

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 agree this is an important issue that deserves additional discussion. As noted, we have partially addressed these concerns on line 527 but have not explicitly described the “end member” problem. We have added text to this section that elaborates these points.

‘A useful approach for conceptualizing instances when weighted regression may be most appropriate can be described using an ‘end member’ paradigm. These end members can be considered the terminal points of the water mass that is horizontally advected past the DO sensor. One end member may be characterized by tidal marshes during high tide, whereas the second end member may be characterized by oxygenated oceanic waters during ebb tide. An important distinction is made between end members that are influenced by actual biological processes that occur at each terminus and those that are not, where the former case is more likely a common occurrence at locations with large tidal amplitudes. In the latter case, variation in dissolved oxygen at the DO sensor is solely related to physical advection and weighted regression should work as intended provided collinearity with the solar cycle is at a minimum. For other sites where biological processes affect the DO at the “end members”, the tidal characteristics of a site may be the primary factor that affects the ability of weighted regression to isolate the true biological signal. Diurnal or semidiurnal tides that

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 (DO_{adv}) is associated with the highest tide measurement (H). This is in fact slack tide when very little advection is occurring.

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.

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

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 k_a value. This influence is in addition to the advection term and can be significant.

5) The different scales on Fig 6 and 7 for DO are confusing. 6) It would be beneficial to readers if the code for the analysis was distributed with the published paper.

Reviewer 2:

Review of Improving estimates of ecosystem metabolism by reducing effects of tidal advection on dissolved oxygen time series by Beck et al., submitted to L&O Methods, 2015

Summary: Time-series dissolved oxygen data from remote monitoring stations and observing arrays has been applied for several decades now to estimate rates of ecosystem metabolism in a variety of marine and aquatic systems. In estuarine or coastal systems subject to significant physical forcings such as tidal advection which operates over the same daily timescale as the target biological processes the deconvolution of simultaneously encoded physical and biological signals remains a critical obstacle. Signal processing techniques (e.g., Kalman or band-pass filtering, or various applications of the autocorrelation function) can be and have been used to accomplish this deconvolution, but these techniques exclude signals in data based strictly on their periodicity. Thus, the risk in these existing methods of throwing the baby out with the bathwater i.e., discarding desired biological signals along with the nuisance physical signals is very high. Beck et al. instead apply a weighted regression method to filter out tidal signals in time-series data by targeting those tidal trends specifically. The method appears generally successful at (1) reducing unwanted artifacts from tidal advection and (2) reducing overall uncertainties in ecosystem metabolic estimates.

General impression and recommendation: This is fine work that deserves publication in L&O Methods, and has the potential to bring significant change to a sub-discipline that has grown rapidly in the past 20 years with the advent of high-frequency DO time-series data. Despite small advances here and there in data analysis techniques, many of these recent estimates of ecosystem metabolism have essentially rested on minimal updates to the original methods of Gaarder and Gran (1927) or Odum (1956). The method the

authors propose here is (in the environmental sciences, at least) new and innovative. As such, the work represents a significant advance in the state of the art. I particularly liked the authors method of developing and testing their method with simulated datasets, and their sensitivity analyses are robust.

However, I do have some questions and concerns of moderate significance; these revolve less around the authors method for filtering out tidal signals, which I found very well-explained, than around how they then used their method to calculate estimates ecosystem metabolism. In addition, the authors find using their own metrics that the method does not perform well at all in situations (or ecosystems) where the solar cycle and period of tidal height change are highly collinear, or where observational uncertainties are high and the tidal advection is weak. These caveats are certainly addressed in the manuscript, but I am concerned the magnitude of this sometimes poor performance is underrepresented in the abstract (fixable with a few edits).

The manuscript is generally well-written and the figures are generally clear, though some captions are missing information and I recommend the authors change some of their notation to avoid confusion (see specific comments below).

Overall, I very much liked the paper, and I am excited to see when and if the cottage industry that makes these measurements incorporates this elegant new method. The multimedia feature was a very useful addition to the paper, and helped me significantly in understanding the effect of varying various parameters in the authors method on the result. However, 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!

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.

(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). The authors adapted the Hirsch et al. WRTDS method for a laminar or channel flow case (rivers and streams) for use in estuaries, where the regime is certainly not channelized or laminar. How was the method used to filter data from multiple depths simultaneously in the four case study systems? 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 this was the case, are the authors making the assumption that the water column in which the series of DO measurements was made at each of these stations was vertically homogeneous with respect to the processes and properties of interest over the entire period of measurement? Do the authors assume that the parcels of water flowing by the sensor in each case were vertically homogenous with respect to DO and also with respect to the strength of the biological processes (photosynthesis, respiration) that determine DO, even as the tide changed? I can envision reasonable justifications for each of these, but the authors must address them. 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? The method has a serious limitation if it cant be used for a water column deeper than a few meters.

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 $\text{g O}_2 \text{ m}^{-3} \text{ hr}^{-1}$) to the depth- integrated (areal) estimates (in $\text{g O}_2 \text{ m}^{-2} \text{ t}^{-1}$) that appear in all of their figures and tables?

(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.

(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?).

(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.

(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.

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?).

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

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

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.

p. 3, line 26: timescales

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

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.

p. 4, line 46: Reliable estimates

p. 5, line 67: advection, leading to

p. 6, line 81: example illustrates

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.

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

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

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?

p. 9, lines 139-140 ff: 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.

p. 12, line 196: What is the justification for using this particular range of O₂ concentrations to generate the simulated data set?

p. 12, line 204: lunar

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

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?

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 Thibault et al's equation A11) used in each case? The Thibault 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.

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

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.

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.

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

Discussion generally: The discussion read less like a discussion than a simple summary of the rest of the paper. Perhaps a separate discussion isn't really necessary?

p. 23, lines 440-441: remove some variation; the method doesn't remove all variation

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.

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

p. 28: Will code be available publicly somewhere?

Comments on figures and tables:

Figure 1 ff: 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 day's 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. I've not put either of these very eloquently, but I assume the authors know what I'm speaking of. Some brief acknowledgement in the methods section about the representation of the biological signal as a perfect curve like this would be instructive.

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 $W\ m^{-2}$?

Figure 8 ff: 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 isn't 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.

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

References cited in review:

Gaarder, T. and H. H. Gran. 1927. Investigations of the production of plankton in the Oslo Fjord. Rapports et Procs-Verbaux des Runions du Conseil Permanent International pour l'Exploration de la Mer 42: 1-48.

Odum, H. T. 1956. Primary production in flowing waters. Limnology and Oceanography 1(2): 102-117