Response to 2.nd round review:

We thank the reviewer and the editor for valuable comments and hope that our responses to the concerns will make the manuscript publishable in biogeosciences.

MAJOR COMENTS 1. One major concern is that authors use the computational models initially , to set-up the background and aims for the subsequent experiments. However, I see there is a huge gap in between the computational models output and what it was simulated in the lab experiments.

 # In the first review round it was pointed out that we needed to compare our model for dispersal with other models. In the latest version this was included in the M&M. This part is now re-written and moved to the discussion (4.1. Dispersal model and experimental design). We hope that this will make the part of the M&M linking the dispersion model to the experimental condition clearer.

Especially, regarding the carbonate system, speciation, pH and precipitation. Has not pH increase undesirably fast? And so, it is possible that uncontrolled $CaCO_3$ precipitation could lower the CO_2 sequestration efficiency of the approach? Theoretically $Ca(OH)_2$, and $(Mg(OH)_2)$ should dissolve rapidly in the ocean surface , but I am not sure this the case. I have the impression that the ratios dissolution /precipitation were not controlled to check for reactivity and spontaneous precipitation in seawater. In short, were these hydroxides well dissolved? In addition, and regarding the commercially sourced material, was the carbonate content measured before the experiment to check for carbonation? I mean, how can you limit carbonation being present within the hydroxides? I have serious difficulties concealing the results obtained from de lab experiments with the just commented above. Could your results be generalised, and compared to the simulated computer models and extrapolated to general conclusions that can have relevance for the application of OAE in the real-world? I am not sure…

 # At 100 mg/L of $Mg(OH)_2$ and $Ca(OH)_2$, there is the possibility that some uncontrolled $CaCO_3$ precipitation could have occurred. But following the 1 hr "dispersal phase", the 10,000x dilution resulted in 10 mg/L $Mg(OH)_2$ and 12.7 mg/L $Ca(OH)2$ which would result in omega-aragonite and omega-calcite saturation states that would not result in uncontrolled $CaCO_3$ precipitation. However, carbonate chemistry was not comprehensively measured in these experiments. Neither was the carbonate content of the commercially sourced materials. The dilutions, however, were designed after model dispersal results, so similar carbonate chemistry conditions should be expected in a real-world OAE deployment from a ship into the open ocean.

 # To discuss eventual precipitation the the following are added to the discussion:

 "For example, in the dispersal model scenario used for designing the experiments in current study, a 1/10,000 dilution after 1 hour results in a final concentration of $Mg(OH)2$ and $Ca(OH)2$ of 10 and 12.7 mg/L, respectively. At these concentrations, both alkaline materials are expected to fully dissolve for optimal CO_2 uptake while also not resulting in elevated calcium carbonate saturation states leading to "runaway" secondary precipitation of calcium carbonate (e.g., secondary precipitation was observed at $\Omega Ar > 7$ for $Ca(OH)2$ on the timescale of 4-5 h; Moras et al., 2022). Still it cannot be excluded that some uncontrolled $CaCO_3$ precipitation could have occurred at 100 mgL^{-1} of $Mg(OH)_2$ and 127 mgL^{-1} $Ca(OH)_2$ during the initial 1 h of exposure in the present study."

2. I have the impression that the use of the double staining method with Fluorescein Diacetate (FDA) and 5-chloromethylfluorescein diacetate (CMFDA) have been used lacking rigour (just my opinion and experience). It is important to note that fluorescent stains must be validated in each different experiment before use by analyzing the optimal dye loading concentration and loading kinetics for

each specific monoculture and /or phytoplankton community, to avoid sub-optimal fluorescence or saturated fluorescence reaching the laser detectors of flow cytometers or epifluorescence microscopes. This essential step seems to have been omitted previously to the start of the experiment. The concentration used by the authors in this experiment is 2 orders of magnitude higher (2.5 mM final concertation) than working concentrations widely used in different marine coastal and open oceanic waters , as well as in lab cultures by well stablished protocols and SOPs (please check references at the end of the review). Attending to this, the lack of mortality can be an artefact due to the spill over of green fluorescence due to excess oversaturated signal, which commonly occurs when the fluorescent dye concentration has not been customised for every cell type. For example, for CMFDA long-term staining (more than about 3 days) or the use of rapidly dividing cells, 5–25 µM dye is required. Less dye (0.5–5 µM) is usually needed for shorter experiments, such as viability assays in cultures and about 20 µM in natural populations (always final concentrations) as it is the case in this study. To maintain normal cellular physiology and reduce potential artefacts, as already mentioned, the dye concentration must be kept as low as possible. The effects of overloading may not be apparent, hence, to check for this, a cell death stain must be used in combination in the same set of aliquots containing the cells aim of study (please check references at the end of the review).

