

## ***Interactive comment on “Projected 21st century decrease in marine productivity: a multi-model analysis” by M. Steinacher et al.***

**M. Steinacher et al.**

steinacher@climate.unibe.ch

Received and published: 5 February 2010

We would like to thank both reviewers and Andreas Schmittner for their constructive comments. We have followed their suggestions in most cases. Please find below the detailed response (bold/italic font) to the reviewers' comments (normal font). All references to pages, lines, and figures in our answer refer to the submitted discussion paper.

Additionally, results from a very recent simulation with a fourth model (NCAR CCSM3) have been added in order to improve the manuscript. The notation 'CSM1.4' instead of 'NCAR' is used in the revised manuscript when referring to the NCAR CSM1.4-carbon model to distinguish it from the NCAR CCSM3-BEC model.

C4334

For your convenience a version of the manuscript is provided as supplement where all the changes made during the revision are highlighted (without figures). There you also find all references used in this comment.

### **Anonymous Referee 1**

Specific Comments:

- Page 7937, line 11-12 – With ‘The MPIM model, and to a lesser degree, the NCAR model, suffer from a too strong iron limitation compared to the real ocean.’ it is unclear to what aspects of iron limitation the authors are referring here. Do they mean that the concentrations of iron are too low? That the half saturation value was too high? That the formulation of iron limitation was ill-posed? In any case, more specific support for the statement would be helpful given the number of times this point is necessarily returned to when describing the climate sensitivities of and differences between these models

**This is discussed in Schneider et al. (2008). We have added the following explanation on page 7937, line 12:**

*In the MPIM model, overall iron limitation is caused by the combination of low aeolian deposition and, more importantly, a high half-saturation value for iron. In the CSM1.4 model, iron appears to be too strongly scavenged, especially in the subtropical Pacific, also resulting in too strong iron limitation (Schneider et al., 2008). It remains difficult for any model to represent the iron cycle with its intricate coupling between physical transport, spatial and temporal varying iron sources by dust deposition (e.g. Mahowald et al., 2006) and sediments (e.g. de Baar et al., 1995; Dulaiova et al., 2009), iron sinks by particle scavenging, complexation by organic ligands (e.g. Parekh et al., 2008) and ecosystem and remineralization processes (Boyd et al., 2007).*

C4335

Also, the specific problems of the MPIM model with respect to iron are discussed in the revised MPIM model description (section 2.1.2). Further, we have added a new paragraph to the discussion section where the sensitivity of the IPSL and MPIM models to changes in iron input from dust is discussed.

- Page 7942, line 20 – this temperature functionality is odd for two reasons, first, the shape is convex down instead of the usual Eppley or Q10 convex up, secondly, it severely curtails productivity in the coldest waters. What is the justification for this? The Doney et al (2006) simply references HAMOCC which seems insufficient.

**This temperature functionality was used by Maier-Reimer (1993) for HAMOCC3 without giving any details about his choice and it was later adopted for the NCAR CSM1.4-carbon model. In these models, the biogeochemical cycling is represented by relatively simple equations with few parameters. The temperature functionality was presumably chosen as an empirical formula to get a realistic large scale nutrient distribution. Unlike the Eppley curve, it therefore comprises not only the direct temperature effect on local phytoplankton growth but also other temperature-related effects that affect the large scale pattern of NPP. A reference to Maier-Reimer (1993) has been inserted at page 7920, line 17, and the following explanation has been added:**

*This empirical parameterization is intended to model the large-scale nutrient utilization by marine ecosystems. For example, the temperature function, together with iron limitation, forces a low productivity and nutrient utilization in water that is colder than about 2°C. On the other hand productivity depends only weakly on temperature in warmer waters. The temperature factor increases by less than two for a temperature increase from 4°C to 34°C; this may be compared to an increase by a factor of seven in the IPSL model and no temperature-dependent growth rates in the MPIM model.*

**The temperature dependencies of all models are now given in the methods section and they are also discussed in the light of the results of Schmittner et al.**

C4336

**(2008) in a new paragraph in the discussion section (see also response to the short comment by A. Schmittner).**

- Page 7944, lines 10-14 – If the bias in MPIM were due to too intense iron limitation, one would expect to see too much nitrate at the surface, is this the case? Also, what is the explanation of the IPSL low bias? My guess is that the real answer is that none of these models can adequately represent the microbial loop and the role of cyanobacteria and other pico and nanoplankton.

