EBASE manuscript

Marcus W Beck1,✉, Jill M Arriola2, Maria Hermann2, and Raymond G Najjar2

1 Tampa Bay Estuary Program, St. Petersburg, Florida 33701 USA  
2 Pennsylvania State University, University Park, Pennsylvania 16802 USA

✉ Correspondence: [Marcus W Beck <mbeck@tbep.org>](mailto:mbeck@tbep.org)

## Abstract

*Key words*:

## 1 Introduction

* Overview of metabolism - importance
* Existing methods: Odum/WtRegDO, BASEmetab, references in Grace et al. (2015), applications to lakes and streams
* Applications to estuaries - challenges
* Goals and objectives

## 2 Materials and Procedures

The Estuarine BAyesian Single station Estimation (EBASE) method provides a simple approach to quantify metabolic parameters from continuous dissolved oxygen time series. The EBASE method is a modified version of the BASEmetab mass balance equation described in Grace et al. (2015). The original equation is:

where the terms in the square brackets are production (P), respiration (R), and gas exchange (D), respectively, modeled as a function of the change in dissolved oxygen () per unit time of the input data (). Five unknown parameters are estimated by the model: , , , , and . The required inputs are the time series of surface irradiance (, mol/m/s), water temperature (, C), atmospheric pressure (atm), diel oxygen (, mmol/m), and oxygen saturation (, mmol/m). Dissolved oxygen at saturation is calculated from the input data as a function of water temperature, salinity, and barometric pressure. Several changes, outlined and justified below, were made to [Equation 1](#eq-base) for EBASE.

### 2.1 Modification of the gas exchange parameterization

The underlying idea behind most gas exchange parameterizations for metabolism estimates is that the transfer of gases of relatively low solubility, like oxygen, is proportional to the concentration gradient in a thin film of water at the air–water interface. It is further assumed that the water at the top of the film is saturated with respect to the atmosphere, so that the concentration gradient across the thin film can be approximated as the difference between oxygen concentration measured in the surface water layer, , and the saturation (or equilibrium) concentration , which is usually calculated as a function of water temperature and salinity (the negative sign is needed because positive flux is set to be from water to air):

The proportionality parameter, , usually called the gas transfer velocity because it has the dimensions of length over time, increases as the water column in the vicinity of the air–water interface gets more turbulent. As such, is usually modeled as a function of measurable variables that can potentially predict turbulent mixing of the surface water, such as wind speed, current velocity, and water depth (Ho et al. 2016). also increases with temperature because as the random motion of gas and water molecules in the thin film speeds up, the film thins out and the transfer of molecules across the film becomes easier. This dependence of on temperature is usually modeled with the Schmidt number, which combines information on the viscosity of the fluid and the diffusivity of the gas. The Schmidt number, in turn, is often parameterized as a gas-specific polynomial function of water temperature and salinity.

Instead of one of the traditional wind-based parameterization used in oceanography (e.g., Wanninkhof 2014), Grace et al. (2015) uses the following for the gas transfer in BASEmetab:

where is the average water temperature over each 24-hour period in the diel oxygen time series and is the optimization parameter, often called the reaeration coefficient, which has dimensions of inverse time. [Equation 3](#eq-gasexbase) indicates that increases with temperature and depth. The temperature dependence is consistent with the known decrease of with the Schmidt number, but the depth dependence is difficult to justify. One expects that, for a given current speed, would increase as the water depth decreases because turbulence generated at the sediment–water interface by currents would more easily reach the surface (Ho et al. 2016).

Based on the above differences, the first modification to [Equation 1](#eq-base) for EBASE is use of a wind-based parameterization for gas exchange that is more appropriate for estuaries (Wanninkhof 2014):

where is the wind speed at 10-m height, is water column depth (m, as a fixed parameter or time series from a pressure sensor) and is the Schmidt number defined as the ratio of the kinematic viscosity of water to the molecular diffusivity of the gas for oxygen (from water temperature and salinity using the polynomial fit of Wanninkhof 2014). The parameter () must be estimated from the data.

