*We thank all three reviewers for providing constructive comments that have helped to improve the manuscript. Our responses to each of these comments are provided below. Note that line number refer to those in the original manuscript.*

### Reviewer 1

This manuscript describes an advancement in the computation of aquatic estuarine metabolism (i.e., primary production and reparation) using open-water dissolved oxygen (DO) time series. The authors adapted a Bayesian model originally developed for non-tidal, lotic systems (i.e., BASE) to 1) now work with time series that have missing dissolved oxygen observations and 2) can accommodate longer time periods for metabolism estimates. The latter could produce non- real estimates of metabolism in earlier methods while the former is achieved through linear interpolation during pre-processing and estimate removal by a user defined threshold during post-processing. The authors describe in sufficient detail how to operationalize the model and provide a comparison with a historically more standard arithmetic-based model that has been used in earlier studies of estuarine metabolism. Furthermore, the authors provide an R package “EBASE” to allow others to estimate estuarine metabolism using this method advancement; a package I am sure the coastal community will find useful. Overall, this paper was well written and the advancement significant. This paper should be considered for publication in L&O Methods following incorporation of the comments below.

*Response: text goes here*

There were two studies cited in the introduction (Line 98) as having used Bayesian techniques to estimate aquatic metabolism in estuaries. Please elaborate on how the method described in this study is consistent with and/or different from those in the cited studies.

*Response: text goes here*

The authors describe pitfalls in their model, particularly when it comes to estimating the gas transfer velocity. Additionally, the authors provide evidence that sometimes the right answer (i.e., modeled DO time series, lines 349-352) can be achieved through incorrect processes. Similar issues have been described and addressed in the stream metabolism literature which was not cited or discussed in this study, please see Appling et al. 2018 - https://doi.org/10.1002/2017JG004140). Consider adding to the discussion how EBASE might incorporate these advancements to address these issues.

*Response: text goes here*

Detiding (i.e., removing the influence of tidal cycles on time series data) the DO time series substantially improved model estimates and therefore was recommended as a step prior to applying the model. However, it was not clear how the authors detided their time series (they used Weighted Regression but it is unclear how this method removes tidal influence from the predicted time series – see lines 400-402). Please add additional information on the detiding approach and which variables were detided (i.e., was only DO detided? Wouldn’t all water-related input variables, such as water temperature and salinity, need to be detided?).

*Response: text goes here*

How would your results be impacted by using different observed data collection frequencies such as every 5-min, 15-min, 1 hour?

*Response: text goes here*

In my use of Bayesian models for estimating metabolism, with each model run you can get different metabolism estimates. This is due to the prior distribution space that the model can operate in. Is this something you have noticed with your model? If so, could you add a sentence somewhere explaining that a user should expect different modeled estimates with each model run?

*Response: text goes here*

Line 32-36: Estuaries are also significantly influenced by their freshwater end-members, consider adding ‘river’ to ‘…interface between land, river, ocean, and atmosphere’. Similarly consider incorporating rivers into ‘…transported from land to rivers to ocean…’.

*Response: Added.*

Line 47: Consider adding how materials are lost to the atmosphere, the sediments, and the ocean. For example, “…that are lost to the atmosphere (i.e., outgassing), the sediments (i.e., burial), and the ocean (i.e., advection).”.

*Response: Added.*

Line 48-50: Consider adding the underwater eddy covariance method to the list of methods for estimating aquatic metabolism. See Berg et al. 2022 (https://doi.org/10.1146/annurev-marine-042121-012329)

*Response: Added to the list, including the citation, “…use of inert gases, and aquatic eddy covariance (Kemp and Testa 2011; Staehr et al. 2011; Berg et al. 2022).*

Line 80: As written, it seems ‘assumptions’ should be singular, not plural (…can violate an assumption of the method…).

*Response: Corrected.*

Line 82: I appreciate that the sign of P and R estimates in the literature has been applied non- uniformly. However, I believe the convention is that P is positive (DO is produced) while R is negative (DO is consumed). Could you change the physically impossible statement to state negative P and positive R? Or add a sentence about why both values are described with a positive sign – this is somewhat described on line 236.

