Dear referee,

Thank you for your comments on our manuscript. We appreciate the time and effort that you have dedicated to providing your valuable feedback. Here are our point-by-point responses to these comments and concerns.

General comment from Reviewer 1:

This paper presents an interesting and ambitious experiment examining the biological, chemical, and biogeochemical effects of adding olivine (Mg2SiO4) or steel slag (primarily CaO) to a Tasmanian estuarine plankton community. A particular strength of the experimental design is the use of complex natural plankton communities in relatively large-volume mesocosms, which improves the ecological relevance of the experiment. The biological sampling regime included bacteria and zooplankton in addition to the main focus on the phytoplankton community, and chemical measurements like carbonate chemistry, nutrients and dissolved and particulate trace metals all add additional valuable dimensions to this study. The findings of this incubation experiment will be of great interest to researchers seeking to understand the impacts of proposed OAE mitigation strategies on coastal marine ecosystems.

Response: We thank Reviewer 1 for his kind comments.

Major comment1:

One important qualification of this study is that Guo et al. must necessarily deal with something of an "apples and oranges" issue with the two alkalinity sources they used. CaO is well known to be a far more concentrated source of alkalinity and dissolves much more readily and rapidly than olivine, as their results also show. This means that it is virtually impossible to make rigorous, quantitative "apples to apples" comparisons between slag and olivine by using equivalent levels of added alkalinity, adding identical weights or volumes of total mineral particulates, adding the same amount of mineral-associated trace metals or nutrients, or by using uniform levels of any other property they have in common to make comparisons. Thus, amounts of each mineral added are necessarily rather arbitrary. This means that direct comparisons between the treatments are highly context-dependent, in that a somewhat different experimental design (i.e., adding more or less of either component) is likely to yield quite different relative outcomes for the chemistry, and perhaps the biology too. This doesn't compromise the value of their experiments, since the general trends in each treatment are still well worth presenting and considering, but it does suggest that interpreting the results in terms of direct quantitative comparisons between the two addition treatments should be done cautiously, and be explicitly qualified in the text.

Response: Excellent comment, thank you and we agree with your assessment. The amounts of olivine and slag powder added in the treatments were significantly different resulting in the issue of a quantitative comparison, as you describe. Our original goal was to yield somewhat similar amounts of detectable alkalinity enhancement in the dissolved phase from olivine and slag addition. However, olivine was much less efficient in releasing alkalinity as we expected so that even a 50-fold higher addition of olivine (in mass) did not compensate for this difference. Therefore, our discussion mainly relates the observed environmental effects with the alkalinity enhancement achieved. We added a paragraph section 4.2 of the Discussion pointing out this "apples and oranges" issue to the reader.

Specific comments:

Comment2: Abstract, lines 22-26: Results for aluminum, manganese and nickel are described here, but many readers will also be looking for the iron results. Perhaps briefly include these in the Abstract too?

Response: We added a sentence in the abstract describing dissolved Fe concentrations. "After 21 days, no significant difference was found in dissolved iron concentrations (>100 nmol L^{-1}) in different the treatments and the control."

Comment3: Methods, Figure 1 and lines 123-126: Using heat belts wrapped around the base of the mesocosms is an interesting and innovative way to control temperatures and set up convective circulation to help keep the plankton community in suspension. Was there a persistent vertical thermal gradient inside the mesocosms? It is also notable that temperatures were fairly variable in some of the mesocosms, with differences of up to 2-2.5C between replicates in some treatments on several days (Fig 2b). Is it possible that this affected replicability of some of the biological parameters?

Response: Thanks for pointing this out. The convective circulation methods were established and described in more detail in a recent paper by Ferderer et al., (2022). Convective mixing was monitored using a food dye, and the water colour usually become consistent 20mins after the food dye was added. The time lapse video of the convective mixing test can be accessed online at https://doi.org/10.5446/55861 (Federer, 2021). Yes, this could have affected the replicability of the treatments/control and may have added noise. We added a statement emphasizing that the convective mixing method could have introduced noise in the biological response data (Line 354-356). However, on average there was no statistically significant difference in temperature between control/treatments during the experiment.

