This paper describes a set of long (> 500 y) simulations with an EMIC, with and without phytoplankton absorption of solar radiation and the associated ocean heating, and argues that this represents an important and mostly neglected process in the climate system. What is novel and interesting is that (1) phytoplankton biomass tends to increase rather than decrease under enhanced greenhouse forcing, (2) the effect is not monotonic with increasing emissions, and tends to be damped or reversed in the highest emission scenario examined, and (3) the effect on atmospheric CO2 concentration appears to be large.

I believe that this is an important experiment that deserves to be published, but the paper is not well written and requires, at least, major revision. Possibly it would be better if the editors declined the paper and returned it to the authors so that they could take the time required to craft a more substantial contribution. The English is adequate, but it would be best if the authors could find a native-speaker colleague to give it a thorough English editing before resubmission. The title could be revised to be more specific about what the actual content of the paper is; the present wording is fairly generic and uninformative.

Major points:

(1) The authors do not make a lot of effort to explain the mechanisms underlying the differences observed between the Light Absorption (hereafter LA) case and the non-LA case. Combined with the inadequacies of the model description, this makes it difficult to credit some of the more dramatic claims made, in particular the much higher atmospheric CO2 concentration in the LA case.

I estimate that by temperature-dependence of CO2 solubility alone, an increase in the global mean ocean temperature of 1 C would increase atmosphere CO2 by ~14 PgC. In the experiments shown here the increase in SST (which would be an extreme upper limit for the change in the ocean mean) is <1 C (Figs 4+6, Table D1). A weaker biological pump could add another ~5 PgC, assuming Delta-DIC/Delta-PO4 = 106 and a net increase of ~0.1 mmol P m^-3 (Table C1) over a surface layer 100 m thick. The atmospheric CO2 increase shown here is extremely large by comparison, ranging from ~75 PgC in RCP2.6
to ~250 Pg in RCP8.5, assuming that it takes ~2 PgC as CO2 to increase the atmosphere concentration by 1 ppm (Figure 7).

Maybe these are simplistic, static calculations. The experiments are long and the ocean is continuously overturning, so maybe the explanation lies in upwelling of the existing ocean inventory of DIC into a surface ocean that is getting warmer and therefore outgassing more CO2 to the atmosphere. A problem with this hypothesis is that RCP8.5 has by far the largest amount of excess atmospheric CO2 associated with LA (Figure 7), but the smallest increase in SST (Figure 4, Table D4). However, RCP8.5 also has the largest cumulative CO2 uptake (both LA and non-LA), so maybe this could be explained by greater outgassing of anthropogenic CO2 taken up earlier. Possibly the authors could consider comparing the 3D distribution of DIC at the beginning and the end of the experiments, or making a map of net outgassing of CO2. At any rate, I think more effort to identify the underlying mechanisms is warranted. It could also help if they cited some previous publications that demonstrate that this very coarse-resolution model can produce at least approximately realistic ocean upwelling.

There are some important details missing from the model description. I understand that all of the submodels are previously published, but key details that are directly relevant to the results presented should be briefly reiterated. Most disturbingly, the assertion that the ocean biology model operates in a static fashion at each grid point, without ocean transport of the related tracers, appears from nowhere in the Results (198), whereas the Methods appears to say the opposite (90). One could imagine, for example, that increased upwelling of nutrients causes large increases in phytoplankton biomass at the grid points where upwelling occurs (with the additional heat being advected away), in a way that might not occur if the phytoplankton were also being advected by ocean currents. Could they show some profiles of how chlorophyll concentration evolves over time? Do they remain within the range of historical observed values? Or at least within the realm of plausibility? The paper also appears not to state whether chlorophyll concentration in the non-LA expts is given a constant, nonzero value, or is assumed to be 0 (k=k_w).

There is no description of parameterizations of phytoplankton photoacclimation or chlorophyll synthesis or degradation. As I understand it, the model has prognostic phytoplankton C, P, and chlorophyll (117). So there is no photoacclimation per se: both C and Chl are prognostic. But a brief description of the phytoplankton growth and chlorophyll synthesis model is warranted, and a statement of whether there is any loss of chlorophyll independent of grazing or other loss of cells. Chlorophyll synthesis requires N and Fe, but not P. I assume they use a fixed N/P ratio to estimate the dependence of chlorophyll synthesis on N; whether it also depends on Fe is not stated. Nor is it stated whether the C/P and C/Fe ratios are fixed or variable; I assume that C/Fe is fixed and C/P variable, as there is prognostic phytoplankton P but not Fe. This should be clearly explained and ratios used (where they are fixed) stated.