**Yes, the MPIM model overestimates nitrate at the surface almost everywhere except for the Southern Ocean. The pattern of the difference between simulated nitrate and observation-based estimates from the World Ocean Atlas is similar to that of phosphate (Fig. 4 in Schneider et al. 2008) except for the Arctic Ocean, where nitrate is significantly biased high but phosphate is not. The following sentence has been added on page 7944, line 11:**

*This is supported by the fact that surface nitrate concentrations are largely overestimated by this model.*

**We agree that the representation of the microbial loop is limited in these models and that this may be one reason for the low bias. We have extended the analysis in regard to the microbial loop in section 3.1, added a new figure (Fig. 3), and also extended the discussion in that respect (see response to Referee 2).**

- Page 7944, line 20 – Filtering a potential comparison of modeled chlorophyll and temperature with satellite estimates through the Behrenfeld (VGPM) algorithm seems inappropriate here. As the authors note on lines 24-26, the satellite algorithms are highly uncertain. It is extremely misleading to treat VGPM as a standard, unless one is dealing exclusively with the North Atlantic region to which that algorithm was calibrated. Comparisons with field 14C PP estimates show the VGPM algorithm to almost always underestimate productivity in regions outside of the North Atlantic. Better would be to compare to an ensemble of satellite PP algorithms, and better yet to chlorophyll

C4337

and temperature themselves, if the NCAR model could be used to derive a chl value consistent with its estimate of PP from the 'biomass proxy' described on page 7946, line 25.

**We agree that VGPM is only one of many possible targets for a data-model comparison and that comparing to chlorophyll would be the best choice. However, we see no reasonable way to derive chlorophyll concentrations from the NCAR CSM1.4 model output. We have changed 'observations' to 'observation-based estimates' when referring to the satellite derived data at several places in the manuscript. Further, we have added a paragraph to the methods section (subsection 2.2, page 7943) where we point out the uncertainties in the satellite estimates. As suggested by the reviewer, we explicitly state that we use the VGPM data as an illustrative example:**

*As a point of reference and following Schneider et al. (2008), we utilize throughout this study satellite-based estimates obtained with the Behrenfeld algorithm (VGPM; Behrenfeld and Falkowski, 1997b; Behrenfeld et al., 2006) for data-model comparison and to compute skill-score weighted multi-model averages. The satellite-derived estimates have uncertainties. For example, Carr et al. (2006) report that global PP estimates from twenty-four ocean-color-based models range over a factor of two. On a more positive side, ocean-color-based models agree with respect to the spatial pattern of chlorophyll distributions and correlations among the resulting fields are typically high. Given these substantial uncertainties in satellite-based productivity data, the comparison of model results with one single satellite-based data set should be viewed as an illustrative example.*

- Page 7945, line 28 – Why use globally averaged air temperature rather than SST as one would surmise the regression to be even better?... okay, on Page 7958, line 21 the authors justify this choice as a means to facilitate comparability with other metrics in order "to account for the different climate sensitivities of the models." It would be good to have this kind of explanation in this earlier section.

C4338

**The sentence on page 7945, line 28 has been changed to include this explanation:**

*A linear regression between global PP and global mean surface air temperature is used to normalize PP changes with respect to climate change in order to account for the different climate sensitivities of the models (Fig. 1c).*

- Page 7947, lines 16-17 – I'm not sure what the authors are getting at by saying that, "On the other hand, reduced nutrient concentrations in combination with increased export are indicative of a sustained nutrient input into the euphotic zone." I think this sentence could be eliminated unless there is an additional point the authors are trying to make beyond the one in the sentence that follows this one.

**We agree that this sentence is not needed and it has been eliminated. The idea was to make clear that we separate regions with reduced nutrient concentrations into two regimes, depending on whether EP increased (sustained nutrient input from the thermocline) or decreased (reduced nutrient input) at the same time.**

- Page 7947, lines 22-24 – To say that nutrient supply in the North Atlantic is 'linked' the thermohaline circulation is far too vague. Is it simply that both depend on the occurrence of deep wintertime convection which decreases, or is it that the waters that supplant those exiting in the thermohaline circulation supply the nutrients? The sentence that follows suggests the former, but is also vague.