Comment4: Lines 185-187. The authors should be commended for acknowledging that they are not presenting trace metal or phosphate results from several samples they considered contaminated, a workaround which has often been used in the trace metal literature. However, throwing out 7 of the 36 samples (~23%) is an unusually high proportion of the total. It would perhaps be a good idea to add a small table showing these excluded measurements in the SI so readers can judge the merits of this decision for themselves. Along these lines, it would be good to know what precautions were taken to facilitate trace metal clean water collections, incubations and sampling. Other than some brief description of acid-washing supplies and equipment, no specific precautions are described in the Methods section. I agree with the authors that in situ trace metals in this (quite contaminated) estuary are naturally elevated, and obviously the mineral additions push these even higher, but levels of some easily-contaminated metals like Fe or Zn could still be accidentally significantly increased by sub-optimal experimental protocols.

Response: We agree that the proportion of samples we excluded were relatively high compared with other research, and there could be contamination introduced from the sub-optimal experimental protocols. Sampling the microcosm in the temperature control room with potential contaminations coming from the air is the most likely source of contaminations in the dissolved trace metal samples (added in line 191-192). We added a table in the supplementary material (Tabel S1) showing all the raw data including these excluded values. Please note that in previous manuscript, we removed microcosm 5 and 7 on day one because their P concentrations were much higher than other samples excluding the outliers. The P concentrations from these two microcosms measured by ICP-MS were also higher than the P concentrations measured by the spectrometer. However, these values fall into the IQR zone which means they are not outliers determined by IQR methods. So we decide to keep them in the revised manuscript.

Comment5: Lines 203-209: The flow cytometer is an excellent way to enumerate single-cell phytoplankton or stained bacteria. However, it doesn't work at all well to count chain-forming, very large or very spiny species like many diatoms and some dinoflagellates, groups which tend to be quite prominent in the coastal ocean. Were any other methods (microscopy, flow cam) used to assess the abundance of these often important groups that are not easily counted with flow cytometry?

Response: Unfortunately, we did not use another method to assess the abundance of these large phytoplankton groups. We added a statement noting the potential underrepresented large microeukaryotes abundance in the result (line 471).

Comment6: Line 237: Please add a citation to the original paper presenting the oxalate wash cell surface-wash method (Tovar-Sanchez et al. 2003, Marine Chemistry 82).

Response: Thanks. We have added the citation (line 245).

Comment7: Line 245: The CHN analyses would have yielded numbers for PON as well as the POC discussed here in the Methods and presented later. Are these PON data interesting and potentially worth presenting (possibly in the SI)? The PON values would also allow calculation of changes in whole plankton community C:N ratios, which could be worth examining too. On the subject of ratios, it might be interesting to normalize the BSi values to the POC and the PON to get an idea of the relative degree of silicification of the communities, instead of presenting BSi only as volume-normalized values in Fig 5.

Response: Thank you for your comment. We took samples for POC, PON and C:N ratios, but a large part (day 12 to 21) of our POC/PON data were lost due to system failure (the auto-sampler "ate" these samples). The C:N ratio was calculated for the remaining data, but no significant difference was found between treatments and the control (see Figure 1 below). PON was lower during day 2-5 in the olivine treatment (see Figure 2). In addition, the BSi data during day 2 and 4 from the olivine treatment was removed due to particle influence so we didn't normalize the BSi to POC or PON. Considering these data were not complete and the manuscript is already quite crowded with many plots, we decided not to include PON, C:N or BSi:POC in the manuscript.

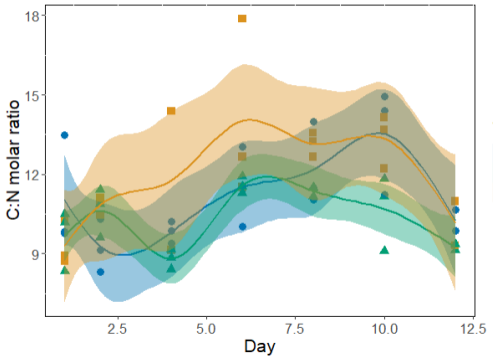

Fig. 1. The C:N ratios in microcosms.

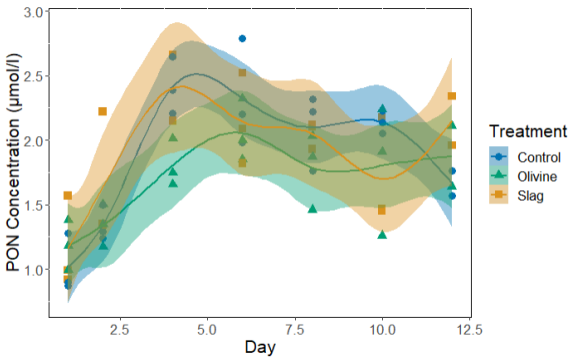

Fig. 2. The particulate organic nitrogen concentration

Comment8: Line 257: I am concerned about the accuracy of the zooplankton abundance measurements made using the self-made plankton net, which apparently had a diameter of only 1.5cm. Zooplankton tend to be patchily distributed, as they discuss later, and such limited volume sampling is likely be especially problematic for larger, low-abundance groups like larvaceans and copepods. The latter also have issues with active net avoidance that may be quite difficult to deal with using such a small collecting aperture.