The remaining variables in [Equation 4](#eq-basewan) are defined as in [Equation 1](#eq-base), with the exception that is calculated using only temperature (C) and salinity (psu), removing the need for barometric pressure. Methods from García and Gordon (1992) are used as follows:

where is thr solubility of dissolved oxygen per volume of seawater at the given temperature and (kg/m) is seawater density at zero elevation given temperature and salinity (Millero and Poisson 1981). is calculated using a polynomial equation:

where the coefficients are = 5.80818, = 3.20684, = 4.11890, = 4.93845, = 1.01567, = 1.41575, = -7.01211e-3, = -7.25958e-3, = -7.93334e-3, = -5.54491e-3, and = -1.32412e-7. is temperature scaled as input to [Equation 6](#eq-co) defined as:

### 2.2 Modification of the production and respiration parameterization

The second modification to [Equation 1](#eq-base) for EBASE is a simplification to remove the and parameters from the production (P) and respiration (R) components, respectively, of the mass balance equation. The BASEmetab method provides options to estimate both and from the data or to set as fixed values defined in the model file (, ). For EBASE, the temperature dependency of respiration is removed by setting = 1. The parameter, which is dimensionless, is also set to 1 as in Holtgrieve et al. (2010). Removing these components mproved model stability by creating more reasonable estimates for the other parameters. This further simplifies [Equation 4](#eq-basewan) to the final form for EBASE:

As for BASEmetab, the metabolic estimates in EBASE are defined by the change in dissolved oxygen over the time step, , where gross production is provided by , respiration is provided by , and gas exchange is provided by the remainder (as in Wanninkhof 2014). Units for each of the metabolic estimates are .

The remaining three parameters (light efficiency, ), (respiration, ), and () are estimated by likelihood given the observed data and prior distributions for the parameters. The JAGS software (Plummer et al. 2003) is used with the EBASE R package (described below, Beck et al. 2022; R Core Team 2022) to estimate the unknown parameters using Markov Chain Monte Carlo simulations in a Bayesian framework. At each time step, the change in oxygen concentration between time steps is calculated from the equation using model inputs and parameter guesses, and then a finite difference approximation is used to estimate modeled oxygen concentration. Estimations proceed using a forward in time integration where the dissolved oxygen at each time step is informed by parameter estimates from the preceding time steps. The estimated dissolved oxygen concentration is also returned, which can be compared to observed as one measure of model performance.

### 2.3 Bayesian prior

The metabolism estimates and their parameters returned by the Bayesian routine implemented in JAGS are affected by the prior distributions assigned to each. As for the BASEmetab model, relatively uninformed prior distributions following a normal Gaussian distribution are used by default, although the priors can of course be changed based on previous knowledge of parameters specific to an ecosystem or as informed by other metabolic modeling approaches. The default priors were based on those that returned reasonable parameter estimates using a forward-feeding oxygen mass balance equation call FWOXY (ForWard OXYgen, Arriola and Herrmann 2022). For FWOXY, a dissolved oxygen time series is simulated by providing known parameter values in an equation similar to [Equation 8](#eq-ebase). All of the physical parameters are held constant, e.g., wind speed, temperature, and salinity. The other required inputs are the , , and parameters, which are set by the user. An initial oxygen concentration is provided and the simulated time series is created by forward addition of the inputs. As such, the dissolved oxygen output simulated by FWOXY as change per unit time were modelled by EBASE using the constant physical parameters as input. Reasonable prior distributions by EBASE were identified (Equations [9](#eq-aprior), [10](#eq-rprior), and [11](#eq-bprior)) based on the ability to return the known parameters from FWOXY:

The notation is JAGS model syntax indicating the distributions are truncated to a minimum (left) and maximum (right) value defined in the parentheses. All prior distributions were constrained to positive values based on known physical constraints and requirements of the model formula in [Equation 8](#eq-ebase). Additionally, the prior distribution for was constrained to an upper limit of 0.504 (~twice the default mean). Wanninkhof (2014) states that the parameter typically does not vary by more than 20% and the standard deviation of the default prior is sufficiently high to expand beyond this range. Initial development of EBASE also showed that using an undefined upper limit of led to unstable and unreasonable parameter estimates. The density distributions for the default priors in EBASE are shown in [Figure 1](#fig-priorplot).

