

Interactive comment on “MAREDAT: towards a World Ocean Atlas of MARine Ecosystem DATa” by E. T. Buitenhuis et al.

Anonymous Referee #1

Received and published: 13 December 2012

Very important summary of a major and largely successful effort to compile and make data available on plankton functional types. For the most part, well organized and well written, a useful introduction to and synthesis of the other 11 papers / datasets.

I would have liked a summary / recommendations paragraph. Perhaps a few sentences on the successes: steps forward from prior situation, amount and types of data compiled, community response, utility of the MAREDAT data sets for green ocean modeling. Then perhaps a sentence or two on the weaknesses: fundamental lack of open ocean data; lack of compliance or cooperation by data holders - this no doubt varied within the community for each PFT - but did it limit the success for some PFT? Next steps: how biological oceanographers could collect samples and document data to improve utility for such compilation efforts; priorities for data gathering (or data rescue); which

C284

PFT groups most need improvement; time scale and processes for next iterations? I did not get a sense of the overall quality of the product - quite good I suspect - or of the success of the data gathering effort.

Specific comments

Page 3, line 5: “products with coverage of all biological components of the global ocean”. I think the authors mean all planktonic components? ‘All biological components’ implies a larger survey of fish and other larger organisms?

Page 3, line 18: “More data will be needed” From where? New data? Existing data but unavailable?

Page 4, lines 3-8: Confusion arises here. Report of a 2009 kick-off meeting by Le Que ^{re} and Pe ^{asant} 2009. But in the next sentence, a quote evidently from the meeting cites Le Quere et al. 2005). Perhaps it should say (roughly) the meeting endorsed the earlier goals specified by LeQuere “key plankton ...”.

Page 4, line 28: “have been included in the data products in this special issue. Nanophytoplankton” Should read ‘included in other data products in this special issue. For example, nanoplankton ...’?

Page 5, lines 4-7: “However, the most reliable way to prevent double counting and achieve a consistent dataset would be to measure the biomass of all phytoplankton groups in the same samples in transects that cross all ocean basins.” Do the authors recommend a biomass or carbon measurement of the total biomass in a given water or net sample? Or, systematic biomass (and carbon) measurements of all PFT groups, from which they can assemble better composite measurements? If they mean the latter, but uncertainty about nanoplankton definition and identification remains, how will subsequent measurements improve the biomass estimates for the nanoplankton PFT?

Page 5, line 13: Here, the “sum of all relevant phyla” represents a way to check the size

C285

overlap issues among the zooplankton PFTs? But, can groups actually do that check reliably?

Page 5, line 28: "datasets that are filled by interpolation". The authors mean that WOA achieves data coverage across all grid points by interpolation? Using this more precise description will help the reader later understand why the authors have not applied interpolation to the PFT data.

Page 6, line 2: Reference the CARINA or SOCAT products in ESSD?

Page 6, line 3: "all biological components of the global ocean ecosystem". Again, this phrase appears to imply a greater scope (fish, etc.) than intended.

Page 6, lines 8-9: "We accept that using a 4D grid will not yet provide enough information to furnish filled (interpolated) datasets." Dropping the time dimension (the 4th D) would not in fact make much difference when the real problem remains very sparse spatial coverage? E.g. even if the authors used only X, Y, and Z data (or even only X and Y data), they still would not have nearly enough global data coverage to satisfy any interpolation scheme? Thinking of Fig 1 in the macrozooplankton paper (Moriarty et al) where even for abundance, we really do not have enough data for interpolation through the central ocean empty spaces? According to Table 2 here, macrozooplankton actually has the best grid coverage, so all other PFTs have even less data? Main point - scarcity of data, more than 4D grid, prevents useful interpolation?

Page 6, line 15: What about the compiled data used in this paper? Also available? E.g. the data behind Table 2?

Page 7, line 2: I do not understand why cell volume matters for abundance data for the diazotrophs. As I remember that paper, the largest issue had to do with abundance to biomass conversions, wherein size (or volume) played a large role in the uncertainty in biomass, not in abundance?

Page 7, line 11-15. The authors should publish the microzooplankton corrections! But,

C286

have the authors and editors identified the best way to do this correction, and subsequent corrections and improvements to all the MAREDAT data sets? Should each correction have a least its own description and doi? How will authors / ESSD handle future corrections and improvements?

Page 7, line 18: "all collated data". Does this statement imply that the MAREDAT datasets include some 'un-collated' data, collected by each team's solicitation process but then collated or otherwise quality-controlled by the authors themselves? Of, have they only used 'collated' data? If so, what exactly does that term imply?

Page 8, line 19: "justified" YES! Even necessary?

