justify.

Changed. We have 1) qualified this statement to reflect specifically the *simulation* of grazing:

Line 26: “Yet the way in which we simulate zooplankton grazing in marine biogeochemical (BGC) models remains a large gap in our representation of marine carbon cycling.”

2) Supported it with references in the immediately following paragraph:

Lines 29: “Zooplankton species exhibit a considerable functional diversity and their relative distribution varies widely (Brandao et al., 2021; Benedetti et al., 2023; Brun et al., 2015, 2016; Barton et al., 2013; Kjørboe 2018). However, marine biogeochemical (BGC) models represent the integrated behaviour of all zooplankton using a strictly limited number of zooplankton groups (Fennel et al., 2022), especially in computationally expensive IPCC coupled-climate models that typically include only 1-2 groups (Kearney et al., 2021; Seferian et al., 2020). This leads to a substantial uncertainty in how modellers implicitly represent complex communities and their impact marine carbon cycling (Le Quéré et al., 2005; Sailley et al., 2013; Laufkotter et al., 2015, 2016; Petrik et al., 2022; Vallina et al., 2011; Gentleman et al., 2008; Chenillat et al., 2021; Prowe et al., 2012; Vallina et al., 2014; Dutkiewicz et al., 2020; Dupont et al., 2022; Karakus et al., 2021, Negrete-garcia et al., 2022; Hashioka et al., 2013; Truscott 1994).”

3) Detailed the findings of those references in the revised ‘Results and Discussion’ section:

Line 93: “Variability in the structure of the food-web and grazing formulation substantially impact ecosystem stability (Vallina et al., 2011; Gentleman et al., 2008), plankton diversity (Chenillat et al., 2021; Prowe et al., 2012; Vallina et al., 2014; Dutkiewicz et al., 2020), NPP (Laufkötter et al., 2015; Prowe et al., 2012), EP (Dupont et al., 2022; Laufkötter et al., 2016; Karakus et al., 2021; Negrete-garcia et al., 2022), regional (Chenillat et al., 2021) and global (Sailley et al., 2013; Le Quéré et al., 2005; Prowe et al., 2012) phytoplankton biomass and chlorophyll distributions, and the phenology of phytoplankton of blooms (Hashioka et al., 2013; Karakus et al., 2022, Truscott 1994).”

13. Yet grazing dynamics are poorly constrained (7) and difficult to model (8), hampering our understanding of the ocean’s – Referencing utterly insufficient, look at works by Vallina (multiple), Sailley et al., Ecological Modelling, 2013, Chenillat et al., PLOS One, 2021 (and references therein); Dutkiewicz et al., Biogeosciences, 2020 and so on. Also have a look at Séférian et al., Current Climate Change Reports, 2020 on improvements in model complexity between CMIP5 and CMIP6 models.

Added. This specific sentence has been removed from the revised text; however, the general argument that grazing dynamics are poorly constrained and difficult to model has been supported in the 3rd and 4th (see below) and 2nd (see **Line 29**, quoted in response to **Comment 12**) paragraphs of the Introduction, respectively. All suggested references (and others) have been included.

Line 35: “Substantial observational work over the last two decades has helped describe the global distribution of zooplankton and their grazing dynamics, but remains limited as a means to adequately constrain grazing in global models. Global, in-situ distribution of zooplankton biomass (Buitenhuis et al., 2006, 2013; Moriarty et al., 2013; Moriarty and O’Brien, 2013) are relatively patchy and must reconcile uncertainties across disparate acoustic, optical, and net-based methodologies (Lombard et al., 2019; Everett et al., 2017). Progress has been made towards estimating zooplankton biomass from space but current products are either limited to identifying swarms of specific species (Basedow et al., 2019) or heavily reliant on statistical inference from chlorophyll fields (Druon et al., 2019; Stromberg et al., 2009; Heneghan et al., 2020).

Moreover, grazing pressure on the phytoplankton community, defined as the phytoplankton specific loss rate to grazing (Prowe et al., 2012), varies with both zooplankton biomass and the specific grazing rate of zooplankton. However, empirical estimates of the parameters that describe how specific grazing rates vary with prey abundance span three orders of magnitude, varying with zooplankton size, species, and age (Hansen et al., 1997; Hirst et al., 2003; Rohr et al., 2022). This means that detailed information about community composition must accompany bulk estimates of biomass to estimate grazing pressure. Field-based dilution experiments can help average over this variability, providing estimates of community-averaged grazing pressure (Morrow et al., 2018; Landry et al., 2009, Landry et al., 2008; Landry et al., 2000), but these are highly localized in space and time and can be biased by nonlinearities in the functional response (Sandhu et al., 2019), trophic cascades (Calbet et al., 2008; 2011) and the presence of mixotrophs (Calbet et al., 2012). Review studies have helped constrain the range of reasonable values, but are limited to broad basin-scale averages and often exclusively target the role of microzooplankton (Schmoker et al., 2013; Calbet and Landry, 2004; Landry and Calbet, 2004). Measuring the contribution of larger, mesozooplankton in the field typically requires gut content analysis and various assumptions about metabolic rates (Morrow et al., 2018; Calbet et al., 2001). Finally, give modellers a good snapshot of which functional groups belong where, but not time evolving global data sets. Collectively, the observational record provides a strong baseline understanding of zooplankton grazing dynamics but cannot be used in the same way as high spatial and temporal resolution, satellite-derived chlorophyll and NPP products to tune global models.”

14. Yet marine biogeochemical (BGC) models represent the integrated behavior of all zooplankton using a strictly limited number of zooplankton groups, especially in computationally expensive IPCC coupled-climate models. – Cite Le Quéré et al., Global Change Biology, 2005 on PFTs, Laufkötter et al., Biogeosciences, 2015, 2016, Sailley et al., Ecological Modelling, 2013 and so on.

Changed. References have been added. See **Line 29**, quoted in response to **Comment 12**.

15. Further, while chlorophyll, NPP, and phytoplankton biomass can be estimated from space, global in-situ zooplankton observations cannot – Not true, see recent papers cited in general comments above on the detectability of zooplankton from space. Rephrase.

Changed. This have been revised to:

Line 39: “Progress has been made towards estimating zooplankton biomass from space but these products are currently either limited to identifying swarms of specific zooplankton species (Basedow et al., 2019) or heavily reliant on statistical inference from chlorophyll fields (Druon et al., 2019; Stromberg et al., 2009; Heneghan et al., 2020).”

16. making simulated zooplankton biomass much more difficult to validate against spatially-resolved global data sets (22). – Not a discussion paper anymore, and see also Sailley et al. 2013 and think about the availability of further global data sets now available for different size classes and taxonomic groups, e.g. used in Buitenhuis et al., Global Biogeochemical Cycles, 2019

Changed. We have:

- 1) Provided a more nuanced introduction zooplankton observations (see **Line 35**, quoted in response to **Comment 13**).
- 2) Updated the reference to Moriarty and O’Brien (2013).
- 3) Incorporated the MAREDAT data set (as well as many other observational data sets) used by Buitenhuis et al., 2019 (i.e. Figure 4 therein) throughout the revised manuscript to both refine the Prescribed Grazing Index and compare emergent model output to the observations (see **Section 4.4**; specifically, **Line 695**, quoted in response to **Comment 4**).

17. rooted primarily in theoretical predator-prey ecology and small-scale, in-vitro, laboratory experiments (23) – Referencing not up to date with literature on the multiple trait collections now used to validate models.

Changed. This sentence has been removed and replaced with a much more nuanced introduction to zooplankton/grazing observations (see **Line 35**, quoted in response to **Comment 13**).

18. While these values can be estimated empirically (26, 27), measured parameter values span over three orders of magnitude (23), vary with zooplankton size, species, and age (26), and are not likely representative of grazing dynamics in the open ocean (28). – (28) seems wrong reference here.

Changed:

Line 42: “Grazing pressure, defined as the phytoplankton specific loss rate to grazing (Prowe et al., 2012), varies with both zooplankton biomass and the specific grazing rate of zooplankton. However, empirical estimates of the parameters that describe how specific grazing rates change with prey abundance span three orders of magnitude, varying with zooplankton size, species, and age (Hansen et al., 1997; Hirst et al., 2003, Rohr et al., 2022).”

19. Unsurprisingly, Rohr et al. (23) surveyed 40 BGC models, dating back to foundational NPZD box-models (29, 30) and found that the values used for K and G both span over two orders of magnitude. – You are not the first to survey different models, see e.g. early attempts in Sailley et al., Ecological Modelling 2013, or the work by Hashioka et al., Biogeosciences, 2013 on the quantification on top-down versus bottom-up pressures to explain bloom dynamics.

Changed:

Line 59: “This leaves a large degree of uncertainty in how grazing dynamics should be represented, with models differing largely in their associated qualitative and quantitative assumptions (Petrik et al., 2022; Kearney et al., 2021; Rohr et al., 2022; Sailley et al., 2013, Hashioka et al., 2013).”

20. how the distribution of prey modifies grazing behavior and influences the integrated grazing pressure exerted on the full phytoplankton community – references needed, see e.g. the work by Vallina et al., and Vallina and Le Quéré.

Changed. This sentence has now been removed; however, both Vallina et al., (2014) and Vallina and Le Quéré (2011) have now been referenced earlier in the Introduction and incorporated elsewhere throughout the Results and Discussion (see, for example, **Line 29 and 93**, quoted in response to **Comment 12**).

21. Despite the uncertainty associated with the representation of grazing, there has been no global study to quantify the sensitivity of marine carbon cycling to the degree of grazing pressure prescribed in coupled ocean-BGC models. – Not true, see Sailley et al., Ecological Modelling, 2013 and Laufkötter et al., Biogeosciences, 2015, 2016. And Vallina et al., Prog. Oceanog. 2014 on the effect of the grazing formulation on diversity and biomass distribution. Just because your metric was not used, this doesn't mean that the community hasn't looked at the role of zooplankton or grazing, wouldn't you agree?

Changed. This sentence has been removed and all references have been integrated into the Introduction and the rest the manuscript (see, for example, **Line 29 and 93**, quoted in response to **Comment 12**).

22. Together, this could have contributed to the persistent uncertainty surrounding NPP (32) and zooplankton biomass (8) in future climate projections – see also Laufkötter et al., Biogeosciences, 2015.

Changed. References now included:

Line 61: “This could contribute substantially to the persistent uncertainty surrounding global and regional estimates of NPP (Tagliabue et al., 2021, Laufkötter et al., 2015), EP (Laufkötter et al., 2016, Fu et al., 2016) and zooplankton biomass (Petrík et al., 2022) in future climate projections.”

23. the lack of adequate global, in-situ grazing observations to constrain them (35) - Really? What about all the important and extensive work on grazing rates that there is? Multiple papers by Calbet? Landry? Ohman? Compilations made by marine ecosystem modellers for each new zooplankton FT included? See also references in Barton et al., Ecology Letters, 2013.

Changed. This sentence has been removed and replaced with a more nuance description of available observations, including Morrow et al. (2018), Landry et al. (2009), Landry et al. (2008), Landry et al., (2000), Schmoker et al. (2013), Calbet and Landry (2004), Barton et al. (2013), Brun et al. (2016), and Benedetti et al. (2023). See **Line 35**, quoted in response to **Comment 13**.

Moreover, we have now leveraged large reviews of grazing pressure by Schmoker, Hernandex-Leon, & Calbet (2013), Calbert and Landry (2004) and Calbet (2001) to derive global and zonal estimates of observed grazing pressure (and other rates terms) with which to compare the emergent properties of the CMIP6 simulations (see **Figures 4 and 5; Section 4.4**). Thank you for this great suggestion.

Line 383: “The observed mean global grazing pressure (0.34 d^{-1} (0.24, 0.62); Figure 5A) was computed from field measurements of microzooplankton (Calbet et al., 2004; Landry et al., 2004; Schmoker et al., 2013) and mesozooplankton

(Calbet et al., 2001) phytoplankton specific grazing mortality rates (i.e. grazing pressure) estimated primarily using shipboard dilution experiments and gut content analysis, respectively. Field measurements for microzooplankton from 110 studies were compiled by Schmoker et al. (2013), extending the work of Calbet et al. (2004), and grouped into bio-geographic provinces as defined by Longhurst et al. (1995). Schmoker et al. (2013) report the median, 25th and 75th percentile grazing pressure (referred to as the grazing mortality rate) for each province. We estimate the global mean grazing pressure from microzooplankton using an area weighted average across all open ocean provinces, with uncertainty propagated using the 25th and 75th percentiles for each province. Field measurements for mesozooplankton from 27 studies were compiled and used to fit a log-linear relationship between NPP and total autotrophic ingestion by mesozooplankton (Calbet et al., 2001). We use this relationship along with the median NPP in each province reported in Table 2 of Schmoker et al. (2013) to further estimate the associated total mesozooplankton autotrophic ingestion rate in each province, with uncertainty estimated using the standard error of the regression parameters reported by Calbet et al. (2001). We then compute the area weighted global mean mesozooplankton bulk grazing rate and divide by the global phytoplankton inventory (see above) to estimate the global mesozooplankton grazing pressure, add it to microzooplankton grazing pressure to reach total grazing pressure, and propagate uncertainty accordingly. Note, grazing pressure from mesozooplankton accounts for roughly 10% of total grazing pressure.

Schmoker et al. (2013) also report the total autotrophic losses to microzooplankton grazing ($PgC\ yr^{-1}$) allowing use to estimate the globally-integrated bulk phytoplankton grazing mortality rate ($28.0\ PgC\ yr^{-1}$ (23.2, 54.4); Figure 5K) in a similar fashion, again using the relationship from Calbet et al. (2004) for the mesozooplankton contribution. Dividing by globally-integrated zooplankton biomass from the MAREDAT dataset gives us zooplankton specific grazing rates ($0.23\ d^{-1}$ (0.12, 0.61); Figure 5B). Dividing by globally-integrated NPP gives us an estimate of global ecosystem efficiency (56 % (46, 108); Figure 5N). Note, an ecosystem efficiency over 100% indicates that uncertainties in the observations exceed the theoretical bounds. Additionally, we can multiply the globally-integrated bulk phytoplankton grazing mortality rate by an estimate of assimilation efficiency to reach GSP ($7.9\ PgC\ yr^{-1}$ (19.6, 38.4); Figure 5L). We assume assimilation efficiency to be 0.7, as typically assumed by the ecological modelling community (Heneghan et al., 1984; Kiorboe et al., 2018; Abe et al., 2013; Montagnes et al., 2012), with an uncertainty range of 0.3-0.9 (Landry et al., 1984; Kiorboe et al., 2018; Abe et al., 2013; Montagnes et al., 2012)."

Line 836: "Next, zonal estimates and uncertainty for bulk grazing mortality, grazing pressure, and specific grazing rates are computed identically to the global values described above, except now sorting the provinces of Schmoker et al. (2013) into the zonal bands described above and pairing them with the corresponding MAREDAT biomass zonal bins. Note, Schmoker et al. (2013) do not distinguish the northern from southern westerlies, so grazing data is identical for each. However, we use zonally resolved biomass data to compute specific rates where required. Finally, zonal estimate of phytoplankton specific division rates are now computed using zonal estimates of Schmoker et al. (2013)".

We also now cite the bio-geographies of Barton et al (2013) and others in the revised 'Moving Forward' Discussion section'.

Line 383: "Adequately constraining grazing may require a new generation of observational products that leverage higher resolution, time-evolving remote sensing and in-situ BGC Argo float-based data sets. While these platforms cannot directly measure grazing, both satellites and floats products have been used to infer net community production from biomass (Rohr et al., 2017; Li et al., 2021) and oxygen budgets (Su et al., 2022; Arteaga et al., 2019). Combined with observationally-based estimates of phytoplankton biomass (Westberry et al., 2008), NPP (Westberry et al., 2008; Yang et al., 2021), and the assumption that phytoplankton losses are dominated by grazing (Mojica et al., 2021; Deppeler et al., 2017), grazing pressure could then be estimated at much finer space and time scales than ship-board dilution experiments (Schmoker et al., 2013; Calbet et al., 2004). Further, pairing these with rough estimates of zooplankton biomass, either from statistical models (Heneghan et al., 2020; Petrick et al., 2022) or emerging remote sensing products (Basedow et al., 2019; Druon et al., 2019, Stromberg et al., 2009), could yield specific grazing rate terms, enabling modellers to better tune zooplankton turnover (i.e. the balance of growth vs. mortality).

Together, such products will not only help constrain the mean state of grazing at global and regional scales, but could also complement existing bio-geographical resources to constrain how grazing dynamics qualitatively vary regionally with community composition. That is, by plotting the evolution of observed grazing rates against phytoplankton biomass, one can empirically define the community-averaged apparent functional response at regional and even seasonal scales. Spatio-temporal variability in this curve will improve our understanding of the distribution (Barton et al., 2013; Brun et al., 2016; Benedetti et al., 2023), seasonal succession (Brun et al., 2019; Kenitz et al., 2017), and migration (Richardson et al., 2008; Atkinson et al., 2019) of zooplankton species, which can then be compared to the associated variability in the emergent apparent functional response simulated in models (see "Supplemental Discussion 1"). Such comparisons will enable modellers to determine if explicit competition in their assumed food-webs are reasonably approximating natural variability in community-averaged grazing dynamics, or if increasing trophic complexity by including, for example, macrozooplankton (Le Quere et al., 2016), salps

(Luo et al., 2020), larvaceans, euphausiids, chaetognaths, jellyfish (Heneghan et al., 2020; 2023), pteropods (Buitenhuis et al., 2019), heterotrophic protists (de Vargas et al., 2015) or Rhizarians (Biard et al., 2016; Guidi et al., 2016) might be required.

Moreover, as stratification under climate change likely leads to greater light availability but fewer nutrients, the phytoplankton community may shift toward smaller phytoplankton with faster nutrient uptake but higher light requirements (Portner et al., 2019; Bopp et al., 2005). This may increasingly favour rapidly-grazing microzooplankton or gelatinous filter feeders, which can ingest smaller phytoplankton size classes (Conley et al., 2018; Heneghan et al., 2023), over crustacean mesozooplankton, which generally graze slower but can consume much larger prey (Everett et al., 2021; Barton et al., 2013). Considering these groups serve critically different roles in trophic transfer through prey preference (Biard et al., 2016) and carbon export through diel-vertical migration and (Bianchi et al., 2013; Steinberg and Landry, 2017; Archibald et al., 2019; Aumont et al., 2018) the formation of rapidly-sinking fecal pellets and carcasses (Luo et al., 2020, Clerc et al., 2023), it is essential that models include enough trophic diversity to capture present-day and emergent changes to community composition and its associated traits (Heneghan et al., 2023)."

24. A better understanding of zooplankton grazing and its influence on marine carbon cycling could substantially improve estimates of past, present, and future oceanic carbon sequestration and transfer to higher trophic levels – please cite relevant work, from global to regional modeling and from all over the marine ecosystem modeling world – Dutkiewicz, Le Quéré, Hashioka, Bopp, Aumont, and many more. Interesting in this context may also be the discussion in Nissen et al., Biogeosciences, 2021.

Changed. Reference have been included in the revised text:

Line 66: "One of the most important mechanistic uncertainties in IPCC-class marine BGC models is the representation of zooplankton grazing (Dupont et al., 2022; Gehlen et al., 2006; Hashioka et al., 2013; Laufkotter et al., 2015, 2016; Le Quéré et al., 2016; Negrete-Garcia et al., 2022, Nissen et al., 2021; Prowe et al., 2012, Vallina 2014; Dutkiewicz et al., 2020)."

Additionally, based on Nissen et al., Biogeosciences, 2021, we have included *Phaeocystis* in the observed global median estimate of the 'Large Phytoplankton' PFT (in addition to diatoms) used to compute the Prescribed Grazing Index (see Table S1; Section 4.2).

Line 569: "*Phaeocystis*, which no surveyed model includes but compete with diatoms for nutrients, particularly in the Southern Ocean (Nissen et al., 2021), are included in the "large phytoplankton" pool."

Results

25. 2.1 Models - Here, I lack a discussion of sub-optimal grazing and the KTW formulation by Vallina et al., Prog. Oceanog. 2014 in the context of advantages and disadvantages of each formulation. None of the models here account for the fact, that overall grazing decreases as the number of prey increases, which is unlikely to reflect the behavior of zooplankton in nature.

Changed: We have rewritten Section 2.1 (now predominately found in Methods Section 4.1) to include a more thorough discussion of the advantages and disadvantages of each model grazing formulation, their place in the literature, their ecological assumptions, and their numerical implications. This includes a discussion of which models employ active-switching and thus can exhibit sub-optimal grazing (i.e. BFM5.2, COBALTv2) and how the KTW formulation has been proposed to permit both active switching and maximal feeding by decrease $K_{1/2}$ as prey becomes more evenly distributed (see immediately below).

Additionally, we discuss the appearance of sub-optimal feed in the Apparent Functional Response (see 'Supplemental Discussion 1').

Line 444: “The representation of grazing in IPCC CMIP6 BGC models varies widely (Fig 1; Table 1; see also Kearney et al., 2021), as does their representation of phytoplankton and zooplankton groups. At one end of the spectrum, iHAMOCC (Tjiputra et al., 2020), CMOC (Zahariev et al., 2008) and WOMBAT (Law et al., 2017) include only one zooplankton and one phytoplankton group. In each, the zooplankton specific grazing rate, g (d^{-1}), is assumed to increase monotonically with the phytoplankton concentration, P (mmolC m^{-3}), before eventually saturating (Gentleman et al., 2008). This relationship, $g(P)$, is known as the single-prey functional response and is typically represented with a type II or III response curve (Gentleman et al., 2003; Rohr et al., 2022). Both curves can be described with the same two parameters. The saturation grazing rate, g_{max} (d^{-1}), describes the food-replete specific grazing rate, where as the half-saturation concentration, $K_{1/2}$ (mmolC m^{-3}) specifies roughly where the curve begins to saturate (Fig. S1). All else equal, both increasing g_{max} or decreasing $K_{1/2}$ increase zooplankton specific grazing rates.

iHAMOCC uses a type II response, which linearly increases with prey density below $K_{1/2}$, when prey is scarce (rectangular hyperbola; Fig S1). A type II response is typically what has been measured in empirical laboratory experiments (Hansen et al., 1997; Hirst et al., 2003) and is thus likely to be representative of individual zooplankton grazing in a well-mixed environment. However, the type II response can contribute to dynamically unstable solutions (Gentleman et al., 2008), requiring iHAMOCC to include a small prey threshold (below which grazing stops) to prevent phytoplankton from going extinct. WOMBAT and CMOC use a type III response, which is concave upward below $K_{1/2}$ (sigmoidal; Fig S1), thereby offering prey refuge and population stability at low prey concentrations (Gentleman et al., 2008; Oaten et al., 1975). Ecologically, the type III response can be justified as the implicit representation of more complex behavioral changes such as active prey switching (Prowse et al., 2012; Gentleman et al., 2013), which has been observed in multiple zooplankton species (Strom et al., 2000; Kiorboe et al., 1996; Saiz et al., 1995; Kiorboe et al., 2018), or as the mean state of a spatially patchy ocean (Morozov et al., 2010; Rohr et al., 2022), which relatively coarse global models must implicitly average across. Despite qualitatively identical grazing formulations, differences in the parameter selection in CMOC and WOMBAT (particularly $K_{1/2}$; Table 1) mean that zooplankton specific grazing rates on phytoplankton concentrations $< 1 \text{ mmolC m}^{-3}$ are nearly 30-times faster in CMOC than WOMBAT.

OECO-v2 (Hajima et al., 2020) and MARBL (Long et al., 2021) use a slightly more-complex ecosystem structure, resolving multiple groups of phytoplankton but just one generic group of zooplankton that grazes independently on all phytoplankton groups. Both models still employ a single-prey response, in that the presence of one prey option does not influence the zooplankton specific rate at which the other is grazed, g (d^{-1}). However, because the biomass of the generic zooplankton group, Z (mmolC m^{-3}) increases with grazing on any single phytoplankton group, so does the grazing pressure, $gp = gZ/P$ (d^{-1}) on all phytoplankton groups, largely eliminating the possibility of refuge for relatively scarce prey (Sailley et al., 2013). In turn, this structure can obtain an effective food web that is functionally similar to including only one phytoplankton group (Sailley et al., 2013), in which strong top-down controls work to prevent species co-existence and stifle large blooms (Hashioka et al., 2013). This is particularly relevant in MARBL, which uses a type II response, a very small $K_{1/2}$ value, and requires relatively large prey thresholds to prevent extinction (Table 1), whereas OECO-v2 is able to offer prey refuge in the form of a type III response with a relatively large $K_{1/2}$ value.

CanOE (Christianet et al., 2021) is slightly more complex, with two phytoplankton and two zooplankton groups. Here, each zooplankton group grazes independently, pair-wise, on one unique phytoplankton group, with the large zooplankton group additionally grazing on the small zooplankton group. Unlike MARBL and CanOE, in which the indirect interaction strength (Montoya et al., 2009) between phytoplankton groups is always negative, trophic cascades can now lead to positive interaction strength between phytoplankton groups (Sailley et al., 2013; Le Quere et al., 2016). That is, large phytoplankton growth can relieve grazing pressure on small phytoplankton by increasing the biomass of large zooplankton which in turn graze down the biomass of small zooplankton, the sole predator of small phytoplankton.

Increasing in complexity, MEDUSA2.1 (Yool et al., 2021); PISCESv2 (Aumont et al., 2015); BFM5.2 (Lovato et al., 2022), and COBALTv2 (Stock et al., 2020) all include at least 2 zooplankton groups, each with multiple prey. In MEDUSA2.1 and PISCESv2, two types of zooplankton split their grazing effort between two types of phytoplankton in addition to detritus and cannibalism from large to small zooplankton, but both do so with fixed prey preferences. This means that the abundance of one prey option does not influence the desirability of another, such that when prey preferences are equal grazing effort is proportional to the relative distribution of prey (Vallina et al., 2014). Using fixed prey preferences, known as passive-prey switching, ensures maximal feeding (i.e. adding prey biomass, all else equal, always leads to an increase in total ingestion) but does not provide prey refuge for relatively less abundant prey (Prowse et al., 2012). In turn, to help prevent extinction and provide some form of prey refuge, MEDUSA2.1 uses a type III (sigmoidal) response and PISCESv2 includes a relatively strong prey threshold which decreases the apparent prey concentration by half when the total prey concentrations falls below 0.6 mmolC m^{-3} .”

