Response to reviewer comments "Improving estimates of ecosystem metabolism from dissolved oxygen time series", by M. W. Beck, M. C. Murrell, and J. D. Hagy III

The authors wish to thank the reviewers and associate editor for providing thoughtful comments on our manuscript. We have provided our response to each of these comments below, indicated in italics. Line numbers refer to the original manuscript. The reviewer comments have been shortened for brevity.

Reviewer 1:

I hope the authors can address this difficult question: does the model work better in diurnal tidal systems because the errors between the day and night conditions on the end-members cancel each other out? This is as opposed to the mixed tides where the errors can be heavily skewed towards respiration or production during the high high tide and thus lead to the large inaccuracies noted for Elkhorn Slough. This is an important point, because if the weighted regression approach is generating offsetting errors then it is not really working.

I think that the caveats described above have been acknowledged by the authors, such as starting on line 527, but I would recommend framing the issues associated with the end member problem in a very straightforward way.

We suggest that the model can work well in any tidal regime provided that the correlation between tidal changes and the solar cycle is low at the time scale of interest. As we acknowledged, this condition was often not met in the case of Elkhorn Slough, leading to significant problems in applying the open water method with a single DO time series regardless of the method used. Critically, it is possible to know by analyzing the correlation of tides and the solar cycle that our method would not be a good choice before applying it to Elkhorn Slough. The 2014 paper by Nidzieko et al. illustrated the complex and interesting dynamics there and showed that they are better explored with a more detailed data set than our method is intended to require. It is useful to point out that simply ignoring advection is also not a good option in estuaries like Elkhorn Slough, unless perhaps you are only interested in the means at the annual or longer time scale.

Our method works by statistically separating DO variability associated with tide height from DO variability associated with the solar cycle. We think that it is unlikely that we would have obtained the results that we did by generating offsetting errors.

Although we mentioned these points previously (e.g., lines 431, 521) we added additional explanation based on the reviewers comments. Specifically, we added text to this section to emphasize how and why the model can effectively characterize and remove the tidal signal, emphasizing that the model can work well in a variety of tidal regimes.

'Tidal currents in estuaries can move water across a mosaic of habitat types that likely have different biological rates. For example, the magnitudes of production and respiration, and the net metabolism may differ across salinity gradients, between open water and

shallow water habitats, and inside and outside of tidal marshes and seagrass habitats. These complex estuarine seascapes are especially prevalent at sites like those in the NERRS network. The interaction between spatial differences in metabolic processes and tidal advection moving water masses among habitats contributes to tidal artifacts in open water metabolism calculations. Advection alone, such as in an open coastal environment might actually generate little distortion of the metabolic signal within the DO record.

The objective of the weighted regression method is the remove or reduce variations associated with tides, allowing for an improved estimate of the average metabolic rates between the "end members" that define the tidal excursion about the DO sensor. The method works because in most tidal regimes, tidal cycles pregress fully relative to the solar cycle within several weeks. Pregression occurs quickly with semidiurnal tides and slightly more slowly with diurnal tides. At time scales longer than this period, DO variations associated with metabolism can be separated statistically from DO variations associated with tides. Locally-weighted regression is necessary because both the average metabolic rates associated with the solar cycle and the horizontal DO gradients created by spatial differences in metabolism are variable in time. The temporal scale and the choice of window width can be based on the time required for pregression of tides relative to the solar cycle.

Problems may arise in applying the method in certain situations, apparently most common with mixed tides, where complete phase pregression occurs across a seasonal time scales. For example, Nidzieko et al. (2014) showed that the higher high tide, which floods high marsh habitats, can occur at night in Elkhorn Slough throughout summer causing both real and apparent changes in metabolism. Large errors in standard open water metabolism calculations result because tide is generally rising during afternoon and falling in the early hours of the day, amplifying the true metabolic signal. Because the tide and solar cycle are correlated, however, our method ascribes all variability to the tide, resulting in equally unsatisfactory and large underestimates of rates. Weighted regression could be used, but only at a very long time scale, where the benefits of using it are not as great. As Nidzieko et al. (2014) showed, these complex patterns are best quantified and understood with the benefit of a more detailed data set.'