### 2.4 EBASE R package

For ease of application, an R package was developed to implement the EBASE methods described above (Beck et al. 2022). A detailed web page (<https://fawda123.github.io/EBASE/>) describes the required data inputs for estimating the metabolic parameters using the core R functions that pass inputs to the JAGS software. A JAGS model file is included with the package that implements [Equation 8](#eq-ebase), with options to supply a custom model file as needed. Users can also specify alternative prior distributions from those in Equations [9](#eq-aprior), [10](#eq-rprior), and [11](#eq-bprior) for each parameter. Functions are available for viewing model results, including a plot of the modelled dissolved oxygen with the observed, a time series plot of the primary metabolic estimates (P, R, and D), and a time series plot of the credible intervals for the , , and parameters. An additional option also allows the user to specify the model optimization period, where the time period can vary from a minimum of one day to the maximum of the entire length of the time series. This differs from the BASEmetab approach where the optimization period is set at one day. Sensitivity of the EBASE method to different optimization periods is described in [Section 3.1](#sec-cmpmod). All results provided herein were generated using the EBASE R package.

## 3 Assessment

### 3.1 EBASE comparison with modelled observations

A preliminary assessment of the ability of EBASE to produce reasonable parameter estimates for , , and ([Equation 8](#eq-ebase)) was conducted using a simulated time series created with FWOXY. These assessments evaluated if 1) EBASE estimates similar values for the known parameters, and 2) how the estimated values change with different model configurations. The simulated time series used inputs from an approximate one year of continuous data from Apalachicola Bay, Florida. The input data were 2012 observations from February 22nd to December 19th at the Cat Point water quality monitoring station (29.7021 N, -84.8802 W) and included 15 minute observations of dissolved oxygen (mg/L), water temperature (C), and salinity (psu). Continuous meteorological observations from the nearby East Bay station (29.7909 N, -84.8834 W, ~10km north of Cat Point) were combined with the water quality data at the same time step and included air temperature (C), wind speed (m/s), and PAR (mmol/m as total per 15 minute observation). An oxygen time series was generated from Fwoxy using the inputs from Apalachicola Bay ([Figure 2](#fig-simapa)a), such that the simulated dissolved oxygen was created as close as possible to the observed given the inputs. Known values for and parameters were returned for each time step of the simulated data, as well as all metabolic estimates for production, respiration, and gas exchange ([Figure 2](#fig-simapa)b and c). The only parameter that could not be simulated by Fwoxy was , which was set to 0.251 .

Several configurations of the EBASE model can influence how the parameters and metabolic rates are estimated. Two types of sensitivity analyses were conducted to evaluate changes in the results returned by EBASE, where in all cases, the results were compared to the known values in the simulated time series. First, results were compared to the known values after changing characteristics of their prior distributions. Second, results were compared using a model period of optimization for one day and one week of observations. Details of these analyses are as follows.

The standard prior distributions for the EBASE parameters follow normal distributions with defined means and standard deviations (Equations [9](#eq-aprior), [10](#eq-rprior), and [11](#eq-bprior)). The sensitivity analysis evaluated changes in the distributions by varying the standard deviations from small to large to test the effect of more or less constrained ranges, respectively, on the priors. For each parameter, low, medium, and high values for the standard deviation were assessed as 10%, 100%, and 1000% of the default values. Each unique combination of low, medium, and high values for the standard deviation of each parameter were evaluated, creating 3 = 27 different combinations of prior distributions that were used with the simulated time series in EBASE.

The optimization period used for a time series determines how much data are used to estimate the three EBASE parameters (, , and ). Unlike BASEmetab, the EBASE software allows flexibility in the period of time used for optimization, with a minimum period of time being one day and a maximum period of being the length of the time series. For a chosen optimization period, a single estimate for and is returned by the model, whereas number of observations equal to the number of time steps in each optimization period are returned for the parameter (e.g., 96 values for a 24 hour time period if the sampling interval is 15 minutes). Short periods of observation may return unstable parameter estimates if forcing factors (e.g., wind, temperature) that affect estuarine biogeochemical rates occur at time scales longer the chosen period of optimization for EBASE. As such, potentially longer optimization periods may return more stable and robust estimates given that more data are used for parameter estimation. An optimization period of one day was compared to results using an optimization period of seven days for the simulated one-year time series. Each optimization period was evaluated for every unique combination of prior distributions, totaling 54 (27 times the two optimization periods) unique analyses for the assessment of changing standard deviations of all parameters.

