

*Response to review from Dr. Erik Smith, “Improving estimates of ecosystem metabolism from dissolved oxygen time series”, by M. W. Beck, M. C. Murrell, and J. D. Hagy III*

Line by line response to reviewer comments are provided in italics. Line numbers refer to the original manuscript.

This manuscript presents a technique that attempts to remove the influence of advective oxygen transport on the estimation of net ecosystem metabolism rates computed from time series of dissolved oxygen (DO). There are a growing number of in situ data sondes (equipped with the new and robust optical DO sensors) being continuously deployed in coastal waters throughout the world. The use of these time-series to quantify and investigate the magnitude and variability in ecosystem metabolic rates continues to be an active area of research, having relevance to both basic and applied science questions. Of course a major challenge for the approach in coastal and estuarine environments is that, unlike lakes, tidal advection can greatly influence DO dynamics. As such, there is a great need for tools that can help improve estimates of net ecosystem metabolism from DO time series that include DO variability associated with tidal advection. The weighted-regression technique presented in this manuscript clearly represents a significant step forward in addressing this need this need and should make a significant contribution to the field.

While the approach presented does not entirely solve this challenge in all estuarine time series, I greatly appreciated that the manuscript clearly described how to determine from the method whether specific DO time-series are simply not amenable to net ecosystem metabolism computations (even with the approach presented here). That too is extremely valuable.

While I had to read this manuscript a few times to really feel comfortable with it, I must say that I found it relatively easy to follow despite being very computationally heavy. Of course, I still would not have the ability to do any of the computations needed to apply the technique on my own after reading this manuscript, but the manuscript certainly succeeded in conveying the need for the technique and what it can/cant do if properly applied.

A couple of issues/questions I still have after digesting the manuscript:

1. From discrete measurements in the North Inlet estuary, metabolic rates, especially respiration, tend to be significantly higher on ebb tides, compared to flood tides (due to ebb waters being enriched in substrates, compared to flood waters). I am sure this holds for many estuaries, as well. I still cannot figure out if the technique presented allows for biological rates to vary over the course of tidal cycles or how this affects choice of averaging windows and resulting tidallycorrected DO signals. That is, if the time series includes a DO component produced by biologically-driven DO change that vary as a function of tide (on top of the changes that vary as a function of the PAR cycle) in addition to the physically driven tidal changes in DO, what is there an effect on the filtered time series and resulting NEM calculations?

*This is a complicated issue, particularly for a model that is basically a multiple linear regression at its most basic level. My guess is that any metabolic activity that is correlated with the tide is diminished using the weighted regression, whether it's simply a result of moving water masses or as a result of real metabolic changes as in the case of ebb waters being enriched in substrates. The two cases would be statistically indistinguishable and considered a tidal effect by the model, and therefore would be removed. The relative contributions of each are not discriminated between using our model. I suppose a larger discussion that should occur before applying the model is the extent to which the DO signal is influenced by both. Locations where tidal influences are large and the water is relatively deep (as in the Vierra Mouth station at Elkhorn Slough) are clear examples where physical movement is likely the dominant source of variation in the DO signal. The method has real value in such instances. Application to other stations where this is less clear should be carefully evaluated, particularly with respect to changing window widths on the results. The following was added to the comments/recommendations section.*

*'Finally, DO time series that include components of metabolism related to diurnal cycling as well as true biological effects related to the tide should be evaluated with caution. Specifically, increased respiration may be observed during ebb tides as a result of increased enrichment of substrates (Sasaki et al. 2009). In such cases, variation in the metabolic signal is correlated and directly related to the tide but is not related to physical movement of water masses. The extent to which the model discriminates between tidally-induced variation in true metabolism as compared to water movement is not clear. Accordingly, the model should not be used without critical evaluation of site-level characteristics that suggest physical movement of water masses is a primary source of noise in a DO signal. Locations where tidal influences are large and the water is relatively deep (as in the Vierra Mouth station at Elkhorn Slough) are clear examples where physical movement is likely the dominant source of variation in the DO signal. Application to stations where this is less clear should be carefully evaluated, particularly with respect to changing window widths on the results.'*

2. I am not sure I follow the reasoning behind assuming that reduced variation of metabolism estimates after the filtering procedure provides greater confidence of the estimates (line 428 and elsewhere). Substantial short-term variability in Pg and R are well documented from discrete measurements and perhaps this variability is just being filtered out by the procedure.

*This is a valid concern that we discussed at length during the development of this technique. This is a potential issue when the solar cycle is correlated with tidal height changes such that the two are quantitatively indistinguishable. Figure 9 provides a visualization of these occurrences and when the method may not work as expected. Removing variation in the DO time series related to biology is possible when the two are correlated. When the two are not correlated, it is statistically unlikely to remove variation attributed to true metabolism. In both cases, a reduction in standard error would be expected and this may not be the most appropriate metric of performance. This is why we relied on four different performance measures. We feel that the comments and recommendations section adequately addresses*

these concerns. The following was also added.