Specific questions or corrections:

1) This may be a naive question - but is tidal height an accurate proxy for advection? It is intuitive that the higher the tidal amplitude the more advection that will occur over the period. However, it is also true that the fastest currents associated with tides occur in the mid-tide range, for example on a large ebb tide the highest currents (and presumably a significant part of the advection term) occur when the tide is actually near the mid-range in tide height. Equation 13 seems to contradict this fact, where the highest advection (DOadv) is associated with the highest tide measurement (H). This is in fact slack tide when very little advection is occurring.

We recognize that tide height is not the same measure as advection, and specifically, that advective currents can vary in both magnitude and phase relative to tide height. Equation

13 describes our method for computing a simulated component of the DO time series associated with advection. This equation describes the DO changes, not the magnitude of the tidal current as the reviewers question suggests. We decided that equation 13 actually provided excessive detail which confused more than it informed, and therefore we eliminated the equation from the text. The correct interpretation is that the DOadv "end members" occur at either high or low tide when the water mass and associated DO gradient is maximally displaced either "up-estuary" or "down-estuary." Our approach assumes that this occurs around high tide and low tide, respectively, such that DOadv is linearly related to H (i.e., our eq. 1). Because of this, our model is most applicable in the case of a standing wave, which is reasonable for the types of estuarine sites where this approach is most applicable.

Fortunately, our weighted regression method only requires a variable that is statistically correlated with DO changes associated with displacement of the water mass by advection. We need not predict either the actual current speeds or the length of the tidal excursion, which we describe in eq. 10 and 11 only to conceptualize the processes that we model empirically. We added additional text to make these points more clear, beginning on line 123. We also revised lines 204-213 in the methods to clarify differences between how we generated the simulated data and how what we expected in the actual time series.

Line 123: 'Tidal height is used as a proxy for advection because this measurement is widely available, easy to measure, and meets the model requirements, which is a variable that is correlated locally in time with DO variations attributable to advection. Importantly, tide height need not predict the actual magnitude of advection. This is important because, across estuarine locations, the relationship between tide height and tidal currents can be variable in both magnitude and phase, depending on local characteristics. The weighted regression model is well-suited for the situation in estuaries because it empirically fits the local relationship between tide height and DO, such that no a priori knowledge of the relationship between tide height and either the magnitude of advection or the horizontal gradient in DO is needed.'

Line 204: 'A tidal series was simulated based on the principal lunar semidiurnal (M2) tide with a period of 12.42 hours (Foreman and Henry 1989). The amplitude was set to 1 meter and centered at 4 meters. The tidal time series simulated DO changes with advection, DO_{adv} (eq. (7) and Fig. 1). Conceptually, this vector represents the rate of change in DO as a function of horizontal water movement from tidal advection such that:

$$\frac{\delta DO_{adv}}{\delta t} = \frac{\delta DO}{\delta x} \cdot \frac{\delta x}{\delta t} \qquad (10)$$

where the change in DO from advection over time t is the product of the change in DO over the change in tidal excursion x and the change in tidal excursion over time. The simulated tidal time series was used to create DO_{adv} as a simple sine wave as in eq. (8). This time series was centered at zero to simulate increasing or decreasing DO with tidal height change. The final time series for observed DO was the sum of biological DO and advection DO (eq. (4) and Fig. 1). For time series from the NERRS monitoring sites, changes in

DO associated with advection are modeled as linear function of tidal height as indicated by the parameter β_2 in eq.(1). We can express the tidal component of DO variations alone as

$$DO_{adv} = \beta_2 H \tag{11}$$

,

2) The symbol for uncertainty in equation (6) and (7) and in the text on page 11 is not the same as the symbol used in figure 1, and both are different than at the bottom of page 13.