*Response: Yes, we agree the sign has not been consistently described in the literature. We have added clarifying text at the end of the paragraph: “Note that R is often expressed as a negative value indicating consumption of dissolved oxygen. Here and throughout, we express correct values of R as positive values given the metabolic model described below.”*

Line 96: Add the word statistical between ‘Bayesian’ and ‘techniques’ (i.e., Bayesian statistical techniques…)

*Response: Added.*

Line 98: These two cited studies – Ciavatta et al. 2008; Tassone and Bukaveckas 2019 - have used Bayesian approaches to estimate estuarine metabolism; what makes them limited? How is the approach described in this manuscript different from these two earlier studies? Have those earlier two studies informed the model selection/development described in this manuscript?

*Response: text goes here*

Line 106: Curious why your method builds off BASE when there has been an updated Bayesian technique for stream metabolism that more accurately estimates uncertainty and reduces equifinality - please see Appling et al. (2018) cited above. Additionally, one of the sources cited above (Tassone and Bukaveckas 2019) used the updated Bayesian technique to estimate single- station estuarine metabolism.

*Response: text goes here*

Line 126: Presumably, EBASE produces daily metabolic estimates. What constitutes a day in EBASE? Is it from midnight to midnight the following day or from 4 am to 4 am?

*Response: The “metabolic” day in EBASE is midnight to midnight the following day. Clarifying text was added to line 182. In our experience, very similar estimates are obtained whether the day is defined midnight to midnight or some other period. This is especially true if the optimization period is greater than one day, as suggested in model application, i.e., lines 317 - 320, 523-525.*

Line 129-131: I normally see water column depth represented as Z not H. Should clarify if the values of P, R, and D are positive, negative or both.

*Response: The water column depth notation was changed to Z (including the documentation and code in EBASE). Clarification was added for P, R, and D on line 132: “Both and are expressed as positive values, whereas D can be positive (outgassing) or negative (ingassing).”*

Line 135: Consider making C = Cobs if it represents the observed DO concentration. If that is not what it represents, then will need to be described in text.

*Response: Yes, C represents observed dissolved oxygen concentration (as noted on line 132). We have retained C instead of using Cobs for simplicity in the model equations. Further, the addition of the model input/output table as requested by another reviewer helps clarify the input data.*

Line 181: Specifying the model optimization period is an advancement of the BASE model but is not described in text. What does ‘optimization’ mean? What are the implications of using 1 day, 7 days, 30 days, or the entire time series length on the output? If you use 7 days, do you only get weekly estimates of metabolism?

*Response: This was clarified in the text.*

*Line 181: “First, users can specify the model optimization period as the amount of time over which the unknown parameters are estimated from the Bayesian prior distributions (i.e., , , and , see below).”*

*Line 184: “Note that only one estimate for each of , , and is returned for each optimization period as defined by the model, although and can vary continuously with each timestep given the dependence on other inputs that can vary across the time series (e.g., and wind).”*

Line 185: On line 546 it is stated that linear interpolation is used, that should be described here. Furthermore, the threshold discussion on line 547-551 should be mentioned here. Lastly, on line 550-551, a 12 hour threshold should be further discussed. Why 12- hours? Why not 6 considering anything longer than 6 and you would be missing half of a semidiurnal tidal cycle.

*Response: 12 hours was chosen as a default value to cover approximately half of a diurnal cycle. Users may want to consider smaller gaps if tidal cycles are having a large influence on the observed time series. Our default value was chosen under the assumption that tidal effects, although they may be present, are likely not the dominant signal in the data. If the latter is expected, detiding is strongly encouraged. We have added text to this effect.*

*Line 186: “All gaps are linearly interpolated to allow the model estimation to proceed continuously and any results for an optimization period with gaps larger than 12 hours are removed, eliminating the need to pre- and post-process the data when using EBASE. The user can vary the criteria for removing results (e.g., 6 hours as a more conservative criteria).”*

*Line 555: “More conservative gaps can be chosen if, for example, semidiurnal tidal signals operating at time scales less than 12 hours are expected to influence the results. However, detiding is strongly encouraged in these cases and the default gap size of 12 hours that describes the diurnal oxygen signal is likely appropriate after detiding.”*

*Also note that data for an entire optimization period are removed if any gaps greater than the chosen gap size are present, providing an additional safety measure that the results are not based on inaccurately interpolated inputs.*

Line 188: With your improvements to the original BASE method, could EBASE be used to estimate metabolism in non-tidal, lotic ecosystems?