 # We thank the reviewer for making us aware about a typo regarding dye concentrations. It should be µM instead of mM. This is changed in the manuscript. Regarding to validation, the FDA/CMFDA method has been used by us for viability staining of Tetraselmis sp. since 2016 when it was compulsory method for USCG testing of BWMS. We perform in-house reproducibility testing of operators to ensure performance quality acceptance.

In addition, the CMFDA fluorescent probe is well retained in living cells through several generations. The probe is transferred to daughter cells but are not transferred to adjacent cells in a population. Cells loaded with the CMFDA fluorescent probes display fluorescence for at least 72 hours and exhibit ideal tracking dye properties—they are stable, nontoxic at working concentrations well retained in cells, and brightly fluorescent at physiological pH. Therefore, assessing cell viability with this fluorescent probe is not entirely accurate since daughter cells can be metabolically non-viable and yet, show green fluorescence. In this case, cell viability is overestimated (please check references at the end of the review). Regarding FDA I have similar concerns except for the transfer of dye to the daughter cells that is not the case with FDA.

#The samples were stained 10 minutes prior to counting. Counting was performed within a maximum of 45 minutes. The staining of dead daughter cell would not occur within this time frame. In addition, elevated background green fluorescence would occur long before any dead cell showed elevated fluorescence.

3. I must also comment that I miss the detailed staining protocol for cells and fluorescent microscopy quantification. How were cells harvested: by centrifugation and aspirate the supernatant? By filtration? Were they resuspended in pre-warmed or RT working Solution? Gently or vortexed? For how long were cells incubated with the dyes? Were cells centrifuged to remove the excess dye working solution? Was culture media added and the labelled cell dispensed onto slide or into a chamber-wells for imaging? For how long were they imaged using the appropriate emission and excitation filters under the scope? The point being, if someone is to reproduce your experiment, not sure that would be possible with the insufficient information provided… The kinetics and loading curves set-up for the

optimal dye concentration and time of incubation should be provided in supplemental material, or at least mention in the text that they are available to reviewers in case they would like to check on them (as it is my case for example).

 # The paragraph regarding staining has been rewritten and now states: "The density of living Tetraselmis suecica was determined using the double staining method with fluorescein diacetate (FDA) and 5-chloromethylfluorescein diacetate (CMFDA) (NSF, 2010). This double staining method, FDA/CMFDA, is based on the validation work of the US Navy Research Laboratory to distinguish between living and dead cells after disinfection by a ballast water treatment (Steinberg et al., 2011). This viability method is the only one recognized by both the International Maritime Organization (IMO) and the United States Coast Guard (USCG) for approval of ballast water discharge from 70,000 commercial ships at a global scale (USCG, 2012, IMO, 2018).

The following staining protocol was used: A 2.5 mM CMFDA stock solution was prepared by dissolving 1 mg of CMFDA in 0.86 ml DMSO (Dimethyslsulphoxide). It was then divided into 50 µl batches and stored at -20 °C. The 5 mM FDA stock solution was prepared by dissolving 10 mg FDA in 4.8 ml DMSO. The FDA stock solution was divided into 100 µl batches and stored at -20 °C. For each analysis, a 4 ml subsample was collected and 4 µl of 10% HCl was added, bringing the pH back to approximately 8 prior to staining. 4 µl of each stock solution was added to each subsample, resulting in final concentrations of 2.5 µM CMFDA and 5µM FDA. The subsamples were then incubated in darkness for 10 minutes, after which they were loaded into 1 ml Sedgewick-Rafter counting chambers etched with 1-mm 2 grids, with 1000 fields of view (FOV). Chambers were examined at 100x magnification using compound epifluorescent microscopes with standard blue light excitation (480 nm) and green bandpass emission (530 nm) filters. Furthermore, FOVs were counted until a minimum of 100 viable cells was observed in each camber. Cells in 3 replicate chambers were counted for each sample. For samples with <100 cells/ml, cells in up to six chambers were counted. *Samples were counted within a 45-minute period after incubation Cells numbers in sample were set to zero If no cells were observed in six chambers. Viable cells in samples (%) vere calculated as 100 * cell number in treated sample/cell numbers is control sample (without alkaline at each specific day, in this case). Average cell numbers were presented in tables in the MS."*

4. More details on negative and positive controls choice would be desirable. Were general procedural negative controls were done? Is this the NaOH teratement

 # Negative controls were unamended seawater which are described at the end of section 2.1, 3rd paragraph as well as in Fig. 1. Positive controls were the NaOH treatments described in section 2.1, 8th paragraph. The negative controls were with no additions and represent phytoplankton growth/physiology under normal conditions. The positive controls received an extreme high pH treatment that is intended to have an extreme negative impact on phytoplankton growth/physiology.