26. 2.2 Grazing pressure – Metric used is ill-defined and ill-motivated. Why not simply calculate the actual grazing rates offline, since you know the grazing functions, and have prey fields and

temperature as outputs in CMIP6? What is gained by the very coarse global averaging? And what are you actually even averaging?

Changed. We have revised the entire manuscript to incorporate an extensive analysis of emergent specific grazing rates (and grazing pressure) calculated offline using the monthly mean prey fields, zooplankton biomass, bacteria, POC, and temperature fields, exactly as suggested. This was a great idea and we believe has thoroughly improved the manuscript. Thank you.

However, we still believe it is important to include what is now referred to as the Prescribed Grazing Index (PGI). Note, this value is now computed using much more realistic assumptions about the standardized prey field – the observed global, median concentration of each PFT (see Table S1).

While the PGI deliberately collapses many modes of variability into a single value, in the revised manuscript we have now shown that it does a very good job of explaining variability in the emergent grazing dynamics, at global, regional and seasonal scales (see **Section 2.4**, our response to **Comment 5**, and further discussion below).

27. Rather than saying you use some unspecified ‘mean’, let us know if you used monthly or annual fields to average, if you are using surface or depth integrals, etc. This very important information is partly (!) given in the Methods, but must be stated in the main text, since else it is not possible to evaluate the suitability of the metric used. And I do not believe I learn so much from that metric, actually. It seems to me that it is only useful to be able to determine the range within which you fit WOMBAT. So for me it is collapsing complexity and variance into one single variable that is only useful for comparison with an NPZD model, which, in my view, doesn’t account for the necessary complexity in zooplankton to even assess the role of higher trophic levels.

Changed. We now describe in detail the observed global median field (0-200 m) used to evaluate the Prescribed Grazing Index (see our response to **Comment 4**).

Moreover, as you suggest, it is conceivable (if not expected) that after collapsing so much complexity into a single number, two models with very different *PGIs* could simulate very similar emergent grazing dynamics via the competing dynamical influence of differences in their food-web structure, active vs passive switching, trophic complexity, capacity for prey refuge, etc. Even models with large differences in emergent specific grazing rates could be tuned to apply similar grazing pressure if sufficiently compensated by differences in mortality and thus the size of the zooplankton population.

This hypothesis is now directly stated and explicitly tested:

Line 93: “Variability in the structure of the food-web and grazing formulation substantially impact ecosystem stability (Vallina et al., 2011; Gentleman et al., 2008), plankton diversity (Chenillat et al., 2021; Prowe et al., 2012; Vallina et al., 2014; Dutkiewicz et al., 2020), NPP (Laufkotter et al., 2015; Prowe et al., 2012), EP (Dupont et al., 2022; Laufkotter et al., 2016; Karakus et al., 2021; Negrete-garcia et al., 2022), regional (Chenillat et al., 2021) and global (Sailley et al., 2013; Le Quéré et al., 2005; Prowe et al., 2012) phytoplankton biomass and chlorophyll distributions, and the phenology of phytoplankton of blooms (Hashioka et al., 2013; Karakus et al., 2022, Truscott 1994). However, while this work has highlighted the influence of qualitative differences in the formulation of grazing, such as trophic complexity (Le Quere et al., 2016; Sailley et al., 2013; Dutkiewicz et al., 2020) and capacity for prey refuge (Chenillat et al., 2021; Prowe et al., 2012; Vallina et al., 2014), less attention has been paid to large differences in the magnitude of grazing rates assumed across models. This may be because differences in the structure of the food-web, value of parameters, shape of the functional response, and determination of prey preferences all exert a multivariate, often competing, influence on emergent grazing dynamics, making it difficult to discern how fast the average zooplankton is assumed to graze. We thus introduce a new diagnostic metric, the Prescribed Grazing Index (*PGI*) to approximate the magnitude of grazing rates prescribed by the innate properties of a given formulation.”

Line 143: “Although decreasing the complexity of the food-web should generally increase the strength of predator-prey interactions (Vallina et al., 2011) and thus increase the *PGI* (as appears true in MARBL and CMOC), OECO-v2 and WOMBAT have the second and third lowest *PGI* despite having the simplest food-webs. This suggests that the influence of qualitative differences in food-web structure can, in some cases, be dominated by quantitative differences in parameter values (Sailley et al., 2013), which vary dramatically (Rohr et al., 2022). Nevertheless, the *PGI* reduces many modes of variability into a single variable. It is plausible then that the emergent dynamics in models do not reflect the large differences implied by their *PGI*. To test the hypothesis that variability in the *PGI* can either capture or dominate the influence of other qualitative aspects of the grazing formulation, such as trophic complexity, prey-switching preferences, non-maximal feeding, and/or prey refuge (Prowse et al., 2012; Le QuereR et al., 2016; Vallina et al., 2014; Sailley et al., 2013), we assess the variability in emergent grazing dynamics (Section 2.3) and determine if it can be explained by variability in the *PGI* (Section 2.4).”

Line 358: “Qualitative differences in the structure of the food web (Sailley et al., 2013; Prowse et al., 2012; Vallina et al., 2011), degree of trophic complexity (Le Quere et al., 2016; Hashioka et al., 2013), sensitivity to temperature (Laufkotter et al., 2015), and capacity for prey refuge (Chenillat et al., 2021; Nissen et al., 2018; Prowse et al., 2012; Vallina et al., 2014) all exert a strong top-down control on NPP and subsequent carbon cycling. Ideally, models that leverage these distinctions to increase biodiversity (Chenillat et al., 2021; Prowse et al., 2012; Vallina et al., 2014; Dutkiewicz et al., 2020), prevent competitive exclusion (Adjou et al., 2012), improve ecosystem stability (Vallina et al., 2011; Gentleman et al., 2008), or replicate species succession (Prowse et al., 2012) and bloom phenology (Hashioka et al., 2013; Karakus et al., 2022), ought also tune their parameters to converge on a relatively constrained degree of bulk grazing pressure. However, this does not appear to be happening. Large variability in emergent grazing dynamics is well explained by the global, regional, and seasonal scales, despite the *PGI* reducing many modes of qualitative variability into a single value. In turn, models with substantial qualitative differences in their grazing formulation and food web structure but similar *PGI* can yield similar zonal distributions of emergent grazing rates (i.e. MEDUSA2.1 vs OECO-v2; Figure 4B) and/or grazing pressure (i.e. CanOE vs PISCESv2; Figure 4E). Together, this highlights how far CMIP6 models remain from converging on a first-order magnitude of zooplankton specific grazing rates, leading to very large uncertainty in the representation of grazing pressure throughout the global ocean.”

28. We define Grazing Pressure, $g(P)$ ($d-1$), as the mean rate at which phytoplankton biomass would be lost to grazing per unit zooplankton in a standardized prey field. – how does the ‘standardized prey field’ reflect nature? Phytoplankton are highly diverse, coexistence of multiple types of plankton is standard, and distributions are patchy and highly variable. How do you reconcile this with your choice of ‘one value fits all’ test fields?

Changed. The standardized prey field is now defined by the observed global median values of each PFT (see Table S1 in our response to **Comment 4**) is not intended to perfectly reflect nature but rather allow us to quantify an innate property of the magnitude of the grazing formulation in a single dimension. In the revised manuscript, we reconcile this diagnostic with the vast heterogeneity of the real and simulated ocean by showing that it is well correlated with emergent grazing dynamics at global, regional, and seasonal scales (see **Section 2.4** and our response to **Comment 5**). This is possible, despite the diversity in the simulated plankton fields, because the first-order differences in the magnitude of the grazing formulation across models are so large that they can be captured with such a simple, coarse, and highly-reduced, metric.

We thank you for pushing us to show this explicitly, which would not have been possible without the offline computation of emergent dynamics you suggested.

29. Grazing Pressure can be used to compare the relative strength of grazing assumed in models with disparate formulations to that implied by empirically estimated functional response curves. – Yes, and so could the actual grazing equations. Why is your approach better?

Changed. In the revised manuscript we:

1. Use the ‘actual grazing equations’ (i.e. Table 1) are used to calculate both, the Prescribed Grazing Index and emergent grazing dynamics (see our Response to **Comment 4**).
2. Use the ‘actual grazing equations’ (i.e. Table 1) and zooplankton mortality equations (i.e. Table S4) are used to compute the emergent grazing dynamics offline.

Line 849: “We analysed the output from historical simulations of 11 CMIP6 models (Table S2). Model (82-92) has been made publicly available by the Earth System Grid Federation (ESGF) and was downloaded from the Lawrence Livermore National Laboratory data node (<https://aims2.llnl.gov/search>). Note, three variables from ACCESS-ESM1.5 (phytoplankton biomass, zooplankton biomass, NPP) were not available on the ESGF-LNLL node and were downloaded directly from the Australian National Computing Infrastructure. For each model we considered monthly climatologies created from the historic period from 1850 to 1900. Considering years up to 1900 provides a long enough time series to average out ENSO variability but stops before there is a substantial anthropogenic forcing, as we are interested in the non-transient dynamics stemming from structural differences in the grazing formulation between models. We used the pre-1990 historical simulation instead of the pre-industrial control because more variables of interest were available across all models. For each model, we downloaded monthly averaged column integrated NPP, export through 100 m, and all depth resolved phytoplankton and zooplankton plankton fields. Depth resolved detritus, bacteria, temperature, and oxygen fields were additionally downloaded when required for offline rate computations. Any exceptions are discussed below.

We then calculated diagnostic rate terms offline, pixel-by-pixel, using a given model's specified equations and parameters (Table 1; Table S4) evaluated with the depth resolved predator and prey biomass fields in addition to depth resolved temperature and oxygen fields where required. The bulk phytoplankton grazing mortality rate ($\$G\$, $mmolC\ m^{-3}\ d^{-1}$), defined as the bulk rate at which all phytoplankton biomass is lost to zooplankton grazing, was calculated by integrating grazing losses from all phytoplankton pools to all zooplankton pools....”$

See Section 4.5 for full text, Equations 2-7, and discussion of any missing biomass output.

3. Compare the empirically determined functional response curves to the *PGI* (rather than emergent dynamics) because both can be thought of as innate properties of a given species or model. We also now compare the observed ship-board observations of grazing to simulated grazing (see our response to **Comment 23, Figure 4**) to compare emergent properties. We think it is useful to consider both.

30. Here, we compute Grazing Pressure as the mean of two end-member scenarios, both with the same standardized, mean, global phytoplankton concentration as in the simple scenario described above, but now with two different distributions. In one ‘segregated’ scenario plankton are assumed to be completely heterogeneously distributed into unique niches with no competition, while in one ‘mixed’ scenario plankton are assumed to be completely homogeneously distributed throughout the ocean. – I don’t see why these two cases are representative of global ocean marine ecosystems. And your explanation below does not convince me. Communities with a very high dominance of one PFT all others occur, perhaps, in bloom-situations at mid latitudes. Why is this interesting? For the majority of the globe and during most times, phytoplankton coexist and/or closely succeed one another, with huge differences in the communities between seasons. You bring up theoretical scenarios, but then you still want to make inferences for the carbon cycle variables as projected by models that aim to represent real plankton systems. You lose me here as to how these two scenarios are relevant beyond a merely theoretical (and thus not easily applicable) knowledge-gain.

Changed. In the revised manuscript we have a) replaced the unrealistic end-member scenarios with the observed global median field (see our response to **Comment 3**) and b) show how the *PGI* correlates with diverse, emergent simulated environments (see our response to **Comment 5**).

31. Experimental estimates of Grazing Pressure can be determined from the functional response curves fit empirically to > 60 laboratory dilution experiments reviewed in the literature (26, 27). –

Experiment absolutely insufficiently explained (here and in Methods). Which organisms are being studied, under which conditions? Here, the prey species matter just as much as the zooplankton predators. List needed.

Change. A more detailed description of the empirical review studies by Hansen et al (1997) is provided along with a more detailed description of how the *PGI* is computed from their empirically derived parameters, and an accompanying Table. Note, as discussed in **Section 4.4** (quote immediately) below, temperature (by us) and prey (by Hansen et al. (1997)) have been controlled for.

Line 643: “The single-prey functional response curve describing how zooplankton specific grazing rates increase with prey abundance can be estimated empirically via laboratory dilution experiments. In these studies, specific grazing rates are measured at different prey concentrations and then typically fit to a type II response function. The “empirical” estimations of the *PGI* in Figure 1B are computed by evaluating empirically estimated functional response curves as if they described the grazing formulation of a theoretical model with 1 pPFT and 1 zPFT. We consider 107 empirically-estimated functional response curves from over 60 species of zooplankton compiled in a review of laboratory work by Hansen et al (1997). They include seven zPFTs: nanoflagellates, dinoflagellates, ciliates, rotifers, meroplankton larvae, cladocerans, and copepods (Table S2). To control for prey quality, Hansen et al (1997) applied a selection criterion to exclude all studies in which dilution experiments were conducted with prey of sub-optimal size, such that the empirical parameters they compare are all assumed to constitute grazing on optimal prey, as is the implicit assumption for models with a single pPFT and zPFT. Temperature, predator volume, and empirically derived $g_{\max}$ and $K_{1/2}$ values from each experiment are reported in Table 3 of Hansen et al (1997). Here, we convert $g_{\max}$ values to units of d^{-1} and control for temperature using a Q_{10} value of 2.8, the mean value reported by Hansen et al (1997). To be consistent with model estimates of the *PGI* (Table S1), all $g_{\max}$ are converted to a temperature of 15.3°C using the relationship

$$g_{\max,15.3} = g_{\max} Q_{10}^{\hat{((15.3-T)/10)}}$$

where $g_{\max,15.3}$ is the temperature dependent saturation grazing rate at 15.3 °C and $g_{\max}$ is the value measured at an experimental temperature of T . $K_{1/2}$ values are converted from biovolume to $mmolC\ m^{-3}$ using the reported carbon density of $0.12\ gC\ cm^{-3}$, which is consistent with the range of carbon densities in phytoplankton (Menden-deuer et al., 2000). All empirical parameters were fit to a type II response. Empirical estimates of the *PGI* are then derived by evaluating the measured single-prey functional response curve for each species at a prey concentration of $0.79\ mmolC\ m^{-3}$, as described above. Note, as per the model evaluated *PGI*, this is an estimate of how fast each zooplankton species would graze on its optimal prey at the global median total phytoplankton concentration. It is not an estimate of how fast each actually graze in the open ocean, but rather a quantification of the innate properties of their empirically defined functional response curves.

The probability density function for the *PGI* of all 107 experiments if overlaid in Fig 1B. The *PGI* reported for each zPFT (Fig 1B, Table S2) is computed using the median $g_{\max}$ and $K_{1/2}$ values for all included species. zPFTs are further grouped into microzooplankton ($<1e4\ um^3$) and mesozooplankton ($>1e4\ um^3$) with their *PGI* computed identically. These coarser groupings do not account for differences between the sample size of each species and their actual global distribution, but do give a general sense of the difference in the *PGI* for broadly categorized micro- and mesozooplankton. We also consider a theoretical model with 2 zPFTs and 2 pPFT, with zooplankton grazing prescribed separately based on the empirically described micro- and mesozooplankton parameters (Table S2). The *PGI* here does take into account the actual distribution of micro- and mesozooplankton (Table S1).

Finally, note that this is not inclusive of all important zooplankton species or functional groups. For example, large photosymbiotic Rhizaria may constitute 50% or all mesozooplankton biomass (Biard et al., 2016) and help explain why oligotrophic food web may be shorter than predicted by classical ecology (Everett et al., 2021). Moreover, neither macrozooplankton nor gelatinous filter feeders are included. Macrozooplankton typically graze slowly on smaller zooplankton but can regulate grazing pressure on diatoms and other pPFTs through carnivorous trophic cascades (Le Quere et al., 2016). On the other hand, larvaceans and salps are extremely abundant mesozooplankton that can filter bacteria and picophytoplankton, unavailable to crustaceans. They generally grow an order of magnitude faster than crustacean zooplankton (Sutherland et al., 2022) and have been shown to play an important role in carbon export via the production of rapidly-sinking carcasses and fecal pellets (Luo et al., 2020, Clerc et al., 2023). Never the less, the empirical values included here are meant to provide context for the two PFTs most typically included CMIP6-class BGC models, not provide an exhaustive list of empirical parameters for all known zooplankton species. Note, Salps (Luo et al., 2020; Clerc et al., 2023) and macrozooplankton (Le Quere et al., 2016; Dutkiewicz et al., 2020) have been routinely included in targeted global simulations, but their inclusion in high resolution climate-ensembles is not yet standard.”

Table S3:

	Zooplankton Group	Studies	Volume (μm^3)	g_{max} (d^{-1})	$K_{1/2}$ (mmolC m^{-3})	<i>PGI</i> (d^{-1})
	Nanoflagellates	18	5.9e1	8.36	23.5	.270
	Dinoflagellates	7	6.3e3	1.63	25.0	.050
	Ciliates	14	2e4	2.76	6.30	.307
	Rotifers	6	2.9e5	2.24	25.2	.068
	Microzooplankton	45	7.1e3	3.71	14.0	.117
	Meroplankton Larvae	16	5.4e6	.12	7.1	.012
	Cladocerans	32	3.1e7	.65	35.4	.014
	Copepods	14	1.4e8	1.44	24.7	.045
	Mesozooplankton	62	2.8e7	.58	21.0	.009

32. These empirical estimates of Grazing Pressure differ by over three orders of magnitude, ranging from $< 0.003 \text{ d}^{-1}$ to $> 3 \text{ d}^{-1}$ and vary widely with zooplankton species and functional group. – yes, as I said above. We need a supplementary table where you show what these 60 experiments really studied.

Changed. We have included a supplementary table (see Table S3 directly above) that reports the sample size and median volume and parameter values for each zPFT. We also now directly reference Table 3 of Hansen et al (1997), where all 107 experiments are documented, but did not copy it in its entirety as it is 3 pages. We do, however, provide a much more detailed description of our derived products and their review (see our response to **Comment 31** above).

33. When zooplankton are grouped into crustacean mesozooplankton (Fig. 1A; orange) and microzooplankton (Fig. - Why? Rationale behind the grouping? Here the issue is that many organisms previously considered microzooplankton (e.g. heterotrophic dinoflagellates) are mixotrophs, that mesozooplankton previously only contained copepods, which are abundant, but new data shows other zooplankton functional groups are closely associated with export (Rhizaria for example, Guidi et al., Nature, 2016), and that macrozooplankton has been shown to be very important in high latitude systems (e.g. Le Quéré et al., Biogeosciences, 2016; Hashioka et al. 201X).

Changed. The empirical PGI is now reported for 9 PFTs (see below). Missing PFTs (i.e. Rhizaria and macrozooplankton) are discussed in the Methods (see our response to **Comment 31** above). We only include empirical measurements from a single review to keep things as consistent as possible, and believe enough PFTs are include to illustrate the uncertainty in model values.

Line 106: “The Prescribed Grazing Index (*PGI*) is defined as the specific rate at which zooplankton in a given model would graze in a standardized plankton community (see “Methods”). It is computed by evaluating each model's grazing formulation (Table 1) at the long-term mean sea surface temperature and median observed global concentration of phytoplankton and zooplankton PFTs (Table S1). The *PGI* in surveyed CMIP6 models ranges from 0.009 d^{-1} in MEDUSA2.1 to 0.637 d^{-1} in MARBL (Figure 1B, Table 1). This means that zooplankton in MARBL would graze ~70-times faster on the observed median plankton community than those in MEDUSA2.1. Even when ignoring the two most extreme models on either side, the *PGI* range is still over an order of magnitude.

Even when ignoring the two most extreme models on either side, the *PGI* range is still over an order of magnitude. To put this in context, we considered the functional response curves measured empirically from 107 different laboratory dilution experiments on >60 species of zooplankton (Hansen et al., 1997) and computed the *PGI* for a theoretical model including one zPFT grazing on one pPFT as described by each set of empirical parameters (see “Methods”). We then sorted the empirical experiments into functional groups and computed the *PGI* for 10 theoretical models, each including 1-2 zPFTs with grazing described by their median empirically measured parameter values (Table S2; see “Methods”). Most models (7 out of 10) have a

PGI larger than that of a theoretical two zPFTs model parameterized using the median empirical meso- and microzooplankton measurements (0.06 d^{-1} ; see “Methods”). This is consistent with the fact that the apparent half saturation concentration for grazing ($K_{1/2}$) for the mean state of a patchy grid cell should be lower than any individual zooplankton grazing in it (Rohr et al., 2022). However, it is inconsistent with the fact that laboratory experiments are highly idealized and consider grazing on exclusively optimal prey. It is thus unclear which bias should dominate. More concerning is the spread of *PGI* in models relative to the empirical estimates.

Ostensibly, inter-model differences should reflect uncertainty in the mean state of their selected functional groups, but not uncertainty in the much larger, full range of species-level diversity. Yet, the *PGI* range across all surveyed CMIP6 models, is roughly as large as the middle 80% of all surveyed zooplankton species ($0.005\text{--}0.55\text{ d}^{-1}$). Only 5 models have a *PGI* within the middle 50% of over 60 species (see Table 3 of Hansen et al., 1997). On one hand, three models (MEDUSA, OECO-v2 and WOMBAT) have a *PGI* equivalent to or lower than a model with only one zPFT parameterized using the median empirical mesozooplankton parameters. On the other, three models (BFM5.3, CMOC and MARBL) have a *PGI* equivalent to or larger than a model with only one zPFT parameterized using the median empirical microzooplankton parameters (Figure 1B, Table S2). While well-mixed, experiments do not exactly represent the dynamics required to recreate the mean state of a patchy ocean, it is concerning that some models imply something statistically similar to an ocean filled entirely with very slow grazing meroplankton larvae (MEDUSA2.1, OECO-v2) and others an ocean filled entirely with very rapidly-grazing ciliates (MARBL, CMOC).”

34. the range of Grazing Pressure assumed in surveyed CMIP6 models spans $84\times - 84\times$ what? Revise.

Changed. The ‘x’ notation has been dropped throughout.

Line 110: “The *PGI* in surveyed CMIP6 models ranges from 0.009 d^{-1} in MEDUSA2.1 to 0.637 d^{-1} in MARBL (Figure 1B, Table 1). This means that zooplankton in MARBL would graze ~70-times faster on the observed median plankton community than those in MEDUSA2.1”

35. substantially less Grazing Pressure than would be expected from an ocean populated exclusively with slow-grazing crustacean mesozooplankton. – why is this your reference here? Role of macrozooplankton? Also, reference needed for mesozooplankton growth rates.

all of which assume Grazing Pressure similar or greater than would be expected from an ocean populated exclusively with fast-grazing microzooplankton. – As above. References needed.

Changed. The revised version of this sentence now references Figure 1B, which is described and referenced in detail in the revised Methods (see our response to **Comment 31** above).

Line 138: “While well-mixed, in-vitro, empirical experiments do not exactly represent dynamics required to recreate the mean state of a patchy ocean, it is concerning that some models seem to imply something statistically similar to an ocean filled entirely with very slow grazing meroplankton larvae (MEDUSA2.1, OECO-v2) and others an ocean filled entirely with very rapidly-grazing ciliates (MARBL, CMOC) (Figure 1B).”

36. models, there does not appear to be a relationship between Grazing Pressure and model complexity. – Claim not backed up by evidence.

Changed. This claim has been clarified by more nuance:

Line 143: “Although decreasing the complexity of the food-web should generally increase the strength of predator-prey interactions (Vallina et al., 2011) and thus increase the *PGI* (as appears true in MARBL and CMOC), OECO-v2 and WOMBAT have the second and third lowest *PGI* despite having the simplest food-webs. This suggests that the influence of qualitative differences in food-web structure can, in some cases, be dominated by quantitative differences in parameter values (Sailley et al., 2013), which vary dramatically (Rohr et al., 2022). Nevertheless, the *PGI* reduces many modes of variability into a single variable. It is plausible then that the emergent dynamics in models do not reflect the large differences implied by their *PGI*. To test the hypothesis that variability in the *PGI* can either capture or dominate the well-established influence of other qualitative aspects of the grazing formulation, such as trophic complexity, prey-switching preferences, non-maximal feeding,

and/or prey refuge (Prowse et al., 2012; Le Quere et al., 2016; Vallina et al., 2014; Sailley et al., 2013), we next assess the variability in the emergent grazing dynamics across models (see Section 2.3) and its relationship to the *PGI* (see Section 2.4).”

37. And even if this were true for different models, what about the same model run with different trophic complexity?

Because the *PGI* explicitly includes/captures differences in the trophic complexity (albeit in a highly reduced manner), we think an otherwise identical model with different trophic complexity would have a different *PGI*, and presumably different emergent grazing dynamics too. Will differences in the magnitude of different grazing formulations be large enough that the reduced *PGI* can reflect differences in their emergent dynamics? We would argue that our results suggest they do.

38. Complexity has been shown to affect food web structure substantially, see Le Quéré et al., 2016, Prowse et al. 2012, Vallina et al. 2014. Can it be that you don’t see this in your metric, because models are highly tuned to get NPP and export right, regardless of their structure? And if so, what would that imply for your statement? In my experience, the addition of prey species decreases grazing pressure in these models (unless KTW is used, see Vallina and Le Quéré, 20XX), and the addition of higher trophic levels strongly affects biomass partitioning in the lower trophic levels (Prowse, Vallina), with tri-trophic interactions strongly altering zooplankton grazing pressure across all zooplankton functional groups represented (Buitenhuis et al., 2019).]

Changed. We have now tested this hypothesis directly by comparing the *PGI* to emergent grazing pressure (see our response to **Comment 5**). Thank you for posing it. While we agree it would make sense for models to tune the parameterization of the grazing formulation to compensate for the reduction in interaction strength associated with increasing food-web complexity, and thus achieve similar NPP while resolving a more complex community composition/biodiversity, this does not appear to be happening. Instead, large variability in the *PGI* appears to drive large uncertainty in emergent grazing pressure which appears to be compensated for from the bottom-up, via the tuning of phytoplankton growth rates (see Section 2.4, quoted in our response to **Comment 5**).