*Response: text goes here*

Line 248-252: How was the NERRS data accessed? Also be sure to cite the NERRS system-wide monitoring program, see link on NERRS data citation here - https://cdmo.baruch.sc.edu/data/citation.cfm

*Response: The NERRS data were accessed through the CDMO, citation was added.*

Line 267: Consider rephrasing – “…has a structure similar to, yet noisier than, that of light efficiency…”.

*Response: Rephrased.*

Line 270-271: Wouldn’t one expect the opposite? During periods of high productivity, ingassing should occur given the relatively high(er) rates of primary production. Then during periods of low productivity (i.e., outside of the growing season) that outgassing would dominate as respiration of allochthonous material continues to occur while primarily production would be light limited (assuming the example dataset used is in a temperate environment).

*Response: text goes here*

Line 276-278: As pointed to earlier in the manuscript, it is unclear what is meant by the optimization period. Why are the results in Figure 3 averaged by the optimization period? Why not show the daily estimates? I liked seeing the fit of the EBASE estimates on the synthetic time series however, I would like to see the fit of a linear regression of the synthetic (observed) estimates vs. the EBASE (predicted) estimates.

*Response: text goes here*

Line 313: This explanation of optimization needs to come much earlier in the text and be expanded upon.

*Response: text goes here*

Line 324-327: I’m not sure if I follow correctly - are you seeking daily metabolic estimates? If so, are the 7 & 30 day optimizations using a forward looking moving window approach that advances by one-day (similar to a rolling average)?

*Response: text goes here*

Line 473-477: It is stated that “Bayesian…methods have not been rigorously explored for estimating metabolic rates in estuaries” however, I point back to the two cited studies in the introduction (line 98) that did use Bayesian techniques to estimate estuarine metabolism. Consider rewording this sentence and/or expand on why those earlier studies are not considered rigorous.

*Response: text goes here*

Line 581: Maybe I missed it but how many MCMC simulations were used? How would a user change the number of MCMC simulations in the EBASE package? Can you provide a recommendation for the number of simulations to use or at least how many were used in this analysis?

*Response: text goes here*

Figures 3 & 5: These figures have a gray background while the rest are white. Please make the background color of the figures consistent across figures.

*Response: Figures 3 & 5 were changed to white background.*

Figure 7: Consider adding points to the lines in the left column plots. You may need to make the thickness of the lines smaller.

*Response: Points were added, line thickness reduced in the left column plots.*

### Reviewer 2

Beck et al. developed a new method for estimating metabolism in estuaries using continuous dissolved oxygen data. Metabolism is a key component of ecosystems, yet it is inherently difficult to estimate, especially in estuaries due to tidal currents. Their method is based on a stream Bayesian approach, and the authors expanded it by adjusting the time interval over which model parameters are estimated. The longer intervals allow for estimation over tidal cycles, which acts as noise in the time series, and allows for improved model estimation. The authors also compared their model to a more conventional approach (i.e., Odum model) and show the model’s improved estimation, specifically avoiding impossible results, such as negative production.

The model is written in the programming language R, making it widely available to the public. The online documentation and vignettes are clear and show the users key features and examples of the package. This is a nice bonus to go along with a well-written manuscript and expand the impact of the products.

I enjoyed reading the manuscript. It is technical where it needs to be, but also provides general context for the problem and the rationale for the technique. I was tripped up during a few of the sections, and I provide some suggestions that might improve the flow of the article. I’m also wondering if this model approach is only applicable to estuaries, or if it could be useful for other aquatic ecosystems? Overall, the authors did a great job. Nice work!

*Response: text goes here*

#### General comments

Table of model parameters. There are several parameters, most are relatively simple (single letter), and they appear on figures. I think it would be valuable to have a table. This could include, name, acronym, units, and any other relevant information (source/equation/etc.). I can also see organizing this table by inputs and outputs (L146) to help future users identify their input data and expected outputs.

*Response: We have added a new table showing model inputs/outputs, parameter names, acronyms, and units.*

Is EBASE only applicable to estuaries? It seems that the main change from BASE is the inclusion of a wider temporal window to estimate model parameters and changes in the b parameterization. Can these modifications also improve upon metabolism estimates in other systems (lakes, streams, etc.)? While I see the goal of improving metabolism in estuaries, there may be wider applicability to rivers, lakes, ocean, or other coastal areas. If so, mentioning what kinds of systems (big lakes, deep lakes, big rivers, complex hydrodynamics, etc.?) would be most appropriate for the model and potential get improved estimates over other metabolism approaches. This may expand the impact of this paper.