**The wintertime MLD is indeed reduced by about 45% in the North Atlantic. However, we are investigating a coupled system which makes it difficult to attribute certain effects to individual processes. We have changed the wording to reflect this better:**

*Nutrients are used up more efficiently, and PP decreases likely in response to less surface-to-deep exchange, which is linked to a reduction in the North Atlantic thermohaline circulation (Frölicher et al., 2009) and a reduced deep wintertime convection.*

C4339

- Page 7947, line 26 – In describing why light limitation increases, by ‘changes in cloudiness and changes in MLD’, do the authors mean ‘increases in cloudiness and increases in MLD’? If so, why does MLD increase under intensified stratification. . .
- wind mixing? Or do the authors mean that increased cloudiness overwhelms the decrease in MLD in increasing limitation?

**The dominant factor here is the large MLD decrease. It changes from a depth below the compensation depth  $z_c$  to a depth above  $z_c$ . The decrease in MLD affects both the light limitation factor in the production zone and the mixed layer scaling factor, which we treat together as ‘light limitation’ because of the specific parametrization of PP in this model (Eq. 4). The mixed layer scaling factor implicitly extends the production zone to the base of the mixed layer when the MLD exceeds the compensation depth. The effect of the decrease in MLD on the combined factors and thus on PP is a stronger limitation. The changes in cloudiness is not significant here and we have revised the statement:**

*The model also simulates an increase in light limitation, mainly caused by the decrease in mixed layer depth, and a somewhat stronger limitation by iron in the east and by phosphate in the west.*

- Page line 16 – While I am very supportive of this objective statistical approach, again, the idea that VGPM provides a robust estimate of PP (assuming that B&F alone was used) is inappropriate. The only justification I can think of for the use of B&F is that all of these models are so strongly biased that anything is better than nothing, and suggests that the authors lack any confidence in the model PP formulations. The authors should either be using chlorophyll directly and avoid filtering through algorithms such as VGPM that necessarily parameterize variability induced by nutrient effects on PP through temperature and general down-regulation or creating an ensemble pp algorithm with a variety of those used in Carr et al., 2006 such as was done in Sarmiento et al. (2004).

C4340

**As stated above, we use VGPM only as an illustrative example. We declare this in the methods section (see above) as well as in the discussion:**

*We have applied regional model skill metrics as weights in the computation of multi-model means. Here, we have used the satellite-based PP estimates (average of annual mean PP for the period 1998 to 2005) of Behrenfeld et al. (2006) as an example target against which the performance of individual models is assessed; in the future it might be useful to compare models to the ensemble of satellite-based reconstructions of PP and chlorophyll given their uncertainties. Other metrics, such as how well the models reproduce current surface nutrient distributions, could be used as additional targets.*

## Anonymous Referee 2

General comments:

An essential difference between PP and EP is the microbial loop, which contributes to PP but not to EP. In order to understand why “mechanistic” models predict a decline in EP and “empirical” models predict an increase in PP, I would first look at the response of the microbial loop to global warming. It is problematic that the mechanistic analysis is based on results of a model (the only one among the three) that does not include any representation of the microbial loop.

**We don’t agree that the mechanistic analysis is based on results of the NCAR CSM1.4-carbon model only (see response to major concern 3 below). An analysis of IPSL, MPIM, and in the revised manuscript also of NCAR CCSM3 results is provided as well, although in these highly coupled systems the attribution is more complicated. We have extended our analysis in regard to the microbial loop: A new paragraph has been added to section 3.1 (page 7945, line 2) where the e-ratio as a measure for regenerated production in the IPSL and MPIM models is discussed. Following Laws et al. (2000), the e-ratio as a function of**

C4341

temperature is compared to field data. Further, projected changes in the e-ratio are discussed at the end of section 3.1 (page 7946, line 18) and a new figure has been added that shows these results (Fig. 3). The following sentence has been added to the introduction:

*Large scale biogeochemical models often lack an explicit representation of the microbial loop. The energy and nutrient flows initiated by bacterial consumption of dissolved organic matter and grazing by bacterivores (Azam et al., 1983) are represented by a decay function for dissolved organic matter. The decay of dissolved organic matter releases nutrients which are in turn available for plankton consumption.*