39. While I do not argue that the total grazing pressure, as you calculate it, here may not depend on complexity, as far as you can diagnose it from model with a generally very low PFT complexity, additional complexity has been shown to affect, e.g., surface chlorophyll-a and NPP (Le Quéré et al. 2016).

Revised. We have now included explicit discussion of how additional complexity (in terms of the number of PFTs included, food web connections, and their functional response curves) can influence surface phytoplankton dynamics (see Section 4.1, quoted in our response to **Comment 25**).

Additionally, by evaluating the *PGI* and emergent dynamics across the entire zooplankton and phytoplankton community, both capture differences in PFT complexity. Albeit, the *PGI* does so in a highly reduced manner. However, we now test the hypothesis that variability emergent grazing dynamics can be explained by variability in the *PGI*, despite its simplicity (see Section 2.4, quoted in our response to **Comment 5** as well as our response to **Comment 27**).

40. everything from slow-grazing krill to fast-grazing ciliates – krill is usually considered macrozooplankton

Changed. Krill have been replaced in this comparison with meroplankton larvae, for which we now provide an explicit estimate of their empirically derived *PGI* (see Figure 1; Table S3).

41. satellites, it is unclear what the isolated effect of such dramatic differences in the assumed grazing behavior is on marine carbon cycling. – No. See Laufkötter et al. 2015, 2016. And these models aren't just tuned to fit NPP, they are also tuned to fit EP.

Changed. This sentence has been removed.

42. global, mean annual NPP (NPP; Fig 1B - green) declines rapidly across a narrow range of plausible Grazing Pressure - Well, what is plausible here? Fig. 1A shows a very wide range?

Increasing Grazing Pressure monotonically decreases NPP from a state of complete nutrient utilization to nearly zero. - Absolutely unsurprising. If the model didn't do this, there would be a mistake in your code...

Changed. This sentence has been replaced with:

Line 284: "Unsurprisingly, globally integrated NPP (Figure 6A) and EP (Figure 6C) collapse from complete nutrient utilization to nearly nothing as the *PGI* increases. However, surprisingly this collapse occurs across roughly the same range of *PGI* used in CMIP6 models, further highlighting the degree to which the *PGI* is under constrained."

43. maximum that is determined by the 'bottom-up' parameters that describe phytoplankton growth and is largely invariant to the grazing formulation. – As above. This is exactly as it should be!

Changed. This sentence has been removed.

44. S4K), leading to high depth-integrated phytoplankton biomass – Obvious.

Changed. This sentence has been removed.

45. Pressure, nutrient stress is entirely relieved but NPP approaches zero. – Completely obvious from the equations. What is your target audience? Someone who has never worked with such a model before?

Changed. This sentence has been removed.

46. The decline in export efficiency is driven primarily by a non-linear reduction in particle aggregation, which scales quadratically with each biomass pool (Fig. S3), such that less biomass sinks less efficiently. – In. Your. Model. Please have a look at Laufkötter et al., 2016 as to how other models funnel carbon through the food web. This is not the case for all models, and does not depend on the grazing pressure solely. What about the ballast effect, btw? Has been shown to strongly modify particle export patterns, see e.g. Nissen et al. 202X.

General comment here: Why focus on your own model here only, when we know how export works for a range of the other models already?

changed. We have revised our discussion of export efficiency to both clarify the analysis specifically applies to WOMBAT and also generalized to other models:

Line 293: “Moreover, export efficiency in WOMBAT monotonically declines with *PGI* (Figure S8M). This is driven by a non-linear reduction in particle formation, which scales with the square of each biomass pool (Figure S7), such that less biomass exports less efficiently. Notably though, export efficiency declines even as total biomass initially increases (Figure S8F). This cannot be explained by a decrease in particle formation through sloppy feeding or a decrease in PFT-specific aggregation rates as GSP and the ratio of zooplankton (which aggregate more efficiently in WOMBAT; Figure S7) to phytoplankton both increase. Instead, this decline is dominated by a reduction in the skewness of the biomass distribution as the Z:P ratio approaches 1. This reduces the sum of the squares of the two biomass pools (Figure S8I) and in turn community integrated particle formation. Thus, so long as particle formation occurs non-linearly (Burd et al., 2013), trophic controls on the distribution of biomass can regulate export efficiency despite competing assumptions about the differential rate of particle formation in PFTs with implicitly different size and composition (Laufkotter et al., 2016). This is consistent with the reduction in export efficiency associated with spreading out biomass across longer food-chains chains (Negrete-garcia et al., 2022) and means that a shift to more efficient exporters may not necessarily increase export efficiency if accompanied by a decrease in the skewness of the biomass distribution.”

47. If Grazing Pressure is too low, zooplankton starve because grazing rates are too slow to sustain their biomass. By contrast, if Grazing Pressure is too high, zooplankton starve because overgrazing depletes the food supply before zooplankton biomass has an opportunity to accumulate. Similarly, - Again it feels like you are just explaining the model equations to us, not your findings.

Revised. We have tried to better describe our findings beyond simply stating the obvious.

Line 309: “While grazing pressure (Figure S8L) and ecosystem efficiency (Figure S8M) monotonically increase with the *PGI*, zooplankton specific grazing rates (Figure S8K), zooplankton biomass (Figure S8E) and GSP (Fig 7B) all peak at intermediate *PGI* then decline. This means a larger *PGI* can actually lead to slower emergent grazing rates and less secondary production if overgrazing stifles phytoplankton biomass before it has a chance to bloom. It also demonstrates how NPP is not necessarily correlated with the bulk transfer of mass and energy up the food web. Thus, while future increases in temperature-limited grazing reduce NPP, even during improving phytoplankton growth conditions (Laufkotter et al., 2015), it is not clear how this will translate to higher trophic levels. Likely, different BGC models will yield qualitatively different projections of secondary production based on their initial *PGI* and on what side of the “overgrazing” inflection point they begin. This is particularly problematic for ecosystem (Butenschon et al., 2016) and fisheries models (Tittensor et al., 2021), which are typically driven with a one-way forcing from NPP and assume NPP and GSP have an exclusively positive relationship (Tittensor et al., 2018; 2021), further stressing the need for two-way coupling with primary producers to forecast fisheries stocks accurately (Kearney et al., 2012).”

48. contour), regionally integrated NPP, EP, and GSP all behave similarly to the global mean (Fig. 2D, H, - strange, since it is usually the largest surface areas that determine the global mean? And what we usually see in these models in bioregionalisation attempts is a very clear separation between tropical and extra-tropical regions, that appears in marine ecosystem structure, but also global biogeochemical cycling (consult the many papers on bioregionalisation).

Clarified. ‘Eutrophic’ regions are defined broadly here (see contours on Figure 7), such that the area of the eutrophic swath (which mirrors the global mean) is larger than the oligotrophic one. So, we are in agreement, I believe. The regional size difference is now noted:

Figure 7 Caption: “The boundary between the smaller ‘oligotrophic’ and larger ‘eutrophic’ regions plotted in Figure 6 our contoured in black. The boundary between ‘oligotrophic’ and ‘eutrophic’ regions is contoured in black on each map and defined by the region where GSP is reduced by more than 100% when using the smaller, relative to intermediate, *PGI* (i.e. in panel D)”

Note, these regional delineations are relatively consistent with various other aspects of bioregionalization throughout the literature. e.g. traits of copepods (Brun et al., 2016; Figure 3); simulated and observed Chlorophyll distributions (Le Quere 2015, Figure 4), simulated size-class

diversity (Dutkiewicz 2020; Figure 7), NPP (Laufkötter 2015; Figure 3), zooplankton community composition (Everett et al., 2022; Figure 2)

49. to match observed NPP, which has been estimated from remote sensing at roughly 44–67 PgC/yr (42). – see newer estimates in Buitenhuis et al., Global Biogeochemical Cycles, 2013.

Changed. Our globally integrated NPP estimate has been updated to 53 PgC yr⁻¹ with an uncertainty window of 46–60 PgC yr⁻¹ following the even more recent work of Johnson et al., 2021. This is described in detail in the new ‘Description of observations’ section and compared to other estimates, including Buitenhuis et al. (2013).

Line 780: “Globally integrated NPP (53 PgC yr⁻¹ (46, 60); Figure 5I) is reported per (Johnson et al., 2021), derived from ARGO-float measured deil-cycles in oxygen draw down. This range is consistent with observationally-constrained model estimates (51 PgC yr⁻¹ from Carr et al., 2006; 58 +/- 7 PgC yr⁻¹ from Buitenhuis et al., 2013) and the climatologic (2012–2018) mean Carbon-based Productivity Model (CbPM) remote sensing product (51.1 PgC yr⁻¹, Westberry et al., 2008) when integrated between 60S–60N (as for the models). However, the Vertically Generalized Productivity Model (VGPM) (Behrenfeld et al., 1997) estimates a lower climatologic mean (2012–2018; 60S–60N) of 42.7 PgC yr⁻¹ and some observationally-constrained model estimates reach as low as ~35 PgC yr⁻¹ (Carr et al., 2006). Note, integrating either remote sensing product from 90S–90N only increases the NPP by 1–2 PgC yr⁻¹ and is thus reasonable to compare to full global estimates from (Johnson et al., 2021; Carr et al., 2006; Buitenhuis et al., 2013).”

50. to match observed NPP, which has been estimated from remote sensing at roughly 44–67 PgC/yr (42). – As above. 84x what? And redundant with above.

Changed. The ‘x’ notation has been removed throughout.

51. This is less a reflection of any particular model’s skill, but rather a testament to the extreme sensitivity of carbon cycling to grazing, which remains poorly constrained by observations. – Too easy – please cite relevant work. Have a look at Barton et al., Ecology Letters, 2013 and references therein, e.g. Hansen et al., 1997. And look at the seminal works of Calbet, L&O, 2001, Calbet and Landry, L&O, 2004, as well as Steinberg and Landry, Ann Rev. Mar Sci., 2017 and Landry & Calbet, ICES Journal, 2004.

Changed. This sentence has been removed and all references have been incorporated. In particular, the work of Hansen et al., 1997 is directly incorporated into the observational comparison of the Prescribed Grazing Index (see our response to **Comment 31; Figure 1**) and the works of Calbet and Landry (2004) and Landry & Calbet (2004), as updated by Schmoker et al., 2013, have been directly incorporated into the observational comparison of the emergent grazing dynamics (see our response to **Comments 23; Figures 4 & 5**)

52. pathways, the internal dynamics can yield substantial differences in the efficiency of carbon subsequently transported to depth and higher trophic levels. – Yes, but export is usually also tuned.

Changed: EP tuning has been addressed throughout:

Line 63: “Marine BGC model are tuned to observations of NPP and EP but still remain notoriously under-constrained and over-parameterized parameterized (Schartau et al., 2017; Ward et al., 2013; Doney et al., 1999).”

Line 273: “In short, large differences in the *PGI*, and thus grazing pressure, appear to be tuned out in terms of historical NPP via faster nutrient uptake, but models with similar NPP can have very different standing stocks of phytoplankton (Figure 5E), zooplankton (Figure 5F), total biomass (Figure 5G) and zooplankton to phytoplankton ratios (Figure 5H).

Line 344: “Importantly, though, the primary driver of elevated EP was a larger zooplankton population (Figure 8E) and zooplankton mortality was not tuned in this experiment. Despite nearly identical total biomass in each ensemble (Figure 8F) accommodated by a 38% reduction in phytoplankton biomass (Figure 8D), the ratio of zooplankton to phytoplankton grew by 151% (Figure 8H). This lead to a dramatic skewing of the biomass distribution, an increase in the sum of squares of the biomass pools (Figure 8I), and an increase in the efficiency of particle formation. Conceivably, Zooplankton mortality could be tuned up (alongside a larger *PGI*) to reign in the size of the zooplankton population and constrain EP without modifying grazing pressure or NPP. However, such tuning: a) yields large differences in the size and distribution of biomass (Figure S9) and b) is difficult because zooplankton predation is typically resolved implicitly (Mitra et al., 2009). This later point is evidenced by the inadequate tuning of zooplankton mortality across CMIP6 models (Section 2.3.1) and the established benefit of explicitly adding higher trophic levels (Le Quere et al., 2016).”

Line 373: “In turn, even if models are tuned to the same NPP and EP, they must do so with very different-sized biomass pools and turnover rates to accommodate such large differences in grazing pressure (Figure S9). However, to constrain EP with different-sized biomass pools, models must further differ in their particle aggregation, ballasting, and/or remineralization rates (Laufkötter et al., 2016). Yet each of these terms have different temperature sensitivities, meaning they will respond differently to a warming ocean. this helps explain the large uncertainty in projection of EP (Fu et al., 2016) despite fairly well constrained historical reconstructions (Figures 4; 5J). Therefore, constraining uncertainty in the *PGI*, and in turn emergent grazing pressure and the distribution and abundance of biomass, appears to be a necessary and urgent next step towards constraining predictions of future marine carbon sequestration and transfer to higher trophic levels.”

53. Methods”) and found the sensitivity of all variables to Grazing Pressure increases dramatically when a type II response is employed – see here also Nissen et al., Biogeosciences, 2018.

Changed. This reference has been included:

Line 461: “WOMBAT and CMOC use a type III response, which is concave upward below $K_{1/2}$ (sigmoidal; Figure S1), thereby offering prey refuge and population stability at low prey concentrations (Gentleman et al., 2008; Oaten et al., 1975) and increasing co-existence in the phytoplankton community (Nissen et al., 2018).”

54. lines). For example, the collapse in NPP from 40 to 10 PgC/yr requires a roughly 32x change in Grazing Pressure when using a type III response, but only a 6x change when using a type II. – 32x and 6x what? Reference?

Changed. This text has been revised to:

Line 288: “For example, running WOMBAT (Law et al., 2017) with a *PGI* as large as MARBL (Long et al., 2021) causes NPP to drop by nearly 40 PgC yr⁻¹, all else being equal. Worse, while the range of *PGI* used in CMIP6 models spans nearly two orders of magnitude (70-fold), just doubling the *PGI* at intermediate values (from 0.0125 to 0.25 d⁻¹) can drive a roughly 8 PgC yr⁻¹ drop in NPP, roughly 15% of observed estimates (Figure 5O; see “Methods”).”

55. simulations (45), suggesting that predictions of future ocean carbon sequestration, artificially enhanced or not, are extremely sensitive to how fast we assume zooplankton are grazing. – Yes, but we also tune export. You did not.

Changed. See our response to **Comment 52**.

56. We did not explicitly simulate any change in ocean temperature; however, another way to interpret these results is as the isolated effect of temperature-modified grazing rates on carbon cycling in a warming ocean. – What? Rephrase.

Changed. This entire section has been removed.

57. by roughly two-fold for every 10 °C of warming (i.e. $Q_{10} = 2$, see (8)) – Yes, but models now use

size-specific Q10 functions, see Le Quéré et al., Biogeosciences, 2016.

Changed. This entire section has been removed.

58. by roughly two-fold for every 10 °C of warming (i.e. $Q_{10} = 2$, see (8)). Moving across an intermediate range of values, a 2x increase in Grazing Pressure from $\log_2(g(P)) = -4$ to -3 leads to a 7.6, 1.6, and 1.83 PgC/yr drop in NPP, EP, and GSP, respectively. Not sure I understand this correctly. Why would you assume a ten (!) degree warming here for this estimate? What is your ‘intermediate range’? Be specific.

Changed. This entire section has been removed. However, elsewhere where we reference an ‘intermediate range’, this range has been specified. For example:

Line 289: “Worse, while the range of *PGI* used in CMIP6 models spans nearly two orders of magnitude (70-fold), just doubling the *PGI* at intermediate values (from 0.0125 to 0.25 d⁻¹) can drive a roughly 8 PgC yr⁻¹ drop in NPP, roughly 15% of observed estimates (Figure 5O; see ‘‘ Methods’’).”

59. translates to a 2.7%, 2.9%, and 1.5% decline in productivity per degree of warming, - Not really valid, as many other metabolic rates also scale with the same Q10, and among these, phytoplankton growth is one of them. Caveat this finding.

Changed. This entire section has been removed.

60. perspective, that is nearly another PgC/yr that would need to be removed to counteract the feedback of warming on zooplankton grazing rates by – I don’t buy this, since phytoplankton growth scales with similar temperature dependencies.

Changed. This entire section has been removed.

61. From a fisheries perspective, GSP declines by 1.5%/ °C despite a 1.45%/ °C increase in ecosystem efficiency. – Same here – what are the assumptions here? That only grazing scales with temperature? We know that this isn’t true. And lack of precision in how calculations are described.

Changed. This entire section has been removed.

62. The whole climate change topic also has data-based hypotheses to be considered: Brun et al., Nature Ecol. Evol., 2019, for example. Please include in your discussion.

Changed. This entire section has been removed.

63. Crustacean mesozooplankton dominant current estimates of global zooplankton biomass (22, 56), but they are not the only important functional group – dominate instead of dominant, important group for what? And please cite relevant work such as Buitenhuis et al., ESSD, 2013 and Le Quéré et al., Global Change Biology, 2005.

Changed. This sentence has been removed and references to other PFTs and their functional role have been incorporated throughout. See, for example, **Line 383 quoted** in our response to **Comment 23**.

64. Therefore, it is reasonable to consider that the best first order estimates of global mean Grazing

65. Pressure in models are those that fall between what is predicted empirically for an ocean of relatively slow-grazing mesozooplankton and faster-growing microzooplankton, i.e. a mix of meso- and microzooplankton. – This seems obvious. But can you cite the relevant sources? Evidence on relative biomasses of different zooplankton functional groups (Bar-On et al., Science, 2018; Buitenhuis et al., ESSD, 2013)?

Changed. This sentence has been removed. However, we have now incorporated observations of the relative biomasses of different zooplankton functional groups from Buitenhuis et al., ESSD, 2013 and the entire MAREDAT data set extensively in our revised definition of the Prescribed Grazing Index and derived observational estimates included in **Figures 2, 4, 5 and S4-6**. See our response to **Comment 3**.

66. may be adequately resolving the variability and the associated functional diversity therein. – Cite appropriate references, e.g. Benedetti et al., Ecography, 2022, Brun et al. 2016 (data base), Barton et al. Ecology Letters, 2013 and so on.

Changed. This sentence has been removed but these references have been incorporated throughout (See, for example, **Line 383 quoted** in our response to **Comment 23**).

67. Mesozooplankton, which generally graze much slower than microzooplankton (Fig. 1A), are able to consume much larger prey, such as dinoflagellates and diatom assemblages - What? Many copepod species are omni- or carnivores, see Bendetti et al. 2015, Benedetti et al., Ecography, 2022.

Changed. This sentence has been revised as to not imply that mesozooplankton cannot also be omnivores or carnivores:

Line 410: “This may increasingly favor rapidly-grazing microzooplankton or gelatinous filter feeders, which can ingest smaller phytoplankton size classes (Conley et al., 2018; Heneghan et al., 2023), over crustacean mesozooplankton, which generally graze slower but can consume much larger prey (Everett et al., 2021; Barton et al., 2013).”

68. In turn, mesozooplankton and microzooplankton tend to dominate different niches, thriving in more and less productive biomes, respectively (58, 59). – Bad referencing. And is this true? Check the (biomass) data to support this claim.

Changed. This sentence has been removed.

69. This plays an critical role in regulating the biological pump but is rarely resolved (55). – NEMURO had DVM a decade ago. See also Aumont et al., Global Biogeochemical Cycles, 2018, Bianchi et al., Global Biogeochemical Cycles 2013 and multiple others.

Changed. References to Aumont et al. (2018), Bianchi et al. (2013) and Steinberg and Landry (2017) have been added.

Line 413: “Considering these groups serve critically different roles in trophic transfer through prey preference (Biard et al., 2016) and carbon export through diel-vertical migration and (Bianchi et al., 2013; Steinberg and Landry, 2017; Archibald et al., 2019, Aumont et al., 2018) the formation of rapidly-sinking fecal pellets and carcasses (Luo et al., 2020; Clerc et al., 2023), it is essential that models include enough trophic diversity to capture present-day and emergent changes to community composition and its associated traits (Heneghan et al., 2023).”

70. Other groups could be equally important to resolve. – Cite community recommendations.

For example, larvaceans and salps are extremely abundant mesozooplankton that can filter bacteria and picophytoplankton, unavailable to crustaceans, and generally grow an order of magnitude faster than crustacean zooplankton – But then look at de Vargas et al., Science, 2016 and Guidi et al., Nature, 2016. Lots of other and unexpected groups associated with carbon cycle dynamics, e.g. rhizaria.

Changed:

Line 402: “Such comparisons will enable modellers to determine if explicit competition in their assumed food-webs are reasonably approximating natural variability in community-averaged grazing dynamics, or if increasing trophic complexity by including, for example, macrozooplankton (Le Quere et al., 2016), salps (Luo et al., 2020), larvaceans, euphausiids, chaetognaths, jellyfish (Heneghan et al., 2020; 2023), pteropods (Buitenhuis et al., 2019), heterotrophic protists (de Vargas et al., 2015) or Rhizarians (Biard et al., 2016; Guidi et al., 2016) might be required.”

71. In turn, improving the representation of zooplankton grazing could result in dramatic improvements in marine BGC models and our predictions of future ocean states. – Agree, but you are not the first to diagnose this. Cite relevant references.

Changed. This sentence has been removed but the relevant references are now cited in a similar sentence in the introduction:

Line 66: “One of the most important mechanistic uncertainties in IPCC-class marine BGC models is the representation of zooplankton grazing (Dupont et al., 2022; Gehlen et al., 2006; Hashioka et al., 2013; Laufkötter et al., 2015, 2016; Le Quéré et al., 2016; Negrete-Garcia et al., 2022; Nissen et al., 2021; Prowe et al., 2012, Vallina 2014; Dutkiewicz et al., 2020).”

72. As stratification under climate change likely leads to greater light availability but fewer nutrients, the phytoplankton community may shift toward smaller phytoplankton with faster nutrient uptake but higher light requirements (53). – Already found in Bopp et al., GRL, 2005 (!).

Changed. Reference to Bopp et al. (2005) added.

Line 407: “Moreover, as stratification under climate change likely leads to greater light availability but fewer nutrients, the phytoplankton community may shift toward smaller phytoplankton with faster nutrient uptake but higher light requirements (Portner et al., 2019; Bopp et al., 2005).”

73. but we believe further attention to grazing could substantively improve simulated carbon cycling and projections of both future climate scenarios and global food production. – Cite relevant references (see above). And redundant with previous.

Changed. This sentence has been removed and the relevant reference have been added. See for example:

Line 66: “One of the most important mechanistic uncertainties in IPCC-class marine BGC models is the representation of zooplankton grazing (Dupont et al., 2022; Gehlen et al., 2006; Hashioka et al., 2013; Laufkötter et al., 2015, 2016; Le Quéré et al., 2016; Negrete-Garcia et al., 2022; Nissen et al., 2021; Prowe et al., 2012, Vallina 2014; Dutkiewicz et al., 2020).”

74. With increasing computational capacity, recent BGC model development has largely focused on expanding the complexity of the phytoplankton community (65). – I disagree, see Buitenhuis et al., Global Biogeochemical Cycles, 2019, Henson et al., Science Advances, 2021 (see model structure in

supplement), Clerc et al. 2021 and many more.

Changed. This sentence has been removed.

75. However, the functional traits and population dynamics of the mean phytoplankton community may actually be more sensitivity to variability in how they are grazed, rather than how they grow (66) – cite Hashioka, Nissen, multiple papers by Vallina, Prowe and many more! This has been shown!

Changed. This sentence has been removed but these references have all been incorporated extensively.

Methods

76. S3. We did not use more a complex model, with multiple plankton pools or preferential feeding, as we believe that we must first better understand the impact of grazing in a relatively simple global BGC model. – This, in my view is a flawed decision, probably based on a lack of overview over the literature, as highlighted in my general comment above. We know a lot more about the effect of zooplankton than what you suggest in these models, and you are not taking into account what we know to a sufficient degree. A more thorough literature search will allow you to revise the focus of your work, to better place your paper into the context of the current literature, and to identify the elements of novelty that your study brings in a more convincing way.

Changed. We now explicitly consider the emergent dynamics of 11 CMIP6 models (Table S2).

Table S2:

	BGC Model	Earth System Model	Simulation	Model ID	Institution (Location)	Personal Communication
	iHAMMOCC (64)	NorESM2-ESM-LM (64)	Historical (82)	r2i1p1f1	NORC (NOR)	Jerry Tjiputra Jörg Schwinger
	CMOC (65)	CanESM5 (69)	Historical (83)	r1i1p2f1	ECCC (CAN)	Jim Chirstian
	WOMBAT (66)	ACCESS ESM1.5 (66)	Historical (84)	r1i1p1f1	CSIRO (CAN)	Andrew Lenton Mathew Chamberlain
	OECO-v2 (67)	MIROC-ES2L (67)	Historical (85)	r1i1p1f2	JAMSTEC (JAP)	-
	MARBL (68)	CESM2 (78)	Historical (86)	r1i1p1f1	NCAR (USA)	Matt C. Long
	CanOE (69)	CanESM5 (69)	Historical (87)	r2i1p2f1	ECCC (CAN)	Jim Christian
	MEDUSA2.1 (70)	UKESM1-0-LL (79)	Historical (88)	r1i1p1f2	Met Office, NERC (UK)	-
	PISCESv2 (71)	IPSL-CM6A-LR (80)	Historical (89)	r10i1p1f1	IPSL (FR)	-
	PISCESv2 (71)	CNRM-ESM2.1 (79)	Historical (90)	r1i1p1f2	CNMR (FR)	-
	BFM5.2 (72)	CMCC-ESM2 (72)	Historical (91)	r1i1p1f1	CMCC (ITA)	Tomas Lovato
	COBALTv2 (73)	GFDL4.1 (81)	Historical (92)	r1i1p1f1	GFDL (USA)	Charlie Stock

Throughout we consider their complexities and test the hypothesis that large difference in the magnitude of their grazing formulation, as expressed by the PGI, can explain variability in their emergent dynamics. See our response to **Comments 4, 5, 25, and 27**.

77. Finally, note that WOMBAT quantifies biomass in nitrogen units and uses disk parameters to describe the functional response – I do not know what disk parameters are? Reference needed.

and then from Michaelis-Menten parameters into disk parameters (e – Lack of equations and references.