*Response: text goes here*

Model parameter b.

I’m less familiar with this term, and the authors could do a little bit more to explain what it is. On L143, they say that it is “fairly well known for open for open-ocean conditions”, but they never actually define the term. Looking at Eq 2, it seems to relate to “gas transfer efficiency” given that it is a coefficient on the wind-speed/viscosity term. That being said, I’m also a little confused on how Eq 3 is integrated into the model framework. Eq. 3 seems to relate more closely with streams as water depth is included as a scaler for kw. So somehow the two models (one more typical for lentic systems, the other more typical for lotic systems) are integrated together. In theory this makes sense in an estuary as both wind and flow likely affect gas transfer.

Diving into eq 3, higher water levels usually indicate higher flows and thus higher gas transfer velocity (Kw). I’m not exactly sure how temperature plays into this equation, and if it is appropriate for an estuary. It seems to suggest warm periods within a day have higher exchange rates, due to differences in kinetics at the water surface. However warmer temperatures may also reflect increased stratification and less vertical mixing (small lakes). There may be something to clean from changes in temperature that drive changes in turbulence, but simply using warmer = more exchange seems like an oversimplification. Eq 3 also includes a term related to water depth (H). In flowing waters this equation suggests that Kw increases with water level, likely a result of high-water levels = higher velocities = higher turbulence = higher kw. In an estuary, I’m not sure this relationship truly holds as water depth is not as good of a surrogate for velocity. Highest water level may be at slack tide and thus minimal water movements. It could be that change in water depth (assuming the tidal currents are strong relative to the river inputs) would be a better term to reflect instantaneous discharge. Or possibly flow could be included as a possible input parameter. But I realize this term is not typically available.

All told, perhaps the issue with poorly constraining b is related to assumptions in the equations used to model gas exchange in an estuary. The authors do a good job explaining that the model parameter b was difficult to constrain in the example dataset. This is understandable, and my comments are mainly my mental exploration of ways that might improve that aspect of the model. I’m guessing the authors have thought extensively about this problem, and I’m mainly offering ideas. Perhaps the authors could include some other ideas for why the b parameter is difficult to constrain? Or explore additional datasets to understand when/where it can be estimated more accurately? For example, if the model was applied to an artificial dataset from a stream, lake, or other type of estuary does it perform better? If so, what is different about the system, which might help with tuning the model? Would any of these help understand the parameterization/limitation of the current model, and or identify systems it is most appropriate?

*Response: text goes here*

#### Specific comments

L26: which open-source software? If only one (i.e., R), then can list it here.

*Response: text goes here*

L40: Organic matter produced in the estuary has other outcomes. Re-respired, consumed and incorporated into biota, exported to the ocean, etc.

*Response: text goes here*

L83: which computed rates were negative? P, R, or both? L128: mean depth?

*Response: text goes here*

L133: Thanks for providing the units. I am curious about area and time. I’m assuming that area is surface water area and time is the interval of the dissolved oxygen data. If so, perhaps expand on these so the author knows what the temporal and horizontal spatial scales are.

*Response: text goes here*

L134: “R is constant”. I believe R is allowed to vary across days, but R is constant within a day. Please specify.

*Response: text goes here*

L143: what is b? and for what time interval is b constant? This ends up being an output (Figure 3) and it is confusing to say that it is constant. So maybe it is constant within a day? Is this a standard parameter for gas exchange? As with the other parameters, providing some type of ‘name’ would be helpful, especially if provided in a table.

*Response: text goes here*

L164: The equation implies that Kw would increase as water level (H) increases. It seems the authors have this statement inversed or talking about measured reaeration (K). I believe this relates to streams where increasing water levels relate to higher water velocities, and thus higher turbulence and higher gas transfer velocity (kw). See General Comments.

*Response: text goes here*

L189: Thanks for putting together the web page. I’m curious if the data in this manuscript are included within the package itself, so it is possible to regenerate the results entirely? I’m also wondering how stable this website is and if any online information that is described in the manuscript will change? I’m not that familiar with these online reference materials, but I’m thinking of a static version (tag) that will stay with the paper so a reader 10 years from now could access the exact same materials that go along with this paper.