**Also, the following paragraph has been added to the discussion:**

*None of the models used here explicitly represents bacterial pools. The microbial loop describing the energy and nutrient flow initiated by bacterial consumption of dissolved organic matter and grazing by bacterivores is implicitly represented in the models. Dissolved organic matter is assumed to decay and released nutrients are then available to fuel productivity. Three of the models show a lower global PP than observation-based estimates and one might be tempted to link to the low productivity to the missing explicit representation of bacteria. However, the PP of 49 GtC yr<sup>-1</sup> simulated by the CCSM3 model falls well within the satellite-based range of 35 to 70 GtC yr<sup>-1</sup> and the PP of the IPSL model is with 34 GtC yr<sup>-1</sup> only slightly lower than the satellite-based range. As already discussed, PP in the MPIM model is too strongly limited by iron and the simple empirical formulation of productivity in the CSM1.4 is biased low by design. Taken together, this suggests that the missing explicit representation of bacteria does not necessarily cause an underestimation of PP.*

A second concern is that the authors do not discuss the possible role of iron in causing changes in EP and PP. What do the models assume about the evolution of aeolian iron supply during the 21st century? As shown by a previous analysis of the same models (Schneider et al., Biogeosciences 2008), at least two of the models are severely limited

C4342

by iron (leading to way too high surface phosphate values e.g. in the Pacific). Similar to earlier studies of predicted changes in EP, the mechanistic investigation focuses on nutrient supply from below. If iron supply from above is relevant and production is iron limited, changes in stratification or mixed layer depth may have opposite effects: Iron added to a shallower mixed layer may have a better chance to end up in biomass than iron added to a deeper and darker mixed layer. Is this relevant?

**Changes in aeolian iron supply during the 21st century are not considered in this study. We have added the following sentence to the methods section at page 7943, line 22:**

*Dust deposition fields were kept at a constant climatology in all experiments.*

**Further, we discuss the role of iron more extensively in the introduction (please see response to reviewer 1), in the MPIM model description, in a new paragraph in the discussion section on page 7957, line 29, and in the response to the comment regarding line 13 of page 7950 below.**

Major concerns:

1. The paper pretends that there is a "general" conflict between a recent empirical-statistical approach (Sarmiento et al., 2004) and what the authors call "mechanistic" models. As A. Schmittner in his comment rightly points out, this conflict is less general than stated in the paper and may just be a coincidence of the particular choice of models used here. The interesting question is why different "mechanistic" models can predict opposite changes in primary production in response to global warming. The authors ignore this question by referring mostly to changes in "productivity" without clarifying that they look more at new production rather than primary production (see point 2 below).

**We have changed the wording in this respect (see responses to the specific comment below and to the short comment by A. Schmittner).**

C4343

2. The terminology is confusing. Although a definition of primary production (PP) and export production (EP) is given at the end of chapter 1, the repeated use of "marine productivity" or "productivity" makes it difficult to follow the paper. It doesn't help that one of the models (NCAR) represents "the carbon flux associated with net nutrient uptake". Depending on the decay time scale of the semi-labile dissolved organic matter (not given in the paper), this should be close to net community production or new production (NP). In the NCAR model there is essentially no representation of the microbial loop. Without any attempted justification stating that the NCAR model's "PP" "is a reasonable proxy for the time and space variability of PP" and "for reasons of simplicity...is considered here as PP" (line 9-10, page 7943) does not make it such.

**We now use either the complete notations 'net primary production' and 'POC export' or the corresponding abbreviations 'PP' and 'EP' throughout the manuscript. The abbreviations are explained in the abstract, in the introduction, and at the beginning of the discussion. Further, the decay time scale of the semi-labile dissolved organic matter of half a year is now mentioned in the NCAR CSM1.4 model description (page 7943, line 5) and the sentence at line 7f. has been reformulated and extended with a reference to a comparison of the NCAR CSM1.4 PP with the SeaWiFS dataset:**

*It appears to be a reasonable proxy for the time and space variability of PP if somewhat underestimating the absolute magnitude (Schneider et al., 2008).*

3. The entire analysis of the mechanisms of shifts in PP (section 3.2) is based on results of the NCAR model. It is therefore rather an analysis of mechanisms of shifts in NP, similar to previous studies of changes in EP, except for the possible change in the export of dissolved organic matter (which is not commented on in the paper). It is not an analysis of changes in PP, and there is no justification given why "the mechanisms identified for the NCAR model are also key for the productivity changes in the IPSL and MPIM model" (line 20, page 7948) if "productivity" is to be equated with "primary productivity".

