Earth Syst. Sci. Data Discuss., 5, C114–C119, 2012
www.earth-syst-sci-data-discuss.net/5/C114/2012/

Earth System Science Data Discussions

ESSDD
5, C114–C119, 2012

Interactive Comment

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

Y.-W. Luo et al.

yluo@whoi.edu

Received and published: 22 July 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.

Authors: We appreciate Dr. Gruber's evaluation and constructive comments. The detailed responses to his comments are listed below.

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 t remove them from the data, I suggest to write "flagged" instead.

Authors: We agree. To clarify, in this paragraph, we emphasized that we use "Chauvenet's criterion to flag suspicious outliers", not to remove these outliers. We also added another paragraph to clarify that even a data point was flagged, we have to check its specific sampling conditions: "All the data points not flagged by the Chau-

ESSDD
5, C114–C119, 2012

venet's criterion are accepted. However, not all the suspicious outliers flagged by the Chauvenet's criterion have to be rejected, as our datasets are not strictly log-normal distributed and the distribution estimated from the existing samples may not well represent the true distribution, especially at the sites with unusual environmental conditions. Thus for each flagged outlier, we evaluate whether its extraordinary high value is reasonable or spurious based on the specific sampling environmental conditions and/or the discussion with the original data contributor."

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

Authors: revised to "By applying Chauvenet's criterion, there are only 8 data points flagged, including 1 volumetric Trichodesmium cell count (Figure 2a), 1 volumetric Trichodesmium N2 fixation (Figure 2b), 2 volumetric N2 fixation rates by UCYN (Figure 2c), 1 volumetric whole seawater N2 fixation rate (Figure 2d) and 3 volumetric UCYN-B nifH genes (Figure 2e)."

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

Authors: agreed and corrected.

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.

Authors: agreed and revised to "The cell count data demonstrate that Trichodesmium is the dominant diazotrophs, except that the abundance of Calothrix can be as large as Trichodesmium (Table 5)."

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.

Authors: we revised to "which is at the low end of the current geochemical estimates ... ".

Authors: As described in the first paragraph of section 2.2, the error range for a geometric mean was based on the standard error for the log-transformed data, and then back-transformed exponentially, which resulted in the error range between "geometric mean multiplied by geometric standard error" and "geometric mean divided by geometric standard error" (geometric standard error: EXP(standard error of log-transformed data)). To clarify, we added one section at the end of the first paragraph of section 2.2: "Thus in this study, the error range for a geometric mean is represented as between and ." Also, we added text where the error range of a geometric mean first appeared in this paper: "Note that as described in Section 2.2, the error range for a geometric mean in this study is represented as between the geometric mean multiplied/divided by the corresponding geometric standard error."

Authors: Note that the standard error can be small with large sample numbers, even if the standard deviation is relative large.

Authors: We did assume that each 3x3 pixel is independent.

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.

ESSDD

5, C114–C119, 2012

Full Screen / Esc

Printer-friendly Version

Interactive Discussion

Discussion Paper

Authors: added following discussion (section 3.6) and Figure 10 for analyzing data distributions in different basins: "It is also a notable issue that the difference between the geometric and the arithmetic mean N2 fixation rate is larger in the North Atlantic than those in the other basins (Table 6). A strict log-normal distribution is symmetric in the log space and its geometric mean locates at the peak of the distribution, i.e., at the highest probability. However, the depth-integrated total N2 fixation rates in our database, after binned on 3ËŽ×3ËŽ grid, are not strictly log-normally distributed in each basin (Figure 10). The North Atlantic has highest spatial coverage, and the N2 fixation rates in this basin differ in 6 orders of magnitude from 10-3–103 μmol N m-2 d-1 (Figure 10). Its distribution is not symmetric, with more values observed in the left (lower) side. Thus the geometric mean N2 fixation rate for the North Atlantic is one order of magnitude lower than the peak of the distribution. The arithmetic mean in this basin is more than one order of magnitude higher than the geometric mean. In the North and South Pacific, the N2 fixation rates only differ in 2 orders of magnitude, without any data lower than 10 μmol N m-2 d-1 (Figure 10). Thus the difference between the geometric and arithmetic means is smaller in the Pacific than the North Atlantic. Although the geometric mean N2 fixation rate in the North Atlantic is about one order of magnitude lower than those in the North and South Pacific, the peaks of the distributions are actually much closer in these basins (Figure 10, also see Table 6). Compared to the North Atlantic, the Pacific is not intensively sampled for N2 fixation. Thus we need more samples with wider spatial coverage from the Pacific to better evaluate the globe N2 fixation rate."

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.

Authors: we thank Dr. Gruber's suggestions. We combined these two sections and

ESSDD

5, C114–C119, 2012

Interactive Comment

tables, which greatly saved the redundancy and made the paper more concise.

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