Reviewer 2, Henry Bittig

Below the review is reproduced in black font and our responses interspersed in blue.

Reviewer Comments:

In this paper by Gordon et al., the authors present and discuss (1) a method to determine effective response times from consecutive up- and downcasts, and (2) how to discern between physically- and biologically-driven diel variations in O_2 observations and the limits to derive gross primary production GPP and respiration R from them.

The paper is structured logically and is written excellently. The math around the time response correction is particularly well presented, and the time response part is applicable not only to O_2 optode sensors and BGC-Argo floats, but to any sensor on any profiling platform. The discussion around GPP/R estimation shows a high degree of critical assessment of sensor data accuracy, which is a good example and should happen more often. I recommend to publish this paper with minor revisions.

Response: Thank you, we appreciate the positive and constructive comments

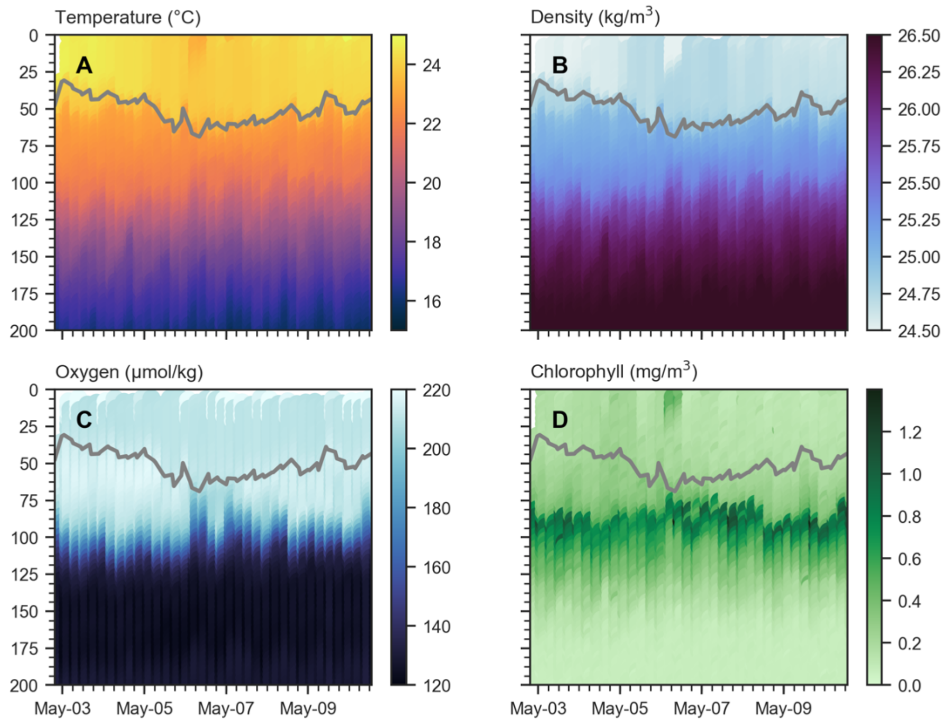
Comments:

(1) The authors refer to Bittig et al. (2014), who studied oxygen optode time response and found flow and temperature to be the main factors modulating response times. As the authors write, flow around the optode modulates the water boundary layer thickness through which O_2 has to diffuse, thus slowing down sensor response. The authors discuss the impact of flow and conclude that flow variations are of second order importance in their application and that they can assume a uniform (or at least common) flow regime and thus one effective response time. Given the variations in profiling speed, varying over an order of magnitude, this could be argued. But the desire for simplicity and the results give justification to this approach.

Response: We didn't state this as an assumption but agree that implicitly we have assumed it. We would like to explicitly acknowledge this in the revised manuscript.

What the authors do not discuss and do not include by using a uniform effective τ is the variation in response time induced by temperature. Bittig et al. (2014) show that a temperature change from 5 °C to 25 °C (deep vs. surface waters in the Gulf of Mexico) can reduce the response time by 33 % within the same profile. Could that explain some part of (a) the uncertainty in the calculation of one uniform, effective response time for the entire profile (e.g., table 2) and (b) leading to incomplete correction of the time response thus adding uncertainty/bias in the O_2 gradient region and preventing the GPP/R analysis in part 4?

Response: We are only focused on the euphotic zone here, hence don't think the temperature change across the oxycline (<3 degrees C; see figure pasted below) would have a noticeable effect. However, this point is relevant when the correction is applied for the whole water column. We will emphasize this in the Discussion in the revised manuscript.



By using a simple two-layer diffusional model, Bittig et al. (2014) show that the temperature effect can be removed, leaving one parameter (the boundary layer thickness) to characterize the temperature-dependent response time at a given flow regime.

Can the authors (I) modify their approach to not neglect the temperature influence on response time? Bittig and Körtzinger (2017) provide the data of Bittig et al. (2014) as look-up table (LUT) in their supplementary material (T_lL_tau_3830_4330.dat[1]). The authors could either use this LUT to find the corresponding boundary layer thick- ness for their effective response time at a certain temperature, and modify the response time applied for the correction according to the LUT's temperature dependence. Or, they could optimize for an effective boundary layer thicknes l_L instead and apply the response time for correction according to the LUT.