Nor is there any description of surface solar irradiance or of radiative transfer in the atmosphere. It is stated that the incoming shortwave varies seasonally (140), but there is nothing about its geographic distribution (for example, top-of-atmosphere irradiance might be calculated from astronomical formulae and atmospheric attenuation assumed constant). They should also specify the fraction of total solar irradiance that is assumed to be shortwave/longwave, as only the former is affected by phytoplankton absorption. The energy balance atmosphere presumably has a submodel for radiative transfer (e.g., how does upwelling/downwelling longwave radiation vary as a function of atmospheric CO2 concentration). The climate changes are strongly dependent on this, so at least a brief description is warranted.
Other less immediately relevant processes that could use a brief description include carbon chemistry and gas exchange (106-107), the wind data used and the calculation of the wind stress and the drag coefficient (the non-dynamical energy balance atmosphere requires that wind speed and wind stress at the ocean surface be specified), and the dependence of ocean vertical mixing on stratification.

Finally, the description of the spinup and the experimental design is confusing. First they spun up the model for 10000 years with BIOGEM but not ECOGEM "to have a realistic distribution of nutrients" (142-143). Then there is possibly a further spinup with ECOGEM turned on, before the historical/RCP experiments are launched, but the description is confusing and I can't really tell what was done. Why would spinning up the model produce a realistic distribution of nutrients if there is no biological pump? I would have spun it up for a further 2000-3000 years with all of the biological processes active. Nor is it stated how they know that the system is in steady-state at 2500 (149).

(3) The Introduction is a grab-bag of literature citations intended to provide the impression that there is a broad consensus that phytoplankton biomass has declined over the historical period and is likely to decline further in the enhanced-greenhouse future. In my opinion this assertion is nowhere near as robust as the authors imply and gives the main premise of the paper a "straw man" quality. In Kwiatkowski et al., the average decline in NPP is only 3% by 2100, in the highest emissions scenario. Bopp et al (2022; 10.5194/bg-2021-320) suggest that phytoplankton biomass may be a more robust diagnostic than NPP (this paper is still in the Discussion stage, but the authors should at least take a look at it). Boyce et al 2010 drew some rather vigorous criticism (www.nature.com/articles/nature09953). Boyce et al (2014, 10.1016/j.pocean.2014.01.004) address some of these criticisms and should certainly be cited here. They claim that the basic conclusion that a long term secular (downward) trend is detectable remains sound, but this conclusion remains controversial and I think that the authors of the current contribution should treat it a bit more skeptically. The results of Polovina et al and McClain et al represent too short time series to be inferred to represent long-term secular trends, and should be discussed in the context of the difficulty of separating such trends from natural variability (e.g., 10.5194/bg-7-621-2010, 10.1029/2019GB006453). Behrenfeld et al show a statistical relationship between chlorophyll and stratification in the historical record of observed climate variability (mainly ENSO); their extrapolation of this to changes expected under anthropogenic warming is quite speculative (in any case, how can observations (21) tell us what will happen in the future?) I have not read Sonntag or Paulsen, and I can't say I find the synopses offered here very illuminating. As these are PhD theses rather than journal articles it is important to summarize their findings clearly, as the original text may not be accessible to the reader.

Minor points:

Terminology regarding IPCC and the RCPs (57-62): It is a common misconception that CMIPs/RCPs/SSPs are 'commissioned' or 'solicited' or 'approved' by IPCC. Proper citation format for IPCC Assessment (or other) Reports is given in the reports, but citing these in the present context is unnecessary. It is better to just cite Moss et al 2010 (10.1038/nature08823) for the RCPs and Taylor et al 2012 (10.1175/BAMS-D-11-00094.1) for CMIP5. Referring to scenarios as predictions (61) should be avoided (as should referring to scenario-based climate projections as predictions, e.g., 16, 270).

I don't think Section 3 is necessary, and it could be folded into the Results. I think this result is worth showing (although it might be better treated as Supplementary). But I think it is overreaching to say that it by itself 'validates' the model setup (and by
implication all of the submodels that affect results shown in this paper). The wording should be a bit more tentative and simply describe what was actually tested against what.