C4344

**There must be a misunderstanding and we are puzzled by this comment. For the analysis of the IPSL, MPIM, and CCSM3 model, we did not assume that the mechanisms identified for the NCAR CSM1.4 model are the same for these models as well. Our analysis of these models (sections 3.2.2. and 3.2.3) is based on (i) regional changes (Figures 4 and 5), (ii) local changes (Figures 7 and 8), (iii) and correlations for a suite of variables (Figure 6), and (iv) the known model equations. These analyses indicate that the mechanisms identified for the NCAR CSM1.4 are also operating in the IPSL, MPIM, and CCSM3 models that include ecosystem models and representations of different algae functional types. We believe it is a strength of our approach that we analyze both a model with a simple representation of productivity (NCAR CSM1.4) and climate models that include different ecosystem modules. The simple formulation applied in CSM1.4 permits for a quantitative attribution of changes in PP and EP to individual driving factors, whereas the climate-biogeochemical-ecosystem models allow us to study the coupling between forcing and response in a more elaborated and complex setting. The sentence on page 7948, line 20 has been changed to reflect this better:**

*Next, we will show that the mechanisms identified for the CSM1.4 model are also key for the productivity changes in the IPSL, MPIM, and CCSM3 models. Namely, we find that a reduced nutrient input related to enhanced stratification, reduced MLD, and a slowed circulation tends to decrease PP and EP under transient global warming not only in the CSM1.4, but also in the other three models.*

4. As shown by Schneider et al. (2008), two of the three models do not (or much too weakly) show the observed anti-correlation of "PP" with temperature on interannual time scales. Why then should one have confidence into the "PP" response to global warming? Also shown in the Schneider et al. (2008) paper is that two of the models have way too high surface phosphate concentrations in large areas of the Pacific (0.5-1.0 mmol m<sup>-3</sup> too high in the MPIM model, 1.0-1.5 mmol m<sup>-3</sup> too high in the NCAR

C4345

model). Doesn't this indicate the the models' response to spatial differences in nutrient and/or light supply is wrong? Why should this result in any reasonable response to temporal changes in nutrient/light supply?

**We agree that the models are far from being perfect and the deficiencies are discussed in various parts of the manuscript. This is exactly the reason for proposing the skill-score weighted multi-model mean approach for this type of problem.**

Specific comments:

p.7935,l.5: According to my knowledge, the first study coupling a marine ecosystem model to an oceanic GCM was Sarmiento et al. (1993) and Fasham et al. (1993).

**We have added references to Sarmiento et al. (1993) and Fasham et al. (1993) on page 7935, line 5.**

l.22: include "from SOME mechanistic models", though one may argue whether any of the empirical ecosystem models are mechanistic at all....

**Done. We have changed this sentence to:**

*Our interest is further fueled by the contradicting projections for global PP from some "mechanistic" models, as used here, and a recent statistical model approach (Sarmiento et al., 2004).*

l.24ff . What is "marine global productivity"? Many of the models cited here do not include an ecosystem model and cannot make any statement about primary production. I would agree that there is a general decrease in simulated export production, but since you refer to "marine global productivity and organic matter export" you seem to refer to something else. p.7936,l.1: again, the statement that "global productivity and export" are reduced is not supported by the references given.

**This passage has been formulated more precisely:**

C4346

*A general finding across the hierarchy of mechanistic models is that global EP decreases in 21st century global warming simulations (Klepper and De Haan, 1995; Maier-Reimer et al., 1996; Joos et al., 1999; Matear and Hirst, 1999; Plattner et al., 2001; Bopp et al., 2001; Fung et al., 2005; Schmittner et al., 2008; Frölicher et al., 2009). Increased stratification and a slowed thermohaline circulation in response to surface warming and freshening cause a decrease in the delivery of nutrients to the surface. As a consequence, global EP and in some models also PP is reduced.*

l.5 Here "productivity" probably refers to "new production" (in C units?) or "export production"

**We have revised the terminology:**

*In the simpler models, EP (or some approximation of PP) is tied to the availability of nutrients...*

p.7937, l.21/22 : "the use if a multi-model ensemble increases the robustness of the results". This remains to be shown. At closer look none of the models is very good. You may get a more robust result from averaging many bad models, but, if anything, it will remain a bad result.

