

Review of “**New constraints on biological production and mixing processes in the South China Sea from triple isotope composition of dissolved oxygen**” by Hana Jurikova, Osamu Abe, Fuh-Kwo Shiah and Mao-Chang Liang (Biogeosciences Discussions, 2021, manuscript bg-2020-448)

Summary: Jurikova et al provide measurements from the SEATS station in the South China Sea of O₂ triple isotopes (Oct 2013, Aug 2014 and Apr 2015) and O₂/Ar (Aug and Apr only). They quantify rates of NP and GP and discuss seasonal changes. The topic of the manuscript is relevant to the Biogeosciences audience, as indicated by the companion paper Jurikova et al. (2016). I am always interested to read papers with triple oxygen isotope data from a new region and to see new research groups conducting these very challenging and valuable measurements.

Summary of major comments:

Below I summarize the major issues that I believe should be addressed before acceptance. More details on each point follows on subsequent pages.

- 1) Although the article focuses on the calculated GP and NP rates, there uncertainties in these rates are not quantified. Comprehensive error analysis of the calculated rates is required. The manuscript should acknowledge that they cannot correct for the effects of complex physical processes (vertical mixing, lateral advection) and non-steady state on their mass balance. The authors state that the system was net heterotrophic in Aug 2014 (L16); in fact, the NP rate was effectively 0 considering the uncertainty. I think the article would be more useful to the broader community and future investigators if they also described how future studies could constrain the largest uncertainties in the NP and GP estimates.
- 2) The authors describe the use of “ $^{17}\Delta$ of deep O₂ as a valuable novel conservative tracer for probing mixing processes” (L22) but simple calculations and previously published papers have shown that $^{17}\Delta$ is nonconservative in the subsurface when the effects of mixing and respiration considered together (Nicholson et al., 2014).
- 3) The authors report $^{17}\Delta$ values for the thermocline that are “much higher than any previously reported values” and very high values in the deep ocean at low O₂ levels. I am concerned these results may be an analytical artifact related to size/pressure effects (nonlinearities in the mass spectrometer response when the sample and standard contain different amounts of gas). The authors should describe the calibration procedures and directly address the possible uncertainties.
- 4) The authors are not in compliance with the journal’s data policy. More details are required on the methods for calculating GP and NP so that they can be reproduced by others. The mixed layer depth calculations may need to be modified.

Detailed comments:

First, a comprehensive error analysis of the dataset is required. In my opinion, the claim in the abstract of the system being net heterotrophic in Aug is not supported by the analysis. Looking at Table 1 and the supplementary data, the O_2/Ar disequilibrium at the surface in Aug 2014 was $\leq 0.5\%$ (the stated analytical uncertainty, L120-124), so they cannot conclude whether the system is net autotrophic or heterotrophic.

Beyond the analytical uncertainty, they have not acknowledged many other larger sources of uncertainty, such as the gas transfer velocity (e.g., the parameterization choice, and the method of time integration), the conversion from O_2 to C units, mixed layer depth, and potential impact of physical processes (e.g., vertical mixing and lateral advection), and non-steady state dynamics on their calculated rates.

Indeed, corrections for physical processes and non-steady state would be very difficult to estimate given these given they only have profiles at one location, and only on one or two consecutive dates in each season. However, the 20-30 per meg difference in $^{17}\Delta$ on consecutive days in April and August at 5-10 m depth, as well as the broad range of literature on circulation in this region demonstrates the system is very dynamic and that a mixed layer NP and GP estimates based on the current dataset will have significant uncertainties which should be acknowledged.

On L150 they state "Assuming the mixed layer is at steady state" – I think it would be appropriate to add "and there is not significant entrainment/upwelling of low- O_2 subsurface water into the mixed layer, nor lateral advection from adjacent waters."

On L280 they state "Furthermore, we note that the calculated production rates should be considered as minimum production only. At SEATS the euphotic zone was persistently deeper than the mixed layer during our sampling (Fig. 3), which may lead to an underestimation of the rates." A similar statement is in the caption of Table 1. I think this would be better framed by stating that you are calculating mixed layer production rates. Mixed layer GP will indeed be less than total water column GP if production occurs below the mixed layer. However, their calculated mixed layer productivity may underestimate the true mixed layer NP values due to mixing/entrainment of low- O_2 waters into the mixed layer, and overestimate the true mixed layer GP values due to mixing/entrainment of high- $^{17}\Delta$ waters into the mixed layer.

I think the article will be of more use to future investigators if they describe how future studies could be designed to better constrain the largest uncertainties in the NP and GP estimates.

Second, the authors incorrectly claim several times that $^{17}\Delta$ in the subsurface is conservative and only affected by mixing (e.g. L22, 72, 322, 363). For example, L322:

“Due to the conservative behaviour of O_2 in a parcel of deep water, where it may no longer be influenced by air–sea gas exchange or photosynthesis, the $^{17}\Delta$ could also present a valuable tracer for deep water mixing processes, since any variations in $^{17}\Delta$ can only result from mixing of waters with different $^{17}\Delta$.”