Response: Thank you for your valuable comment. We apologize for the error in our manuscript regarding the zooplankton net size unit. The correct specifications are as follows: "20cm in height and 15cm in width with a 210µm mesh size," instead of “20mm in height and 15mm in width”. We corrected the manuscript accordingly.

Comment9: Lines 340-345: This text probably belongs in the Statistics section of the Methods, not the Results.

Response: Agreed. We moved this section to the Methods.

Comment10: Lines 390-392 and Fig. S3: It is interesting that the olivine released Cu to the seawater, but it is then puzzling that Cu was not reported to be present in the mineral stock used in Table 1.

Response: Thank you for your comment. It is likely that Cu was not probably detected by the SEM method used in the analysis of the mineral stocks applied here. This of course does not mean it was not present especially the olivine rock we used in the experiment contains many other particles from the quarry. Nevertheless, to avoid confusion, we change the description of dissolved Cu concentration by deleting the sentence “the olivine treatment released Cu into the seawater” (Line 410).

Comment11: Fig. 3: It is surprising that adding only 2 grams of slag to a 53L mesocosm can enrich seawater concentrations of phosphate and silicate to this extent. Although these elevated nutrient additions didn’t seem to have much of an effect on the biota in this N-limited experiment, in other regimes where P or Si are scarce this could definitely have a large impact. This issue should be discussed somewhere in the Discussion.

Response: We agree. We discussed the implications of P release from slag in section 4.2.1 (line 584-599).

Comment12: Lines 385- 405: This section on comparative trace metal releases from both alkalinity sources is a good example of my major general comment above. The relative amounts of metals released from each treatment are clearly a function of the relative amount of each material that was chosen to be added at the beginning: If (for instance) 3g slag had been added instead of 2g, or 50g olivine instead of 100g, the relative release results would likely look quite different. These results as presented are certainly valid, but are very context-dependent in that they only apply to the specific concentrations of each mineral that was added here. One way to deal with this issue would have been to test a range of concentrations of each of these two mineral sources, and examine the data in light of these two gradients. However, I recognize that in a large volume mesocosm experiment this many treatments is usually not practical. I do think some prominent qualifying text in the Discussion is needed to point this important caveat out to readers, though.

Response: Thank you, please find our response in relation to your major comment1. We also added a sentence at the beginning of the paragraph emphasizing the amount of minerals added were 50 times different. (Line 405-407)

Comment13: Line 419: Fig. 5 needs letters to differentiate the panels.

Response: Thanks, letters are mentioned in brackets.

Comment14: Lines 458-465: The trends in biovolume for these groups seem to be quite different from those of cell numbers reported in the previous paragraph. For instance, no differences in picoeukaryote, cyanobacteria or cryptophyte biovolumes were observed in any of the treatments

relative to the control, whereas these groups were sometimes significantly higher in the olivine treatment when assessed using cell numbers. Why is this- were there changes in cell diameters and volumes in the treatments? The f.c. results should be able to show this, if so.

Response: Thank you for your comment. The data presented in the manuscript are biovolume proportion, which is the proportion of biovolume of a certain phytoplankton type in all phytoplankton types. We have added a sentence in methods describing this calculation (Line 222-225). Because the cell size of different phytoplankton types differs up to 10,000-fold, the changes in cell count of some small phytoplankton types are not obvious in biovolume proportion results. The trends in biovolume and cell counts of each phytoplankton type (Fig. 3) are similar and that's why we did not include the latter in the figure.