Changing the variance of each of the prior distributions affected the ability of EBASE to return the known parameters from the simulated time series. [Figure 3](#fig-apasumdat) shows a summary of the comparisons of the simulated data with EBASE for each unique combination of prior distributions and two optimization periods of one and seven days. The results are compared using R-squared () values from the linear fit between the simulated data and results from EBASE. Darker colors represent lower values and can be used to assess how changing the variance of the priors affects the ability of EBASE to estimate the true parameter.

Several conclusions can be made from the results in [Figure 3](#fig-apasumdat) based on the patterns of the values. Overall, the model was able to estimate the parameter with reasonable precision regardless of any of the prior distributions, whereas most of the variation was observed in the and parameters. Changing the variance of the parameter had little effect on any of the comparisons. Smaller variance of the prior distribution for the parameter generally contributed to lower for all parameters, with the precision being worse for the 7-day optimization period (e.g., the parameter). Larger variance of the prior distribution for the parameter generally contributed to lower only for the gas exchange estimate (D), although this effect was minimized with the 7-day optimization period. As such, the results suggest that the default prior distributions used in EBASE may be appropriate in some cases, but increasing the prior variance for and and decreasing the variance for will likely produce more optimal solutions.

### 3.2 EBASE comparison with real observations

Metabolic estimates from EBASE were further compared to those from existing methods to benchmark and evaluate potential differences from alternative model formulations used by each approach. The same one year continuous time series described above for water quality and weather data from Apalachicola Bay was used to compare metabolic estimates from EBASE, BASEmetab, and the Odum open-water method (Odum 1956). The Odum method used an approach described in Murrell et al. (2018) created for estuarine application that used a gas-exchange approach described in Thébault et al. (2008). For better comparison with the other methods, the gas exchange parmaterization in the Odum method was modified to use the Wanninkhof (2014) approach (as available in the WtRegDO R package, Beck 2021). The input data were provided to each method after converting the units as required (e.g., BASEmetab requires PAR as mol m s). Water column depth at Cat Point is approximately 1.8 m and all daily metabolic estimates returned by each method were converted to volumetric rates by dividing by the depth (mol O m d). Each comparison also evaluated metabolic estimates using the observed dissolved oxygen time series and a detided time series using weighted regression (Beck et al. 2015). The latter comparison provided an assessment of metabolic rates that are expected to be minimally influenced by tidal advection. All metabolic estimates were obtained using the default settings for each of the three methods, with an exception that the prior distribution for the reaeration coefficient () in BASEmetab was informed by results from the Odum output. Initial assessments indicated that unstable and inaccurate results are obtained from BASEmetab using an uninformed prior distribution for . All other prior distributions for BASEmetab and EBASE were set as the software default. A one-day model optimization period was used for EBASE.

The common set of metabolic parameters returned by the Odum, BASEmetab, and EBASE methods are net ecosystem metabolism, gross production, respiration, and gas exchange. Pairwise comparisons of the daily estimates returned by each method were evaluated using simple summary statistics including Pearson correlation coefficients and linear regression fits to evaluate potential biases in each method. For the latter analysis, differences in intercept estimates less than or greater than zero suggested a bias in magnitude, whereas differences in slope estimates less than or greater than one suggested a bias that depended on relative value of each observation (e.g., bias may differ between lower or higher values). Comparisons of the slope and estimates values to zero and one, respectively, were based on confidence intervals from the standard errors of each estimate. Results were assessed separately for estimates based on the observed or detided dissolved oxygen time series.

[Figure 4](#fig-apacmpfig) and Table provide a comparison of the EBASE metabolic estimates with those from the Odum and BASEmetab methods.

