Earth Syst. Sci. Data Discuss.,
https://doi.org/10.5194/essd-2018-47-RC2, 2018

Open Access Earth System Science Data Discussions

Interactive comment on "A novel inter-comparison of nutrient analysis at sea: recommendations to enhance comparability of open ocean nutrient data" *by* Triona McGrath et al.

Anonymous Referee #2

Received and published: 31 May 2018

Review of McGrath et al. Novel nutrient comparison

The authors present data from a GO-SHIP cruise where they have the relatively unique opportunity to have two nutrient autoanalyzers on the same cruise. The manuscript focuses on not only the QC procedures but also how the two sets of data compare. The overall conclusions are that 1) levels of analytical precision are better (ie., lower CV) when samples are run ashore than at sea; 2) instrument responses are non-linear over the entire concentration range of the ocean from surface to depth and thus multiple CRMs are needed over this range; and 3) by better tracking of QC information in meta datafiles this will improve intercomparability of datasets generated by different people.

Printer-friendly version

Discussion paper

In summary this is an interesting paper, but really brings no new understanding to the body of knowledge.

Major concerns with the manuscript. 1. Only half the dataset is available. Unless I am missing something only the data from the MI analysis is present at the links given in the manuscript. The text suggests that, and looking closely at the XL file seems to confirm that. So actually this violates one of the primary review criteria of data accessibility. 2. How does this relatively unique comparison improve our ability to assess change in ocean nutrients over time? The authors conclude that having QC data available will improve intercomparability of at-sea generated datasets. It would have been nice to see an actual example of how they viewed this as working. As it is, I look at the control charts and see that there is IMMENSE day-to-day variability that is masked by the way they present the average CVs and RPDs in the paper. These control charts suggest that differences of 10% might need to happen between two datasets to say there is a change/difference between them. This seems rather large to me, especially for deep ocean nutrients. If one looks at ocean time-series, for example BATS, their deep nutrient data over 30 years hasn't changed anywhere near that much. 3. Focus of the recommendations for future work. As a nutrient chemist myself, the authors focus on recommendations that should be self-evident and thus seem to have limited value. Furthermore the recommendations focus on the idea that QC standards will 'fix' everything. Perhaps they will but the authors don't show that. Rather, I'm curious why the authors didn't make more recommendations about the importance of real procedural differences. Like splitting up your sample run into two (or more) components that are linear rather than trying to force a non-linear fitted function to the calibration data. I recognize that a single run is easier, but in the process are there compromises in data quality being made at both 'ends' (ie., shallow and deep). The authors try to demonstrate this in Table 5, so why not have a recommendation that in reprocessing data you only use the portion of the calibration curve around your data that is linear. I fear without focus on the actual methods, not just CRMS, that we will not get closer in our incomparability.

Minor concerns: 1. Given the other differences in Table 1, the authors should give development temperatures as well. Or if they are identical state as such. 2. Given that the authors label data as less than the LOQ or LOD in the datafile, perhaps including in the figures a dashed line representing the same values. 3. Perhaps I missed it, but the DAL method runs air through Cd column until reduction efficiency drops to 95%. How is the 5% loss of NO3 factored into the calculations? If it isn't now, would including it improve comparability given that 5% is one-half of the UAL/LAL range. 4. Line 454, I think they mean precision not accuracy. 5. Line 466-467. The authors comment on the difference between weighing standards and pipetting standards. Given they must know the error associated with both of those methods, could they not quantify that difference and thus provide more support for was is 'random' (and uncontrollable) error vs. error in controllable aspects of their methods. 6. Related to above. Line 469-470 – most of the random error (3% in the lab compared to 4-5% at sea) is not due to adverse conditions on the ship. Is this really random error? And if so what does that mean for detecting long term change in ocean nutrients. Why do some comparison exercises get 1%. The impacts of this on seeing change in the ocean need to be discussed. 7. Line 515: high quality shallow nutrient data is important. Geochemists use it to calculate parameters like N*. And thus our interpretation of ocean function is directly related to the quality of the measurements. This gets back to my earlier comment on recommendations that focus on methods as well as CRMs. 8. Table 1: concentrations of chemicals are needed not masses. 9. Table 3: a 1% measurement error is 2 decimal places. Can the values really be certified to 3 decimal places? 10. Figure 4: might be helpful to put a Zscore envelope around one of the analytical measurements to see where differences between datasets would be flagged 11. Figure 5: for the deeper data (>400m), in TOxN there is bias towards DAL measurements, but in PO4, it's a bias to MI. This isn't really discussed, is there a reason for this? 12. Some considerations for additional recommendations: a. Sample tube size, bigger tubes will minimize contamination effect during sampling. Is this also part of the difference between the two methods? b. Bottle type for freezing, see Dore et al. 1996 c. Is

Printer-friendly version

Discussion paper

Z-score =2 really OK or should we try to get it better?

Interactive comment on Earth Syst. Sci. Data Discuss., https://doi.org/10.5194/essd-2018-47, 2018.