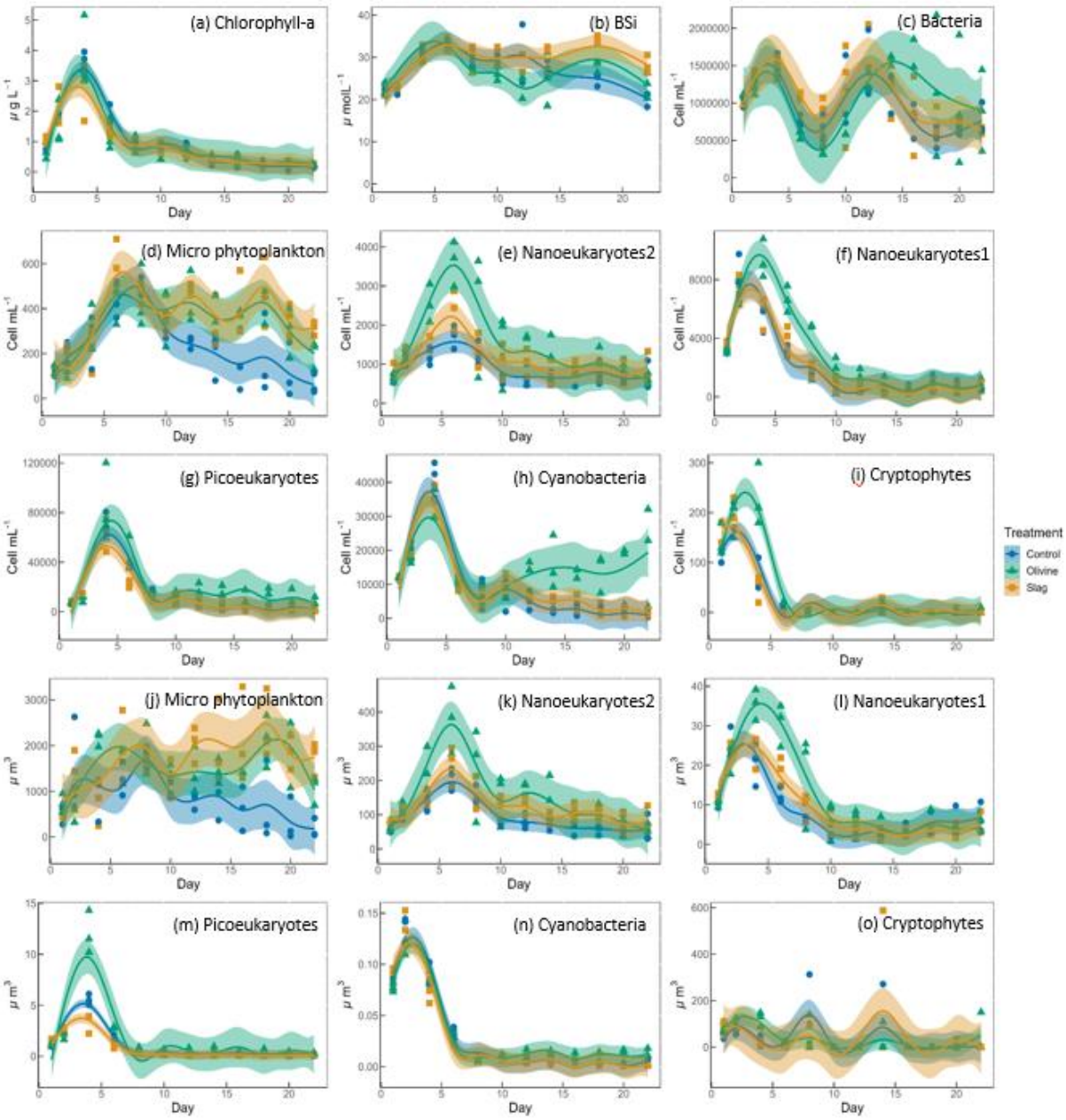

Fig.3. The Temporal development of chlorophyll-a concentration (chl-a), BSi, and different eukaryotic and bacterial plankton groups as determined with flow cytometry. (a) chlorophyll-a; (b) BSi; cell concentrations of (c) heterotrophic bacteria, (d) microphytoplankton, (e) nanoeukaryotes2, (f)

nanoeukaryotes1 (g) picoeukaryotes, (h) cyanobacteria, and (i) cryptophytes; biovolume of (j) microphytoplankton, (k) nanoeukaryotes2, (l) nanoeukaryotes1 (m) picoeukaryotes, (n) cyanobacteria, and (o) cryptophytes. The figure data points represent the raw data, and the fitted curve is the generalized additive model. The shaded area represents the 95 % confidence interval.

Comment15: Lines 493-512: The much more noisy data for zooplankton than for phytoplankton is likely driven partly by greater patchiness of the former, as suggested here. I suggest this may have also been exacerbated by sampling error from the very small diameter plankton net used to make the collections, as detailed above.

Response: Thank you and apologies, there was a typo in our description of the zooplankton net. The net is actually 20cm in height and 15cm in width with a 210µm mesh size while the microcosm is around 510cm tall and 35cm wide. Therefore we think the zooplankton net size is suitable for these microcosms sampling.

