Response to the comments and questions of Conrad Pilditch (Reviewer#2)

We thank Conrad Pilditch for the review that helped improving this manuscript. We addressed all questions and comments, and our responses and actions are listed point by point below.

Overall this is a very interesting paper and one I enjoyed reading very much. It tests whether adding additional O2 sensors to an eddy-covariance instrument improves aliasing in data due to the separation of velocity and O2 sampling locations. Given the increasing use of eddy covariance measurements to estimate benthic primary production/respiration at scale technological improvements are timely and welcome. I am very supportive of this paper however there are a number of elements that if considered in revision could improve the focus and clarity of the manuscript.

Response: We appreciate the positive comments of the reviewer and that the manuscript was perceived as timely.

My main comment addresses the multiple elements to this paper. It seems to bounce around between a confirmation of the fact that shallow water permeable carbonate sands are hot spots of benthic primary production and organic matter processing and testing whether additional O2 sensors improves the precision of flux measurement. Given the paper is submitted as a technical note it could be improved (and shortened) by focusing on the increase in performance of adding additional sensors. The sections of the paper discussing the high production/respiration rates of carbonate sands confirms previous studies (see Fig 6) and in my opinion distracts from the method which is generalizable to many systems.

Response: We appreciate this comment that is similar to the one phrased by Reviewer#1, and in the revised manuscript we now shifted the focus more onto the method and removed some of the ecological interpretations. These changes affect the abstract, introduction and discussion sections. Sections addressing carbonate sand were shortened and placed after the sections addressing the technology. The changes are also described in the responses to Reviewer#1.

If using eddy-covariance techniques I would want to know exactly what gains could be made by adding sensors as the additionally increases costs and/or may reduce the ability to spatial replicate units. These trade-offs are import – is it more important to increase the precision at one location or potentially increase the number of locations at which flux measurements are made to assess spatial variability? So the questions I would like answered explicitly are what improvements are made my adding sensors in terms of precision and do these improvements vary with hydrodynamic setting (eg. uniform steady flow vs more wave dominated flows), do these improvements really matter in system with high natural variability in fluxes and what other conditions/settings need testing to confirm the value of sensor additions. Revising the manuscript (mainly editing the Introduction/Discussion) with this comment in mind I think would result in a much more assessable paper with a tighter focus. In short make the technical note more about the method than the system in which it was tested.

Response: (P12L252). We added a paragraph and table that summarizes the characteristics of the new 3OEC and the 2OEC and highlights the advantages of the new instrument. The table documents that adding sensors increases the cost of the instrument relatively little (~10%) and the effort for setting up the new instrument is very similar to the effort needed for a conventional instrument. The discussion now is more focused on the new method.

Specific Comments

Ln 160 Please provide more detail on what data was used in t-test comparing the 3OS and 2OS. The DF indicates 7 data points – were these the average of the 15 min blocks across the four sample dates? I am assuming that both the 3OEC and 2OES systems were synched so perhaps a better test may have been a paired t-test where you ask whether the difference between the data is <> than 0.

Response (P8L192). We followed the suggestion by the reviewer and now apply a paired t-test for comparing the 3OEC and 2OEC fluxes. According to this test, the nighttime fluxes between the two instruments were significantly different, while the daytime fluxes were not. We changed the text accordingly which now reads (P7L191): "Average 3OEC daytime fluxes were 7% lower and nighttime fluxes 38% lower than the respective 2OEC fluxes (day: 5.6 ± 0.8(SE) night: -3.9 ± 0.5(SE) mmol

m^{-2} h^{-1}). The difference in the nighttime fluxes between the two instruments was statistically significant (p= 0.04685, p(x≤T) = 0.02342, T =-3.268, DF=3), while the difference in daytime fluxes was not (p=0.08077, p(x≤T) = 0.9596, T = 2.5944, DF=3)."

Line 165 Two sentence paragraph that does not make sense on its own

Response (P11L237). Thank you for pointing this out. The paragraph was rewritten and now reads: "Yet, fluxes scaled with the average unidirectional bottom current velocity, which slowed during the deployment week (~30 mmol m-2 h-1 flux increase or decrease per m s-1 flow decrease, Fig. 5a), and not with significant wave height (R2<0.04, Fig. 2ab) that increased during the study except the last day.".