### 3.3 Application to representative NERR sites

## 4 Discussion

## 5 Comments and Recommendations

## Acknowledgments

## Figures

Figure 1: Fwoxy comparisons

Figure 2: Apalachicola comparisons

Figure 3: Sensitivity analyses

Figure 4: NERRS application

|  |
| --- |
| Figure 1: The default prior distributions for the , , and parameters used in EBASE. All priors are normal distributions and truncated to positive values. The parameter is also truncated to a maximum of 0.504. |

|  |
| --- |
| Figure 2: Simulated time series using the FWOXY model for one year of continuous data at Apalachicola Bay. Simulated input to EBASE was the (a) dissolved oxygen time series, with results compared to the (b) simulated parameter and (c) simulatd metabolic estimates for production (P), respiration (R), and gas exchange (D). |

|  |
| --- |
| Figure 3: Comparison of EBASE results with known metabolic parameters from a simulated time series. Results are compared for unique combination of prior distributions, where the variances are evaluated from low (L), medium (M), to high (H) for a one-day and seven-day model optimization period. The metabolic results for dissolved oxygen (DO), production (P), respiration (R), gas exchange (D), and the parameter are evaluated based on the value of the linear fit between the EBASE results and simulated values. The parameter is not evaluated because it is a fixed constant for the simulated data. |

|  |
| --- |
| Figure 4: Comparisons of daily metabolic estimates from the Odum, BASEmetab, and EBASE methods for one year of continuous monitoring data at Apalachicola Bay. Results for net ecosystem metabolism (NEM), gross production (P), respiration (R), and gas exchange (D) are compared using the (a) observed and (b) detided dissolved oxygen time series as input to each model. All values are volumetric as mmol O m d and are grouped by the dry or wet season. Black lines show the 1:1 comparison, the blue line shows the linea regression fit, and the Pearson correlation values are shown in parentheses for each facet title. |

## Tables

Table 1: EBASE parameters and outputs

Table 2: Summary stats of Fwoxy comparison

Table 3: Summary stats of Apalachicola comparison

Table 4: Key parameters form NERRS application

**Table** **:** Summary statisics of daily metabolic estimates comparing Odum and BASEmetab to EBASE for net ecosystem metabolism (NEM), gross production (P), respiration (R), and gas exchange (D). Separate comparisons are made for estimates based on observed and detided dissolved oxygen. Summary statistics include Pearson correlation coefficients and intercept, slope, and residual standard deviation of a linear fit between the same metabolic estimates obtained from each method. Significance of the p-values for the intercept and slope indicates if the estimate is different from zero or one, respectively.

| Dissolved Oxygen | Comparison | Estimate | NA | Intercept | Slope | NA |
| --- | --- | --- | --- | --- | --- | --- |
| Observed |  |  |  |  |  |  |
|  | Odum v EBASE | NEM | 0.23\*\* | -11.4\*\* | 0.23\*\* | 26.26 |
|  |  | P | 0.28\*\* | 15.23\*\* | 0.56\*\* | 43.69 |
|  |  | R | 0.11 | -23.14\* | 0.65 | 54.12 |
|  |  | D | 0.74\*\* | 5.15\*\* | 0.65\*\* | 6.14 |
|  | BASEmetab v EBASE | NEM | 0.89\*\* | -8.14\*\* | 1.11\*\* | 15.12 |
|  |  | P | 0.47\*\* | 23.23\*\* | 0.69\*\* | 29.48 |
|  |  | R | 0.59\*\* | 22.38\*\* | 2.98\*\* | 36.93 |
|  |  | D | 0.57\*\* | 9.29\*\* | 0.48\*\* | 7.12 |
| Detided |  |  |  |  |  |  |
|  | Odum v EBASE | NEM | 0.74\*\* | -4.93\*\* | 1.03 | 7.77 |
|  |  | P | 0.52\*\* | 7.46\*\* | 0.93 | 10.73 |
|  |  | R | 0.53\*\* | -14.03\*\* | 0.87 | 15.72 |
|  |  | D | 0.8\*\* | 1.89\*\* | 1.04 | 6.66 |
|  | BASEmetab v EBASE | NEM | 0.59\*\* | -7.16\*\* | 0.52\*\* | 5.91 |
|  |  | P | 0.87\*\* | 2.79\*\* | 1.01 | 3.95 |
|  |  | R | 0.74\*\* | -9.54\*\* | 0.75\*\* | 7.59 |
|  |  | D | 0.54\*\* | 9.12\*\* | 0.42\*\* | 5.60 |
| \* p < 0.05, \*\* p < 0.005 | | | | | | |

## References

Arriola, J. M., and M. Herrmann. 2022. Fwoxy: Forward model of estuary dissolved oxygen time series,.