**As stated above (answer to major concern 5), that's why we have applied a more advanced method than just averaging the models.**

p.7939 top: A description of the controls of PP and any explicit temperature dependences, as discussed for the MPIM and NCAR models, should also be given for the IPSL model. If I remember correctly, some heterotrophic processes also depend on temperature in this model. If this is the case, this should be pointed out (also for the other models). p.7941, eq.2,3: The text prior to the equations suggests that  $P_{sil} + P_{car}=1$ , but this is not the case for equations 2 and 3.

**The descriptions of the IPSL (section 2.1.1) and MPIM (section 2.1.2) models in the methods section have been updated in that respect. The temperature**

C4347

dependencies of all models are now given in the methods section and they are also discussed in the light of the results of Schmittner et al. (2008) in a new paragraph in the discussion section.

p.7942, eq.4 : What is the reasoning behind this temperature function? Since this paper is about the effects of global warming, the reader should be able to understand this formulation which does not seem to be commonly accepted standard.

**Please see answer to Referee 1 (specific comment 2).**

p.7943, I.25: What do you mean by "Gaussian interpolation"? Is there some spatial scale larger than the grid spacing involved?

**As the models have different and variable grid cell sizes the variables have to be regressed onto a common grid in order to be able to compare the results. This has been done by calculating the weighted average of the model grid cells within a radius of 4° around the centers of the 1°x1° grid cells. As weights the Gaussian weight function has been used with a mapping scale of 2°. We have stated this more precisely in the manuscript:**

*For analysis, all variables have been regressed onto a common 1°x1° grid using a Gaussian weighted average of the data points within a radius of 4° with a mapping scale of 2°.*

I. 25ff: I do not understand this. First you use the control run to detrend model results (i.e. by subtracting the control run results from the global warming results?). Do you then, in a second step, detrend T, S, and nutrients? Does this mean you detrend the difference global warming minus control? Why should one do this? Also, in the case of the NCAR model you detrend T, S, and nutrients, but not EP and PP. Doesn't this affect your analysis that assumes consistent T, S, nutrients, EP, and PP fields?

**Obviously, we do not directly subtract results from the control run from those from global warming simulations. This would bias high the variability (interan-**

C4348

**nual to decadal) for any output variable. The trends in T, S, and nutrients are very small for the euphotic zone in the NCAR CSM1.4 model, while somewhat more substantial trends are found in the deep ocean in this model. As a measure of caution, we have detrended the entire 3-D fields for these variables and for oxygen for an earlier analysis on variability in marine oxygen concentration (Frölicher et al., 2009). The detrending does not affect the overall consistency between the T, S, nutrients and EP and PP fields in the CSM1.4 model. The century-scale detrending does also not alter the spatial or seasonal-to-decadal variability nor does it remove global warming trends. We provide now more technical explanations as other readers might also not be familiar with coupled Earth System model analyses:**

*Control simulations in which CO<sub>2</sub> emissions are set to zero and other forcings are set to constant preindustrial levels are used to remove century-scale model drifts for each grid point and for each calendar month (Frölicher et al., 2009). Affected are the three-dimensional distribution of temperature, salinity, and nutrient concentrations in the IPSL and CSM1.4 models, as well as PP and EP in IPSL. For these variables, detrended values from the scenario simulations are used for analysis. We note that trends in surface values are small in the CSM1.4.*

p. 7945, I.23: How do you know this is the MAIN reason? Bopp et al. (2005) have "only" shown that this is ONE reason.

**We have modified this statement:**

*The reasons for this decoupling of PP and EP in the IPSL model are a shift from diatoms and zooplankton to the smaller nanophytoplankton and the increased recycling of nutrients and carbon in the surface ocean (Bopp et al., 2005).*

p.7946, I.26: Isn't the "biomass proxy derived from phosphate and iron concentrations" just the phosphorus equivalent of the limiting nutrient? Under what conditions does biomass equal the nutrient concentration?

C4349

**Yes, in this simple phosphate-based model, the phosphorus equivalent of the limiting nutrient is used to determine the biomass turnover. The same arguments with respect to the parametrization in these simple models as for the temperature functionality also apply here (see response to Referee 1, specific comment 2).**

p.7947, l.15: Does the "unrealistically strong iron limitation" change during the simulated 21st century?

