Earth Syst. Sci. Data Discuss., 5, C67–C69, 2012
www.earth-syst-sci-data-discuss.net/5/C67/2012/

Earth System Science Data Discussions

Interactive comment on "Database of diazotrophs in global ocean: abundances, biomass and nitrogen fixation rates" *by* Y.-W. Luo et al.

Prof GRUBER (Referee)

nicolas.gruber@env.ethz.ch

Received and published: 27 June 2012

Assessment: Marine nitrogen fixation is not only a key process of the global N cycle, but also an important local contributor to the availability of fixed N in regions where the physical supply of this limiting nutrient is small. In response, investigators around the globe have been investigating the diazotrophic organisms that are capable of undertaking this process by enumerating their abundance, determining their taxonomy, and by undertaking rate measurements. Marine ecosystem and biogeochemical modelers have begun to incorporate N-fixation into their models by simulationg the abundance of the diazotrophs and how their N-fixation rates depend on environmental factors. Despite this large interest and the clear need of modelers to evaluate their predictions, no global, harmonized data collection of nitrogen fixers existed so far. Thus, this detailed

and carefully executed data compilation effort by Luo and colleagues is an urgently needed and very important contribution to the field. By combining biomass and taxonomic information from several methods and by incorporating also rate measurements, this compilation is as exhaustive as it can be.

The sources of the data, the procedures employed to harmonize the data and to quality control them, as well as the methods used to convert the cell counts and qPCR results into carbon biomass are well described, justified, and transparent. The data are well displayed, and the characteristics of the data set well described. The authors go even beyond the expected by computing global means for the biomass and the globally integrated rates of N-fixation.

Recommendation: This in an excellent and important paper that could be published pretty much as is. I have a few minor comments that the authors may want to take into consideration before preparing the final version.

Minor comments: page 60, lines 7-26: Quality control. Chauvenet′s criterion is certainly a reasonable choice, but it is based on certain assumptions about the distribution of the data and how well the sample distribution represents the "true" distribution of the data. I suggest to be more explicit about this here. This is particularly relevant, since the discussion later on suggests that some of the "outliers" are real, and simply reflect very specific situations that are rare and associated with certain locations.

page 60, lines 26: "are removed". As you actually don′t remove them from the data, I suggest to write "flagged" instead.

page 63, lines 6-15: Difficult to read. I am sure that this can be written more concisely.

page 65, line 5: "Continuous" I suggest to write "long-term sustained" instead. Sampling once per month is not really continuous.

page 66, lines 7-11: Sentence is unclear. I suggest to reformulate. Simply say that Trichodesmium dominates except in a few occasions where the biomass of Calothrix can be as large as that of Trichodesmium.

page 69, line 22: "which is consistent with the current geochemical estimates". I wouldn't say that this is consistent. Rather, I would say that this is at the low end of all recently published estimates. The error range seems very small, particularly when considering that the arithmetic mean gives such a different result. This begs the question how the uncertainty of these estimates were computed. Did the authors assume that each 3x3 pixel is independent? Please elaborate.

page 70, line 4-6: difference between arithmetic and geometric means in the North Atlantic. The presence of such a large difference in this ocean basin, compared to the other basins is puzzling me. This must indicate that the distribution in the North Atlantic must be quite a bit more skewed than in the rest of the ocean. Why should this be the case? Since the North Atlantic is one of the basins with the highest numbers of samples, this is actually a source of concern, i.e., perhaps the other basins have a similar distribution as the North Atlantic, but the limited sampling has not revealed this yet. I recommend that the authors discuss this puzzling finding in more detail than presently done.

page 71, section 3.6: I suggest to move this paragraph to the section where the biomass conversion factors are described the first time, i.e., 2.3, and then discuss the implications of these uncertainties also right when the results are discussed first. This also applies to Table 2 and 9, which can be easily combined into one table.

Tables 2 and 9: see previous comment. I suggest to merge these two tables.

Nicolas Gruber, June 2012

Interactive comment on Earth Syst. Sci. Data Discuss., 5, 47, 2012.