Changed. This has been revised to:

Line 1036: "... then from Michaelis-Menten parameters into disk parameters (e and $g_{\max}$) using $e = g_{\max}/K_{1/2}^2$ for a type III response and $e = g_{\max}/K_{1/2}$ for a type II response. For detailed information on the difference between the two mathematically equivalent parameter schemes see Rohr et al. (2022)."

78. by a model, we define Grazing Pressure, $g(P)$ (d⁻¹), as the mean rate at which phytoplankton biomass would be lost to grazing per unit zooplankton in a standardized prey field. - this is an extremely poor description of the main variable of your entire work. What are you averaging over? Surface only or depth-integrated? Annual or monthly? Which years and time period, or single year? Resolution?

Changed. We have completely revised the definition and description of the Prescribed Grazing Index. See our response to **Comment 4**.

79. The key is describing the standardized prey field such that it remains as standardized as possible across different model configurations, with different numbers of phytoplankton, zooplankton and prey groups (see Table 1). To start, we assume a constant, standardized, mean, global, phytoplankton concentration, P mmolC/m³, which is inclusive of all integrated phytoplankton groups in a given model. – as above. – First sentence is an tautology. And again, what are you calculating? Estimates made per model grid point? Annual? Monthly or daily average?

Changed. See our response to **Comment 4**.

80. As such, P must be chosen to be generally representative of the global mean phytoplankton concentration that might be expected to occur across all models being compared. - So what was chosen? And I disagree – the global mean isn't representative. You have oligotrophic and eutrophic regions, with each covering different fractions of the global surface ocean. And in the mid-latitudes, you have huge seasonal differences. So how can this one number be representative of anything. Uncertainty analysis/range bars for real open ocean conditions would be helpful, and may convince me that your metric is actually useful to describe CMIP models that attempt to describe real marine ecosystems.

Changed. We now:

- a. Use the globally observed median abundance (and distribution) of phytoplankton to evaluate the PGI (See our response to **Comment 4**).
- b. Provide uncertainty bars on the PGI (i.e. Fig 5; See our response to **Comment 4**).
- c. Show the PGI is well correlated with the globally averaged emergent grazing pressure and specific grazing rates (Fig 5A, B; See our response to **Comment 5**)
- d. Show the PGI is well correlated with the zonally averaged grazing pressure and specific grazing rates at most latitudes during both winter and summer (Fig 4C, F; See our response to **Comment 5**).

81. To take into account multiple zooplankton, Grazing Pressure is averaged across all zooplankton groups, which are assumed to exist in equal proportions. - Why? Differences between the zooplankton groups are vital for global biogeochemical cycling (Si, N, CaCO₃, DMS,...). And these groups do not

exist in similar proportions. Why don't you use the actual equations and simulated biomass ratios at each grid point to study the grazing pressure across models?

Changed. We have made it explicit that:

1. We do now use the actual equations (Table 1), both to compute the PGI and the emergent grazing dynamics (See our response to **Comment 4**).
2. The explicitly simulated biomass ratios at each grid point are used to compute the emergent grazing terms (see Section 4.5, quoted in part in **Comment 29**)
3. The observed global median ratios (Table S1) are now used to evaluate the PGI (See our response to **Comment 4**).

82. as the mean of two end-member scenarios, both with the same standardized, mean, global, phytoplankton concentration as in the simple scenario described above, but now with two different distributions. – The mean of two very different scenarios is statistically not very telling, and the mean scenario across ensemble members with totally different assumptions is likely to be neither representative of one, nor of the other scenario, and, more importantly, not representative for the real ocean either. This seems nonsensical to be. Need a far more exhaustive supplementary to back up your methodological decisions.

Changed. We now compute the PGI using the observed global median biomass distributions and have completely removed the 'end-member scenario' analysis. See our response to **Comment 4**.

83. In models where non-plankton prey (i.e. detritus, bacteria) is available, all other prey types are assumed to exist in equal concentrations to phytoplankton, that is P. – Arbitrary and utterly unmotivated assumption. You could have chosen these to be Z. And which models and which non-plankton prey species did you encounter in your work?

Changed. We now compute the PGI using the observed global median biomass distributions for plankton as well as bacteria and POC. Model food webs are details in Figure 1 and Table 1 and described in Section 3.1. Briefly, MEDUSA2.1 and PISCESv2 include POC as prey and BFM5.2 and COBALTv2 include bacteria. Both are accounted for in computation of emergent rates as well as the PGI. See our response to **Comment 4**.

84. Neither distribution is realistic, but together they offer useful bounds on how fast zooplankton could be expected to graze down phytoplankton in given model, without having to integrate it. – If they offer bounds, and both aren't realistic, why (a) were they chosen, and (b) why would you average them? And what don't you want to 'integrate'? Need to motivate, and justify and test and evaluate your methodological choices.

Changed. We now compute the PGI using the observed global median biomass distributions and have completely removed the 'end-member scenario' analysis. The sensitivity of this metric to the prey field it is evaluated with is tested in Section 4.3 (See our response to **Comment 4**) and its ability to predict variability in emergent properties of the models is tested in Section 2.4 (See our response to **Comment 5**)

Note, by 'integrate' we were distinguishing emergent properties of a given simulation from innate properties of the model structure. Hopefully the new methodological description (see above) is much clearer.

85. That is, at a given location we assume one phytoplankton and one zooplankton have out competed all other plankton groups, driving them to extinction. – Essentially, it is exactly this behaviour that any such model I have seen tries to avoid at all costs... I am at a loss why this would be your underlying assumption? This is a complete simplification of the food-web, with known impacts on biomass dynamics, see references above.

Changed. These assumptions have been removed in favour of the observed global median. See our response to **Comment 4**.

86. That is, the concentrations of the dominant phytoplankton, as well as any non-plankton prey options, are equal to P, while all other phytoplankton or zooplankton concentrations are assumed equal to 0. – Why? Then, it could be that grazing pressure is assessed based on grazing on detritus and one prey species alone? How is that representative of real ocean conditions?

Changed. These assumptions have been removed in favour of the observed global median. See our response to **Comment 4**.

87. In both scenarios, to take into account model differences in temperature limitation Grazing Pressure is computed at 16°C , roughly the global, pre-industrial mean sea surface temperature (79). - Surface ocean mean temperature is 16°C ? Which depth range was taken? Why? Zooplankton live in upper few 100 m of the ocean, phytoplankton mostly in upper 100 m... so which temperature is relevant here? Justify your assumptions.

Changed. This has been clarified to:

Line 717: “Temperature and detritus are also needed to compute the *PGI* for some models, but are not included in the MAREDAT database. In models in which grazing is temperature-limited, rates are computed the global (-60°S - 60°N) mean long-term average (1854-2022) sea surface temperature from the NOAA ERSSTv5 product (Huang et al., 2017).”

Line 572: “Note that while these values are deliberately agnostic to any specific model run, thereby providing a standardized field to obviate other inter-model differences and isolate innate properties of the grazing formulation, they compare well to the inter-model mean of the median simulated biomass fields (Table S1). In both models and observations, the concentration of biomass increases from total zooplankton to bacteria to total phytoplankton to detritus, and the relative distributions are generally consistent. Finally, the zooplankton biomass-weighted temperature from the models (i.e. the temperature the median simulated zooplankton experiences) is close to the standardized observed surface temperature used (15.30 and 15.61°C , respectively). Together, this confirms that the standardized observed values are representative of the median values simulated across models.”

Note, all offline, emergent grazing rates were computed using the model-specific, depth resolved, temperature fields and are completely agnostic to this assumption, which is only relevant to the computation of the *PGI*.

88. Here, we set P equal to the mean phytoplankton concentration simulated across all simulations conducted in our study, $P = 0.56 \text{ mmolC/m}^3$. – What environment is this representative for?

Changed. We now use the MAREDAT observed global median of 0.79 mmolC m^3 .

In Figure 4, we now show that the *PGI*, evaluated at the global median, is a strong predictor of emergent grazing dynamics in many environments, with very different phytoplankton concentrations

(Figure S4H). For example, there is still a statistically significant relationship between the PGI and emergent specific grazing rates at relatively high latitudes during the winter (Figure 4F, left panel).

89. This is appropriate for comparing the influence of Grazing Pressure between our runs; however, we note that this value is not necessarily representative of the mean phytoplankton concentration in other simulations or the values observed from space. – So. Why. Was. A. Nonsensical. Value. Chosen???

Changed. This sentence has been removed. But what we now show is that even though the phytoplankton concentration the PGI is evaluated with is not broadly representative of all times/places, the influence of the PGI largely is. This is due to the dominate influence of large differences in the ‘magnitude’ of the grazing formulation across models. See our response to **Comments 4 and 5**.

90. These values were calculated using roughly 18 years (July 2002 - April 2021) of data from the Carbon based Productivity Model (42). - Why this model? Justify your choices.

Changed. We now use the MAREDAT estimate and support it citing its proximity to the remote sensing (CbPM) value.

Line 709: “... estimates of total phytoplankton ($.79 \text{ mmolC/m}^3$) are generally consistent with independent of estimates of the median phytoplankton biomass ($.97 \text{ mmolC/m}^3$) estimated from 18 years (July 2002 - April 2021) of remote sensing data from the Carbon-based Productivity Model (Westberry et al., 2008)”

91. S2). It is important to reiterate that the absolute value Grazing Pressure for a given model is of little consequence and necessarily varies with the value of P. – The repeated statements that you methodological choices ‘do not matter’ and ‘have little consequence’ and ‘aren’t supposed to represent the real world’ (see throughout manuscript) do not increase my confidence in your analysis. To me, this means that perhaps you haven’t sufficiently tested this and you lack the ambition to understand how real system works, since very little robustness analysis is presented in your supplement, and you also state repeatedly that your findings aren’t relevant for real-world problems. Yet, you claim we learn more about CMIP models and their capacity to simulate future carbon cycle characteristics.

Here were attempting to highlight the importance of the relative range of in the PGI across models, rather than the absolute value of a given model. But more importantly, we have heavily revised the analysis and hope you find it much more well founded and useful. While we would argue some misunderstandings throughout may have presented as a lack ambition (rather than an actual lack of ambition), we never-the-less sincerely thank you for pushing us to go much further with this analysis and think it has greatly improved the results.

92. Fig 1: Figure must be stand-alone. Revise figure caption and color schemes. We do not know how the empirical estimates have been calculated (need to describe this in your Methods). We do not know what was measured, not what was simulated, and neither by which model. B must tell us what we simulated by which model(s). Colors unsuitable for the many people affected by red-green blindness. And I would like to see variability envelopes around each of your dots.

Figure 1 has been entirely revised and its components now appear the revised Figure 1 and 6.

In Figure 1:

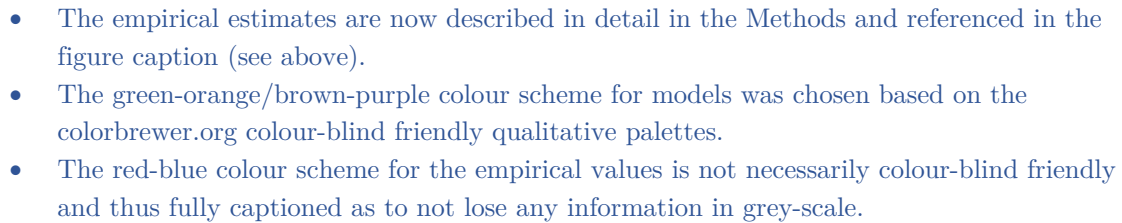


Figure 1: Variability in the formulation of grazing. **A)** The marine food webs represented in 10 CMIP6-class BGC models are presented in clockwise-order of increasing complexity. Aside the name of each BGC model is the number of parameters required to describe grazing. Grazing relationships (arrows) are solid for single prey responses, dashed for multi-prey responses with fixed preferences, and dotted for multi-prey responses with active switching. Red arrows indicate temperature sensitivity. Green, red/orange, and purple color schemes refer to models with 1, 2, or actively switching zooplankton. Hue scales qualitatively with complexity. PFTs have been generalized into small (nano-, small, non-diatom, or nanoflagellate), large (large, diatoms), or diazotrophs for phytoplankton (P) and small (micro-, small), medium (meso-, medium), or large (macro-, large) for zooplankton (Z). Bacteria (B) and Detritus (D) are included when available as prey. Note, model specific nomenclature for PFTs is included in Table 1. **B)** The Prescribed Grazing Index (*PGI*), an innate property of the grazing formulation, is plotted for all models below the *PGI* associated with empirical estimates of the functional response for 7 zPFTs reported in the review of (36). We further include three theoretical model configurations: two with a single zPFT described based on the median empirical parameters for meso- (dark red) and microzooplankton (blue blue), and one with 2 zPFTs including both (see “Methods”). Overlaid is the PDF of *PGI* for all 107 survey experiments.

Changed. In Figure 6:

- NPP, GSP, and EP are separated into separate panels to be colour-blind friendly
- Uncertainty envelopes have been added to illustrate variability in the size of the seasonal cycle
- Error bars have been added to indicate the sensitivity of the PGI to the prey field with which it is evaluated.
- A second set of panels is included to illustrate regional variability
- Note, the seasonal uncertainty bounds and PGI error bars are omitted from the lower panels showing the regional differences because they are qualitatively consistent with the global panels above and made the figure very cluttered/difficult to interpret.
- The Methods are referenced for a detailed description of each simulation.

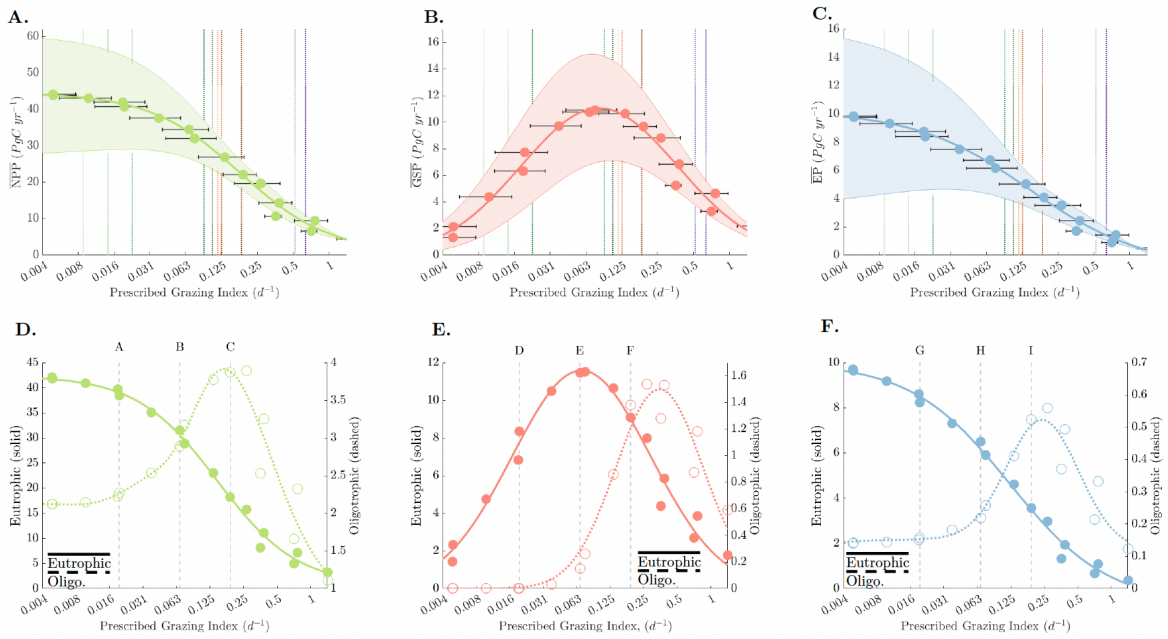


Figure 6: Global, annually averaged **A)** NPP, **B)** GSP, and **C)** EP from each simulation of WOMBAT (see “Methods”) is plotted against the *PGI* of the corresponding model. Horizontal error bars refer to the precision of the *PGI*, per Figure 5. The vertical uncertainty window is proportional to the amplitude the seasonal cycle, computed as $\pm 25\%$ of the difference between the annual max and mean. Vertical dashed lines refer to the *PGI* of 11 CMIP6 models, color-coated per Table 1. Regional, annually averaged **D)** NPP, **E)** GSP, and **F)** EP are plotted separately for eutrophic (left axis) and oligotrophic (right axis) regions. The boundary between ‘oligotrophic’ and ‘eutrophic’ regions is contoured in black on the maps in Figure 7. Vertical dashed lines denoted which runs were plotted in the corresponding panels of Figure 7.

93. Fig. 2: Figure must be stand-alone. Revise figure caption and color schemes. Which models are you using for this figure? Same issue of red-green blindness unfriendly color scheme.

Changed. Figure 2 has been separated into Figure 6 (see above) and figure 7 (see below).

- We retained the red-green colour scheme but ensure that both colours never appear in the same panel, and more generally, work hard to provide either colour-blind friendly palettes or appropriate labelling throughout.
- Caption revised to specify these are from the WOMBAT sensitivity simulations, reference the description of simulations in the Methods, and point to Figure 6 which specifically notes which simulations are being mapped.

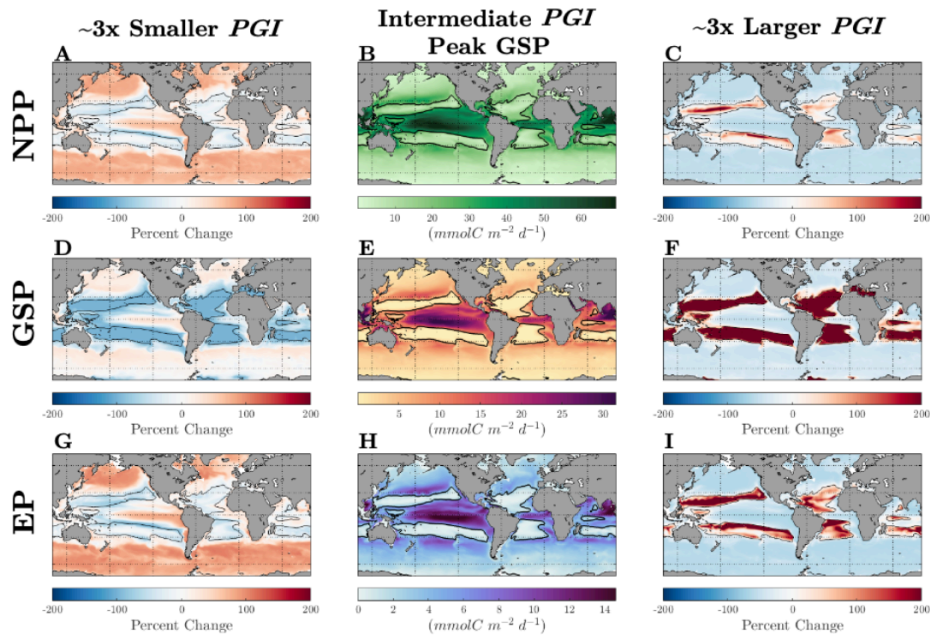


Figure 7: Regional sensitivity of marine carbon cycling to the *PGI*. Maps of the distribution of mean annual NPP (top row), GSP (middle), and EP (bottom) from three simulations from the ‘Slow Turnover’ suite of WOMBAT experiments are plotted (see “Methods”). The specific simulations are noted in Figure 6. In the second column the raw values are reported from the run with roughly peak $\overline{\text{GSP}}$ and an intermediate *PGI*. To the left and right are the percent deviation that occur when a $3.4\times$ smaller (left) and $2.9\times$ larger (right) *PGI* is used.

94. Fig. 3: Why are some axis titles written in red color? This looks like a supplementary figure to me. Color scheme is better.

Changed. Figure 3 was replaced with the more comprehensive Figure 8 per the request of Review 2, who wanted more supplemental aspects to be brought to the main manuscript. While we couldn’t honour both requests, we note that this analysis does now take a backseat to that of the emergent grazing dynamics in Fig 2-5. Also note, now:

- All traces are plotted separately
- Axis titles are all in black and refer to a single variable
- Green and red are both still used, but never on the same panel. No information is lost in grey scale.
- Methods are again referenced for a description of which model runs are being plotted.

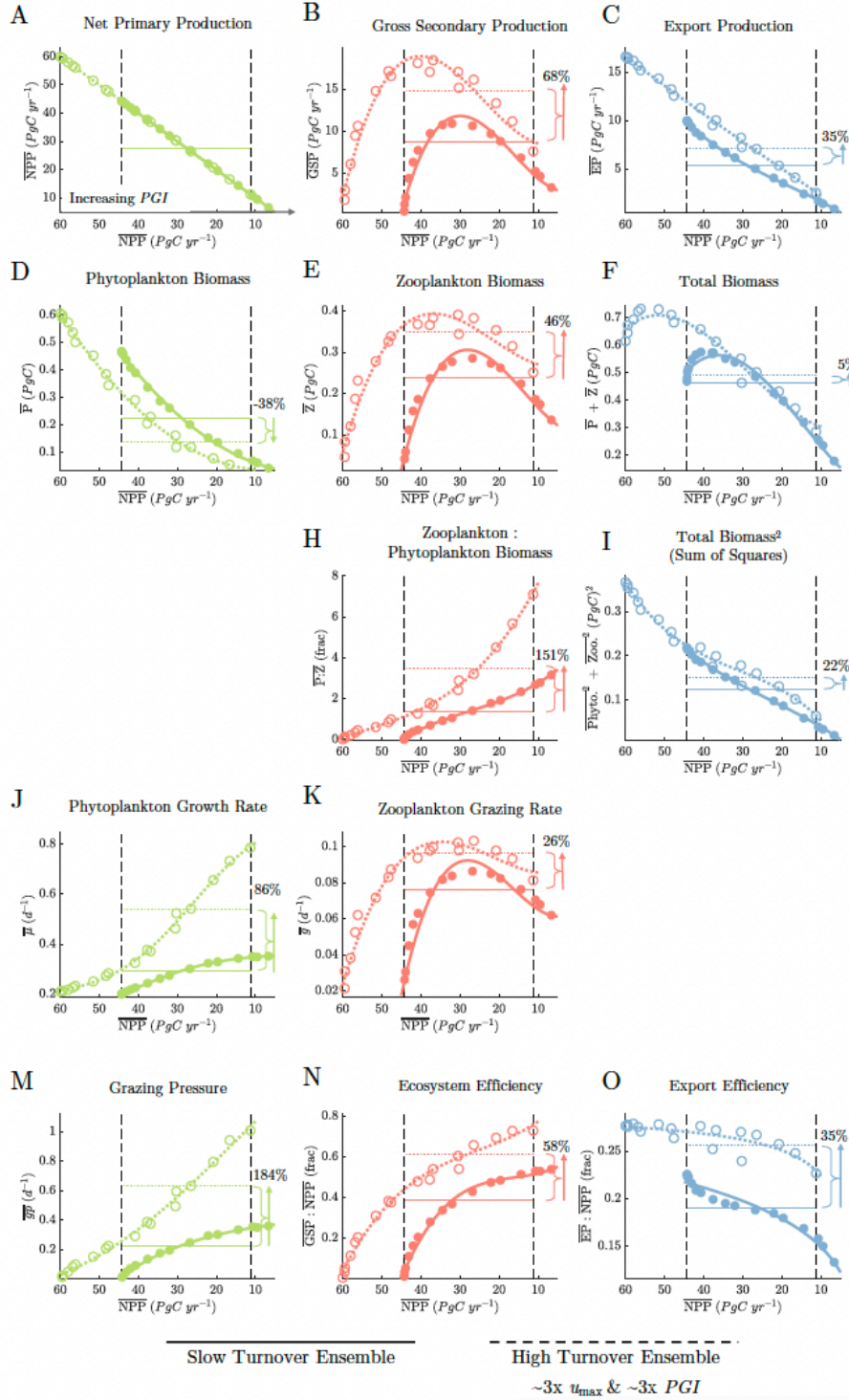


Figure 8: Sensitivity of marine carbon cycling to model tuning. **A-O** Various aspects of global, annually averaged marine carbon cycling are plotted against NPP for two suites of experiments. The ‘Slow Turnover’ WOMBAT experiments (solid lines) are the same simulations shown Figure 6 and S8. The ‘High Turnover’ WOMBAT experiments (dashed line) were parameterized with faster phytoplankton growth rates but are otherwise identical (see “Methods”; Table S4). Dashed vertical line bracket the ensemble of runs used to quantify the difference (adjoining horizontal lines) between ‘Slow Turnover’ and ‘High Turnover’ (adjoining horizontal lines) simulations (see “Methods”). The x-axis is flipped to plot increasing *PGI* from left to right, consistent with Figures 6 and S8.

95. Fig. 4: Again, state which model this figure is based on. Revise color scheme.

Changed. Figure 4 now appears as Supplemental Figure S12 and the caption has been revised to reference Table S3 and the methods where the simulations are described. Again, the colour scheme, was kept but no information is lost in grey-scale.

96. Table 1: Very nice. Revise color scheme. See colorbrewer.org.

Changed. Thank you. Per consultation with colorbrewer we did our best to select a colour-blind friendly palette for the CMIP6 simulations but note that they offer no completely colour-blind safe qualitative schemes for 11 different items. For the food-webs, we ensured they were all appropriately labelled as to lose no information in grey-scale.

97. Additional figs: I'd like to see a schematic that summarises the relevance of your findings with WOMBAT for CMIP models and future projections.

Changed. Instead of extrapolating the WOMBAT analysis to the other models through a schematic we now include 4 Main figures (and 7 supplemental figures) explicitly analysing the emergent properties of the other CMIP models. Figure 1 provides a schematic highlighting the range of different Grazing formulations and their PGIs relative to empirical estimates.

Reviewer #2 (Remarks to the Author):

First and foremost, we would like to sincerely thank Reviewer 2 for their constructive, thoughtful and extremely helpful comments. In particular, we believe their suggestion to investigate the balance between zooplankton mortality/biomass and specific grazing rates in driving grazing pressure has substantially improved the manuscript. Please see below for our point-by-point response to specific comments and note we have numbered your comments for reference in our response.

The authors present an interesting manuscript on an important topic. I find their discussion of the differences between grazing parameterizations in IPCC-class models to be enlightening.

Thank you.

However, I think there are substantial issues with the authors' attempt to simplify the different models into a single metric. As I describe below, I do not find this metric useful.