**In most regions there is no change in iron limitation throughout the simulation. There is a slight increase in iron limitation in the western Pacific around 10°S as well as around 30°S that decreases PP to some extent.**

l.26: what is the relative importance of changes in cloudiness and mixed layer depth?

**We have revised that statement. Please see response to Referee 1.**

p.7948, l.19-20. Why does a correlation suggest that the mechanisms are the same?

**We have reformulated that passage. Please see answer to major concern 3 above.**

p.7950, l.1. Does the Fe:C ratio show a similar change in the exported particles?

**Yes, the Fe:C ratio of POC shows similar changes in these regions.**

l.13: EP decreases because nutrient supply from below goes down. Iron is mostly supplied from above (is it?). Therefore, assuming no changes in aeolian iron supply, I would expect iron concentrations to go up when EP goes down. Why do iron and macro-nutrients show parallel changes here?

**Iron is supplied from above by atmospheric deposition and from below through physical transport (and from sediments in the IPSL and CCSM3 models). The relative contribution is highly variable in space. Therefore we find regions with parallel changes (e.g. large parts of the North Atlantic, North Pacific, Arctic, and Southern Ocean) as well as regions with opposite changes (e.g. around**

C4350

**Australia, in the Atlantic at low latitudes, and in the north and eastern parts of the Indian Ocean). To some extent the regions with opposite changes correspond to regions with a high dust flux, which indicates that atmospheric deposition is the dominant factor there and surface iron accumulates. The statement has been changed to reflect this better:**

*In the CSM1.4 and MPIM model, the iron-to-carbon and other elemental ratios are constant. Generally, surface iron tends to increase in regions with substantial aeolian iron input and increased stratification or reduced mixed layer depth, whereas it tends to decrease in parallel with macro-nutrient concentrations in the surface ocean in regions with little iron input. This leads to an increase in global mean surface iron of 4% in the CCSM3 model, while CSM1.4 and MPIM project a slight decrease of about 2%. In contrast to the IPSL model, these three models all project a decrease in surface iron in the Southern Ocean and in the Arctic.*

p.7951, l.20-21. This is a circular argument: first you say that the mechanisms of the NCAR model can be applied to the IPSL and MPIM models, and then you conclude that the multi-model analysis confirms the results of the NCAR model analysis.

**As explained above in the response to major concern 3 this is not the case.**

p. 7952, l.8-11. Just picking those particular grid points from each model that agree best (or disagree least) with the observational estimates does not mean the any of the models or the models as a class can represent most of the features. Would you apply this to the global warming simulations, you could possibly find a picking scheme that results in a decrease in PP (or EP?). The statement that the mechanisms are the same is again a circular argument, because it was assumed earlier that the mechanisms identified for the NCAR model apply to the other models.

**We don't understand what exactly this comment is getting at. The deficiencies of the models and also of the multi-model mean approach are discussed in the manuscript. Also, we indicate, for example, where none of the model's skill score**

C4351

is higher than 0.5 (dotted areas in Fig. 9).

p. 7953, l.11: what is the reason for using an influence radius of 10 degrees?

As stated later at page 7958, line 17 "the scale length has been selected to be representative for the spatial scale of marine biogeographical provinces ( $\approx 10^\circ$ ); the exact choice of the numerical value is not crucial for our application." We have tested values from  $3^\circ$  to  $50^\circ$  found that the results don't change significantly. We have added an additional statement on this in section 3.3:

*...and  $\rho = 10^\circ$  characterizes the width of the distribution (the distance at which the weight has decreased from one to  $1/\sqrt{e}$ ). We note that the results are not sensitive to the exact choice of  $\rho$ .*

p.7960, l.16-17. The assumed temperature difference is also very different among the models used here. Why does this seem to play such a little role in the "mechanistic models" compared to the large role in the satellite-based estimates?

We don't understand what exactly 'temperature difference' refers to in this comment. The projected temperature increase in the transient simulations is very similar for the IPSL, MPIM, and CCSM3 models, and somewhat lower for the CSM1.4 model. PP limitation by temperature is typically more relevant at high latitudes in these models, while at mid and low latitudes nutrient limitation dominates. In some cases the temperature functions used in the satellite algorithms don't include only the direct effects of temperature on PP, but also indirect effects linked to light or nutrient availability. This is probably one reason for the different sensitivities of the algorithms.

