

Interactive comment on “An evaluation of ocean color model estimates of marine primary productivity in coastal and pelagic regions across the globe” by V. S. Saba et al.

Anonymous Referee #2

Received and published: 3 November 2010

In this study the skill of 21 ocean color models modelling marine net primary productivity (NPP) was assessed by comparing their estimates of depth-integrated NPP to 1156 in situ ^{14}C measurements encompassing ten marine regions. The study has been done applying up to date statistical analyses also focusing on analysing the influence of the uncertainty in the input data. The study shows that on average, the simplest depth/wavelength integrated models performed no worse than the more complex depth/wavelength resolved models and that all models seem to be suited much better for case-1 waters. The paper contributes with it's multiregional assessment of satellite-based NPP models to judging the value of and differences among these models. The manuscript fits well in the scope of the journal of Biogeosciences. The study

C3600

has been done applying up to date statistical analysis and gives useful information for what application these models are appropriate and which are their limitations. The paper is mostly written quite clearly. However, still a few analyses should be added (uncertainty analysis of the satellite SST, investigation of the euphotic depth estimates of the models, include the carbon based models in the PPARR) and a few things should be changed in the paper as outlined below: This would make clearer what are the differences among the models and what factors drive their limitations. Therefore I recommend to publish the paper after the modifications regarding these comments have been applied.

Specific Comments: 1. In Table 2: Models 8 and 9, also 12 and 13, and 20 and 21 are each time both characterised by the same reference. When reading the supplement it becomes clear what the differences between these twin models are. Still it would be much clearer to state already in the paper itself (Table 2) what the difference between these models are. E.g. for Models 8 and 9 (etc.) cite Model 8 with Behrenfeld and Falkowski 1997 and cite Model 9: Behrenfeld and Falkowski 1997, but modified by deriving P_{Bopt} following Eppley (1972)- anyway there is a typo: please change also “Bahrenfeld” to “Behrenfeld” 2. you excluded the carbon based models because the satellite data for particulate backscattering are not available prior to 1997- but the application for the satellite based NPP models are of greatest values using the SeaWiFS data base which starts in 1997. Therefore, I recommend to include these two models in the PPARR for the stations after 1997. 3. Chapter 2.4, page 6758, line 5 ff.: It is not clear what is the scientific basis for the uncertainty assumed for the NPP measurements- can you give more information about why for NPP values less than or equal to $50\text{mgCm}^{-2}\text{day}^{-1}$ $\pm 50\%$ error and for values larger than or equal to $2000\text{mgCm}^{-2}\text{day}^{-1}$ a $\pm 20\%$ error were assumed. 4. Chapter 2.4, page 6758, line 17 ff.: Also for the SST satellite data the real collocated SST (probably different satellite sensors products should be evaluated here) satellite data within the 27×27 km grid window should be extracted and the error should be calculated from these data. It is not clear why (and not appropriate that) for this parameter just an estimate obtained

C3601

by another study for a different data set is used. 5. Chapter 2.4, page 6758, line 19 ff.: What is the reason for comparing the MLD to the TOPS data? Are the TOPS data the products used in all the validated PPR models. If so, please explain. If not this data set is used by all models, please clarify. 6. The numbering of Figure 3 and 4 was mixed up between the description in the text and the given figure caption. Please change! 7. Figure 3: It is not clear from the legend what is the difference between blue and red columns. What is the basis for being the “Best model”? Also the layout has to be improved: the colours and the columns do not match in most of the plots. What is the cause that in the plots for the Med Sea and HOT there are only 20 models (=columns)? It is very difficult to see in the plots which model responds to which column. Please add the x-axis labels at each plot, not just only at the bottom. 8. Chapter 3.2.1, page 6760, line 11 ff.: What is the criterion for the best model? It seems to be more appropriate to consider that models perform best if their Model Efficiency is below 0. Therefore, the text should state in how many cases these models have a Model Efficiency below 0. 9. Page 6761, line 13: The last two sentences are more appropriate to be stated in Chapter 3.3.1 when it is talked about the different kinds of models and their performance. Anyway Fig. 6 should be stated after Fig. 5 is mentioned. 10. Page 6763, line 6 ff.: If Fig. 8 shows that at 20°C generally underestimate NPP at SST, then below 5° the models only rather overestimate NPP. Please change the text. 11. Page 6763, line 14: add here the citation to Fig. 7 because then it becomes much clearer which results show the inverse proportionality. 12. Page 6763, line 20: This result are shown in Fig. 9b not Fig. 9a. Obviously, the figure caption for Fig. 9 a and b was mixed up. Please change! 13. Page 6769, line 26ff. and page 6770, line 12ff: It is not clear to me why it should be difficult to get euphotic zone depths estimates for all tested stations from all the models which participated in this PPARR. Probably in most of these models this is parameter calculated anyway. The paper will benefit if the hypothesis “models overestimate the euphotic depth in case-2 waters while they underestimate it in case-1 waters” is further investigated.

C3602

Interactive comment on Biogeosciences Discuss., 7, 6749, 2010.

C3603