In the revised manuscript we have worked extensively to be construct, refine, support and incorporate this metric. In doing so we have:

1. Changed the name from 'Grazing Pressure' to 'Prescribed Grazing Index' (*PGI*) to more clearly differentiate emergent grazing pressure, defined per Prowe et al., (2012) as the phytoplankton specific loss rate to grazing (i.e. $g(P)Z/P$) from our metric, which is innate to a model's prescribed grazing formulation and intentionally agnostic to all other aspects of the simulation that might influence grazing (i.e. bottom-up dynamics, physics, zooplankton mortality).
2. Refined the definition and derivation of the Prescribed Grazing Index to better reflect realistic community structure by using observed global mean fields, rather than averaging across unrealistic endmember scenarios as in the original manuscript (see Section 4.2)
3. Incorporate an extensive analysis of the emergent grazing dynamics and zooplankton mortality rates calculated offline pixel by pixel, using the monthly mean model output (See Section 2.3, 2.4, & 4.2). These terms were calculated for 50 years of historical output for all 11 CMIP6 using their explicitly described grazing formulation (Table 1) and mortality scheme (Table S3).
4. **Importnatly**, show that despite reducing many dimensions of variability and important qualitative differences in food-web structure, prey refuge, and active/passive switching, into a single metric, the Prescribed Grazing Index is actually a very strong predictor of emergent grazing dynamics. We show that the Prescribed Grazing Index is well correlated with not only inter-model variability in specific grazing rates, but also grazing pressure at globally, regional, and seasonal scales (see Section 2.4).
5. Further discuss the sensitivity and limitations of the metric (see Section 4.3).

Each other these points in discussed in further detail in our response to comments below.

1. Instead, I think it is far more important to focus on the functional responses of zooplankton within each model. For instance, extensive research has shown how the fundamental behavior of models can change with switches between Type II and Type III grazing, switches between Holling's Disk and Ivlev grazing formulations, inclusions of additional trophic levels, and switches between zooplankton losses mediated primarily by linear or quadratic closure terms. These topics are all likely to

fundamentally alter how zooplankton respond to changing bottom-up forcing, and hence will alter the predicted responses of the marine ecosystem to climate change.

We fully agree that qualitative differences in the functional response are incredibly important. To this end, as suggested:

1. We have cited and incorporated much more work highlighting the importance of Type II vs. III functional responses, active vs passive switching, trophic complexity, and closure throughout the manuscript (see Response to R1).
2. We have substantially expanded our description of each models' grazing formulation, working to highlight these distinctions along with their ecological basis and numerical significance (see Section 4.1).
3. **Importantly**, we have now explicitly tested the hypothesis that variability in the *PGI* can either capture or dominate the well-established influence of other qualitative aspects of the grazing formulation and closure scheme. We do this by computing offline grazing terms for all CMIP6 models and comparing the *PGI* to emergent grazing dynamics across models. We find that large variability in the *PGI* appears to dominate both qualitative differences in the functional response and uncertainty in zooplankton mortality rates to drive very large uncertainty in emergent grazing pressure. This is possible, despite the simplicity of our metric, because the first-order differences in the magnitude of the grazing rates prescribed by difference formulations is so large that they can be captured with such a reduced value (see Section 2.3).

Line 143: "Although decreasing the complexity of the food web should generally increase the strength of predator-prey interactions (Vallina et al., 2011) and thus increase the *PGI* (as appears true in MARBL and CMOC), OECO-v2 and WOMBAT have the second and third lowest *PGI* despite having the simplest food-webs. This suggests that the influence of qualitative differences in food-web structure can, in some cases, be dominated by quantitative differences in parameter values (Sailley et al., 2013), which vary dramatically (Rohr et al., 2022). Nevertheless, the *PGI* reduces many modes of variability into a single variable. It is plausible then that the emergent dynamics in models do not reflect the large differences implied by their *PGI*. To test the hypothesis that variability in the *PGI* can either capture or dominate the influence of other qualitative aspects of the grazing formulation, such as trophic complexity, prey-switching preferences, non-maximal feeding, and/or prey refuge (Prowe et al., 2012; Le Quere et al., 2016; Vallina et al., 2014; Sailley et al., 2013), we assess the variability in emergent grazing dynamics (Section 2.3) and determine if it can be explained by variability in the *PGI* (Section 2.4)."

4. We have conducted a fairly substantial analysis on the influence of qualitative differences in the functional response, both using the offline calculations of emergent grazing dynamics in CMIP6 simulations and the WOMBAT sensitivity studies. However, due to space limitations this has been included in the SI (see 'Supplemental Discussion 1').

2. Differences of $g(P)$, however, mostly reflect differences in zooplankton biomass, since most of the models have tuned bulk grazing to approximately match NPP. I elaborate on this issue below.

First and foremost, thank you for bringing up this line of argument. We now explore this idea, and find that differences in $g(P)$ are only weakly related to zooplankton biomass. We are grateful to have been prompted to investigate this line of reasoning and think it has contributed to substantially improving the manuscript.

In the revised manuscript we now look explicitly at the emergent grazing dynamics across CMIP6 models and show that large differences in emergent zooplankton specific grazing rates (i.e. $g(P)$) **are not** adequately balanced by differences in zooplankton biomass (via differences in mortality). In turn, we find that the largest source of uncertainty across 14 major carbon fluxes, pools, and rates, is in fact grazing pressure (i.e. the phytoplankton specific loss rate to grazing -- $g(P)Z/P$). To tune NPP, it appears that models are not tuning mortality to balance $g(P)$ and achieve similar grazing pressure, but are rather tuning phytoplankton division rates to balance large differences in grazing pressure. This argument holds at annual/global scales (see **Section 2.3.1**) as well as smaller zonal/seasonal scales (see **Section 2.3.2**).

Line 168: “In a globally-averaged sense, the large variability in g is not compensated for by competing differences in zooplankton mortality, leading to similarly large variability in the grazing pressure imposed on the phytoplankton population ($gp=gZ/P$). While there is a positive, relationship between g and mean zooplankton specific mortality rates (m , Figure 2A), suggesting that models are tuned to have more efficient grazers die faster, there is not a statistically significant relationship between g and the mean globally-integrated zooplankton biomass (Z) (Figure 2B), further suggesting that the magnitude of this tuning is insufficient to constrain grazing. Instead, variability in mean grazing pressure (gp) appears to be driven by variability in g . That is, gp has no statistically significant relationship with m (Figure 2C) or Z (Figure 2D), but is correlated with g (Figure 2E). In turn, there remains a very large degree of uncertainty in the representation of gp , which increases by over an order of magnitude (116%) from OECO-v2 to MARBL, largely mirroring the uncertainty in g and drifting even further from observed estimates (Figure 2C-E).

In fact, grazing pressure appears to be the largest source of uncertainty in simulated marine carbon cycling. To quantify sources of uncertainty across CMIP6 models, we compared the inter-model coefficient of variation (std/mean) for 14 major carbon fluxes, pools and rates, computed using global, annual median values from each simulation (Figure 3). By this metric, uncertainty in grazing pressure is 15 % larger than the next largest source, specific grazing rates. Notably, uncertainty in specific grazing rates is 60% larger than zooplankton specific mortality rates. Thus, it appears inter-model differences in rate of zooplankton mortality are insufficient to balance much larger differences in the specific rate of grazing. In turn, uncertainty in grazing pressure is at least 3-times larger than that of EP, NPP and export efficiency, and 2-times larger than that of phytoplankton and total plankton biomass. By contrast, uncertainty in phytoplankton specific division rates is substantial, on par with zooplankton mortality, and appears necessary to offset large differences in grazing pressure in order to constrain NPP.”

Figure 2:

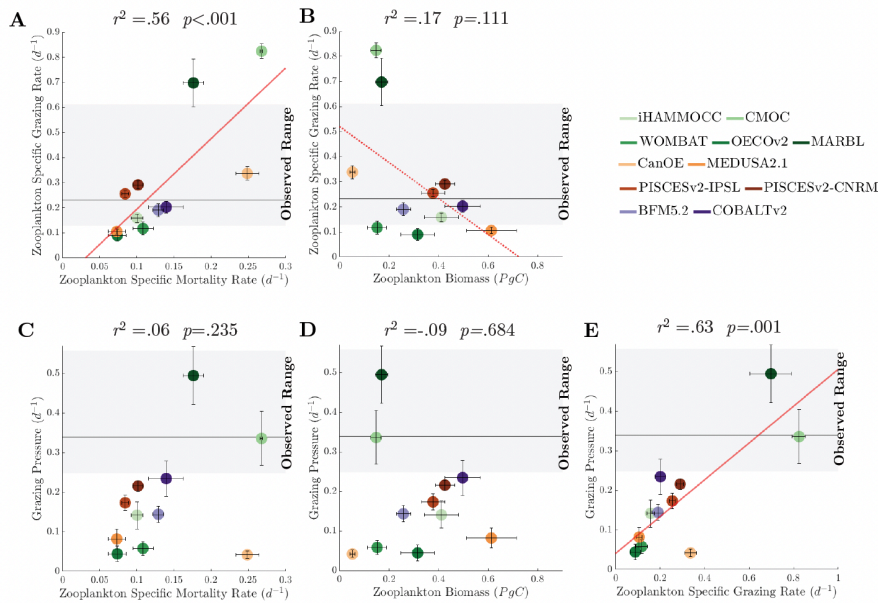


Figure 2: Inter-model Variability in global, biomass weighted, annually averaged **A-B)** zooplankton specific grazing rates and **C-E)** grazing pressure plotted as a function of **A, C)** zooplankton specific mortality rates, **B, D)** biomass and **E)** specific grazing rates. Colors correspond to models in Table 2 and Figure 1. Regression statistics are provided above each plot, with the linear fit include in red if the relationship significant above the 95% (solid) or 80% (dashed) confidence threshold. Vertical and horizontal error bars refer to the amplitude of the seasonal cycle, computed separately in the northern and southern hemisphere, then averaged. The black horizontal line correspond observed values, with uncertainty bounds overlaid in gray (see “Methods”).

Excerpt from Figure 3:

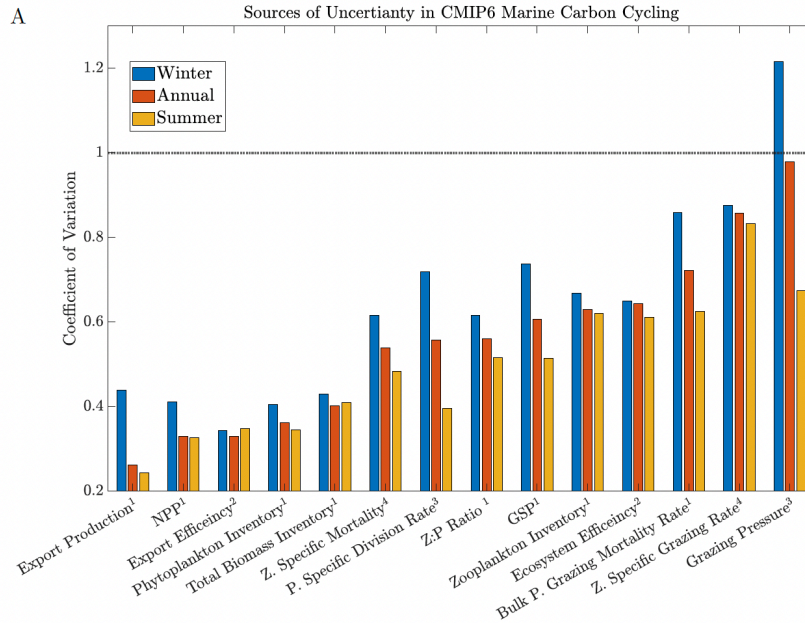


Figure 3: Inter-model uncertainty in marine carbon cycling. **A)** The coefficient of variation (standard deviation divided by the mean) is shown for the annual (orange), winter (blue), and summer (yellow) median of major biological marine carbon fluxes, stocks, and rates. Winter values are computed using the December and June for the northern and southern hemispheres, respectively. Vice versa for summer. Medians are computed over the top 200m of the water column and weighted by either (1) volume, (2) NPP, (3) phytoplankton biomass or (4) zooplankton biomass to account for the distribution of where/when the relevant carbon cycling occurs. Note, qualitatively similar values are computed using weighted-means, instead of medians, and are reported in Figure 5. **B-O)** Hovmöller diagrams (month vs. latitude) of the seasonal and zonal distribution of inter-model uncertainty are provided for each variable. Areas in which the standard deviation is greater than the mean are contoured in black.

However, I do think there is a lot of merit in the authors’ broader analysis of differences between models.

3. Fundamentally, what matters is the functional response of zooplankton to changing phytoplankton concentrations (and in this case “functional response” depends on the interests of the study and likely will include either bulk grazing pressure, zooplankton secondary production, or carbon export efficiency). Unfortunately, the authors’ $(g(P))^-$ metric gives us no information about this response.

Changed. In the revised manuscript we now:

1. Explicitly test the hypothesis that the *PGI* (Previously called Grazing Pressure) can capture emergent properties of grazing dynamics, despite reducing many important modes of variability to a single metric (see **Section 2.4**).
2. Explore the influence of the functional response explicitly by semi-empirically deriving the ‘Apparent Functional Response’ as the evolution of community averaged grazing rates plotted

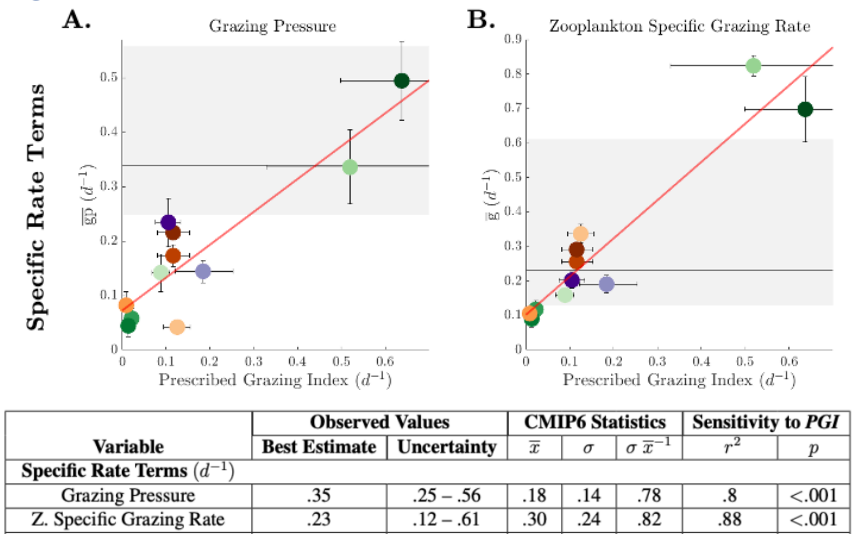
against the total phytoplankton concentration (see ‘Supplemental Discussion 1’; Figure S10-11).

Line 236: “Large differences in emergent grazing dynamics at the global, regional and seasonal scales appear to be driven by large differences in the magnitude of the grazing formulation, as quantified by the *PGI*. Here we investigate the relationship between the *PGI* and emergent grazing dynamics and broader marine carbon cycling in CMIP6 models (Section 2.4.1) and in a controlled sensitivity study (Section 2.4.2). Qualitative differences in the functional response are explored ‘‘Supplemental Discussion 1’’.

Despite reducing many modes of variability to a single value, the *PGI* is a strong predictor of not only global annually averaged specific grazing rates (Figure 5A; $r=.82$), but also grazing pressure (Figure 5B; $r=.78$). In fact, emergent grazing pressure is better correlated with the *PGI* at the global scale than emergent zooplankton mortality (m), biomass (Z) or even specific grazing rates (g) (Figures 2C-E). Moreover, despite substantial spatial-temporal variability in zooplankton biomass (Fig 4H), the *PGI* remains significantly, positively correlated with specific grazing rates (Figure 4F) and grazing pressure (Figure 4C) at most latitudes, during both winter and summer. In turn, even though most NPP and EP disproportionately occur during blooms, when the biomass is much higher than the global median, it appears that understanding how fast that models assume their zooplankton will graze at lower concentrations (as quantified per the *PGI*) is an excellent predictor of the emergent grazing pressure and zooplankton specific grazing rates across the entire dynamical system. That is, the rate of grazing prescribed for low to intermediate biomass stocks appears to be an important control on grazing dynamics across all prey levels via its feedback on the size of the prey population.”

Line 358: “Qualitative differences in the structure of the food web (Sailley et al., 2013; Prowe et al., 2012; Vallina et al., 2011), degree of trophic complexity (Le Quere et al., 2016; Hashioka et al., 2013), sensitivity to temperature (Laufkötter et al., 2015), and capacity for prey refuge (Chenillat et al., 2021; Nissen et al., 2018; Prowe et al., 2012; Vallina et al., 2014) all exert a strong top-down control on NPP and subsequent carbon cycling. Ideally, models that leverage these distinctions to increase biodiversity (Chenillat et al., 2021; Prowe et al., 2012; Vallina et al., 2014; Dutkiewicz et al., 2020), prevent competitive exclusion (Adjou et al., 2012), improve ecosystem stability (Vallina et al., 2011; Gentleman et al., 2008), or replicate species succession (Prowe et al., 2012) and bloom phenology (Hashioka et al., 2013; Karakus et al., 2022), ought also tune their parameters to converge on a relatively constrained degree of bulk grazing pressure. However, this does not appear to be happening. Large variability in emergent grazing dynamics is well explained by the cat global, regional, and seasonal scales, despite the *PGI* reducing many modes of qualitative variability into a single value. In turn, models with substantial qualitative differences in their grazing formulation and food web structure but similar *PGI* can yield similar zonal distributions of emergent grazing rates (i.e. MEDUSA2.1 vs OECO-v2; Figure 4B) and/or grazing pressure (i.e. CanOE vs PISCESv2; Figure 4E). Together, this highlights how far CMIP6 models remain from converging on a first-order magnitude of zooplankton specific grazing rates, leading to very large uncertainty in the representation of grazing pressure throughout the global ocean.”

Modified from Figure 5:



Modified from Figure 5. Inter-model variability in (A) emergent grazing pressure and b) specific grazing rates is well correlated ($r^2=0.8, 0.88$, respectively) with the Prescribed Grazing Index at the global scale

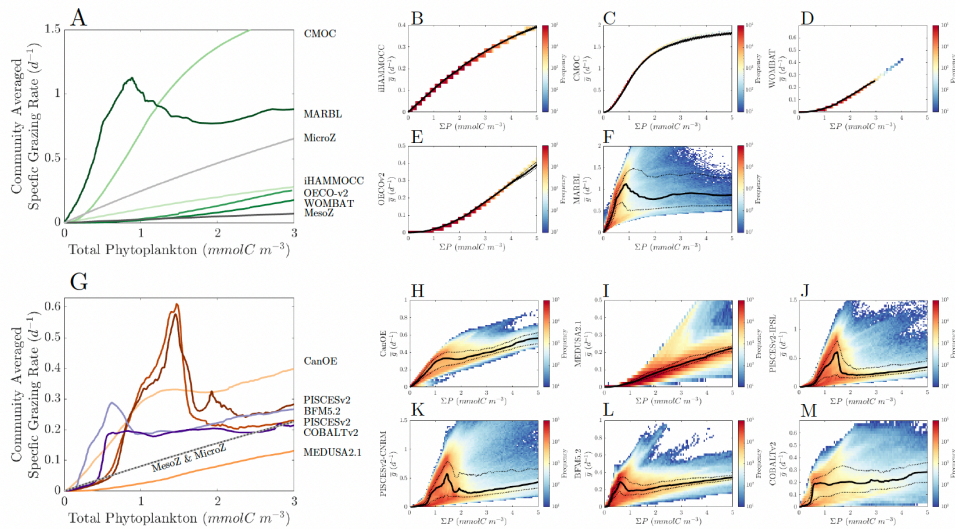


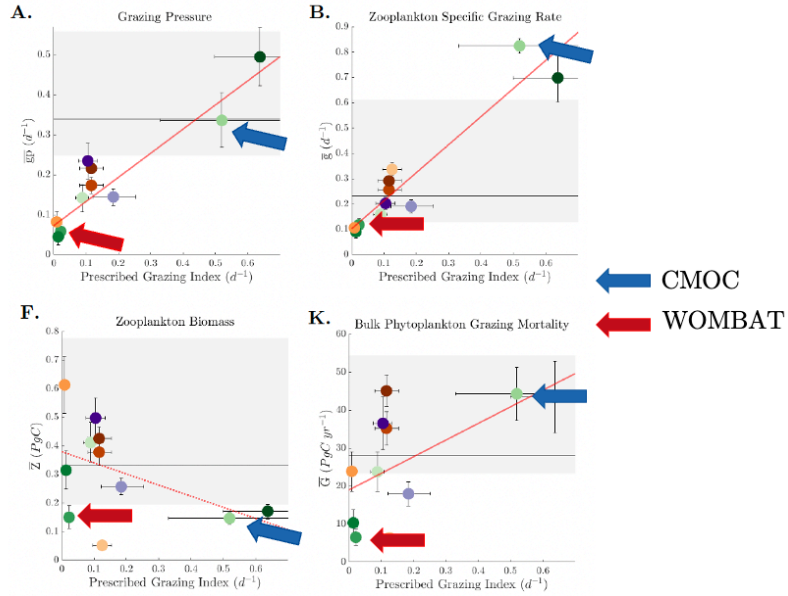
Figure S10: The Apparent Functional Response, as described in ‘Supplemental Discussion 1’ is plotted for models with A-F one or multiple *G-M* zooplankton groups. Sub-panels for each model on the right show the frequency distribution of zooplankton specific grazing rates plotted against total biomass, from which the apparent functional response was compute empirically. All grid cells from 0-200m are included. The Apparent Functional Response is shown as the median ($\pm 25\%$; dashed lines) specific grazing rate occurring in all grid-cells containing each phytoplankton concentration.

4. To a certain extent, this issue is really highlighted by the authors’ analysis with the WOMBAT model. Within the WOMBAT model, increasing $g(P)$ to values found in other model’s causes NPP to crater. This is, of course, because WOMBAT is tuned to behave as if the single zooplankton is a metazoan zooplankton (krill or copepod). As such, it likely obtains a large biomass in the model, with slow turnover rates (turnover rates that are likely modulated mostly by the zooplankton mortality term and/or the zooplankton respiration term). These closure terms are low and hence WOMBAT has high zooplankton biomass leading to low $g(P)$. In another model in which the single zooplankton state variable is tuned to act like a protist, a much higher mortality rate would be prescribed, leading to more rapid turnover, lower biomass and hence a higher $g(P)$ necessary to balance NPP. This is a very simple way of looking at things, but it helps to clarify two issues: first, it is bulk grazing pressure (an emergent property of both $g(P)$ and zooplankton biomass) that matters for NPP

Changed. This is a great point, which led us to compute not only the emergent grazing pressure i.e. $(g(P)Z/P)$ explicitly, offline, for each model (See [Section 4.5](#); [Table 1](#)) but also emergent mortality rates (See [Section 4.5](#); [Table S3](#)). It turns out that inter-model uncertainty in emergent grazing pressure is actually worse (when considering median values) than that of specific grazing rates, highlighting how mortality/closure schemes do not appear to be sufficiently tuned to compensate for large difference in specific grazing rates, which appear driven by large differences in the *PGI* (see our response to [Comments 2](#) and [3](#)).

Note, this is not true for every model. For instance, CMOC, is tuned almost exactly as you described: with very high specific grazing and mortality rates leading to low zooplankton biomass and grazing pressure lower then specific rates (relative to the inter-model mean). However, WOMBAT is a great counter example, with slow specific grazing **and** mortality rates leading to low zooplankton biomass **and** low grazing pressure. Overall, across the 11 survey models, the case you described (i.e. CMOC) appears to be an exception rather than the rule.

Modified from [Figure 5](#):



Modified from Figure 5. Arrows have been added to clearly point out the location of CMOC (blue) and WOMBAT (red).

5. and second the functional response of zooplankton is controlled equally by its grazing rate AND its mortality rate – considering one without considering the other is not particularly helpful.

Changed. Another great point. As discussed above (see our response to **Comment 2**), this has prompted us to not only compute emergent specific grazing rates offline for all 11 models, but also compute zooplankton specific mortality rates (see Table S4). This has allowed us to show how variability in emergent specific grazing rates dominates that of specific mortality rates and leads to very large variability in grazing pressure, which can be well explained by variability in the *PGI* (see our response to **Comment 3**).