Comment16: Lines 549-554: I agree, it is hard to draw a direct cause and effect line between higher Fv/Fm and higher abundance.

Response: Thank you for your comment.

Comment17: Lines 596-609: To my knowledge, no one has shown that Mn or Ni additions can increase photosynthetic efficiency. If so, please cite appropriate references. In addition Mn, like Fe, is typically very abundant in coastal and riverine-influenced waters. I agree that it is puzzling that Fv/Fm increased in the slag and olivine additions despite ambient Fe levels of 100 nM or so, but attributing this response to other metals not known to influence photosynthetic efficiency is quite speculative.

Response: Thank you for your feedback. There are some lab experiments indicating the addition of Ni and Mn can enhance the photosynthetic efficiency, like Fv/Fm. These effects are likely species-specific We have cited relevant literature (Pausch et al., 2019; Balaguer et al., 2022; Guo et al., 2022). We agree that it's unknown why the Fv/Fm increased in the mineral addition treatments, and the only theory we can think of is that the coastal phytoplankton community has a higher trace metal requirement than the lab single strain culture. The higher dissolved trace metal concentrations may have elevated the bioactive trace metal concentrations which are easier to be taken up and utilized by the phytoplankton (discussed in line 646-651).

Comment18: Lines 617-626: Please see my comments above about patchiness of macro zooplankton and possible sampling artifacts, in view of these questions this paragraph on larvacean trends may be overinterpreting the data a bit.

Response: Thank you for your comment. We think part of the confusion probably came from our incorrect description of the zooplankton net in the previous version (we clarified this in Comment8). We agree that the patchy distribution of larvacean, the *Oikopleura* sp., generally brings a large standard error in abundance data, often more than we are used to from phytoplankton data. Nevertheless, we think the interpretation is worth mentioning here because *Oikopleura* is an effective filter feeder, and it is plausible to speculate that the suspended olivine that was highly abundant at the onset of the study slowed down initial growth.

Comment19: Lines 719-730: Again, these concluding statements really apply only to the levels of slag and olivine chosen for this particular experiment, and some suggestion of the need for experiments comparing the two under other concentrations and initial conditions is probably needed. One of the reasons CaO sources like slag might have questionable environmental impacts is that they have the

potential to cause extremely rapid and dramatic swings in the carbonate buffer system- for instance, pH increased ~0.5 units over just 4 days in the slag treatments here (Fig. 2a). This of course should drive a correspondingly large and rapid uptake of CO2 from the atmosphere, but could be potentially problematic for some marine organisms. It is reassuring that the plankton community used here was apparently able to accommodate to these major carbonate system swings with little sign of apparent stress. However, this is not necessarily going to be the case for marine metazoans including many invertebrates and fish, which as any aquarist knows are not very tolerant of rapid water chemistry changes. A few words about the need to examine impacts on other trophic levels of the marine food web in the concluding paragraphs would be in order.

Response: Thank you for your feedback on this issue. We addressed this comment in our response to comment1. Furthermore, we added text to section 4.2 and in the conclusion that a potential environmental concern of slag may be that it is almost too efficient in that it increases pH too rapidly for some species to acclimate. This must remain a hypothesis for now as our data do not allow firm conclusions. See line 784-791.

References:

Balaguer, J., Koch, F., Hassler, C. et al.: Iron and manganese co-limit the growth of two phytoplankton groups dominant at two locations of the Drake Passage. Commun Biol 5, 207, https://doi.org/10.1038/s42003-022-03148-8, 2022.

Guo, J. A., Strzepek, R., Willis, A., Ferderer, A., and Bach, L. T.: Investigating the effect of nickel concentration on phytoplankton growth to assess potential side-effects of ocean alkalinity enhancement, Biogeosciences, 19, 3683-3697, https://doi.org/10.5194/bg-19-3683-2022, 2022.

Pausch, F., Bischof, K., Trimborn, S.: Iron and manganese co-limit growth of the Southern Ocean diatom *Chaetoceros debilis*. PLOS ONE 14, e0221959.. https://doi.org/10.1371/journal.pone.0221959, 2019.

Ferderer, A.: Convective mixing vs no mixing inside two microcosms, TIB [video supplement], https://doi.org/10.5446/55861, 2021.

Ferderer, A., Chase, Z., Kennedy, F., Schulz, K., Bach, L. T.: Assessing the influence of ocean alkalinity enhancement on a coastal phytoplankton community. Biogeosciences, 19, 5375-5399, https://doi.org/10.5194/bg-19-5375-2022, 2022.