We have verified all equations, text, figures, and tables show the same symbol, ε.

3) Line 247: should this be decreased process error?

No, we found that increasing process uncertainty actually improved the ability of the model to isolate the biological signal. Process error is serially correlated and we suspect that this correlation structure is characterized well by the model since time is used as an explanatory variable. This is a minor point because this effect was most pronounced when there was no advective component and observation error was high in the simulated DO time series. In other words, there is no reason to apply weighted regression to an actual time series where advection effects are minimal.

4) Wind speed can have a significant influence on the calculation of DO flux, and therefore should be acknowledged as another source of error not accounted for by using a constant ka value. This influence is in addition to the advection term and can be significant.

Yes, our implementation of the open-water method for estimating metabolism accounts for changes in DO from wind. This was not apparent in the previous text and we have added additional information.

Line 300: 'This coefficient accounts for the influence of meteorological effects, such as wind, on air-sea gas exchange at the surface.'

Additionally, we did not explicitly consider any effect of wind in the simulated time series because it is not relevant for the analysis where the primary objective was to evaluate advective effects. Wind effects could be considered a component of process uncertainty, which was previously included. The following was added for clarification.

Line 186: 'For example, wind events can affect air-sea gas exchange (Ziegler and Benner 1998; Caffrey et al. 2014) such that high wind may contribute to increased process uncertainty. Although this was not an explicit focus of the simulations, wind effects could be considered an implicit component of process uncertainty in addition to the effects of other unmeasured or latent variables that influence DO in a time-dependent manner.'

5) The different scales on Fig 6 and 7 for DO are confusing.

The figures now have the same scale for DO.

6) It would be beneficial to readers if the code for the analysis was distributed with the published paper.

We created a simple R package on GitHub for implementing weighted regression and estimating ecosystem metabolism. A link was added as supplementary information: https://github.com/fawda123/WtRegDO

Reviewer 2:

Do the authors plan to make their R code available? They mention several times how useful it could be to others (a point on which I certainly agree), but it wont be useful unless its available!

We have added a link in the supplementary information for an R package to implement weighted regression and ecosystem metabolism: https://github.com/fawda123/WtRegDO

General comments:

(1) As the authors acknowledge, the practice in physical oceanography of filtering high frequency DO time-series data (or time series of other properties) to isolate certain signals is far from new. As far as I can tell, the major advantage of the authors proposed method over the traditional time-series methods (i.e., Kalman or band-pass filtering, or various applications of the autocorrelation function) is that the authors use weighted regression to create a separate, distinct time series that represents the tidal signal at each station, then use that time series to detrend the observed data. This has the distinct advantage of targeting the tidal signal explicitly, rather than simply filtering out all signals that happen to fall within a certain range of periodicity. The latter approach is fine in coastal or offshore systems, where the magnitudes of the tides (and of production and respiration themselves) are often not so high. But, the new method offers significant advantages for the types of systems in the NERRS network, which are the focus of their study. The authors touch on this distinction briefly in lines 98-100, but I think they need to expand on the point and make it more prominent.

We addressed this issue as part of our response to Reviewer 1. We address this again here in brief. We agree that NERRS sites are characterized by a complex mosaic of habitat types. These habitat differences and associated differences in metabolism generate horizontal DO gradients in estuaries, which interact with tidal currents to complicate interpretation of metabolism from DO time series. As the reviewer notes, the weighted regression approach is ideal for such locations because it explicitly targets the tidal component using a dynamic fitting process. The locally weighted approach is important because the magnitude of DO changes associated with tides is likely to vary through time as a result of changes in metabolism. This variability would complicate application of traditional deconvulution methods. We added text to emphasis the advantage of using

weighted regression in estuarine environments like the NERRS sites.

