Biogeosciences Discuss., 7, C4119–C4124, 2010
www.biogeosciences-discuss.net/7/C4119/2010/

Biogeosciences Discussions

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.

V. S. Saba et al.

vsaba@princeton.edu

Received and published: 29 November 2010

Reviewer #2:

Comment: 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 PBopt following Eppley (1972)- anyway there is a typo: please change also "Bahrenfeld" to

"Behrenfeld."

Response: Excellent suggestion. We have added additional references to models 8-9, and 12-13 to distinguish them. Models 20 and 21 only differ in that one uses SST and one does not and this is already indicated in Table 2.

Comment: 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.

Response: The authors of the CbPM do not wish to include these two models in this analysis due to the poor performance of these models. (See Figure 4 of Saba et al., 2010 for an RMSD comparison of these models with the other satellite-derived PP models in the subtropical Pacific and Atlantic). They believe that this is due to both the sensitivity of these models to uncertainty in MLD input data (Milutinovic et al., 2009) as well as to the use of climatological bbp data as opposed to instantaneous bbp data. Specifically, the CbPM authors feel that because of the differences required to run the carbon-based models, it's not fair to compare them to more traditional PP models. There are three additional inputs required by the CbPM that are not given in the PPARR common dataset (bbp(443), Kd(490), and z_no3). Nearly 1/3 of the records (even the post-1997 period) required a climatological value of bbp to run the CbPM models. These climatological values can be up to 5x smaller or 2x greater than the actual time-varying matched up values. Even for many of the records that did have matchups, they were drawn from monthly imagery and required looking for matches anywhere from 20km to 1 degree away from the location of the in situ PP measurement (15-20% of the data points). This had to do with the fact that many of the data used in this analysis were from places where finding matchups are challenging due to relatively small numbers of clear-sky days, e.g. the Southern Ocean and North Atlantic.

BGD

7, C4119–C4124, 2010

Comment: 3. Chapter 2.4, page 6758, line 5 ff.: It is not clear what is the scientific basis for the uncertainity assumed for the NPP measurements- can you give more information about why for NPP values less than or equal to 50mgCm−2 day−1 ±50% error and for values larger than or equal to 2000mgCm−2 day−1 a ±20% error were assumed.

Response: There is no published citation for these values as they were based on communication with an experienced phytoplankton ecologist (performed many C14 NPP measurements), Walker Smith. These values represent the absolute error in the measurement that is commonly used for any type of field measurement, which is a linear function and thus has higher error at low values (less detectable) and lower error at high values (more detectable).

Comment: 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 27x27 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.

Response: The error between satellite-derived SST and in situ SST of +/- 1 oC is a conservative estimate given that typical error between Pathfinder SST and in situ SST is only about +/- 0.5 oC. This value was chosen to represent the maximum possible uncertainty in SST, rather than the average value. This is now made clearer in the text. The maximum uncertainty was used here because previous studies have shown that the ocean color models that use SST are not very sensitive to changes in SST of a few degrees (Carr et al., 2006; Friedrichs et al. 2009). Also, since the major drivers in our uncertainty analysis were errors in Chl-a and NPP, changing the SST error would not make a substantial difference to our results.

Comment: 5. Chapter 2.4, page 6758, line 19 ff.: What is the reason for comparing the

BGD

7, C4119–C4124, 2010

MLD to the TOPS data? Are the TOPS data the products used in all the validated PPR models. If so, please expain. If not this data set is used by all models, please clarify.

Response: The TOPS data set is typically used by ocean color models that require global MLD input data (i.e. see the Oregon State Productivity Website). By comparing measured or modeled MLD to TOPS, we get an idea of how sensitive ocean color model (those that use MLD) estimates are to the source of MLD. Our results showed that MLD is not a major driver of model-data misfit (as with SST). The exception is the CbPM's, but as discussed above, these needed to be removed from the analysis.

Comment: 6. The numbering of Figure 3 and 4 was mixed up between the description in the text and the given figure caption. Please change!

Response: Fixed.

Comment: 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.

Response: The red versus blue in Figure 3 (old 4) was based on visual inspection. Because we only have one RMSD value for each region for each model, we couldn't run an ANOVA to determine which models had significantly lower RMSD than others (like we were able to do for multiple regions). We have fixed the shortened columns for the Med. Sea and HOT and we have added x-axis labels for every panel instead of just the bottom two. We really appreciate the reviewer's scrutiny here as we somehow overlooked this error before we submitted the figure.

Comment: 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

BGD

7, C4119–C4124, 2010

Model Efficiency is below 0. Therefore, the text should state in how many cases these models have a Model Efficiency below 0.

Response: Excellent suggestion. The criterion for the 'best model' as used in this paragraph is stated to be the "lowest RMSD" (page 6760, line 12). As the reviewer suggests, it is also interesting to look at which models have ME <0 for the most regions, and this is now reported in this paragraph as well.

Comment: 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.

Response: Section 3.3.1 discusses individual model skill and model type across all regions whereas 3.2.2 discusses model performance in terms of bias and variance at individual regions, which is why we feel that the last two sentences are actually more appropriate for 3.2.2. We have removed the citation for Fig. 6.

Comment: 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.

Response: We have revised the text to indicate that the models typically over-estimated NPP at SST below 5 C.

Comment: 11. Page 6763, line 14: add here the citation to Fig. 7 because then it becomes much clearer which results show the inverse proportionality.

Response: Added citation for Figure 7.

Comment: 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!

Response: Fixed.

Comment: 13. Page 6769, line 26ff. and page 6770, line 12ff: It is not clear to me

BGD

7, C4119–C4124, 2010

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.

Response: The reviewer is correct in that it would not have been too difficult for all model participants to have provided their euphotic depths along with their PP estimates. Unfortunately, this idea did not come to us until after the analysis was complete. With the large number of participants involved in this study, it is not feasible to now ask them to go back and calculate these depths, but as we try to infer in the text, this is definitely something we’re keeping in mind for our next PPARR comparison effort.

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