Ocean Sci. Discuss., 11, C1116–C1120, 2014
www.ocean-sci-discuss.net/11/C1116/2014/

Ocean Science
Discussions

Interactive comment on "Technical Note: How long can seawater oxygen samples be stored before titration?" *by* M. Lankhorst et al.

H. Bittig (Referee)

hbittig@geomar.de

Received and published: 2 December 2014

General comments

The authors address an important question of the Winkler oxygen analysis which is the storage time of oxygen samples before titration. This duration is mostly based on experience rather than hard evidence and a thorough, data-based investigation would be a very useful step forward.

However, I have a few serious issues with the approach, the data, and the validity of the conclusions of the present study which are detailed below. The authors do not present the evidence to support their bold claims in the abstract.

At the same time I appreciate the authors' work to raise awareness for this issue and encourage them to extend and solidify their analysis.

Specific comments

Issue 1: Precision to be tested against

The comparison with the oxygen electrodes / optodes is irrelevant. Sensors have a whole lot of issues for themselves, e.g., the functional model to convert the sensor raw data to oxygen, a possible drift with time, the effect of hydrostatic pressure, ... These issues are the main reason for the larger accuracy given in sensor spec sheets in comparison to the oxygen titration. If the authors consult typical calibration sheets, e.g., for individually calibrated AADI 4330/4831 optodes or Sea-Bird SBE63 optodes, they will recognize that the initial calibration accuracy of sensors is usually much higher than claimed from the spec sheets (order of <0.01 mL L^{-1}, i.e., approximately the titration precision).

Since this study aims to assess the effect of sample storage on oxygen *titration*, its results need to be tested against the precision of the titration itself. In this respect, the authors must aim for the 0.5 μmol kg^{-1} / 0.01 mL L^{-1} target: Most of the analysis variation is introduced during sampling while the titration itself is very accurate (e.g., Langdon, 2010). Since the different storage treatments are sampled at the same time, by the same people, and in the same manner, they should be comparable within the "one operator" precision.

The different cruise/laboratory precision listed by Dickson (1996) includes effects of different gear (sample flasks, dispensers, chemical impurities), treatment of reagent blanks/calculations, ... for crossovers of *different* cruises. For multiple samples from one cruise by the same people, as presented here, these factors are the same for all samples. Therefore, the authors can not claim to fall into the less-strict multi-cruise criterion for the comparison of their batches.

Issue 2: Sampling approach and duplicates

As the authors note themselves, their "duplicates" are from separate water bottles and

thus superimposed to environmental variations.
With the current sampling, stored samples from the New Horizon clearly exceed the precision that would be possible with on-board titration (0.01 mL L^{-1}), while the Melville samples are within. Because of the adjacent-bottle-"duplicates" the difference in the results can not be attributed exclusively to the storage conditions. I.e., the data as such does not support the absence or presence of a storage effect.

Moreover, there are no replicates within the same storage treatment, so the precision of the analysis can not be judged independently (and compared to the difference between treatments).

It is not clear to me why the authors chose not to sample multiple times from the same water bottle immediately after another. Samples from the same water bottle would represent true replicates and thus give a better indication of the storage effect without environmental variability (especially for the shallower samples of the New Horizon cruise).
I would argue that the headspace air contamination argument is more related to exposure time than to the actual volume of the headspace, so if the samples are taken in direct succession there is negligible contamination in my experience (see, e.g., CLIVAR cruises with excessive CFCs, He, Ar, ... dissolved gas sampling which are all sampled before O_2 without a serious bias on oxygen). Culberson et al. (1991) describe an intercomparison of different titration methods where a total of 8 samples were drawn from the same 10 L water bottle in quick succession without detectable difference between the first and last samples. Their staggering of treatments vs. drawing order would be a good reference for a potential repetition.

Issue 3: Number of samples
The authors present a total of 43 samples of which 80 % are without problems. This boils down to 9 comparable samples in the most favourable case (NH1–NH2) and 4 in the least favourable case (NH1–NH3).

While I appreciate the care visible in the detection of problems and the detailed presentation of the results, I consider this way too few samples to draw a significant conclusion at all, and do not see how a bold claim such as "The implication of this is that such oxygen samples do not necessarily have to be analyzed while still on the ship; instead, it is possible to transport them ashore for analysis there." can be justified.

Considering the abundance of oxygen titration in oceanography, with a four-digit number of Winkler samples on a typical ocean-going cruise, such a claim must be supported by the required statistical and methodological robustness which I don't see presented here.

Further comments
- Abtract (p. 2448, line 4): Langdon (2010) (cited by the authors..) gives a storage time of many days if the water seal is maintained; the only specification I am aware of though.
- What is the volume of the Erlenmeyer flasks?
- You might consider to replace Fig. 1 with a depiction of a flask to illustrate the water seal?
- Section 3, first paragraph: Your title and abstract suggest a different use of the samples, independent of their original purpose. I suggest dropping the sensor aspect completely (this paragraph and second last of the introduction) since it distracts from the actual question "How long can seawater oxygen samples be stored before titration?".

Concluding remarks
The authors pursue a very interesting and worthy question. However, I suggest to repeat the experimental work on a much larger scale and with
- replicates from the same Niskin water bottle,
- replicates within the same storage treatment,

- comparison against the correct precision, i.e., “one operator” precision that is attainable if samples are titrated directly on board.

Culberson et al. (1991) is an example for a similar, well-documented experiment which might serve as inspiration for a revised sampling approach. On quite a few cruises a “test station” is performed at the beginning where all Niskins of the rosette are closed at the same depth, brought on deck and the water analyzed to test for leaking Niskin bottles. Such a station would be an ideal candidate to take a large number of samples for different storage treatments. However, I did like that the samples also covered a range of oxygen concentrations/saturations. This should be incorporated in a repeated study, too. Finally, I suppose the results presented here are sufficient to drop the NH3/batch3-treatment – the procedure sounds awkward both from a handling and an analytical perspective. This would reduce the complexity simply to the storage time experiment.

Interactive comment on Ocean Sci. Discuss., 11, 2447, 2014.