Line 100: 'The method targets the tidal component as an explicit variable using dynamic model parameters that change through time. This approach makes it possible to estimate and reduce the effect of advection on the DO time series, even if this effect varies through time due to changes in metabolic rates. Local fitting provides an advantage relative to traditional deconvulution methods. The method is most useful in estuaries where tidal effects interact with horizontal differences in metabolism and associated DO concentrations. Although the method could be applied in open ocean environments, it may be less useful because horizontal DO gradients may be less in such systems, such that it could be sufficient to simply neglect advection as has commonly been done in the past.'

(2) I was generally confused, both in the authors application of their filtering method, and in the way they calculated their metabolic estimates, about the role of the third dimension (depth). Did the authors apply the same weight vectors and half-window widths to the data from each depth in a given system? Or, perhaps the authors chose the four systems they did for their case studies because the depth at those particular stations was shallow enough such that only one series of DO data was needed? If the latter case is true (i.e., data from only one depth used at each station), how would the authors propose applying their method in a deeper water column? Could it be?

Yes, we assumed DO was vertically homogenous. This is a reasonable assumption for the case studies since all NERRS sites are shallow, productive estuaries. The method is then of course limited to only shallow water systems, but we argue for broad appeal as shallow estuarine environments occur throughout the world in addition to the 28 reserve systems within NERRS. Additionally, extending the method to stratified systems could be possible although additional data would be needed and the implementation would be much more complex (i.e., accounting for pycnocline exchange rates, etc.). The following was added to clarify these points:

Line 259: 'NERRS is a network of 28 shallow, productive estuary reserves...'

Line 285: 'NERRS sites are typically shallow and vertically-mixed such that one water quality monitor is adequate to characterize DO for the entire water column.'

Line 445: 'The regression method and open-water technique may also have broad appeal for application in estuaries that are shallow and vertically-mixed, such as those within the NERRS. Extension to stratified systems may be possible with additional data and model development.'

Further:

(a) Unless I missed it, what was the method of integration by which the authors got from their volumetric estimates for each time step in equation (14) (specified in g O2 m-3 hr-1) to the depth- integrated (areal) estimates (in g O2 m-2 t-1) that appear in all of their

6

figures and tables?

Text was added for clarification on line 307: 'Finally, volumetric rates were converted to depth-integrated (areal) estimates (g O_2 m^{-2} d^{-1}) by multiplying by the mean water-column depth at the site. Site depth was estimated from the pressure sensor on the water quality sonde, with a half-meter added to account for placement of the sonde just above the bottom.'

(b) I am concerned (or perhaps just confused) about the method the authors chose for deriving daily estimates of P and R from the hourly fluxes calculated using equation (14). The description of the method is not clear. Did the authors (1) calculate the individual hourly fluxes, then group them by day or night, then average them, and then multiply each by the duration of the day or night, or (2) calculate the individual hourly fluxes, then divide them by day or night, then add them together within their respective groups? Both methods have been used in the DO time- series metabolism literature; the second may be better since it doesnt assume the rate of primary production is constant over the course of a given day.

We used an approach identical to Caffrey et al. 2014 (cited on lines 294, 298) to average DO flux in each day/night period, which is then multiplied by different time periods to estimate production and respiration. Fortunately, the two approaches as described would produce an identical result, which we confirmed using a simple spreadsheet comparison. For the purpose of the comment, we have revised the text to clarify the method:

Line 300: 'The diffusion-corrected DO flux estimates as hourly rates of DO change were first averaged during day and night periods for each 24 hour 'metabolic day' in the time series. The 'metabolic day' was considered the period between sunsets on two adjacent calendar days. Respiration was hourly DO flux during night hours and net production was hourly DO flux during day hours. Total respiration (Rt) rates were assumed constant during day and night such that daily rates were calculated as the average DO flux during night hours multiplied by 24. Daily gross production (Pg) was the average DO flux during day hours minus the average daily respiration, multiplied by total sunlight time. Net ecosystem metabolism was gross production (positive) plus total respiration (negative).'