5. I was wondering why in the case of Tetraselmis a flow cytometer was not used… this would have most likely had produced more accurate results as compared to microscope observations. Regarding the scope is not clear how the % of viable cells was calculated nor how representative your sample was. What was N? how many fields of view (FOVs) were counted per slide or well? In each of the independent cultures? If only one slide/independent culture counted, seems not enough to me. Not clear either which statistical analyses has been carried out for this? Epifluorescence (or any microscopically quantification) can lead to artefactual data unless N is large enough (not sure this is

the case, and clarification is needed), or other intercalibrating method is used to contrast with numbers, such as flow cytometry.

 # See above reply to concern number 3

6. Another question that is confusing to me is the lack of standardised methods for measuring cell performance. It is not clear enough why fluorescent probes were used with Tetraselmis, but not for Skeletonema nor for the natural community. Moreover, for Skeletonema fluorescence, authors do not specify which fluorescence was measured? Red 666nm? In which device was this measured? Green with probes? Again, in which device? Along the same line, the rationale for the natural community analysis it is not well understood, nor how the % of viable cells was also calculated. Table 3 and 4 are difficult to understand because it cannot be discriminated to which functional group each % belong, therefore, not sure I see the point for this. I'd like to point out that the methodology description does not suffice to understand how this experiment was performed. The ms. does not have an easy thread-line to be followed. Could it perhaps be better organised?

 # The FDA/CMFDA was also used to count Skeletonema and natural phytoplankton. Fluorescence was measured at 645 nm.

The following are inserted in 2.2 (wet test): " The cell density was determined by FDA and CMFDA double staining and fluorescence at 640 nM in SpectraMax iD3 microplates after approximately 24, 48 and 72 hours (±2h)."

And in 2.3. (Natural assemblage…): "The water quality and algal density was monitored daily in each beaker, using the same methods described in Chapter 2.1. Moreover, cell count and viability were quantified using the same protocol as in 2.1., with florescence measured at 645 nm."

 # Ther were a few typos in table 3. They are now corrected. We hope that this makes the table easier to read. Moreover, table 4 and the paragraph discussing functional groups are omitted from the manuscript.

7. The statistical approach used, does not seem appropriate. First, does data distribution meet the requirements for parametrical tests? Assuming so, T-tests do not capture the variability of the system you might have. Hence, most likely, significant differences are not well resolved, questioning the results.

I would suggest that a oneway ANOVA would be adequate since you have 5 levels of concentrations. Also, in those cases in which time is a continuous variable and not an end-point variable, the right approach could be a split-plot ANOVA in which the fixed factor would be the concentration and the repeated measures factor would be time.

 # We thank the reviewer for pointing out the requirements for parametric tests. We have log transformed the values to obtain similar variation between the treatments at day 6. Our choice to use the T-Test instead of a repeated measures ANOVA, was based mainly on missing cells in the time series and the zero values just after exposure to $CA(OH)_2$ creates problems related to normal distribution. However, we think that the T-test is appropriate for investigating differences between the treatments on day 6. Regarding the comment that this approach questions the results, we do not fully understand which part of the results this test questions? The dynamics of algae cultures exposed to the different

parameters is depicted in Fig 2, and the T-Test clearly show that there are significant treatment effects after six days after exposure (p<0.001).

8. In my opinion the discussion is shallow. It does not get deep insight on explaining the data, nor debating them. Not enough quality for a discussion I'm afraid. The same applies to the conclusions, which are merely descriptive.

 # "We believe that the inclusion of Section 4.1, which discusses the model and experimental design, will enhance the clarity of our conclusions. Still, generally, we hold the view that the degree to which our findings should be contextualized and contrasted with other studies is appropriate. Furthermore, we would like to underscore that reviewer 2 has not presented any concerns pertaining to the discussion or conclusions.