*Line 346: ‘However, reduction in standard error may also occur if variation from biological processes is reduced separately or in addition to physical variation. The extent to which this reduction in variation is related to the former should be minimal, provided that the two are statistically distinguishable. Situations when the phasing of the tidal and solar cycles are correlated could be instances when the two are unable to be separated by the model. Reductions in mean annual values (particularly at Elkhorn Slough) in addition to standard errors suggests that biological variation may have been removed.’*

3. I had a bit of trouble following the weight-of-evidence approach for selection of halfwindow widths described starting on line 301. Not sure I fully understand / was able to follow the rationale for the performance metrics. Perhaps just a bit more detail on the thinking behind these metrics and how they specifically inform/influence the choices made?

*The following was added to the paragraph to more clearly describe selection of half-window widths using the performance metrics.*

*‘The selection of half-window widths for filtering the DO time series was based on an evaluation of results using four performance metrics. In other words, the ‘optimal’ window widths for each case study were those that provided the greatest measure of performance based on all four metrics. This is considered a weight-of-evidence approach given that individual metrics are not absolute indicators of model performance. First, the regression results were evaluated using correlations of DO and metabolism estimates with tidal height before and after application of the model, such that window widths that provided maximum reduction in correlation relative to the observed data were desirable [...]. The second and third performance metrics evaluated changes in the mean and standard error of annual metabolism estimates before and after filtering. [...] Optimum window widths in this context were considered those that maintained the mean values while reducing standard error relative to the observed data. Finally, results were evaluated based on the occurrence of ‘anomalous’ daily production or respiration estimates [...] Optimum window widths for this metric were those that provided the maximum reduction in anomalous values.’*

Some minor specific comments/suggestions:

The title is a bit vague and does not really speak to what is novel about the work. Might think about including specific mention of correcting for tidal advection.

*The title was modified: Improving estimates of ecosystem metabolism by reducing effects of tidal advection in dissolved oxygen time series*

Much of the introductory paragraph does not seem particularly germane to the focus of the paper. I might suggest what is needed to start the Introduction is an argument for why estimates of Pg, Rt and NEM by the open-water technique are advantageous despite caveats and uncertainties of the method (e.g., high resolution sampling captures events

missed by traditional grab sampling methods; bottle measurements are labor intensive and have their own caveats; NEM captures entire ecosystem processes rather than components; etc.)

*The introductory paragraph was rewritten to emphasize the above points.*

*‘Ecosystem metabolism describes the balance between production and respiration processes that create and consume organic matter (Kemp and Testa 2012, Needoba et al. 2012). Light exposure experiments of water samples collected at discrete locations and times have traditionally been used to measure metabolic activity. Although highly controlled and precise, bottle-based methods are labor-intensive and not scalable to describe entire ecosystem rates. Bottle-based methods may also only reliably estimate production and respiration associated with planktonic processes, whereas significant contributions of ecosystem production can arise from other habitats, such as benthic seagrass communities. By contrast, open-water techniques have been increasingly used to estimate whole system metabolism given the availability of long-term, continuous time series of dissolved oxygen (Odum 1956, D’Avanzo et al. 1996). Daily integrated measurements of metabolism represent the balance between daytime production and nighttime respiration attributed to all ecosystem components. Open-water estimates also provide a basis for tracking ecosystem change over time and are more practical for capturing events or evaluating trends as compared to bottle-based approaches. Although metabolic rates vary naturally at different spatial and temporal scales (Ziegler and Benner 1998, Caffrey 2004, Russell and Montagna 2007), anthropogenic nutrient sources are often contributing factors that increase rates of production (Nixon 1995, NRC 2000). Reliable estimates of whole ecosystem metabolism are critical for measuring both background rates of production and potential impacts of human activities on ecosystem condition.’*

Line 49: I would also cite Staehr et al. 2010 paper along with the Kemp and Testa 2012 reference, because it is from the same journal this manuscript will be submitted to, it is a really nice review of the concepts and method details (even if specific to lakes), and it is much more accessible than the book that has the Kemp and Testa reference.

*Citation was added.*

Lines 61-72: I think that the method has been applied rather successfully to lakes for many years, mainly because advection is not a significant problem in lakes. Where the method has problems in lakes is when vertical stratification/stability is an issue. In contrast the greatest difficulty in applying the method to coastal waters is the problem of advection, such that the method has relatively rarely been applied in estuaries (with Caffrey being the exception).

*Lines 61-64 were revised to reflect these points.*

*‘Given this critical challenge, the open-water method has been used with varying success in estuaries influenced by tidal mixing (Caffrey 2004, Russell and Montagna 2007, Caffrey et*

*al. 2014). In contrast, the method has been more successfully applied to water quality time series in lakes, although stratification may limit estimates to specific vertical layers (Staeher et al. 2010, Coloso et al. 2011, Batt and Carpenter 2012).'*

Lines 142-144: These two sentences need verbs.

*Added verbs to these sentences:*

*'Windows for time and hour are used to weight observations based on distance (time) from the center of the window. The window for tidal height is used to weight observations based on the difference from the center as a proportion of the total tidal height range.'*

Line 298: Dont you mean that respiration at night was assumed to be equal to respiration during the day, rather than respiration was assumed constant during the night

*Sentence was revised:*

*'Respiration at night was assumed to be equal to respiration during the day and was subtracted from daily net production estimates to yield gross production'*

Line 313: gross production rather than production just to be explicit.

*Text was changed accordingly.*