Table S4:

	BGC Model; Earth System Model	Zooplankton Group	Mortality Rate	Parameters
	iHAMMOCC NorESM2-ESM-LM	Zoo. (Z)	$r_1 Z' + \alpha Z'^2$ $Z' = \min(0, Z - Z_{\min})$	$r_1 = 0.06$ $\alpha = 0.025$ $Z_{\min} = 0.001$
	CMOC; CanESM5	Zoo. (Z)	$r_1 Z + \alpha Z^2$	$r_1 = 0.25$ $\alpha = 0.015$
	WOMBAT; ACCESS ESM1.5	Zoo. (Z)	$r_1 T_{Lim} Z + \alpha Z^2$ $T_{Lim} = 1.066T$	$r_1 = 0.01$ $\alpha = 0.05$
	OECO-v2; MIROC-ES2L	Zoo. (Z)	$r_1 Z + \alpha Z^2$	$r_1 = 0.03$ $\alpha = 0.03$
	MARBL; CESM2	Zoo. (Z)	$r_1 T_{Lim} Z' + \alpha T_{Lim} Z'^2$ $T_{Lim} = 1.7^{(T-30)/10}$ $Z' = \min(0, Z - Z_{\min})$	$r_1 = 0.1$ $\alpha = 0.4$ $Z_{\min} = f(\text{depth})$ $= 0 - .075$
	CanOE ^A ; CanESM5	Small Zoo. (Z_{sm})	$(r_1 T_{Lim} + r_2) Z_{sm} + \alpha Z_{sm}^2$ $T_{Lim} = e^{-4.5e3(\frac{1}{T+273} - \frac{1}{298.15})}$	$r_1 = f(P_{C:N:Fe})$ $= 0 - .1$ $r_2 = 0.05$ $\alpha = 0.06$
		Large Zoo. (Z_{lg})	$(r_1 T_{Lim} + r_2) Z_{md} + \alpha Z_{md}^2$ $T_{Lim} = e^{-4.5e3(\frac{1}{T+273} - \frac{1}{298.15})}$	$r_1 = f(P_{C:N:Fe})$ $= 0 - .3$ $r_2 = 0.1$ $\alpha = 0.06$
		Assimilation Inefficiency	$(1 - \gamma) g_{Z_{sm}}^{Z_{md}} Z_{md}$	$\gamma = .8$
	MEDUSA2.1; UKESM1-0-LL	Microzoo. (Z_{sm})	$r_1 Z_{sm} + r_2 (\frac{Z_{sm}}{2.8 + Z_{sm}}) Z_{sm}$	$r_1 = 0.02$ $r_2 = 0.1$
		Mesozoo (Z_{md})	$r_1 Z_{md} + r_2 (\frac{Z_{lg}}{4.2 + Z_{md}}) Z_{md}$	$r_1 = 0.02$ $r_2 = 0.2$
		Assimilation Inefficiency	$(1 - \gamma) g_{Z_{sm}}^{Z_{md}} Z_{md}$	$\gamma = .41$
	PISCESv2 ^B ; IPSL-CM6a-LR & CNRM-ESM2.1	Microzoo. (Z_{sm})	$(r_1 + \Delta O_2) T_{Lim} Z_{sm} + \alpha T_{Lim} Z_{sm}^2$ $T_{Lim} = 1.079^T$	$r_1 = 0.03$ $\alpha = 0.004$
		Mesozoo. (Z_{md})	$(r_1 + \Delta O_2) T_{Lim} Z_{md} + \alpha T_{Lim} Z_{md}^2$ $T_{Lim} = 1.079^T$	$r_1 = 0.005$ $\alpha = 0.03$
		Assimilation Inefficiency	$(1 - \gamma) g_{Z_{sm}}^{Z_{md}} Z_{md}$	$\gamma = .35$
	BFM5.2 ^B ; CMCC-ESM2	Microzoo. (Z_{sm})	$(r_1 + r_2 + r_3 \Delta O_2) T_{Lim} Z_{sm}$ $T_{Lim} = 2^{(T-10)/10}$	$r_1 = .01$ $r_2 = .01$ $r_3 = .25$
		Mesozoo. (Z_{md})	$(r_2 + r_3 \Delta O_2) T_{Lim} Z_{md} + \alpha T_{Lim} Z_{md}^2$ $T_{Lim} = 2^{(T-10)/10}$	$r_2 = .02$ $r_3 = .01$ $\alpha = 0.01$
		Assimilation Inefficiency	$(1 - \gamma_{sm}) g_{Z_{sm}}^{Z_{sm}} Z_{sm} + (1 - \gamma_{md}) (g_{Z_{sm}}^{Z_{md}} + g_{Z_{md}}^{Z_{md}}) Z_{md}$	$\gamma_{sm} = .5$ $\gamma_{md} = .6$
	COBALTv2 ^C ; GFDL-ESM4.1	Small Zoo. (Z_{sm})	$r_1 T_{Lim} Z_{sm}$ $T_{Lim} = 2^{(T)/10}$	$r_1 = .08$
		Med. Zoo. (Z_{md})	$r_1 T_{Lim} Z_{md} + g_{Z_{md}}^{HP} (Z_{md} + Z_{lg})$ $T_{Lim} = 2^{(T)/10}$	$r_1 = .008$ $K_{HP} = 8.28$ $g_{HP} = .09$
		Large Zoo. (Z_{lg})	$r_1 T_{Lim} Z_{lg} + g_{Z_{lg}}^{HP} (Z_{md} + Z_{lg})$ $T_{Lim} = 2^{(T)/10}$	$r_1 = .0032$ $K_{HP} = 8.28$ $g_{HP} = .09$
		Assimilation Inefficiency	$(1 - \gamma) g_{Z_{sm}}^{Z_{md}} Z_{md} + (1 - \gamma) g_{Z_{md}}^{Z_{lg}} Z_{lg}$	$\gamma = .4$

Table S3: Bulk zooplankton mortality rates in CMIP6 models. Note, all loss terms which vary with zooplankton biomass, including basal metabolic rates, are included and considered mortality. Loss terms that vary with ingestion are considered assimilation inefficiencies and not included except for in models with carnivorous zooplankton to account for biomass lost to sloppy feeding, egestion, and in some cases respiration during trophic transfer. Biomass that is assimilated via carnivorous grazing is not considered part of the bulk mortality term as it remains in the total zooplankton pool. The g_i^j term is the specific grazing rates between zooplankton groups as described in Table 1. Linear loss rate parameters, r , are reported in units of d^{-1} and quadratic loss rates parameters, α , are reported in units of $m^3 \text{ mmolC}^{-1} d^{-1}$. ^A The temperature limited respiration (r_1) in CanOE drops out when excess carbon is available in prey relative the fixed stoichiometry of the zooplankton. ^B The ΔO_2 terms in PISCESv2 and BFM5.2 increases mortality rates in anoxic conditions and were accounted for but are relatively small. Equations can be found in (?) and (?), respectively. ^C In COBALTv2, medium and large zooplankton are lost to an implicit higher predator (HP) with a specific grazing rates ($g_{Z_j}^{HP}$), computed analogously to zooplankton (see Table 1) using equal preference and parameters K_{HP} and G_{HP} .

5. Issues with using $g(P)$ are further exacerbated when there are multiple zooplankton. The term $g(P)$ really loses meaning when multiple zooplankton trophic levels are present. Consider a model with a single (protistan-like) zooplankton. It will have a high $g(P)$. Now consider a second model that builds on the first by adding an herbivorous mesozooplankton (e.g., a small copepod). Most of the grazing in the model should still be mediated by the protist (70% of NPP is the canonical global value), while

the copepod feeds on roughly 15% of NPP. However, the copepod, with its slow turnover rates, will have much higher biomass than the protist. Hence the global $g(P)$ will be much lower, because bulk grazing will remain similar but total zooplankton biomass will be much higher. Nevertheless, the model will probably still exhibit fairly similar grazing behavior, because most of the grazing is still mediated by that rapid protist. Next consider a third model that adds an omnivorous macrozooplankton (e.g., a large copepod or krill). The protist still does most of the grazing, the small copepod consumes ~10% of NPP, the large copepod consumed ~5%, but the long-lived large copepod adds substantially more biomass and $g(P)$ drops again. These models would all behave in functionally similar ways with respect to top-down impacts on NPP, although $g(P)$ would be very different in the three models.

Changed. By computing the offline emergent grazing pressure, specific grazing rates, and biomass of each model, many with multiple zooplankton groups, we explicitly account for these differences. In turn, the emergent grazing pressure we compute should capture the entire, community integrated, top-down control on NPP. While we agree that models with increasing/decreasing trophic complexity should be tuned generally as you described, this does not appear to be the case, as large uncertainty in emergent grazing pressure persists (see our response to **Comment 2**).

Additionally, by refining the definition of the *PGI* to use observed median plankton fields (see **Table S1; Section 4.2**), such that the grazing rates of meso- and microzooplankton groups are weighted to reflect the global, observed, median biomass distribution, it also now better captures some of these differences.

6. Furthermore, the fact that the authors do not actually calculate $g(P)$ based on emergent biomasses and grazing rates, but instead calculate it based on assuming equal biomasses of all grazers is not at all realistic.

Changed. We now explicitly calculate the emergent grazing dynamics. Our calculations of zooplankton specific grazing rates ($g(P)$) and grazing pressure ($g(P)Z/P$) account for emergent variability in the relative distribution and (in the case of grazing pressure) total abundance of grazers. See **Section 4.5**. For example:

Line 849: “We analysed the output from historical simulations of 11 CMIP6 models (Table S2). Model (82-92) has been made publicly available by the Earth System Grid Federation (ESGF) and was downloaded from the Lawrence Livermore National Laboratory data node (<https://aims2.llnl.gov/search>). Note, three variables from ACCESS-ESM1.5 (phytoplankton biomass, zooplankton biomass, NPP) were not available on the ESGF-LNLL node and were downloaded directly from the Australian National Computing Infrastructure. For each model we considered monthly climatologies created from the historic period from 1850 to 1900. Considering years up to 1900 provides a long enough time series to average out ENSO variability but stops before there is a substantial anthropogenic forcing, as we are interested in the non-transient dynamics stemming from structural differences in the grazing formulation between models. We used the pre-1990 historical simulation instead of the pre-industrial control because more variables of interest were available across all models. For each model, we downloaded monthly averaged column integrated NPP, export through 100 m, and all depth resolved phytoplankton and zooplankton plankton fields. Depth resolved detritus, bacteria, temperature, and oxygen fields were additionally downloaded when required for offline rate computations. Any exceptions are discussed below.”

Table S2:

	BGC Model	Earth System Model	Simulation	Model ID	Institution (Location)	Personal Communication
	iHAMMOCC (64)	NorESM2-ESM-LM (64)	Historical (82)	r2ilp1fl	NORC (NOR)	Jerry Tjiptura Jörg Schwinger
	CMOC (65)	CanESM5 (69)	Historical (83)	r1ilp2fl	ECCC (CAN)	Jim Chirstian
	WOMBAT (66)	ACCESS ESM1.5 (66)	Historical (84)	r1ilp1fl	CSIRO (CAN)	Andrew Lenton Mathew Chamberlain
	OECO-v2 (67)	MIROC-ES2L (67)	Historical (85)	r1ilp1f2	JAMSTEC (JAP)	-
	MARBL (68)	CESM2 (78)	Historical (86)	r1ilp1fl	NCAR (USA)	Matt C. Long
	CanOE (69)	CanESM5 (69)	Historical (87)	r2ilp2fl	ECCC (CAN)	Jim Christian
	MEDUSA2.1 (70)	UKESM1-0-LL (79)	Historical (88)	r1ilp1f2	Met Office, NERC (UK)	-
	PISCESv2 (71)	IPSL-CM6A-LR (80)	Historical (89)	r10ilp1fl	IPSL (FR)	-
	PISCESv2 (71)	CNRM-ESM2.1 (79)	Historical (90)	r1ilp1f2	CNMR (FR)	-
	BFM5.2 (72)	CMCC-ESM2 (72)	Historical (91)	r1ilp1fl	CMCC (ITA)	Tomas Lovato
	COBALTv2 (73)	GFDL4.1 (81)	Historical (92)	r1ilp1fl	GFDL (USA)	Charlie Stock

While we want the Prescribed Grazing Index to only reflect innate, prescribed, attributes of the model formulation, and thus intentionally do not evaluate it with any emergent properties, we have revised the standardized prey field over which is evaluated to be more realistic (see **Table S1; Section 4.2**).

Moreover, we have now shown that the Prescribed Grazing Index is a good predictor of emergent properties (see our response to **Comment 3**). This is likely because the difference in the magnitudes, in a mean sense, of different grazing formulations is so large, that they emerge even when many dimensions of the multi-variate functional response are collapsed into a single dimension.

We stress that this important result (and others discussed above) were not defended in the original manuscript and sincerely thank the reviewer for pushing us to interrogate them.

Other comments:

7. The manuscript had many supplemental figures that were frequently referenced. I found this distracting, because I frequently had to look to the supplement. I would recommend including more of these figures in the manuscript.

Changed. We now include 9 display items in the main text, including Figure 8, which was previously in the SI of the original manuscript. If any other revised supplemental figures stand out over those in the main text, we would be happy to consider a switch.

8. Section 2.5: Description of WOMBAT: Authors state that WOMBAT grazing pressure can be modified by three variables, but then only tell us what two of them are.

Changed. This Sentence now appears in the SI but has been modified to:

Supplemental Discussion 1: “Unlike some of the more complex CMIP6 formulations address above, In WOMBAT, the ‘shape’ of the multivariate grazing formulation can only be modified three ways: the value of two parameters ($K_{1/2}$ and g_{max}) and the selection of a functional response curve (typically Type II or III)”

9. Within the text and Supp. Fig. 2 the authors use different names for the same variables, which is confusing.

Changed. Revised throughout, both for original and new variables. All variables should now be consistent across the WOMBAT description (e.g. **Figure S7**) and offline CMIP6 computations (e.g. **Section 4.5**)

10. While I like Fig. S2, the equations should be simplified. I spent a lot of time going through these equations (and substituting in multiple variables for each other) only to find in the end that they are mostly very standard equations (e.g., GSP is a very typical formulation, but that is not clear without working through the math). There are other issues as well, such as the fact that variables are sometimes written out in strange ways (for instance, it took me a little while to realize that ‘gambio’ was just a variable that is probably normally written as Greek letter gamma subscript bio). Also, in the zooplankton equation, mortality is written as mZ so I assumed it was a linear mortality term with parameter m as the daily mortality rate. About an hour later, I realized that apparently the authors were using the term mZ as a term meaning zooplankton mortality that was defined in the export production section as $m_{\text{ezbio}} * [Z]^2$ (i.e., quadratic mortality). I think I now understand this, but why didn’t they simply write in both the zooplankton and detritus equations: $m * [Z]^2$ so that it would be clear and easily readable? These issues are present throughout the equations and add substantial confusion to what is otherwise a nice figure.

Changed. All points have been resolved in the revised Figure S7 (see immediately below).

Moreover, we’d sincerely like to apologize for the state of the equations before. They were a mess. I suspect this happened because I tried to toe the line between generalizing their syntax and keeping some of the Fortran notation as is (which was useful people in the immediate ACCESS community but obviously not general readers). The equations are now presented using a more intuitive, generalized notation and organized such that the positive contribution to the tendency for each NPZD tracer are listed below the corresponding equations. Hopefully this is all much clearer now. Variable and parameter notation is also now consistent with Table 1, S3, and the rest of the manuscript.

Figure S7:

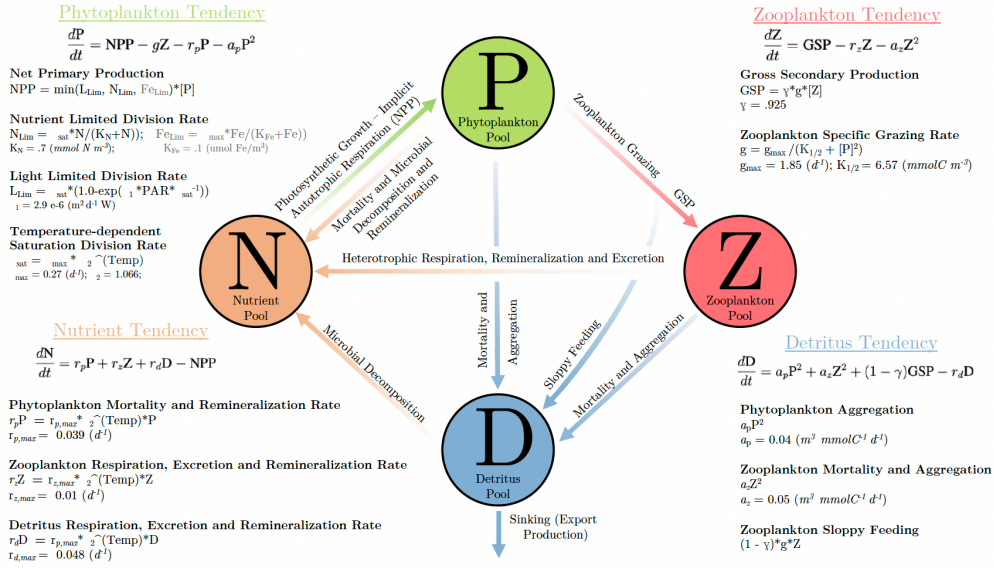


Figure S7: Partial schematic of the control WOMBAT BGC model configuration. Only includes Nutrient (N), Phytoplankton (P), Zooplankton (Z), Detritus (D) state variables. The tendency for each variable is reported beside it, with positive contribution to the tendency detailed below. Note, iron is included in the CMIP6 simulation of WOMBAT analysed here, but not the sensitivity experiments.

11. While I like most of the figures, there are two things that can be improved. First the logarithmic axes would be much more useful if instead of giving the log-transformed values they gave the actual values. For instance, in Figure 1, the x-axes would be far more readable if instead of -9, -8, -7, -6, etc. it was 0.002, 0.004, 0.008, 0.0156, etc. Also, it would be helpful if they had actual legends. Having legends that are kind of hidden in the y-axes was unnecessarily confusing for me, since there is room on the figures for a proper legend.

Changed. All logarithmic axis in Figures 1B, 6, and S8 have been revised accordingly. In general, we've added more legends throughout the figures. As for the specific figures with 'hidden legends' (i.e. original Figure 1), these have now been partitioned (e.g. revised Fig 6—see below) such that legends are no longer necessary.

Figure 6:

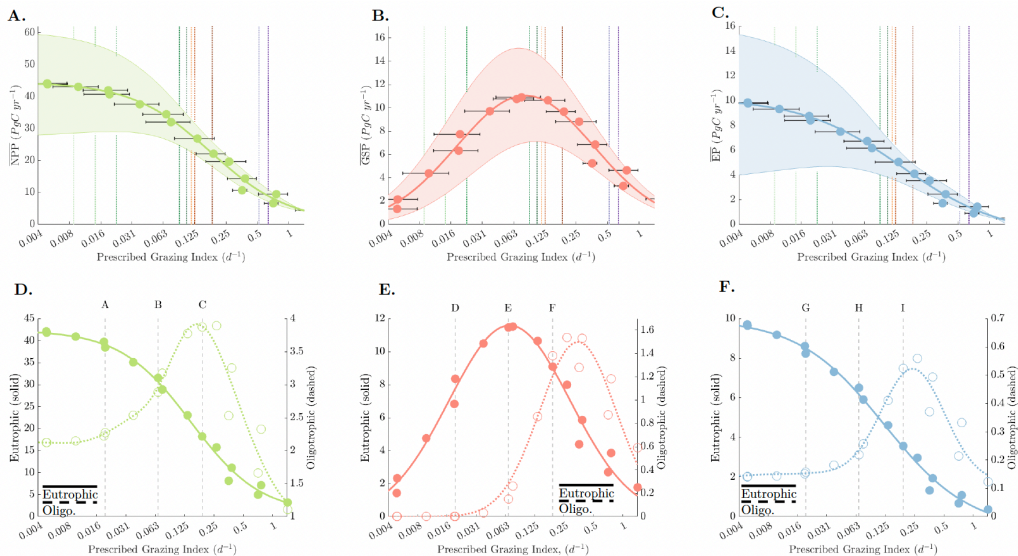


Figure 6: Global, annually averaged **A)** NPP, **B)** GSP, and **C)** EP from each simulation of WOMBAT (see “Methods”) is plotted against the *PGI* of the corresponding model. Horizontal error bars refer to the precision of the *PGI*, per Figure 5. The vertical uncertainty window is proportional to the amplitude the seasonal cycle, computed as $\pm 25\%$ of the difference between the annual max and mean. Vertical dashed lines refer to the *PGI* of 11 CMIP6 models, color-coated per Table 1. Regional, annually averaged **D)** NPP, **E)** GSP, and **F)** EP are plotted separately for eutrophic (left axis) and oligotrophic (right axis) regions. The boundary between ‘oligotrophic’ and ‘eutrophic’ regions is contoured in black on the maps in Figure 7. Vertical dashed lines denoted which runs were plotted in the corresponding panels of Figure 7.

Reviewer #3 (Remarks to the Author):

First and foremost, we would like to thank Reviewer 3 for their constructive and helpful comments. In particular, we believe their suggestion to add a comparison to a more complex BGC model has substantially improved the manuscript. Please see below for our point-by-point response to specific comments and note your comments have been numbered for reference.

Dear authors,

I have now reviewed the submitted manuscript.

Overall the manuscript is well written, interesting, timely but may lack essential aspects dealing with zooplankton dynamics and its representation in models.

In the revised manuscript we have added substantial analysis to better address all aspects of zooplankton dynamics.

Most importantly, we have now explicitly computed offline zooplankton specific grazing rates, grazing pressure, gross secondary production and zooplankton specific mortality rates for 11 CMIP6 models using the documented model code evaluated with archived monthly mean biomass fields from their historical simulations. We are thus now able to directly assess how innate structural differences in the prescribed grazing formulation influence the emergent grazing dynamics of each model.

Additionally, we have incorporated much more of the relevant literature (see **Response to R1**, e.g. **Comment 25** therein), provided a great deal of new observational comparison (see **Comments 3** and **23** in our **Response to R1; Section 4.4**) and, as suggested, included in our analysis a comparison to a more complex model outside of the CMIP6 portfolio (see below).

1. The strength and originality of the work is to propose a unified methodology to inter-compare the impact of modifying zooplankton grazing on BGC models and to compare different formulations and parametrization on export flux and trophic transfer by running simulation in one of the simplest models.

By incorporating the analysis of emergent grazing rates in 11 CMIP6 models we have now dramatically extended the analysis beyond the ‘simple’ simulation in WOMBAT. In the revised manuscript these sensitivity experiments in WOMBAT take a back seat to a detailed description of how large variability in the Prescribed Grazing Index (i.e. the unified metric previously called ‘Grazing Pressure’) drives large variability in emergent grazing pressure (i.e. the simulated phytoplankton specific loss rate to grazing). The WOMBAT work is now used to bolster these results by controlling for other inter-model differences in a relatively simple BGC model, but is no longer the primary focus.

2. BGC models are already further than CMIP model and they should be integrated in this article. Communications Earth & Environment is indeed a journal where the most advanced studies should be addressed.

Changed. We have added a discussion of more complex BGC models, including many of the most advanced studies published in the last several years, and better articulated our reasoning for focusing on CMIP6 models (see below). Moreover, we have added an explicit quantitative analysis of the emergent grazing dynamics simulated in a recent integration of DARWIN to provide a direct comparison between ‘simpler’ CMIP6 class BGC models and the grazing dynamics explicitly

simulated in a more complex food web with nearly 50 plankton groups (see response to subsequent comments).

Finally, we note that of the 11 CMIP6 models analysed, 9 were published in 2020 or later and none before 2017. Additionally, their parsimony is as much a function of balancing large computational demands and under-determined parameterizations (Ward et al., 2013) as it is a weakness. Thus, as far as Earth system modelling is concerned, we would argue they are still the state-of-the-art. Moreover, we now take advantage of their extensive publicly available model output.

Line 524: “Many state-of-the-art BGC and ecosystem models include richer functional diversity (Le Quere et al., 2016; Negrete-Garcia 2022; Buitenhuis et al; 2019), trophic complexity (Heneghan et al; 2020; 2023, Everett et al, 2021), and allometrically scaled-size spectra (Negrete-Garcia 2022, Dutkiewicz, 2015, Ward et al, 2014, Ward et al., 2012, Follett et al., 2022, Chenillat et al., 2021), with some simulations including hundreds of PFTs (Dutkiewicz et al., 2020). Increasing the size and complexity of the food web improves ecosystem stability by increasing resistance to perturbations (Vallina et al., 2011), permits realistic self-regulation of food-chain length (Everett et al., 2021, Negrete-Garcia et al., 2022), replicates the observed biogeography of PFTs and size classes (Heneghan et al., 2020, Ward et al., 2014), and even plays a larger role than iron limitation in shaping high nitrate low chlorophyll conditions in the Southern Ocean (Le Quere et al., 2016). However, increasing complexity can also increase predictive uncertainty if the additional parameter values cannot be constrained by observations. Even relatively simple BGC models, with just 4 plankton groups, can be heavily under-constrained (Ward et al., 2013). Moreover, more complex models are much more computationally expensive, and generally not tractable in IPCC simulations, which require many years, ensemble members, and scenarios to be run at increasingly high, mesoscale, spatial resolution. For this reason, we focus on the current generation of state-of the-art IPCC models, but have included in our analysis, for comparison, the model of (Follett et al., 2022), which is part of the DARWIN Project, run on the MITgcm at 1x1 degree resolution, and includes 47 size-structured PFTs.”

3. I would suggest major revision and hope for the authors to address the main comments.

We have revised this manuscript to include:

- A better framework for the Prescribed Grazing Index (see Section 4.2, 4.3).
- The offline computation of emergent grazing dynamics and zooplankton mortality in 12 global models (see Section 4.5; Table 1, 2, S2)
- Explicit analysis of the emergent grazing dynamics in 11 state-of-the-art CMIP6 models and 1 more complex BGC model (DARWIN) with 47 PFT (see Figures 2-5,S3-6)
- The inclusion of myriad derived observation products for direct comparison with the model dynamics (see Section 4.4; Figures 1,2,4,5, S3-6)
- A much more thorough incorporation/discussion of the contemporary literature throughout (e.g. see Section 4.1).

4. The information that zooplankton is systematically not well represented in BGC models is known for decades and so is not a surprise. Solving this issue is an important aspect in many modelling frameworks.

Many articles have been describing the gaps and propose solutions. They all stress the possible impact on food web transfer and export efficiency. Some new developments are existing but the authors do not report about them in the same way they do for the CMIP.

Changed. Throughout we have incorporated a much richer discussion of the literature discussing the influence of qualitative differences in the grazing formulation and structure of the food web. For example:

Line 93: “Variability in the structure of the food-web and grazing formulation substantially impact ecosystem stability (Vallina et al., 2011; Gentleman et al., 2008), plankton diversity (Chenillat et al., 2021; Prowe et al., 2012;

Vallina et al., 2014; Dutkiewicz et al., 2020), NPP (Laufkötter et al., 2015; Prowe et al., 2012), EP (Dupont et al., 2022; Laufkötter et al., 2016; Karakus et al., 2021, Negrete-garcia et al., 2022), regional (Chenillat et al., 2021) and global (Sailley et al., 2013; Le Quéré et al., 2005; Prowe et al., 2012) phytoplankton biomass and chlorophyll distributions, and the phenology of phytoplankton of blooms (Hashioka et al., 2013; Karakus et al., 2022, Truscott 1994). However, while this work has highlighted the influence of qualitative differences in the formulation of grazing, such as trophic complexity (Le Quere et al., 2016; Sailley et al., 2013; Dutkiewicz et al., 2020) and capacity for prey refuge (Chenillat et al., 2021; Prowe et al., 2012; Vallina et al., 2014), less attention has been paid to large differences in the magnitude of grazing rates assumed across models. This may be because differences in the structure of the food-web, value of parameters, shape of the functional response, and determination of prey preferences all exert a multivariate, often competing, influence on emergent grazing dynamics, making it difficult to discern how fast the average zooplankton is assumed to graze. We thus introduce a new diagnostic metric, the Prescribed Grazing Index (*PGI*) to approximate the magnitude of grazing rates prescribed by the innate properties of a given formulation.”