(3) Notation and terminology:

(a) The use of Pg to denote gross primary production is very confusing, especially since the units of Pg are in g (and Pg also being the SI unit for petagrams). Perhaps something else would be more appropriate? In addition, what is the t in Rt as the notation for respiration? Total? If so, the authors should specify when they introduce their variable (unless they did, and I missed it?).

These are standard terms that have been used extensively in the past to describe gross production and total respiration (see Caffrey et al. 2014, references therein). The above revision in response to the previous comment defines the acronyms accordingly.

(b) The word aggregation is used throughout the text. I gather from my reading of the ms that the authors mean binning or averaging, over some time or spatial scale. Aggregation is not a particularly precise term, because it could mean many different things. Perhaps the authors could consider replacing this word in the ms with a word or phrase that describes what they mean, e.g. time averaging.

We have noted all instances in the text where 'aggregation' is used and revised for clarity, mostly replacing with 'time averaging' or greater detail as needed.

(4) Duplication of presentation. If space is an issue, it seems as though Figs. 3 and 4 present the same data as Tables 1 and 2. I know heatmaps are all the rage right now, but I honestly felt the specific correlation coefficients in the tables told me a lot more than the heatmaps.

We are willing to remove these figures at the editors request but we would prefer they be retained because they do not show identical information. However, we have removed the tables because we feel the information does not provide a meaningful description of the results. Please see our response to the reviewers comment on p. 14, line 237.

Specific comments:

Abstract generally: If the authors could work in a sentence on the significance of their method vis--vis traditional time-series filtering techniques (see my comments above), I think they could increase the impact of their work. Additionally, I would have liked to see at least a few numbers or figures in the abstract (e.g., by what % on average did the method reduce variance in the NEM estimate relative to those calculated from the unfiltered data, or how much did it reduce RMSE?).

The following was added to the abstract: 'The method targets the tidal component using a flexible association while preserving the true biological signal, offering an advantage over traditional deconvulution methods.'

We have decided not to include results in the manuscript. We feel that presenting simple summary numbers without adequate explanation would be more misleading than informative since the model did not work in all cases. The abstract requirements for L&O Methods is word-limited and we feel that we cannot adequately present this information within the allotted space.

p. 3, line 12: integrated by what dimension and over what interval? Depth? Time?

This term was removed as the details are now available in the text.

p. 3, line 22: useful is nonspecific and does the authors work a disservice; can they think of something more specific and impactful?

8

See our response to the first specific comment. We have also revised some of the text for clarity: 'Variability resulting from advection was reduced in all cases, greatly reducing the correlation of DO and metabolism estimates with tides. The model was especially effective when the magnitude of tidal influence was high and correlations between tidal change and the solar cycle were low at the time scales of interest. When tides and the solar cycle were correlated for protracted periods, the model was less useful.'

p. 3, lines 23-24: See my comment above regarding the methods performance when the solar cycle and period of tidal height change were highly collinear. The method yielded estimates at the Elkhorn and Padilla sites which were significantly different from those obtained with the unfiltered (traditional) data. By the authors own admission, neither the filtered nor unfiltered data probably accurately represent what is truly happening at those sites. A sentence specifying those situations in which the method did not perform particularly well would be more honest, rather than simply saying when it generally did work well.

The sentence was modified as follows: 'The model was especially effective when the magnitude of tidal influence was high and correlations between tidal change and the solar cycle were low, whereas collinearity in the latter instance limited model performance.'

p. 3, line 26: timescales

Changed.

p. 4, lines 32-33: Reference to Gaarder and Gran (1927) missing here.

Citation added.

p. 4, lines 42-43: This simply isnt true. There are dozens of reasons bottle incubations are more effective in certain situations, depending on the research objective. Open-water measurements may be more effective for evaluating certain kinds of events or trends, but if the authors are going to make this claim, they should specify what they mean.