Beck, M. W. 2021. WtRegDO: Implement weighted regression on dissolved oxygen time series,.

Beck, M. W., J. D. Hagy III, and M. C. Murrell. 2015. Improving estimates of ecosystem metabolism by reducing effects of tidal advection on dissolved oxygen time series. Limnology and Oceanography: Methods **13**: 731–745. doi:[10.1002/lom3.10062](https://doi.org/10.1002/lom3.10062)

Beck, M. W., M. Herrmann, J. Arriola, R. Najjar, and W. McGillis. 2022. EBASE: Estuarine Bayesian Single-Station Estimation Method for Ecosystem Metabolism,.

García, H. E., and L. I. Gordon. 1992. Oxygen solubility in seawater: Better fitting equations. Limnology and oceanography **37**: 1307–1312. doi:[10.4319/lo.1992.37.6.1307](https://doi.org/10.4319/lo.1992.37.6.1307)

Grace, M. R., D. P. Giling, S. Hladyz, V. Caron, R. M. Thompson, and R. Mac Nally. 2015. Fast processing of diel oxygen curves: Estimating stream metabolism with BASe (BAyesian Single-station Estimation). Limnology and Oceanography: Methods **13**: 103–114. doi:[10.1002/lom3.10011](https://doi.org/10.1002/lom3.10011)

Ho, D. T., N. Coffineau, B. Hickman, N. Chow, T. Koffman, and P. Schlosser. 2016. Influence of current velocity and wind speed on air-water gas exchange in a mangrove estuary. Geophysical Research Letters **43**: 3813–3821. doi:[10.1002/2016GL068727](https://doi.org/10.1002/2016GL068727)

Holtgrieve, G. W., D. E. Schindler, T. A. Branch, and Z. T. A’mar. 2010. Simultaneous quantification of aquatic ecosystem metabolism and reaeration using a Bayesian statistical model of oxygen dynamics. Limnology and Oceanography **55**: 1047–1063. doi:[10.4319/lo.2010.55.3.1047](https://doi.org/10.4319/lo.2010.55.3.1047)

Millero, F. J., and A. Poisson. 1981. International one-atmosphere equation of state of seawater. Deep Sea Research Part A. Oceanographic Research Papers **28**: 625–629. doi:[10.1016/0198-0149(81)90122-9](https://doi.org/10.1016/0198-0149(81)90122-9)

Murrell, M. C., J. M. Caffrey, D. T. Marcovich, M. W. Beck, B. M. Jarvis, and J. D. Hagy. 2018. Seasonal oxygen dynamics in a warm temperate estuary: Effects of hydrologic variability on measurements of primary production, respiration, and net metabolism. Estuaries and Coasts **41**: 690–707. doi:[10.1007/s12237-017-0328-9](https://doi.org/10.1007/s12237-017-0328-9)

Odum, H. T. 1956. Primary production in flowing waters. Limnology and Oceanography **1**: 102–117.

Plummer, M. and others. 2003. JAGS: A program for analysis of Bayesian graphical models using gibbs sampling. *Proceedings of the 3rd international workshop on distributed statistical computing*. Vienna, Austria. 1–10.

R Core Team. 2022. [R: A language and environment for statistical computing](https://www.R-project.org/), R Foundation for Statistical Computing.

Thébault, J., T. S. Schraga, J. E. Cloern, and E. G. Dunlavey. 2008. Primary production and carrying capacity of former salt ponds after reconnection to San Francisco Bay. Wetlands **28**: 841–851. doi:[10.1672/07-190.1](https://doi.org/10.1672/07-190.1)

Wanninkhof, R. 2014. Relationship between wind speed and gas exchange over the ocean revisited. Limnology and Oceanography: Methods **12**: 351–362. doi:[10.4319/lom.2014.12.351](https://doi.org/10.4319/lom.2014.12.351)