5. The different CMIP models have been intercompared for different biological variables including zooplankton. For example in Petrick et al., 2022 (cited by the authors), the authors demonstrate the sensitivity of zooplankton in different models by using a Taylor representation. Using the same framework in addition to their analysis would allow to intercompare based on similar methods. This would be possible in the second section of the articles where the authors runs simulation.

Changed. We have advanced the work of Petrick et al. (2022) by further computing offline specific grazing rates ($g(P)$), grazing pressure ($G(P)Z/P$), and specific mortality rates. Moreover, while Perick et al. (2022) focus only on mesozooplankton, ignoring the smaller zooplankton PFTs in multi-trophic models and single zooplankton models altogether, we consider the integrated dynamics of the entire community in a range of food webs (see **Figure 1; Table 1**).

Regarding the Taylor analysis in Figure 4 of Petrick et al. (2022) they compared simulated estimates of mesozooplankton biomass to statistical estimates from the multiple-linear regression model of Heneghan et al (2020). Currently, we are not aware of any such statistical product for community averaged grazing dynamics for which we could use to reference a similar Taylor analysis. We did, however, add extensive observational comparison through the manuscript. To this end:

1. Added observational estimates and uncertainty bounds of global terms to **Figures 2 and 5**, including a table of observed and model statistics (see below).
2. Added zonally resolved observations with uncertainty bounds to **Figures 4 and S3-S6** (see below).
3. Added a detailed described of how these observation estimates were derived from a collection of sources (see Section 4.4; quoted in our response to **Comments 3 and 23** from **Reviewer #1**).
4. Added a discussion of the models relative to the observations (see below, for example).

Excerpt from **Figure 5**:

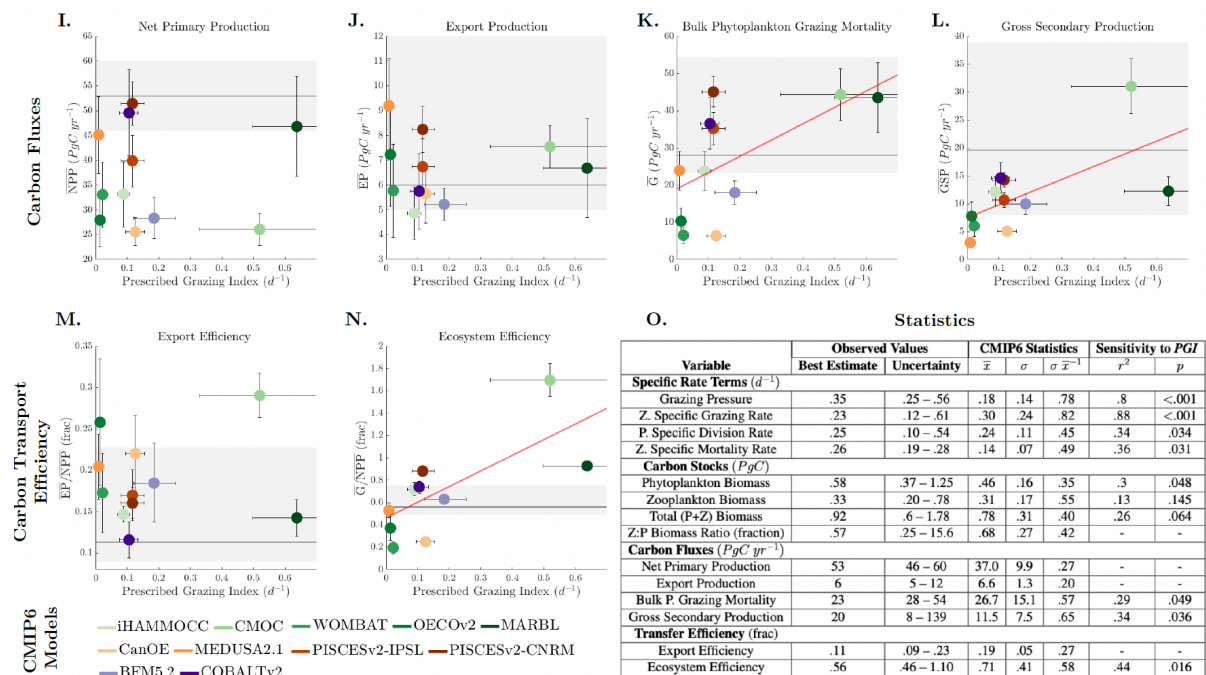


Figure 5: The relationship between global, annually averaged marine carbon cycling and the *PGI*. A-N) Horizontal error bars reflect seasonal variability, as per Figure 2. Vertical error bars reflect the precision of the *PGI* and are computed by evaluating the *PGI* with $\pm 25\%$ total phytoplankton biomass. The black horizontal line corresponds to our best guess of the observed values, with uncertainty bounds overlaid in gray (see “Methods”). Linear regressions are included in red if the relationship is significant above the 95% (solid) or 80% (dashed) confidence threshold. The relationship between global, annually averaged marine carbon cycling and the *PGI*. O) Statistical properties of the observations, model spread, and relationship between the inter-model variability and the *PGI* are reported.

Excerpt from Figure 4:

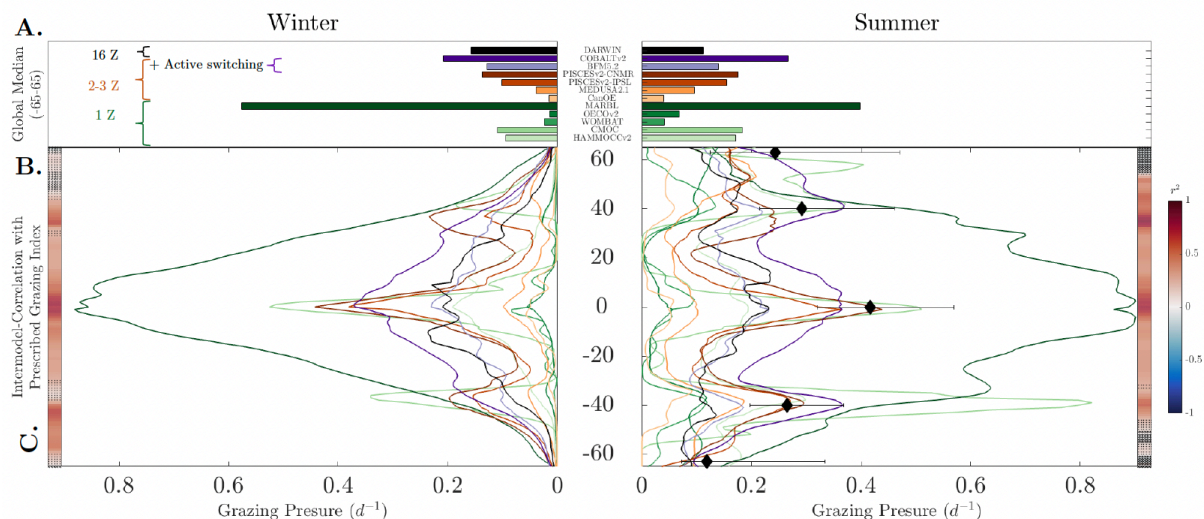


Figure 4: The Distribution of median **A-C**) grazing pressure, **D-F**) zooplankton specific grazing rates, and **G-I**) zooplankton specific mortality rates. **A, D, G**) The global median of each variable is plotted during the climatologic winter (December/June above/below the equator) and summer (vice versa). Medians are computed over the top 200m of the water column and weighted by **A**) phytoplankton or **D, G**) zooplankton biomass. Model colors correspond to the traces below. **B, E, H**) The zonal distributions are plotted for each model using a sliding 3° latitudinal band. **B, E**) Diamonds correspond the median observed estimate with error bars for the 25th and 75th percentiles (see “Methods”). **C, F, I**) On the boarder, the inter-model correlation with the *PGI* is plotted for each zonal band and season, with statistically insignificant relationships x’d out. Lighter x’s indicate a casual, but statistically insignificant, relationship with $.2 > p > .05\%$.

Line 210: “Variability in grazing pressure across models is not well constrained by observations (see “Methods”).

Qualitatively, observations of grazing pressure agree with the zonal distribution simulated in models, peaking at the equator, then tapering off towards the poles (Calbet et al., 2004; Schmoker et al., 2013; Calbet et al., 2001), but the models simulate much lower values and exhibit a much larger spread. In no region do more than 65% of models fall with the middle 50% of zonally-binned, basin-scale observational estimates (Fig 4B; black diamonds). While this underscores how poorly constrained grazing pressure is in models, it is important to remember that field measurements are sparse and subject to substantial sampling biases (Calbet et al., 2001; Schmoker et al., 2013). Thus, a small uncertainty window is probably a consequence of a small sample size, rather than low natural variability (see “Methods”), suggesting that the solution is not simply nudging models toward current observations but also further constraining the observations.”

6. The weak part is that the authors do not include in their quantitative analysis and discuss enough the already existing models that have complex zooplankton representation (for example in size and taxa) and are already improvement relative to CMIP models.

Changed. In the revised manuscript we add quantitative offline analysis of 11 CMIP6 models with a range of trophic complexities (see **Figure 1-5, Section 4.5**). We have also added quantitative analysis of a much more complex BGC model (DARWIN) which we use to assess how well the CMIP6 models implicitly average over the size-spectra and additional taxa explicitly resolved in DARWIN.

Line 933: “Finally, in addition to the 11 CMIP6 models, we included in Figure 4 and Figures S4-S6 the model of (Follett et al., 2022), which is part of the DARWIN Project, run on the MITgcm at 1°x1° resolution, and includes 31 and 16 size structured phytoplankton and zooplankton groups, respectively. Here grazing rates were saved explicitly at monthly resolution and integrated across phytoplankton and zooplankton groups offline. To account for the inclusion of mixotrophs, phytoplankton biomass is considered anything that is a primary producer (i.e. phytoplankton, mixotrophs) while zooplankton biomass is considered anything that consumes primary producers (i.e. herbivorous zooplankton, omnivorous zooplankton, mixotrophs). The bulk phytoplankton grazing mortality rate, *G*, thus include phytoplankton and mixotroph losses to grazing by both zooplankton and mixotrophs. Note, Darwin also includes zooplankton carnivory and 7 types of bacteria. These are not included explicitly in our diagnostics but since they dynamically influence zooplankton biomass and grazing effort, their inclusion, and the broader structure of the food web, is reflected our diagnostics.”

Line 221: “Perhaps a more useful comparison is a more complex BGC model. While no CMIP6 model included more than three zooplankton or phytoplankton groups, we have included in Figure 4 (black trace) the DARWIN model of (Follett et al., 2022), which includes 31 pPFTs and 16 zPFTs. Only BFM5.2, COBALTv2 and MARBL come close ($r^2 > .6$) to matching the zonal variability in grazing pressure simulated in DARWIN during both winter and summer. MARBL, however, exhibits by far the largest magnitude bias, while BFM5.2 has the second smallest. This suggests that active prey switching, only included in COBALTv2 and BFM5.2, is an essential ingredient in implicitly simulating the more diverse ecosystem explicitly resolved in DARWIN (Adjou et al., 2012; Prowe et al., 2012; Vallina et al., 2011; Chenillat et al., 2021). Interestingly though, grazing pressure in DARWIN is achieved via much lower specific grazing rates (Fig 4E) and higher zooplankton biomass (Fig 4H) than in BFM5.3 or COBALTv2. This is likely due to a decline in the direct interaction strength between predator-prey in a more complex food web (Sailley et al., 2013) combined with a reduction in non-linear mortality to explicit and implicit predation, which will be smaller in aggregate when biomass is distributed across more discrete biomass pools.”

7. The prey field of zooplankton is not only phytoplankton but also POC and zooplankton itself.

Changed. This has now been explicitly accounted for in both our Prescribed Grazing Index (see **Section 4.2**) as well as our offline computation of emergent grazing dynamics (see **Section 4.4**) and illustrated in our food web schematics (see **Figure 1, Table 1**).

Line 541: “The Prescribed Grazing Index (*PGI*) is computed as the zooplankton specific grazing rate, integrated across all phytoplankton prey options, prescribed by a given grazing formulation (Table 1) when evaluated at the observed global median phytoplankton and zooplankton plankton distribution and long-term mean sea surface temperature. In models with multiple zPFTs, zooplankton are assumed to be partitioned based on the observed global median distribution (Table S1) and the *PGI* is their biomass-weighted mean specific grazing rate on all phytoplankton prey options. When non-phytoplankton prey is included in the food-web (i.e. bacteria, detritus, or other zooplankton), it is accounted for in so much as it offers refuge for phytoplankton by diverting grazing effort. Mathematically, this is equivalent to evaluating Equation 3, which is used to compute the offline emergent dynamics (see Section 4.5), except here a standardized prey field is used instead of the emergent monthly mean biomass fields. By using a standardized prey field, we can obviate myriad other inter-model differences that drive the emergent distribution and abundance of prey and thus quantify a reduced approximation of the magnitude of grazing rates prescribed by the innate grazing formulation in each model. Accordingly, the *PGI* does not necessarily reflect the emergent grazing dynamics of a given simulation, which further depend on the formulation of bottom-up controls, physics, zooplankton mortality rates, and other aspects of the grazing formulation that cannot be captured with a single value.

The observed global median plankton community used to evaluate the *PGI* is based on the global median picophytoplankton, diatom, *Phaeocystis*, diazotroph, microzooplankton, mesozooplankton, and macrozooplankton concentrations observed across the top 200 m of the water column and compiled by the MARine Ecosystem biomass DATA (MAREDAT) initiative (Buitenhuis et al., 2013). The observed global median particulate organic carbon concentration (Stramski et al., 2022) and bacteria concentration from the upper 200 m (Buitenhuis et al., 2012) are included when available as prey. See ‘Description of Observations’ below. To account for differences in the structure of the food web, some assumptions must be made about how biomass is partitioned in models with different numbers of plankton groups; however, total phytoplankton and zooplankton biomass is kept constant. In models with just one phytoplankton or zooplankton group, total observed phytoplankton and zooplankton biomass is used. In models without diazotrophs or macrozooplankton, those pools are added to the ‘large phytoplankton’ and ‘medium zooplankton’ pools, respectively. *Phaeocystis*, which no surveyed model includes but compete with diatoms for nutrients, particularly in the Southern Ocean (Nissen et al., 2021), are included in the ‘large phytoplankton’ pool.”

Line 696: “The global median standardized prey field used in Prescribed Grazing Index is primarily based on field observations collected by from the MARine Ecosystem biomass DATA (MAREDAT) initiative. MAREDAT is a compilation of global, gridded biomass datasets for a range of PFTs and is described in detail in (Buitenhuis et al., 2013). Here, we use the reported global median epipelagic (0-200 m) concentration of microzooplankton (Buitenhuis et al., 2012), mesozooplankton (O’Brien et al., 2012), macrozooplankton (Moriarty et al., 2012), picophytoplankton (Buitenhuis et al., 2012), diatoms (Le Blanc et al., 2012), diazotroph (Luo et al., 2013), *Phaeocystis* (Vogt et al., 2012) to describe the phytoplankton and zooplankton distribution and in models in which bacteria is available as prey, use global median picoheterotroph concentration (which also include Archaea) (Buitenhuis et al., 2012).”

Line 718: “In models in which detritus is available as prey, detritus is estimated as the global median area-weighted particulate organic carbon (1-200 m) field observations reported in Table 1 of (Stramski et al., 2022) less the global median living biomass concentrations from Table S1.”

Line 850: “We analysed the output from historical simulations of 11 CMIP6 models (Table 2). Model output has been made publicly available by the Earth System Grid Federation (ESGF) and was downloaded from the Lawrence Livermore National Laboratory data node (<https://aims2.llnl.gov/search>). Note, three variables from ACCESS-ESM1.5 (phytoplankton biomass, zooplankton biomass, NPP) were not available on the ESGF-LNLL node and were downloaded directly from the Australian National Computing Infrastructure. For each model we considered monthly climatologies created from the historic period from 1850 to 1900. Considering years up to 1900 provides a long enough time series to average out ENSO variability but stops before there is a substantial anthropogenic forcing, as we are interested in the non-transient dynamics stemming from structural differences in the grazing formulation between models. We used the pre-1990 historical simulation instead of the pre-industrial control because more variables of interest were available across all models. For each model, we directly downloaded monthly averaged column integrated NPP, export through 100 m, and all depth resolved phytoplankton and zooplankton plankton fields. Depth resolved detritus, bacteria, temperature, and oxygen fields were additionally downloaded when required for diagnostic rate computations. Any exceptions are discussed below.

We then calculated diagnostic rate terms offline, pixel-by-pixel, using a given model's specified equations and parameters (Table 1; Table S2) evaluated with the depth resolved predator and prey biomass fields in addition to depth resolved temperature and oxygen fields where required.”

Table S1:

Field	MAREDAT Group	Observed Global Median Prey Field	Mean Modelled Global Median Prey Field	Observational Reference
Phytoplankton ($mmolC\ m^{-3}$)				
Small	Picophytoplankton	0.46	0.37	(28,154)
Large	Diatoms + Phaeocystis	0.33	0.24	(28,155,158)
Diazotrophs	Diazotrophs	2.5e-3	0.02	(28,157)
Total	Pico+Diat +Phae+Diat	0.79	0.61	(28,154-158)
Zooplankton ($mmolC\ m^{-3}$)				
Small	Microzooplankton	0.26	0.17	(28,93)
Medium	Mesozooplankton	0.23	0.16	(28,94)
Large	Macrozooplankton	0.02	0.04	(28,154)
Total	Micro+Meso+Macro	0.51	0.34	(28,93,94,154)
Other Carbon Reservoirs ($mmolC\ m^{-3}$)				
Bacteria	Picoheterotrophs	0.55	0.57	(28,151)
Detritus	-	2.65	1.12	(28,150)
Temperature ($^{\circ}C$)				
SST	-	15.30	15.61	(149)

Figure 1:

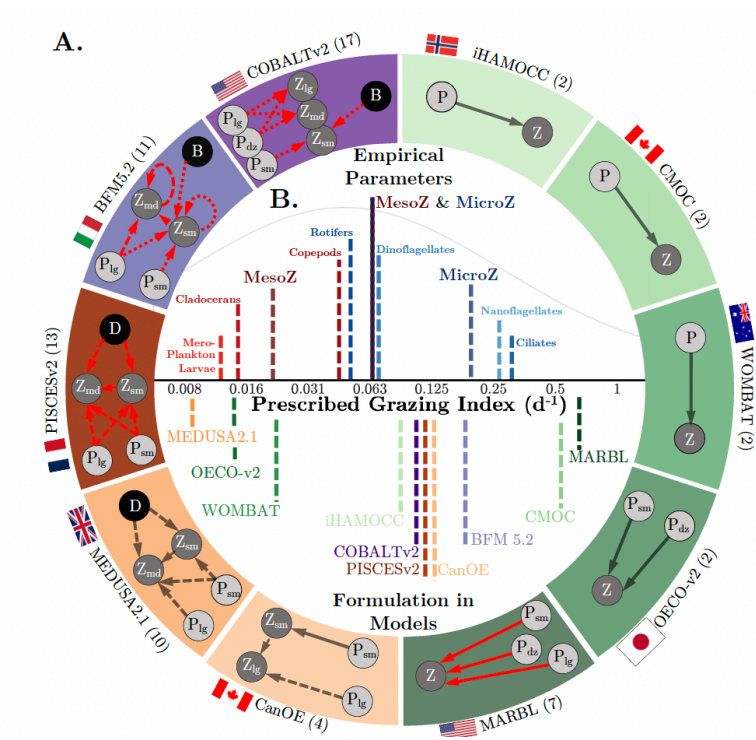


Figure 1: Variability in the formulation of grazing. **A)** The marine food webs represented in 10 CMIP6-class BGC models are presented in clockwise-order of increasing complexity. Aside the name of each BGC model is the number of parameters required to describe grazing. Grazing relationships (arrows) are solid for single prey responses, dashed for multi-prey responses with fixed preferences, and dotted for multi-prey responses with active switching. Red arrows indicate temperature sensitivity. Green, red/orange, and purple color schemes refer to models with 1, 2, or actively switching zooplankton. Hue scales qualitatively with complexity. PFTs have been generalized into small (nano-, small, non-diatom, or nanoflagellate), large (large, diatoms), or diazotrophs for phytoplankton (P) and small (micro-, small), medium (meso-, medium), or large (macro-, large) for zooplankton (Z). Bacteria (B) and Detritus (D) are included when available as prey. Note, model specific nomenclature for PFTs is included in Table 1. **B)** The Prescribed Grazing Index (*PGI*), an innate property of the grazing formulation, is plotted for all models below the *PGI* associated with empirical estimates of the functional response for 7 zPFTs reported in the review of (36). We further include three theoretical model configurations: two with a single zPFT described based on the median empirical parameters for meso- (dark red) and microzooplankton (blue blue), and one with 2 zPFTs including both (see “Methods”). Overlaid is the PDF of *PGI* for all 107 survey experiments.

8. Filter feeding is not the only mechanisms and there is prey selection. Some models have included it but are not used in the quantitative approach.

Changed. We now conduct quantitative analysis on 11 CMIP6 models, each with a diverse range of food webs and prey selection, all of which are detailed in the revised **Section 4.1** (see immediately below).

Note, of the CMIP6 BGC models considered half use a single prey response (no prey selection) and half include multi-prey functional response structure in which there is prey selection. Of these, 3 use fixed prey preferences and 2 (BFM5.2 and COBALTv2) use active switching. These distinctions have been highlighted in **Figure 1** (see above) and **Table 1** with dashed and dotted arrows in the food web.

Finally, prey selectivity (active and passive) is also explicitly accounted for in the Prescribed Grazing Index and offline computation of emergent rate terms (see our Response to **Comment 7**).

Line 444: “The representation of grazing in IPCC CMIP6 BGC models varies widely (Fig 1; Table 1; see also Kearney et al., 2021), as does their representation of phytoplankton and zooplankton groups. At one end of the spectrum, iHAMOCC (Tjiputra et al., 2020), CMOC (Zahariev et al., 2008) and WOMBAT (Law et al., 2017) include only one zooplankton and one phytoplankton group. In each, the zooplankton specific grazing rate, g (d^{-1}), is assumed to increase monotonically with the phytoplankton concentration, P (mmolC m^{-3}), before eventually saturating (Gentleman et al., 2008). This relationship, $g(P)$, is known as the single-prey functional response and is typically represented with a type II or III response curve (Gentleman et al., 2003; Rohr et al., 2022). Both curves can be described with the same two parameters. The saturation grazing rate, g_{max} (d^{-1}), describes the food-replete specific grazing rate, whereas the half-saturation concentration, $K_{1/2}$ (mmolC m^{-3}) specifies roughly where the curve begins to saturate (Fig. S1). All else equal, both increasing g_{max} or decreasing $K_{1/2}$ increase zooplankton specific grazing rates.

iHAMOCC uses a type II response, which linearly increases with prey density below $K_{1/2}$, when prey is scarce (rectangular hyperbola; Fig S1). A type II response is typically what has been measured in empirical laboratory experiments (Hansen et al., 1997; Hirst et al., 2003) and is thus likely to be representative of individual zooplankton grazing in a well-mixed environment. However, the type II response can contribute to dynamically unstable solutions (Gentleman et al., 2008), requiring iHAMOCC to include a small prey threshold (below which grazing stops) to prevent phytoplankton from going extinct. WOMBAT and CMOC use a type III response, which is concave upward below $K_{1/2}$ (sigmoidal; Fig S1), thereby offering prey refuge and population stability at low prey concentrations (Gentleman et al., 2008; Oaten et al., 1975). Ecologically, the type III response can be justified as the implicit representation of more complex behavioral changes such as active prey switching (Prowe et al., 2012; Gentleman et al., 2013), which has been observed in multiple zooplankton species (Strom et al., 2000; Kiorboe et al., 1996; Saiz et al., 1995; Kiorboe et al., 2018), or as the mean state of a spatially patchy ocean (Morozov et al., 2010; Rohr et al., 2022), which relatively coarse global models must implicitly average across. Despite qualitatively identical grazing formulations, differences in the parameter selection in CMOC and WOMBAT (particularly $K_{1/2}$; Table 1) mean that

zooplankton specific grazing rates on phytoplankton concentrations $< 1 \text{ mmolC m}^{-3}$ are nearly 30-times faster in CMOC than WOMBAT.

OECO-v2 (Hajima et al., 2020) and MARBL (Long et al., 2021) use a slightly more-complex ecosystem structure, resolving multiple groups of phytoplankton but just one generic group of zooplankton that grazes independently on all phytoplankton groups. Both models still employ a single-prey response, in that the presence of one prey option does not influence the zooplankton specific rate at which the other is grazed, $g \text{ (d}^{-1}\text{)}$. However, because the biomass of the generic zooplankton group, $Z \text{ (mmolC m}^{-3}\text{)}$ increases with grazing on any single phytoplankton group, so does the grazing pressure, $gp = gZ/P \text{ (d}^{-1}\text{)}$ on all phytoplankton groups, largely eliminating the possibility of refuge for relatively scarce prey (Sailley et al., 2013). In turn, this structure can obtain an effective food web that is functionally similar to including only one phytoplankton group (Sailley et al., 2013), in which strong top-down controls work to prevent species co-existence and stifle large blooms (Hashioka et al., 2013). This is particularly relevant in MARBL, which uses a type II response, a very small $K_{1/2}$ value, and requires relatively large prey thresholds to prevent extinction (Table 1), whereas OECO-v2 is able to offer prey refuge in the form of a type III response with a relatively large $K_{1/2}$ value.