Can the authors then (IIa) discuss whether that reduced the spread in response times per float, and (IIb) whether that reduced RMSEs between up- and downcasts and whether physical imprints on diel O_2 variations (in isopycnal space) are reduced, thus permitting the GPP/R analysis of part 4? If that is not the case, can the authors explain or speculate why?

The authors write that they do not strive to fully characterize and understand the flow around the sensor. Neither do I. The suggestion to optimize for one effective boundary layer thickness instead of one effective response time means only to take the demonstrated temperature-dependence of the response time into account.

Response: These are really excellent suggestions and something we would like to pursue in future work. This analysis is beyond scope of current paper, where we focus on the euphotic zone only, but as we already stated above, we would like to add these thoughts to the discussion section.

(2) Barone et al. (doi: 10.1002/lom3.10340) recently published a work where glider measurements were used to estimate GPP/R from diel O_2 variations in the subtropical North Pacific. The method proposed in that work provides daily GPP/R values and Barone et al. state that the method "resolved variability on time scales of approximately 1 week". Could the authors apply these methods to their data, or comment on Barone et al.'s findings?

Response: Thank you for pointing out this relevant study to us. We will add this in the revision.

(3) Data availability: The float data are not available under the link provided. https://data.gulfresearchinitiative.org/data/R5.x275.000:0002 gives some float data, however, only for the first set of continuous mode profiling during the first couple of days, as far as I can tell. Moreover, they do not include measurement times, which the authors rightfully state as being important.

Response: Thank you for taking the time to look into it. Indeed, the link we provided was only for the subset. This is the proper link to the data in the GRIDC database: https://data.gulfresearchinitiative.org/data/R5.x275.281:0001

Since the link above does not include the raw and metadata that an experienced Argo user would be interested in, we will also prepare the float data in the BGC Argo standard format (one NetCDF file per float with all raw and metadata) and make it available via a permanent data archiving platform (most likely https://zenodo.org/)

(4) The authors provide code to re-apply their approach to determine response times to other or similar data, which is excellent. However, the authors should also comment on the parameters they used. Notably, whether they used the full depth range to find the optimal τ (implicit from the paper, as not stated otherwise), or limited the optimization to the upper O_2 gradient (default depth range of 25-175 dbar in the code).

Response: Agree. We will add this.

(5) The standard deviations of the in-air gains are very high, about an order of magnitude higher than observed usually for other (APEX) floats (see Argo, Bittig et al. 2018, or others)! What's the reason for this large scatter? Do the authors still have confidence in the in-air corrections given these large variations? The optode attachment on a short stalk looks comparable to other (APEX) floats.

Response: Agree, we're looking into this.

Remarks:

- p.3 l.26: "No depth binning was perfomed." Depth resolution should be stated here.

Response: Agree. It was ~5 dbar.

- p.5 table 1: Please verify numbers. E.g., f7941 has about the same N_{cont} as f7939 but 100 profiles more in total. Still about the same start and end date? This does not match the float operation modes.

Response: Yes, because f7939 was out of contact during park-and-profile mode for a while (probably because it was stuck under a fresh surface layer and thus not able to surface and transmit). We're stating that in the text (last paragraph of section 2.1). We could add the float number to the text during the revision.

- Check references. Bittig and Körtzinger (2016) (discussion paper) should be Bittig and Körtzinger (2017) (published paper).

Response: Yes, thank you!

- p.12 l.30: "based on theoretical considerations of flow-dependent boundary layer thickness" Not quite. Response times of Bittig and Körtzinger (2017) are based on an in-situ comparison between two optodes, one with well-defined time response (pumped optode), the other one with variable response time (unpumped) as in this study. Re- sponse time determination was empirical as in this study. The boundary layer thickness as well was empirical and a mere tool to eliminate the temperature-dependence. More- over, the range of response times in the application of Bittig and Körtzinger (2017) is stated as 60-95 s. The range of 70-140 s given in Bittig et al. (2018) covers the global range of possible scenarios, including very cold, polar surface waters with longer re- sponse times. The present setting in the Guld of Mexico is better comparable to the subtropical setting of Bittig and Körtzinger (2017).

Response: Thank you, this will be corrected.

- p.13 l.41: "the impact of this difference on the correction is likely small" Speculation. rather: "unknown"?

Response: Agree. Changed as suggested.

- Figure 8: Why was the mean taken from 25-150 dbar and not from the surface-150 dbar?

Response: We didn't use oxygen measurements when the CTD was turned off. We can add this to the text.

Typos:

- p.12 l.37: -that

Response: OK, removed.

- p.16 l.26: -and

Response: OK, broke up sentence instead of using "and."

- p.21 eq.C2 and C3: Can you check the sign in the numerator?

Response: Sign in C3 was wrong, now corrected.