The sentence was revised: 'Open-water estimates also provide a practical means of capturing events and ecosystem changes over time as compared to discrete sampling events with bottle-based approaches because monitors are present continuously. While bottle-based measurements offer advantages in precision, it is not possible to sustain bottle measurements day after day, making it unlikely that sampling will occur during an unanticipated short-term event. Moreover, more variability is captured in continuous data and statistical power for resolving long-term changes is greater compared to periodic measurements.'

p. 4, line 46: Reliable estimates

Changed.

p. 5, line 67: advection, leading to

Comma added.

p. 6, line 81: example illustrates

Changed.

p. 7, lines 98-100: See my comments above about the methods significance vis--vis traditional detrending methods. An expanded discussion of the difference here is warranted.

See the response to the first general comment, i.e., the additions to the final paragraph of the introduction.

p. 7, lines 101-102: evaluate the ability

Changed.

pp. 7-8: Appropriate place for clarifying role of depth, or why it is being ignored (my comment above).

Yes, see the response to the second general comment. Also, the response to the next comment addresses the role of tidal height/depth vs. advection as a model variable.

p. 8, lines 121-122: I buy the idea, but is there some justification in the literature for using tidal height as a proxy for advection?

See our response to reviewer 1 about the use of tidal height as a proxy for advection, i.e., the expansion of the paragraph on page 8... 'An important distinction between tidal height and advection...'

p. 9, lines 139-140: I was generally confused about why different time vectors were used to represent days and hours (or is it hour of day) separately. Maybe just some clarification of terminology would help.

The weighted regression model evaluates weights associated with distance from the center of the window in days as well as hours within the day. For example, if we care about modelling oxygen at 3 pm, we would weight all other observations close to 3pm in other days with high importance, with further diminishing importance as window distance increases in days from the center. The weights widget in the supplementary information can be used to view the influence of both, in addition to the tidal weights. The citation at the end of the paragraph for the multimedia section should help interpretation. We also revised the line... 'Windows for time (continuous throughout the day) and hour (within each day) are used...'

p. 12, line 196: What is the justification for using this particular range of O2

concentrations to generate the simulated data set?

The values were chosen based on a general approximation of the observed data from the case studies, text was added... 'as a rough approximation from the case studies below'

p. 12, line 204: lunar

Changed.

p. 13 ff: Equations appear to employ the wrong del for the partial derivative, i.e. versus

The symbols were changed to show the correct partial derivative.

p. 14, line 237: The authors are characterizing a corr. coefficient of 0.63 as similar? Or did I miss something? On what specific basis in Tables 1 and 2 is the claim of similarity made?

Considering the reviewers response, we realized that tables 1 and 2 are not helpful because the distributional statistics depended on the simulations that we chose to run and analyze and dont indicate what somebody may expect to get in an analysis of data from the real world. In our revision, we described and explained the patterns in Fig 3/4 and their implications for our method. We think that this is far more informative and thank the reviewer for this question.

Line 234: 'The filtered DO time series were compared to the simulated data to evaluate the ability of weighted regression to quantify and remove the variability due to advection (DOadv), leaving the biological response (DObio) (eq. 4); Fig 3). Comparisons were made using Pearson correlation coefficients and the root mean square error (RMSE). The ability of weighted regression to isolate DObio, indicated by high correlation and low RMSE, was not reduced by the introduction of advection into the simulated time series (i.e., DOobs) because weighted regression was able to quantify and remove this variability. On the other hand, introduction of observation uncertainty (obs) decreased the ability of the model to quantify and remove advection. Perhaps counterintuitively, increased magnitude of advection improved the model performance because a larger advection signal can be fitted empirically despite the presence of random error. When the advection signal was small, it became difficult or impossible to quantify unless there was neither process uncertainty nor observation error. This suggests that if the advection signal is in fact very small, it would be preferable not to quantify it and instead neglect it, as has been done previously.'

p. 17, lines 299-300: Related to my comment about depth, above. To get from a piston velocity to a volumetric reaeration coefficient (isnt this just a depth-corrected piston velocity - why the need for even more jargon?), one must invoke some mixed layer depth. What was the depth (i.e., H in Thbault et als equation A11) used in each case? The Thbault paper seems to contain a not-so-unique method of calculating gas transfer velocity (and air-sea flux), gussied up with new terminology for some of the parameters.