C4352

#### Short comment by A. Schmittner

The authors distinguish between mechanistic models, which show a decrease in both EP and PP and the empirical approach of Sarmiento, which shows an increase. In the introduction they mention the studies by Schmittner et al. (2008) and Oschlies et al. (2008) as "outliers" in the mechanistic model category because these studies also show an increase in PP (albeit they also simulate decreases in EP). The sentence than mentions these two studies implies that the model response is due to pCO<sub>2</sub>-sensitive biotic C:N ratios. However, this is not true. In fact, the Schmittner et al. (2008) model does not use pCO<sub>2</sub> sensitive C:N ratios. This should be corrected.

#### This has been corrected:

*We also note that Schmittner et al. (2008) find a strong increase in primary productivity in 21st century CO<sub>2</sub> scenarios albeit new and export production decrease. The increase in PP in their study results from an exponential dependency of phytoplankton growth rates on temperature (Eppley 1972).*

I suggest including information about how the temperature dependency of phytoplankton growth is dealt with in the various models. For the NCAR model there is a formula given (eq. 4), which I don't understand. Where does this come from? What is the basis for this formula  $(T+2)/(T+10)$ ? It seems to be contrary to the Eppley results and other measurements.

**Please see response to Referee 1. The temperature dependencies of all models are now given in the methods section and they are also discussed in the light of the results of Schmittner et al. (2008) in a new paragraph in the discussion section:**

*The formulation of the temperature dependency of growth rates and other processes vary qualitatively and quantitatively among the four models. An exponential temperature dependency is used for growth rates, microzooplankton grazing, and POC rem-*

C4353

ineralization in the IPSL and CCSM3 models with  $Q_{10}$  values of 1.9 and 1.6, respectively. This is comparable to the temperature dependency for growth rates proposed by Eppley (1972) and applied by Schmittner et al. (2008). In contrast, growth rates are temperature-independent in the MPIM model, thereby assuming that phytoplankton acclimate to local temperature. The temperature limitation for productivity in CSM1.4 corresponds formally to a Michaelis-Menton type formulation (Eq. 2) and has a concave shape in contrast to the exponential-type formulations. In other words, the nominal sensitivity of PP to a temperature change is highest at low temperatures in CSM1.4, in contrast to CCSM3 and IPSL where the sensitivity of growth rates is highest in the warm ocean. Interestingly, global PP is decreasing in all four models under global warming, in contrast to the model of Schmittner et al. (2008). The IPSL model with the highest temperature sensitivity for growth rates and a realistic relationship between the export ratio and temperature (Fig. 3) yields the largest decrease in PP per nominal change in surface temperature (Fig. 2c). Apparently, increasing nutrient limitation is more important in regulating PP on the global scale than the direct temperature effect on growth rates in the IPSL model and in the other models applied here.

The Schmittner et al. study, which is a mechanistic model, agrees with the results from Sarmiento (at least qualitatively). Therefore, there is not a real dichotomy between “mechanistic models” and “empirical models” but the mechanistic models also don’t agree. I think a large reason for this is the treatment of the temperature dependency of growth. (I note that Schmittner et al. reproduce the global NPP estimates of about 40 GtC whereas most of the models used here underestimate it).

**We have changed the wording with regard to the differences between “mechanistic models” and “empirical models”. Further, we note that the CCSM3 model (that has been newly included) shows a similar response as the other models while the simulated global PP of 49 GtC yr<sup>-1</sup> lies well within the satellite-based range of 35 to 70 GtC yr<sup>-1</sup> (see also response to referee 2).**

I also want to express caution when using the satellite estimates as truth. Satellites

C4354

only see the surface, whereas much of the productivity occurs below the surface. To me it is unclear how good satellite estimates represent the real ocean.

**We discuss the uncertainties of the satellite estimates in the manuscript and state that we use the VGPM data only as an illustrative example (see response to referee 1).**

Please also note the supplement to this comment:

<http://www.biogeosciences-discuss.net/6/C4334/2010/bgd-6-C4334-2010-supplement.pdf>

---

Interactive comment on Biogeosciences Discuss., 6, 7933, 2009.

C4355