This statement is unfortunately not true. Please see Nicholson et al. (2014), Figure 6 and section 4.2 which states “while respiration alone does not alter $^{17}\Delta_{dis}$, the tracer is nonconservative when the effects of respiration and mixing are combined”.

The authors’ discussion of $^{17}\Delta$ as a subsurface tracer should be substantially revised. The authors could perform simple calculations to see whether their observed $^{17}\Delta$ values can be explained by mixing and respiration. If the observations can be explained by simple modeling, it would give the reader more confidence that their unprecedented subsurface values are accurate (see below).

Third, the authors report $^{17}\Delta$ values in the subsurface that are “much higher than any previously documented upper ocean values” (e.g. $^{17}\Delta$ of 218 per meg and $\Delta O_2/Ar$ of -48% at 100 m depth on 5 Aug 2014). They also report very high values in the deep ocean (e.g. $^{17}\Delta$ of 215 per meg and $\Delta O_2/Ar$ of -71% on 6 Aug 2014). Modeling results (Nicholson et al., 2014) and the observations of Hendricks (2005) predict $^{17}\Delta$ in low- O_2 waters can actually reach or approach negative values due to the combined effects of mixing and respiration.

Based on my own experience measuring $^{17}\Delta$, I am concerned these highly unusual $^{17}\Delta$ values at O_2 undersaturation are an analytical artifact related to variability in the sample size (number of moles of O_2 in the flask). The methods are not described in this paper but the authors reference Jurikova et al. (2016) where it appears that they applied a correction for the O_2/Ar ratio in the sample (Ar correction), but not the total amount of gas in the sample (size correction).

In their response to the reviewers of the 2016 paper, the authors stated: “We did not perform any corrections due to differential gas depletion between the bellows, as we also did not observe any fractionation (within the current precision).”

I am not sure I understand this statement. The way that researchers typically perform a sample size calibration is by putting the reference standard in both bellows but varying the total amount (volume) of gas in each bellow by adjusting the bellows compression and plotting a calibration curve that spans the full range of

sizes in their sample set. The size correction is discussed in the appendix of Stanley et al. (2010). Note that this size correction also simultaneously corrects for the “pressure baseline” effects described by Yeung et al. (2018). This correction should be performed periodically as it varies with time.

For the system I have used, the size calibration was the largest correction (much bigger than the Ar calibration), whenever the size difference between sample and standard was greater than roughly 10% (i.e., most samples from below the mixed layer). However, this correction will vary with time and between instruments, and it is possible for it to be negligible, even at low O₂ saturation.

The authors need to state in the manuscript whether these size correction tests were done across the full range of sample sizes measured in this dataset, both for the Ar-free and with-Ar methods. If the size corrections were performed with both methods and the offset was negligible, that will give the reader confidence in the data. If they were not performed, then this potential uncertainty should be clearly communicated to the readership base who may not have experience with these very challenging measurements.

It is also worth noting that because the calculated NP and GP values are based on mixed layer data only (they are not corrected for mixing/entrainment of the subsurface waters), any errors due to a lack of size correction may not significantly affect their NP and GP calculations, if the size of sample and reference was similar for all of the mixed layer samples.

On L95 they mention that for the October data (Ar-free method) some samples were run at Nagoya University – did the Nagoya laboratory perform a size calibration? What were the result of the inter-lab comparison? (e.g., what was the mean absolute difference between duplicates run in the two labs)? In general, how many samples were collected at each depth (error bars are not reported on the figures or supplemental data)? What is the precision of replicate water field samples (not lab-equilibrated waters)?

On L120, they report the reproducibility of the equilibrated water samples. What was the mean value of ¹⁷Δ in equilibrated water?

Finally, the authors are not in compliance with the journal’s data policy (https://www.biogeosciences.net/policies/data_policy.html) and have not provided sufficient details on the GP and NP calculation methods. Stating that the CTD data is available by contacting the authors is not acceptable. All data needed to reproduce the GP and NP calculations should be published to a repository capable of issuing a DOI such as PANGAEA or Zenodo.

Authors should archive the following:

- 1) Date and time of each sample (since they discuss in the paper that time of day is important; note the date of the final profile is incorrectly listed as April 25th, 2016; the year should be 2015)
- 2) latitude and longitude
- 3) all CTD data used in the paper: exact depth/pressure, temperature, salinity, O₂, chlorophyll fluorescence for each sample discrete ¹⁷Δ sample, as well as the high-resolution profiles plotted in Fig 4.
- 4) optional: mixed layer depth, wind speed, and k values used to calculate GP/NP and the final GP and NP results, and a description of how each was determined.
- 5) Define in the metadata how ¹⁷Δ is calculated (what lambda is used)