We chose not to change our terminology because we want it to remain identical to that of Caffrey et al. 2014 and references therein, on which our work builds. Since the focus of this paper is on improving the DO time series to remove advection, we dont want to make other changes to the open water method and associated terminology in this paper. In our response to comments elsewhere, we have clarified how we have addressed translation of volumetric rates to areal rates, and in general our treatment of the depth dimension.

p. 17, lines 300-307: The precise method of obtaining the daily ecosystem metabolism estimates was not clear. (See my general comment 2, above.)

See our response to the second general comment.

p. 18, lines 319-321: I think this assumption is valid, but can the authors provide some justification or precedent? One citation would be sufficient.

We too think that this is intuitively true and agree that adding some justification would improve the paper. Since nobody has attempted to remove advection effects from a DO time series for the purpose of open water metabolism calculations before, we dont think that there is a paper that has examined this particular question. To address it here, we utilized our simulations to (a) generate a simulated year-long DO time series with advection artifacts and (b) generate a year-long DO time series with no advection. We then calculated open water metabolism using both time series and calculated a cumulative sum on the differences. The exact period of time required for the means to converge with and without artifacts likely depends on the local regime, particularly the time required for full pregression of tide phase relative to solar cycle. We added a statement to the manuscript to address that we examined this assumption using our simulated data.

'We assumed that annual mean values would not change after filtering because long-term metabolic averages are not likely to be biased from advection. We evaluated this assumption by computing metabolism for a 12 month simulated DO time series with and without advection. These simulations showed that the means converge within the annual time scale, although variance of metabolism was substantially larger for the time series with advection.'

p. 19, lines 350-351: I disagree. These percentages from the filtered data do represent a decrease in the number of anomalous values by a factor of 3, but they are not near zero. In one case, 7.12% of values were still anomalous.

Sentence was revised: 'Anomalous values were substantially reduced for all case studies and were virutally eliminated in two of four of the case studies.'

p. 23, line 427: Particularly confusing use of the word aggregated. Aggregated or averaged by what, and over what time or spatial interval?

Sentence was revised: 'Results for Sapelo Island suggested that time averaged estimates,

either as monthly or annual means, were comparatively unchanged by filtering...'

Discussion generally: The discussion read less like a discussion than a simple summary of the rest of the paper. Perhaps a separate discussion isnt really necessary?

The author guidelines for L&O methods indicate that the discussion should describe the degree to which the method meets defined needs, potential for new insight, comparison with alternative approaches, and new questions that may have been raised (http://onlinelibrary.wiley.com/journal/10.1002/(ISSN)1541-5856/homepage/ForAuthors.html#6). We argue that the current discussion adequately covers these points so we have not removed it from the manuscript.

p. 23, lines 440-441: remove some variation; the method doesnt remove all variation

Changed, though we have used 'a significant amount' instead of 'some'.

p. 26, lines 500-503: This seemed like it (or some similar text) belonged in the methods section, to describe why the sites were selected. In addition, the grammar of this sentence was confusing.

Sentence was revised and moved to line 269 in the assessment section: 'Many daily metabolism estimates were also not interpretable (i.e., 'anomalous') from these sites using standard open water methods.'

p. 27, lines 527-535: This paragraph belongs in the discussion section. Perhaps after the paragraph ending on line 479.

Moved.

p. 28: Will code be available publicly somewhere?

Yes, see the comment above and in response to reviewer 1.