Page 9, line 4: This should refer to Fig and Table 2?

Page 9, lines 12-15. In addition to the regridding already mentioned do readers and users get or need to get more information about the other corrections? Does the doi registry as applied by PANGAEA ensure that users get and understand the corrected dataset?

Page 11, lines 12-16. The text refers to the MAREDAT global HPLC pigment database as the basis for comparison with global abundances and refers to Fig 3, but according to the legend for Fig 3, Fig 3 uses WOA 2005 Chl data, not the Peloquin HPLC data?

Page 12, line 11. The authors attribute the diatom abundance peak at 125m depth (Fig 2a) to sampling bias. Presumably, sampling focused at the base of euphotic zones? Why wouldn't or doesn't such a depth bias also apply to the samples for HPLC pigments, and why therefore does the depth profile of fucoxanthin pigments not show a similar peak?

Page 12, lines 25-26. For macrozooplankton that migrate vertically (over 10s to 100s of meters?), and assuming sampling occurred without day (deep) or night (shallow) biases, wouldn't vertical migration result in a broad vertical distribution, not a single subsurface peak?

C287

Page 13, lines 15-17. Awkward. The “concentrations” are not “quite different from assuming”. Rather, the concentrations will require a re-assessment of the assumption that all deep sea activity derives from picoheterotrophs?

Page 13, line 23. Here we need a careful qualification. The authors interpolated the global biomass data by PFT across the depth bins, for purpose of an integrated biomass estimate - yes? They did not do any horizontal (grid point) interpolation?

Page 13, lines 24-26. Agree with these uncertainty factors. But, can the authors provide any quantitative information about these uncertainties. Do the abundance-to-biomass conversions account for half, or most, of the uncertainty compared to geographic biases. What about data completeness? Do the authors feel that they have gathered and compiled half (50%) of potentially available data sets? 20%? 80%? How would a more complete data set fill out the spatial coverage? Can we get an assessment of uncertainties due to incompleteness as compared to conversions and spatial biases? Even relative statement such as biomass conversions > spatial biases > incomplete datasets? By making such an assessment, the authors could help point the way to the most effective improvement strategies?

Page 14, line 15. But, doesn't the WOA chl dataset also in fact have a coastal bias, just through numbers of samples? And a bloom bias?

Page 15, line 18. But, didn't the authors just finish discussion of coastal biases? Here, we read that the data “are mainly representative of the open ocean environment”.

Page 15, line 19. This reader has a hard time to see Fig 5 as an inverted pyramid.

Page 16, line 10. I think I know what the authors intend with the word “crude”, but they might consider more accurate and descriptive words such as ‘preliminary’ or ‘incomplete’?

Page 16, lines 11-15. Good summary sentences! Do they belong in the abstract, as well?

C288

Page 16, lines 20-28. This section seems to have a paleo-climate focus, almost retrospective. But in the introduction we read about MAREDAT meeting the needs of the DGOM community, with a more prospective function. Please can we get some assessment of the MAREDAT product, as realized and summarized here, with respect to the DGOM requirements?

Page 17, lines 1-10. Only two expeditions, to fill all the gaps exposed by MAREDAT? I think I heard that TARA Oceans already ended, due to funding limitations. What about other data sources? What limits the gathering of additional data - ship time, investigator interest, data access?

Page 17, line 17. Global ocean biomasses of phytoplankton, zooplankton and picoheterotrophs are of roughly equal size?

Page 17, lines 23-24. “including additional observational data and improving the cell to carbon conversion algorithms”. How? On what time scales? New data, or additional existing data not yet included? For all PFT classes? Or, do some have priority over others?

Page 22, line 2. Peloquin et al. now submitted, has a doi?

Page 26, Figure 1. Needs units on the color axis. Not a good choice of colors (see next comment). How many total possible grid boxes for this land mask? Include the number in the legend for Table 1?

Page 27, Figure 2. Very important figure but very poorly presented. The reader needs to go back and forth between complicated figure and complicated legend - awkward and distracting. Very poor color choices, especially mixture of red and green and especially in panel D. See http://geography.uoregon.edu/datagraphics/color_scales.htm for much better color choices; many readers will have problems with this particular color scheme. Add text labels to each line? Make the Figure much bigger?

Page 28, Figure 3. Not particularly readable or useful in this format. Need clearer

C289

panel labels. Unless authors have a good reason for this arrangement, re-arrange the panels with 6 autotroph PFT stacked vertically on left, 6 heterotrophic PFT on right? Clarify the chl or HPLC pigment data used in panel F?

Interactive comment on Earth Syst. Sci. Data Discuss., 5, 1077, 2012.