*Response: text goes here*

L210: Thank you for providing the units for these parameters. I am wondering if that authors considered using negative exponents to ensure the division symbols are properly addressed. For example, 300 mmol/m2/d. While I understand this to be mmol m-2 d-1, days could be mistakenly assumed to be in the numerator of the equation. While unlikely in this situation, the subsequent equations (L218) get even more complicated “(mmol/m2/d)/(W/m2)”, so it could be worth ensuring all the units are clear. Further, this unit has m2 twice, so they could be removed from the final unit (mmol d-1 W-1).

*Response: text goes here*

L237: The upper limit of 0.502 for b. This makes sense to constrain this parameter. I also noticed that many of the estimates of b are close to this value rather than near the synthetic value for the example dataset (Figure 3). I’m wondering if anything within the model could be improved to handle this issue? See general comments.

*Response: text goes here*

L272: I’m curious about the model estimates for kw. The authors show the results for flux (D), but this will vary with concentrations and k. Wondering if there is anything else to glean from those modelled results?

*Response: text goes here*

L301: The units for R are listed as (mmol/m2/d)/(W/m2). Perhaps the denominator is a typo?

*Response: Yes, this was a typo, corrected.*

L541: Wondering if it would be helpful to mention somewhere in this paragraph, “it depends on the question/objective of the project”? For example, if we are interested in annual summations and comparisons among systems, it may be fine to use a longer optimization period. But if we are interested in daily rates, timing of events in relation to storms, floods, etc. we may need to use a shorter optimization period. So the researcher’s question should be considered.

*Response: text goes here*

### Reviewer 3

Beck et al. present a new software package to estimate ecosystem metabolism from time series of dissolved oxygen and accompanying physical variables in estuarine systems. Accurately estimating metabolic rates are critical to understanding impacts on estuaries and their role in material cycling. The work is highly novel as no such statistical approach is currently available for estuaries. The software adapts a package built for estimating metabolism in stream ecosystems using a Bayesian framework, which is highly suited to this task due to the ability to place prior distributions on physical parameters with known physical limitations. I found the paper to be very clear and the choices made in model development, setting of priors, and model validation to be robust and well documented. The improvements made to the model, such as the option to vary the optimization period, are valuable and were demonstrated to be important to accurately estimating rates (and mirror suggestions for dealing with inaccurate K in streams; e.g. Appling et al. 2018 https://dx.doi.org/10.1002/2017JG004140)

In particular, the focus on the importance of accurately capturing the physical variables is a strength of the paper. It is made clear from the outcomes of the synthetic dataset analysis that it is better to use a literature estimate of the parameter b, a constant in the equation that governs air-water exchange, then to let the model estimate inaccurate values for b. The discussion around this issue was appropriate, but I wonder if some more very clear advice could be provided to a potential user that wants to use this package to answer questions about ecosystem metabolism in estuaries but does not have the aim of further refining the modelling approaches. Could they use the estimate from Wanninkhof (2014) as was used in subsequent analyses here? Would an appropriate sensitivity analysis be to vary this estimate by 20% (as mentioned L224) to check if their P and R results are highly sensitive to this parameter? Or is further testing in a range of estuaries required before this can be applied?

*Response: text goes here*

#### Specific comments:

L14: The first use of “unphysical” in the abstract was a little unclear without further detail. Consider a short phrase like “may lead to estimates that are physically impossible” be used to define this first.

*Response: text goes here*

L122: The sentence “We also attempt to constrain the gas transfer velocity using this approach but are unable, at least with the data sets used in this study.“ was unclear. Consider replacing “this approach” with a more precise statement (I assume “this approach” is the Bayesian framework, rather than a reference to the previous sentence on the test with the detided inputs?). I’m also not sure if “constrain” is the right word, as the issue was accurately estimating this parameter.

*Response: text goes here*

L159 & L514: The air-water exchange isn’t depth dependent in BASE. The inputs (and metabolic results) in base are all volumetric and reaeration coefficient (K) is in per day. Depth is not provided to the model but can be used to convert the metabolic estimates afterwards.

*Response: text goes here*

L258: typo in “estimatd”

*Response: text goes here*

L269: Should it be “availability” rather than “variability”? of course, a more variable pool of organic matter could also be correct.

*Response: text goes here*

L352: The “right answer for wrong reasons” is interesting. It could be clearer here that even though DO is estimated accurately, the P and R estimates will be inaccurate in this scenario.-7

*Response: text goes here*