Comments on figures and tables:

Figure 1: One of my few questions about the simulation datasets centers on the way biology was parameterized. I appreciate that modulation in the DO signal due to biological processes was modeled as a purely sinusoidal function for purposes of evaluating the method, but even the purest biological signal will not be sinusoidal. Major differences are (1) that the peak of the biological curve will usually lag each days peak insolation by some number of hours (evident in the authors own data in Fig. 6) and that (2) there is a point of photoinhibition, where the perfect shape of the curve is dented by a decrease in photosynthetic activity before it begins its actual decline. Ive not put either of these very eloquently, but I assume the authors know what Im speaking of. Some brief acknowledgement in the methods section about the representation of the biological signal

13

as a perfect curve like this would be instructive.

We are aware of the dynamics that the reviewer mentions. In thinking about it, we realized that the analysis should work effectively for any phase lag of DOdie relative to PAR since the day/night pattern is merely there for reference. What is critical is that DOdie has a period of 24 hours while the tide has a period that is longer than 12 or 24 hours, causing the phase of the tide to fall increasingly behind the solar cycle, until after a sufficient number of days, the cycles are not correlated.

Text was added in the methods to clarity this point (line 196): 'Although the representation of DO_{die} as a simple sine wave ignores more complex dynamics associated with the photosynthesis-irradiance relationship as applied to the ecosystem, it is sufficient to address the analytical objectives of the simulations.'

Figure 6: What was the threshold PAR intensity used for defining the PAR shading? What are the units of PAR? Millmoles of what? Photons? Microeinsteins (E)? Or W m-2?

The shading indicates day or night based on astronomical sunrise/set times for the location and date, rather than the PAR data. This was clarified in the captions. Units for PAR were also added ($\mu E\ m^{-2}\ s^{-1}$).

Figure 8: One effect of filtering with the authors method is that it appears to dramatically attenuate the interday variation in metabolic parameters evident in unfiltered data. Remember that the discovery of this temporal heterogeneity has been one of most widely celebrated features to emerge from analysis of high-frequency time series (of, e.g., D.O.). What is the implication of the authors method for the existence (or not) of this heterogeneity? Does this mean there really isnt as much temporal heterogeneity in ecosystem respiration and in photosynthesis as this method has suggested? Some discussion would be instructive (perhaps replacing some of the repetitive text in the current discussion); the authors wont have to look far into the literature to find many studies that have remarked on the apparent heterogeneity in unfiltered data.

We agree that this result is intriguing. The short answer, in our view, is that the day-to-day variability is not real, but rather an artifact of advection. For example, our results showed that prior to removing variability due to advection from a time series at Sapelo Island (Fig 8), the interday variation in metabolism was as large as the seasonal variation (see the multimedia widget for Sapelo Island https://beckmw.shinyapps.io/detiding cases/). Based on our understanding of water quality and habitat at Sapelo Island, we cant think of a mechanistic reason for such large interday variations. Conversely, variations in nutrient inputs, phytoplankton biomass, temperature, etc., should generate considerable differences in metabolism rates at longer time scales. In our own study (paper in prep) in Pensacola Bay, we conducted weekly metabolism measurements using BOD bottles at the same location as we had a DO sensor. These results also showed that the bottle based measurements were less variable an observation that was one motivation for this work. We would stop short of saying that there is not day to day variability that weekly or monthly measurements dont

capture. Potentially, after removing advection, one may still see variations that may be related to likely causes, leading to increased understanding of factors that affect metabolic rates in estuaries. We added text in the discussion to address these issues (line 438).

'The application to simulated DO time series with known characteristics and extension to continuous monitoring data from selected NERRS sites suggested the approach can isolate and remove a significant amount of variation in observed DO from tidal change. In application to DO time series from NERRS sites, the method reduces apparent interday variation in metabolic rates. This may indicate that such interday variations result substantially from artifacts of advection, rather than real drivers of metabolic processes. This hypothesis could be investigated further by applying our proposed method while collecting measurements of metabolism and causal factors at a relatively high frequency. The regression method and open-water technique may have broad appeal for application...'

Table 3: Were the metabolic rates given here those calculated using the authors method?

These were based on our methods, without filtering. Text was added to the footnote for clarity: '...estimated using methods described herein without removing advection effects.'