The authors have not stated how the mixed layer depth (MLD) is defined. I believe that the calculated NP on 24 Apr is too low due to an incorrect definition of the MLD. I suggest that they should use an O₂-based mixed layer depth as described in Castro-Morales and Kaiser (2012). This would prevent the authors from including undersaturated O₂ samples in the NP calculation. In April 2015, it appears the authors used the same MLD of 23 m on both sampling dates (Table 1), even though the O₂ profiles on the 24 and 25 Apr are significantly different. On 24 April, ΔO₂/Ar is 1.6% at 5 m, 1.3% at 10 m, and -4.5% at 20 m. Including the 20 m depth sample in the mixed layer leads to an underestimate of NP. They report the mixed layer ΔO₂/Ar = -2% and NP -160 mg C m⁻² d⁻¹ in Table 1, when mixed layer NP should be positive based on the samples at 5 and 10 m depth. On 25 April, ΔO₂/Ar is ~2% (range 1.4 to 2.2%) from 0 to 30 m depth so a MLD of 23 m is reasonable. If they used a shallower MLD on 24 Apr, based on the O₂ profile, then the calculated NP rates on 24 and 25 Apr would be very similar.

On L137, please clarify how “mixed-layer O₂ production time” is calculated and how it differs from mixed-layer O₂ residence time. I have never heard this term before.

Additionally, I did not understand their method of calculating k (gas transfer velocity) and request they provide more details on this. On L133 they state “K was derived from wind speeds measured on the ship using an anemometer and verified against NCEP data” but then on L135 say that K was “averaged over the O₂ residence time in the mixed-layer preceding sampling (16, 7, and 4 days for October 2013, August 2014 and April 2015, respectively). How did they get 16 days worth of data if using the sonic anemometer?”

Please note that the widely-accepted method of calculating k for O₂ mass balances involves a weighted approach accounting for the fraction of the mixed layer ventilated each day rather than a simple average over the residence time of the mixed layer (Reuer et al., 2007; Teeter et al., 2018).

In the data supplement, one of the cruise dates is incorrectly listed as April 25th, 2016; the year should be 2015.

In Figure 4 and elsewhere in the text please specify the time of day each O₂ and triple oxygen isotope profile was collected (which of the fluorescence profiles does it correspond to)? Could differences in the time of day (diel variability) likely to cause some of the differences between consecutive days that they observe?

L 29: replace “unrepresentative” with “highly uncertain” or similar.

Final points:

I hope that the authors find my feedback helpful and constructive. I realize that I have provided many recommendations to improve the manuscript, and that making these changes will not be a simple case of minor revisions that authors often hope to receive. I am suggesting these revisions to ensure that the published manuscript is scientifically accurate and to maximize future reuse of the manuscript and dataset by other researchers, which should hopefully lead to more citations for the authors.

References

- Castro-Morales, K., & Kaiser, J. (2012). Using dissolved oxygen concentrations to determine mixed layer depths in the Bellingshausen Sea. *Ocean Science*, 8(1), 1–10. <https://doi.org/10.5194/os-8-1-2012>
- Hendricks, M. B., Bender, M. L., Barnett, B. A., Strutton, P., & Chavez, F. P. (2005). Triple oxygen isotope composition of dissolved O₂ in the equatorial Pacific: A tracer of mixing, production, and respiration. *Journal of Geophysical Research: Oceans*, 110(12), 1–17. <https://doi.org/10.1029/2004JC002735>
- Jurikova, H., Guha, T., Abe, O., Shiah, F. K., Wang, C. H., & Liang, M. C. (2016). Variations in triple isotope composition of dissolved oxygen and primary production in a subtropical reservoir. *Biogeosciences*, 13(24), 6683–6698. <https://doi.org/10.5194/bg-13-6683-2016>
- Nicholson, D., Stanley, R. H. R., & Doney, S. C. (2014). The triple oxygen isotope tracer of primary productivity in a dynamic ocean model: Triple oxygen isotopes in a global model. *Global Biogeochemical Cycles*, 28(5), 538–552. <https://doi.org/10.1002/2013GB004704>
- Reuer, M., Barnett, B., Bender, M., Falkowski, P., & Hendricks, M. (2007). New estimates of Southern Ocean biological production rates from O₂/Ar ratios and the triple isotope composition of O₂. *Deep Sea Research Part I: Oceanographic Research*, 54, 951–974. <https://doi.org/10.1016/j.dsr.2007.02.007>
- Stanley, R., Kirkpatrick, J., & Cassar, N. (2010). Net community production and gross primary production rates in the western equatorial Pacific. *Global Biogeochemical Cycles*, <https://doi.org/10.1029/2009GB003651>

- Teeter, L., Hamme, R. C., Ianson, D., & Bianucci, L. (2018). Accurate Estimation of Net Community Production From O₂/Ar Measurements. *Global Biogeochemical Cycles*, 32(8), 1163–1181. <https://doi.org/10.1029/2017GB005874>
- Yeung, L. Y., Hayles, J. A., Hu, H., Ash, J. L., & Sun, T. (2018). Scale distortion from pressure baselines as a source of inaccuracy in triple-isotope measurements. *Rapid Communications in Mass Spectrometry*, 32(20), 1811–1821. <https://doi.org/10.1002/rcm.8247>