CanOE (Christianet et al., 2021) is slightly more complex, with two phytoplankton and two zooplankton groups. Here, each zooplankton group grazes independently, pair-wise, on one unique phytoplankton group, with the large zooplankton group additionally grazing on the small zooplankton group. Unlike MARBL and CanOE, in which the indirect interaction strength (Montoya et al., 2009) between phytoplankton groups is always negative, trophic cascades can now lead to positive interaction strength between phytoplankton groups (Sailley et al., 2013; Le Quere et al., 2016}. That is, large phytoplankton growth can relieve grazing pressure on small phytoplankton by increasing the biomass of large zooplankton which in turn graze down the biomass of small zooplankton, the sole predator of small phytoplankton. Increasing in complexity, MEDUSA2.1 (Yool et al., 2021); PISCESv2 (Aumont et al., 2015); BFM5.2 (Lovato et al., 2022), and COBALTv2 (Stock et al., 2020) all include at least 2 zooplankton groups, each with multiple prey. In MEDUSA2.1 and PISCESv2, two types of zooplankton split their grazing effort between two types of phytoplankton in addition to detritus and cannibalism from large to small zooplankton, but both do so with fixed prey preferences. This means that the abundance of one prey option does not influence the desirability of another, such that when prey preferences are equal, grazing effort is proportional to the relative distribution of prey (Vallina et al., 2014). Using fixed prey preferences, known as passive-prey switching, ensures maximal feeding (i.e. adding prey biomass, all else equal, always leads to an increase in total ingestion) but does not provide prey refuge for relatively less abundant prey (Prowe et al., 2012). In turn, to help prevent extinction and provide some form of prey refuge, MEDUSA2.1 uses a type III (sigmoidal) response and PISCESv2 includes a relatively strong prey threshold which decreases the apparent prey concentration by half when the total prey concentration is $< 0.6 \text{ mmolC m}^{-3}$.

The most complex models considered, BFM5.2 and COBALTv2, both include density-dependent prey preferences. This is known as active prey switching and allows zooplankton to redirect grazing effort to relatively more abundant prey, even if it is otherwise less desirable (Fasham et al., 1990). Without invoking behavioural changes, active switching can be understood as a community response, in which one zooplankton group explicitly grazing on n phytoplankton groups implicitly functions as n fully-specialised zooplankton groups growing in abundance in proportion to their preferred prey (Vallina et al., 2014). In BFM5.2, only microzooplankton (but not mesozooplankton) have density-dependent prey preferences, while in COBALTv2, the most complex BGC model considered, three zooplankton groups all graze with density-dependent preferences. Density dependent, active-prey switching, improves ecosystem stability (Oaten et al., 1975), provide prey refuge (Fasham et al., 1990), overcome competitive exclusion (Adjou et al., 2012), increase species diversity (Prowe et al., 2012; Vallina et al., 2011; 2014; Chenillat et al., 2021), and better replicates species succession during phytoplankton blooms (Prowe et al., 2012). However, active-switching also results in the ecologically unlikely phenomena of non-maximal feeding, in which total ingestion can decrease as total prey abundance increases if prey simultaneously become more evenly distributed (Vallina et al., 2014, Gentleman et al., 2008). Vallina et al., 2014 have outlined a formulation for active switching in which the $K_{1/2}$ value for total ingestion decreases as prey become more evenly distributed, such that maximal feeding with active prey switching can be achieved; however, this formulation has not been included in any of the CMIP6 models surveyed here.”

8. for example:

Ward et al., J. Plankton Res. (2013) 0(0): 1–17. doi:10.1093/plankt/fbt097

Heneghan et al., 2020 Ecological modeling <https://doi.org/10.1016/j.ecolmodel.2020.109265>

Gehlen et al., 2006: Biogeosciences, 3, 521–537,

Including these models in their quantitative analysis of sensitivity would really make this paper up to date and would inform about existing improvement. It would also quantify the improvement (or not?)

Included.

1. As detailed in our response to **Comment 6** we have included the model of Follet et al. (2022) in our quantitative analysis. As with the model of Ward et al (2013), the model of Follet et al. (2022) (10.1073/pnas.2110993118) is part of the DARWIN project, run on MITgcm, and was overseen by Stephanie Dutkiewicz and Mick Follows. The model of Follet et al. (2022) also has a comparable number of size-scaled plankton and zooplankton groups to that of Ward et al (2013) (47 vs 55). We chose to analysis the Follet et al. (2022) simulation because a) it is more current and b) Steph still had explicit monthly mean grazing rate and biomass variables archived and was graciously willing to share them with us. This allowed us to explicitly compute community averaged grazing pressure and specific grazing rates across all 47 PFTs and quantitatively compare the zonal distribution in grazing dynamics to models with much simpler trophic structures (see above).
2. We also spoke with Ryan Heneghan about including his ZooMSSv2 model. While ZooMSSv2 does include much richer zooplankton functional diversity it does not include any two-way coupling with primary producers. That is, it is forced with NPP but grazing does not feed-back on. For this reason, we agreed that it was not an appropriate comparison to BGC/Earth system models. We do, however, now reference Ryan’s work and note the need to bridge the gap between BGC and ecosystem/fisheries models.

Line 317: “Likely, different BGC models will yield qualitatively different projections of secondary production based on their initial *PGI* and on what side of the “overgrazing” inflection point they begin. This is particularly problematic for ecosystem (Butenschon et al., 2016) and fisheries models (Tittensor et al., 2021), which are typically driven with a one-way forcing from NPP and assume NPP and GSP have an exclusively positive relationship (Tittensor et al., 2018; 2021), further stressing the need for two-way coupling with primary producers to forecast fisheries stocks accurately (Kearney et al., 2012}).”

3. We now include an explicit quantitative analysis of two CMIP6 models which use PISCESv2 (IPSL and CNMR; see Table S2). PISCESv2 is a more advanced version of the original PISCES model used in Gehlan et al. (2006). Both include in 2 zooplankton groups which graze preferentially on 2 phytoplankton groups and POC (see Figure 1, Table 1). Note, the version used by Gehlan et al (2006) includes a size structured particle ballasting model, but not a size structured plankton community beyond what is implied by the parameterization of the 4 PFTs.

Table S2:

	BGC Model	Earth System Model	Simulation	Model ID	Institution (Location)	Personal Communication
	iHAMMOCC (64)	NorESM2-ESM-LM (64)	Historical (82)	r2i1p1f1	NORC (NOR)	Jerry Tjiputra Jörg Schwinger
	CMOC (65)	CanESM5 (69)	Historical (83)	r1i1p2f1	ECCC (CAN)	Jim Chirstian
	WOMBAT (66)	ACCESS ESM1.5 (66)	Historical (84)	r1i1p1f1	CSIRO (CAN)	Andrew Lenton Mathew Chamberlain
	OECCO-v2 (67)	MIROC-ES2L (67)	Historical (85)	r1i1p1f2	JAMSTEC (JAP)	-
	MARBL (68)	CESM2 (78)	Historical (86)	r1i1p1f1	NCAR (USA)	Matt C. Long
	CanOE (69)	CanESM5 (69)	Historical (87)	r2i1p2f1	ECCC (CAN)	Jim Christian
	MEDUSA2.1 (70)	UKESM1-0-LL (79)	Historical (88)	r1i1p1f2	Met Office, NERC (UK)	-
	PISCESv2 (71)	IPSL-CM6A-LR (80)	Historical (89)	r10i1p1f1	IPSL (FR)	-
	PISCESv2 (71)	CNRM-ESM2.1 (79)	Historical (90)	r1i1p1f2	CNMR (FR)	-
	BFM5.2 (72)	CMCC-ESM2 (72)	Historical (91)	r1i1p1f1	CMCC (ITA)	Tomas Lovato
	COBALTv2 (73)	GFDL4.1 (81)	Historical (92)	r1i1p1f1	GFDL (USA)	Charlie Stock

9. The current version of the manuscript do not really propose and discuss future evolution. They do not really address the possibility of joint co development in observation and modeling. The paper by J. D. Everett, et al., *Frontiers in Marine Science* 4, 77 (2017) is cited but only in the beginning.

Changed. In the revised manuscript we provide a much richer discussion of the current state of observations, which includes and extends far beyond the work Everett et al. (2017) (see below) as well as a “Moving Forward” Section (see below) where we discuss the joint development of observations and modelling.

Line 35: “Substantial observational work over the last two decades has helped describe the global distribution of zooplankton and their grazing dynamics, but remains limited as a means to adequately constrain grazing in global models. In-situ measurements of zooplankton biomass (Buitenhuis et al., 2006, 2013; Moriarty et al., 2013; Moriarty and O’Brien, 2013) are relatively patchy and must reconcile uncertainties across disparate acoustic, optical, and net-based methodologies (Lombard et al., 2019; Everett et al., 2017). Progress has been made towards estimating zooplankton biomass from satellites but current products are either limited to identifying swarms of specific species (Basedow et al., 2019) or heavily reliant on statistical inference from chlorophyll fields (Druon et al., 2019; Stromberg et al., 2009).

Grazing pressure, defined as the phytoplankton specific loss rate to grazing (Prowse et al., 2012), varies with both zooplankton biomass and the specific grazing rate of zooplankton. However, empirical estimates of the parameters that describe how specific grazing rates vary with prey abundance span three orders of magnitude, varying with zooplankton size, species, and age (Hansen et al., 1997; Hirst et al., 2003; Rohr et al., 2022). This means that detailed information about community composition must accompany bulk estimates of biomass to estimate grazing pressure. Field-based dilution experiments can help average over this variability, providing estimates of community averaged grazing pressure (Morrow et al., 2018; Landry et al., 2009; Landry et al., 2008; Landry et al., 2000), but these are highly localized in space and time and can be biased by nonlinearities in the functional response (Sandhu et al., 2019), trophic cascades (Calbet et al., 2008; 2011) and the presence of mixotrophs (Calbet et al., 2012). Review studies have helped constrain the range of reasonable values, but are limited to broad basin-scale averages and often exclusively target microzooplankton (Schmoker et al., 2013; Calbet and Landry, 2004; Landry and Calbet, 2004). Measuring the contribution of larger mesozooplankton in the field typically requires gut content analysis and various assumptions about metabolic rates (Morrow et al., 2018; Calbet et al., 2001). Finally, give a good snapshot of which functional groups belong where, but do not provide time evolving global data sets. Collectively, the observational record provides a strong baseline understanding of zooplankton grazing dynamics but cannot be used in the same way as high spatial and temporal resolution, satellite-derived chlorophyll and NPP products to tune global models.”

Line 358: “Qualitative differences in the structure of the food web (Sailley et al., 2013; Prowse et al., 2012; Vallina et al., 2011), degree of trophic complexity (Le Quere et al., 2016; Hashioka et al., 2013), sensitivity to temperature (Laufkotter et al., 2015), and capacity for prey refuge (Chenillat et al., 2021; Nissen et al., 2018; Prowse et al., 2012; Vallina et al., 2014) all exert a strong top-down control on NPP and subsequent carbon cycling. Ideally, models that leverage these distinctions to increase biodiversity (Chenillat et al., 2021; Prowse et al., 2012; Vallina et al., 2014; Dutkiewicz et al., 2020), prevent competitive exclusion (Adjou et al., 2012), improve ecosystem stability (Vallina et al., 2011; Gentleman et al., 2008), or replicate species succession (Prowse et al., 2012) and bloom phenology (Hashioka et al., 2013; Karakus et al., 2022), ought also tune their parameters to converge on a relatively constrained degree of bulk grazing pressure. However, this does not appear to be happening. Large variability in emergent grazing dynamics is well explained by the fact global, regional, and seasonal scales, despite the *PGI* reducing many modes of qualitative variability into a single value. In turn, models with substantial qualitative differences in their grazing formulation and food web structure but similar *PGI* can yield similar zonal distributions of emergent grazing rates (i.e. MEDUSA2.1 vs OECO-v2; Figure 4B) and/or grazing pressure (i.e. CanOE vs PISCESv2; Figure 4E). Together, this highlights how far CMIP6 models remain from converging on a first-order magnitude of zooplankton specific grazing rates, leading to very large uncertainty in the representation of grazing pressure throughout the global ocean.

In turn, even if models are tuned to the same NPP and EP, they must do so with very different-sized biomass pools and turnover rates to accommodate such large differences in grazing pressure (Figure S9). However, to constrain EP with different-sized biomass pools, models must further differ in their particle aggregation, ballasting, and/or remineralization rates (Laufkotter et al., 2016). Yet each of these terms have different temperature sensitivities, meaning they will respond differently to a warming ocean. this helps explain the large uncertainty in projection of EP (Fu et al., 2016) despite fairly well constrained historical reconstructions (Figures 4; 5J). Therefore, constraining uncertainty in the *PGI*, and in turn emergent grazing pressure and the distribution and abundance of biomass, appears to be a necessary and urgent next step towards constraining predictions of future marine carbon sequestration and transfer to higher trophic levels.”

Adequately constraining grazing may require a new generation of observational products that can leverage higher resolution, time-evolving remote sensing and in-situ BGC Argo float-based data sets. While these platforms cannot directly measure grazing, both satellites and floats products have been used to infer Net Community Production from biomass (Rohr et al., 2017; Li et al., 2021) and oxygen budgets (Su et al., 2022; Arteaga et al., 2019). Combined with observationally-based estimates of phytoplankton biomass (Westberry et al., 2008), NPP (Westberry et al., 2008; Yang et al., 2021), and the assumption that phytoplankton losses are dominated by grazing (Mojica et al., 2021; Deppeler et al., 2017), grazing pressure could then be estimated at much finer space and time scales than ship-board dilution experiments can accommodate (Schmoker et al., 2013; Calbet et al., 2004). Further pairing these with rough estimations of zooplankton biomass, either from statistical models (Heneghan et al., 2020; Petrick et al., 2022) or emerging remote sensing products (Basedow et al., 2019; Druon et al., 2019, Stromberg et al., 2009), could yield specific grazing rate terms, enabling modellers to better tune zooplankton turnover (i.e. the balance of growth vs. mortality).

Together, such products will not only help constrain the mean state of grazing at global and regional scales, but could also complement existing bio-geographical resources to constrain how grazing dynamics qualitatively vary regionally with community composition. That is, by plotting the evolution of observed grazing rates against phytoplankton biomass, one can empirically define the community-averaged apparent functional response at regional and even seasonal scales. Spatio-temporal variability in this curve will improve our understanding of the distribution (Barton et al., 2013; Brun et al., 2016; Benedetti et al., 2023), seasonal succession (Brun et al., 2019; Kenitz et al., 2017), and migration (Richardson et al., 2008; Atkinson et al., 2019) of zooplankton species, which can be then be compared to the associated variability in the emergent apparent functional response simulated in models (see ``Supplemental Discussion 1"). Such comparisons will enable modellers to determine if explicit competition in their assumed food-webs are reasonably approximating natural variability in community-averaged grazing dynamics, or if increasing trophic complexity by including, for example, macrozooplankton (Le Quere et al., 2016), salps (Luo et al., 2020), larvaceans, euphausiids, chaetognaths, jellyfish (Heneghan et al., 2020; 2023), pteropods (Buitenhuis et al., 2019), heterotrophic protists (de Vargas et al., 2015) or Rhizarians (Biard et al., 2016; Guidi et al., 2016) might be required.

Moreover, as stratification under climate change likely leads to greater light availability but fewer nutrients, the phytoplankton community may shift toward smaller phytoplankton with faster nutrient uptake but higher light requirements (Portner et al., 2019; Bopp et al., 2005). This may increasingly favour rapidly-grazing microzooplankton or gelatinous filter feeders, which can ingest smaller phytoplankton size classes (Conley et al., 2018; Heneghan et al., 2023), over crustacean mesozooplankton, which generally graze slower but can consume much larger prey (Everett et al., 2021; Barton et al., 2013). Considering these groups serve critically different roles in trophic transfer through prey preference (Biard et al., 2016) and carbon export through diel-vertical migration and (Bianchi et al., 2013; Steinberg and Landry, 2017; Archibald et al., 2019; Aumont et al., 2018) the formation of rapidly-sinking fecal pellets and carcasses (Luo et al., 2020, Clerc et al., 2023), it is essential that models include enough trophic diversity to capture present-day and emergent changes to community composition and it's associated traits (Heneghan et al., 2023)."

As it stands, grazing dynamics appear to be the largest source of uncertainty in contemporary simulations of the marine carbon cycle. This is understandable given the considerable challenges associated with observing and subsequently modelling zooplankton (Everett et al., 2017; Mitra et al., 2014; Ratnarajah et al., 2023); however, moving forward, improving zooplankton grazing will be critical. This is especially true given BGC models will be at the forefront of informing or deterring the possible implementation of a wide range of ocean-based, potentially biologically-driven, negative emissions technologies (NAS, 2021)."

10. Finally, recent observations have shown that Rhizarians (Biard et al., 2016, Nature Letter, doi:10.1038/nature17652) are important biomass in the ocean and their inclusion in future BGC models should be taken into consideration.

Changed. We now note and discuss the omission of Rhizarians and other zooplankton groups.

Line 680: "Finally, note that this is not an inclusive of all important zooplankton species or functional groups. For example, large photo-symbiotic Rhizaria may constitute 50% or all mesozooplankton biomass (Biard et al., 2016) and help explain why oligotrophic food web may be shorter than predicted by classical ecology (Everett et al., 2021). Moreover, neither macrozooplankton nor gelatinous filter feeders are included. Macrozooplankton typically graze slowly on smaller zooplankton but have can regulating grazing pressure on diatoms and of pPFTs through carnivorous trophic cascades (Le Quere et al., 2016). On the other hand, larvaceans and salps are extremely abundant mesozooplankton that can filter bacteria and picophytoplankton, unavailable to crustaceans. They generally grow an order of magnitude faster than crustacean zooplankton (Sutherland et al., 2022) and have been shown to play an important role in carbon export via the production of rapidly-sinking carcasses and fecal pellets (Luo et al., 2020, Clerc et al., 2023). Never the less, the empirical values included here are meant to provide context for the two PFTs most typically included CMIP6-class BGC models, not provide an exhaustive list of grazing parameters for all known zooplankton species. Note, Salps (Luo et al., 2020; Clerc et al., 2023) and macrozooplankton (Le

Quere et al., 2016; Dutkiewicz et al., 2020) have been routinely included in targeted global simulations, but their inclusion in high resolution climate-ensembles is not yet standard.”

Line 402: “Such comparisons will enable modellers to determine if explicit competition in their assumed food-webs are reasonably approximating natural variability in community-averaged grazing dynamics, or if increasing trophic complexity by including, for example, macrozooplankton (Le Quere et al., 2016), salps (Luo et al., 2020), larvaceans, euphausiids, chaetognaths, jellyfish (Heneghan et al., 2020; 2023), pteropods (Buitenhuis et al., 2019), heterotrophic protists (de Vargas et al., 2015) or Rhizarians (Biard et al., 2016; Guidi et al., 2016) might be required.”

I have also minor comments:

11. - I noticed minor typos. Should be checked as I'm not a native english speaking person.

Changed. We have proofread the revised manuscript several times and hopefully caught and amended all remaining typos in the text and bibliography.

12. - I noticed an unconsistency biobibliographie formatting:
for example reference 58

Specifically, the incomplete reference to Everett et al., (2017) has been corrected. Previously (58), now (32)

12th May 23

Dear Dr Rohr,

Your manuscript titled "Zooplankton grazing is the largest source of uncertainty for marine carbon cycling in CMIP6 IPCC models" has now been seen by our reviewers, whose comments appear below. In light of their advice I am delighted to say that we are happy, in principle, to publish a suitably revised version in Communications Earth & Environment under the open access CC BY license (Creative Commons Attribution v4.0 International License).

We therefore invite you to revise your paper one last time to address the remaining concerns of our reviewers. At the same time we ask that you edit your manuscript to comply with our format requirements and to maximise the accessibility and therefore the impact of your work.

EDITORIAL REQUESTS:

Please review our specific editorial comments and requests regarding your manuscript in the attached "Editorial Requests Table".

*****Please take care to match our formatting and policy requirements. We will check revised manuscript and return manuscripts that do not comply. Such requests will lead to delays. *****

Please outline your response to each request in the right hand column. Please upload the completed table with your manuscript files as a Related Manuscript file.

If you have any questions or concerns about any of our requests, please do not hesitate to contact me.

SUBMISSION INFORMATION:

In order to accept your paper, we require the files listed at the end of the Editorial Requests Table; the list of required files is also available at <https://www.nature.com/documents/commsj-file-checklist.pdf>.

OPEN ACCESS:

Communications Earth & Environment is a fully open access journal. Articles are made freely accessible on publication under a [CC BY license](http://creativecommons.org/licenses/by/4.0) (Creative Commons Attribution 4.0 International License). This license allows maximum dissemination and re-use of open access materials and is preferred by many research funding bodies.

For further information about article processing charges, open access funding, and advice and support from Nature Research, please visit <https://www.nature.com/commsenv/article-processing-charges>

At acceptance, you will be provided with instructions for completing this CC BY license on behalf of all authors. This grants us the necessary permissions to publish your paper. Additionally, you will be asked to declare that all required third party permissions have been obtained, and to provide billing information in order to pay the article-processing charge (APC).

Please use the following link to submit the above items:

[link redacted]

** This url links to your confidential home page and associated information about manuscripts you may have submitted or be reviewing for us. If you wish to forward this email to co-authors, please delete the link to your homepage first **

We hope to hear from you within two weeks; please let us know if you need more time.

Best regards,

Clare Davis, PhD

Senior Editor

Communications Earth & Environment

www.nature.com/commsenv/

@CommsEarth

REVIEWERS' COMMENTS:

Reviewer #1 (Remarks to the Author):

I would like to congratulate the authors on an extremely thorough revision of their work. I am very happy that my comments appear to have been helpful, and it seems to me that all my major concerns have been addressed and all my major comments have been taken into account in a rather extensive revision of the work. So if the editor thinks this work is within the scope of this journal, then I look forward to seeing this paper published in the near future.

Reviewer #2 (Remarks to the Author):

The authors have appropriately addressed my concerns and I believe that this is an interesting study. I have only minor issues to bring to the authors' attention:

The term "ecosystem efficiency" is used in multiple places in the manuscript. Reading through the manuscript, I assumed that ecosystem efficiency was a metric that quantified the efficiency of trophic transfer through the ecosystem (alternately referred to as ecosystem efficiency, trophic efficiency, trophic transfer efficiency, etc. in other manuscripts). This typical definition of ecosystem efficiency is dependent on things like food chain length and organismal growth efficiencies. However, the authors definition is simply grazing divided by net primary production. This is very different from typical definitions of ecosystem efficiency, and since it is not defined until line 890, I would suggest that the authors use a different term that would be less likely to introduce confusion (i.e., "Grazing Percentage" or "Herbivory Fraction" or something along those lines would be more descriptive of the actual metric).

Line 170 (and elsewhere): I would recommend formatting the p in gp as a subscript. Elsewhere in the manuscript, a bar over two letters is meant to mean the mean of the product of those two variables. For this reason, I initially thought that gp was the product of g times p . Making the p a subscript would alleviate this confusion. Alternately, a different variable could be used for gp .

Line 292: I believe that this should reference S8N, not S8M.

Line 314: "reduce can" should be "can reduce"

Line 347: "lead" should be "led"

Line 350: "reign" should be "rein"

Line 353: "later" should be "latter"

Line 377: Capitalize "this"

Line 426: "theri" should be "their"

Line 466: "Cespite" should be "Despite"

Line 512: "zooplantkon" should be "zooplankton"

Line 514: "provide" should be "provides"

Line 515: "overcome" should be "overcomes"

Line 524: "alometrically" should be "allometrically"

Line 669: "if" should be "is"

Line 702: "picohetreotroph" should be "picoheterotroph"

Line 718: grammatical error

Line 768: "use" should be "us"

Line 780: "deil" should be "diel"

Line 808: "micozooplankton" should be "microzooplankton"

Line 944: "is reflected" should be "is reflected in"

Reviewer #3 (Remarks to the Author):

Dear authors,

Thank you for the detailed review of the manuscript.

The introduction, results, discussion have been completed with extensive new results and more complete review of the literature.

This work is now of interest to a wide audience.

The manuscript is now ready for publication.

congratulation for your work.

Response to Reviewer 2

We would first like to thank Reviewer 2 for their careful eye and constructive comments throughout. Please see below for a point-by-point response to all remaining comments. Our Response is in blue.

The term "ecosystem efficiency" is used in multiple places in the manuscript. Reading through the manuscript, I assumed that ecosystem efficiency was a metric that quantified the efficiency of trophic transfer through the ecosystem (alternately referred to as ecosystem efficiency, trophic efficiency, trophic transfer efficiency, etc. in other manuscripts). This typical definition of ecosystem efficiency is dependent on things like food chain length and organismal growth efficiencies. However, the authors definition is simply grazing divided by net primary production. This is very different from typical definitions of ecosystem efficiency, and since it is not defined until line 890, I would suggest that the authors use a different term that would be less likely to introduce confusion (i.e., "Grazing Percentage" or "Herbivory Fraction" or something along those lines would be more descriptive of the actual metric).

We have renamed the 'Ecosystem Efficiency' to 'Grazing Efficiency' to disguise between the transfer of carbon through a larger fraction of the food web. To further clarify we have added the formula for 'Grazing Efficiency' (i.e. G/NPP) in all figures where it is included (3, 5, 9, and S5)

Line 170 (and elsewhere): I would recommend formatting the p in gp as a subscript. Elsewhere in the manuscript, a bar over two letters is meant to mean the mean of the product of those two variables. For this reason, I initially thought that gp was the product of g times p . Making the p a subscript would alleviate this confusion. Alternately, a different variable could be used for gp .

Done. We have reformatted gp as g_p in all figures and the text.

Line 292: I believe that this should reference S8N, not S8M. - Corrected

Line 314: "reduce can" should be "can reduce"- Corrected

Line 347: "lead" should be "led" - Corrected

Line 350: "reign" should be "rein" - Corrected

Line 353: "later" should be "latter" - Corrected

Line 377: Capitalize "this" - Corrected

Line 426: "theri" should be "their" - Corrected

Line 466: "Cespite" should be "Despite" - [Corrected](#)

Line 512: "zooplantkon" should be "zooplankton" - [Corrected](#)

Line 514: "provide" should be "provides" - [Corrected](#)

Line 515: "overcome" should be "overcomes" - [Corrected](#)

Line 524: "alometrically" should be "allometrically" - [Corrected](#)

Line 669: "if" should be "is" - [Corrected](#)

Line 702: "picohetreotroph" should be "picoheterotroph" - [Corrected](#)

Line 718: grammatical error - [Corrected](#)

Line 768: "use" should be "us" - [Corrected](#)

Line 780: "deil" should be "diel" - [Corrected](#)

Line 808: "micozooplankton" should be "microzooplankton" - [Corrected](#)

Line 944: "is reflected" should be "is reflected in" - [Corrected](#)