Fig 3 – Add the p values for the regression statistics and clarify what data is being average for each of the visible data points. If the data represents averages of 15 min intervals then surely there is a variation in mean current velocity between intervals that should be plotted as an error term?

Response (Figure. 5). We added the standard errors for the average velocities and included the following explanation in the legend "The data points indicate the average fluxes calculated for light daytime and dark nighttime periods, separated at 20:00, plotted against the average flow velocity for the respective time periods. The compromised data point from the 16 July 2OEC deployment was excluded from the regression (grey circle). Error bars represent standard error.". The four data points (2OEC night 3 data points) available for the regressions do not allow meaningful regression statistics and we removed the R^2 values. The combined nighttime data of the two instruments produce a statistically significant trend, the daytime data don't. Nevertheless, the consistent and almost identical trends (with respect to slope) observed in the two independently measuring instruments as well as agreeing reports in the literature documenting flux enhancement by flow in permeable sediment, e.g. (Berg et al., 2013;Chipman et al., 2016;McGinnis et al., 2014) support our interpretation that fluxes increased with flow velocity.

I would also like to see what if any difference results from the generated PI curves from the 2OEC and 3OEC systems – are the fits better (r2) are the fitted parameters known with greater precision and does this matter?

Response (P11L244). The 3OEC improves the precision which improves the reliability of the PI curves, however, due to the scatter in the data, the differences between the predicted maximum gross benthic primary production rates were statistically not significantly different. The revised text section now reads: "Improved precision and the generally lower fluxes in the 3OEC were reflected in the community photosynthesis-irradiance (PI) curves (Bernardi et al., 2015). The 3OEC predicted a slightly lower maximum gross benthic primary production (GPP) of 9.9 mmol O_2 m^{-2} h^{-1} (R^2: 0.999) than the 2OEC (10.7 mmol O_2 m^{-2} h^{-1}, R^2: 0.998, Fig. 5b) as well as a lower light utilization efficiency (LUE, ratio between GPP and PAR, 3OEC LUE 12.3% lower than 2OEC LUE at 10 µmol photon m^{-2} s^{-1} and 7.4% lower at 350 µmol photon m^{-2} s^{-1} (Fig. 5c)). Due to the scatter in the data, these differences in GPP maxima and LUE were statistically not significant."

Fig 5 Are there any corrections applied to data in the (a)? That is, is the variation observed between individual sensors a function of sensor performance vs data that has not been corrected for R, S, T & W.

Response (Figure 4). We mention in the legend "(all data in (a) and (b) are STW- corrected)"

Line 180 The order is a little illogical – when looking at Fig 5 c&d I see a T correction was included in the 3OEC data processing – yet in the Introduction it was emphasised that the advantage of using 3 sensors was to avoid needing to do this. It is not until the Discussion that this is discrepancy is explained. I would suggest that Fig 8 & text goes into the Results to explain/justify this correction. Alternatively improve the legend to Fig 5 and point to the Discussion for explanation.

Response (Fig. 3). We followed the suggestion of the reviewer and moved Fig. 8 to the results section (now Fig. 3).

As mentioned above the Discussion might be better focused on the comparisons between sensor configurations. To aid this the authors could consider adding a summary table that summarises the increased in precision and compares this the natural variability and provide some ‘cost-benefit’ analysis for investigators. Are there conditions where a 1 or 2 sensor system may give similar results to a 3 sensor system and where should researchers favour a 3 sensor system?

Response (P15L323). The summary table is a good idea and we added this table as suggested by the reviewer. In addition, we included the following paragraph with the table: “The deployments of the 3OEC demonstrate that the new instrument can improve the precision and reliability of benthic flux measurements. 3OEC fluxes in general were smaller, less variable and had smaller error margins than those produced by the conventional 2OEC eddy covariance instrument that was deployed next to the 3OEC. The advantages of the 3OEC may be most valuable in shallow energetic environments as reflected in the nighttime fluxes recorded by the 3OEC that differed significantly from those measured by the 2OEC. We believe that especially in dynamic settings, the improvements in flux determinations clearly outweigh the downsides associated with the slightly higher complexity of the 3OEC relative to conventional eddy covariance instruments with one or two solute sensors. As summarized in Table 1, the increases in setup time and costs are modest and may be justified by the improvement of quality and reliability of the flux data that can be achieved with the new instrument (Table 1)”