Article

LakeMetabolizer: an R package for estimating lake metabolism from free-water oxygen using diverse statistical models

Luke A. Winslow,[1,2] Jacob A. Zwart,[3]* Ryan D. Batt,[2,4] Hilary A. Dugan,[2] R. Iestyn Woolway,[5,6] Jessica R. Corman, Paul C. Hanson,[2] and Jordan S. Read[1]

[1] *US Geological Survey, Center for Integrated Data Analytics, Middleton, WI, USA*
[2] *Center for Limnology, University of Wisconsin-Madison, Madison, WI, USA*
[3] *Department of Biological Sciences, University of Notre Dame, Notre Dame, IN, USA*
[4] *Department of Ecology, Evolution, and Natural Resources, Rutgers University, New Brunswick, NJ, USA*
[5] *Lake Ecosystems Group, Centre for Ecology & Hydrology, Lancaster, UK*
[6] *Department of Meteorology, University of Reading, Reading, UK*
* Corresponding author: jzwart@nd.edu

Received 2 May 2015; accepted 26 January 2016; published 2 November 2016

Abstract

Metabolism is a fundamental process in ecosystems that crosses multiple scales of organization from individual organisms to whole ecosystems. To improve sharing and reuse of published metabolism models, we developed LakeMetabolizer, an R package for estimating lake metabolism from *in situ* time series of dissolved oxygen, water temperature, and, optionally, additional environmental variables. LakeMetabolizer implements 5 different metabolism models with diverse statistical underpinnings: bookkeeping, ordinary least squares, maximum likelihood, Kalman filter, and Bayesian. Each of these 5 metabolism models can be combined with 1 of 7 models for computing the coefficient of gas exchange across the air–water interface (*k*). LakeMetabolizer also features a variety of supporting functions that compute conversions and implement calculations commonly applied to raw data prior to estimating metabolism (e.g., oxygen saturation and optical conversion models). These tools have been organized into an R package that contains example data, example use-cases, and function documentation. The release package version is available on the Comprehensive R Archive Network (CRAN), and the full open-source GPL-licensed code is freely available for examination and extension online. With this unified, open-source, and freely available package, we hope to improve access and facilitate the application of metabolism in studies and management of lentic ecosystems.

Key words: gas exchange, GLEON, lake metabolism models, open source, R package, sensors

Introduction

Metabolism is a fundamental ecological process that occurs at scales ranging from individual organisms to whole ecosystems (Brown et al. 2004). Whole-ecosystem metabolism cannot be measured directly, but rather represents the balance between carbon fixation (gross primary production [GPP]) and biological carbon oxidation (ecosystem respiration [R]) in an ecosystem. The difference between GPP and R is termed net ecosystem production (NEP) and is used to delineate heterotrophic systems (negative NEP) from autotrophic systems (positive NEP). At an ecosystem scale, metabolism estimates provide insight into the dynamics of food webs through primary productivity (e.g., Carpenter et al. 1987), energy mobilization ratios (Jansson et al. 2003, Ask et al. 2009), rates of carbon accumulation or loss in an ecosystem (Lovett et al. 2006), global carbon budgets (Field et al. 1998, Cole et al. 2007), and anticipated changes in ecosystem state (Yvon-Durocher et al. 2012, Batt et al. 2013).

Lake metabolism is frequently estimated from free-water dissolved oxygen (DO) concentrations (e.g.,

DOI: 10.5268/IW-6.4.883

Staehr et al. 2010). In surface waters, the day–night dynamics of DO are indicative of lake GPP and R, after accounting for abiotic losses and gains in the water column. The free-water method was made popular by Odum (1956) and has recently become prominent with the advancement of automated sensor technology (Cole et al. 2000, Hanson 2007). Estimating metabolism based on high-frequency DO sensor measurements simplifies data collection, especially at times when manual sampling would be difficult (e.g., during storms, under ice), and it can provide high temporal and spatial resolution (inter- and intra-lake), which allows novel questions regarding whole-lake metabolism (Coloso et al. 2008, Staehr et al. 2012b, Van de Bogert et al. 2012, Solomon et al. 2013).

The value of quantifying lake metabolism and the availability of necessary data has led to a rapid proliferation of computational methodologies for estimating metabolism (Staehr et al. 2010, Hoellein et al. 2013). Although technological advances in automated sensors and the expansion of cross-site collaborations have increased the accessibility of high-frequency DO time series (Porter et al. 2012, Read et al. 2012, Solomon et al. 2013), barriers are presented by the statistics, programming, and multitude of equations used to convert sensor observations into estimates of lake metabolism. These analytical barriers may be overcome by documented, functional computer code designed to estimate lake metabolism from commonly collected sensor data. Making the code free and open source will promote inclusion of metabolism estimates into research studies, foster sensor data standardization, and evolve as scientific methodologies advance. In this manuscript, we introduce a new R package called LakeMetabolizer, developed to streamline established approaches to estimating lake metabolism.

Methods

We present approaches to estimating lake metabolism that attribute changes in DO to biological processes (metabolism) and to physical exchange of oxygen (O_2) across the air–water interface. Metabolism models contain both biological and physical terms. We present 5 statistically distinct metabolism models used to estimate the biological components of GPP, R, and NEP (Table 1). Within each of these 5 models is a term for gas exchange (*k*), which can be calculated from 1 of 7 physical models (Table 2). LakeMetabolizer currently only operates on time series free of missing values and with evenly spaced observations, and the following model descriptions assume complete and regular time series.

Metabolism models

Bookkeeping

The simplest metabolism model included in LakeMetabolizer is known as the “bookkeeping” method (Odum 1956, Cole et al. 2000) and is expressed as *metab.bookkeep*. Bookkeeping is unique among the metabolism models we included because it does not include error terms and does not estimate metabolism parameters from data. In this method, the daily metabolism rate NEP (as O_2 in mg L^{-1} d^{-1}) is calculated as the average of the time discrete NEP_t (mg L^{-1} Δt^{-1}). The bookkeeping method attributes changes in DO between consecutive observations (time elapsed = Δt) to NEP_t and discrete gas exchange:

$$\Delta DO = NEP_{t-1} \times \Delta t + F_{t-1}. \quad (1)$$

Gas exchange (as O_2) over a discrete period (*F*; mg L^{-1} = g m^{-3}) is calculated as

$$F_t = \frac{k_t \times \Delta t}{z_t} \times (O_{s,t} - DO_t), \quad (2)$$

where *k* (m t^{-1}) is the coefficient of gas exchange, *z* is the depth of the surface mixed layer (m), and O_s is the saturated oxygen concentration, calculated from salinity, temperature, and atmospheric pressure. NEP_t is the balance between gross primary production (GPP_t) and respiration (R_t) occurring at time t:

$$NEP_t = GPP_t + R_t\,. \quad (3)$$

During darkness, GPP_t is 0, which allows us to estimate the average rate of O_2 respiration (R_μ; mg L^{-1} Δt^{-1}) from the discrete rates of NEP_t for all observations occurring at night (in the morning before sunrise, and in the evening after sunset):

$$R_\mu = \frac{\sum_{i=1}^{n} \Delta DO_i - F_i}{k\Delta t}, \quad (4)$$

where *k* is the number of nighttime observations. Total nighttime respiration, R_{night} (O_2 in mg L^{-1} d^{-1}), is the product of R_μ and *k*. By making the assumption that the rate of respiration is constant throughout the 24-hour period, the total daytime respiration (R_{day}; mg L^{-1} d^{-1}) can similarly be calculated as the product of R_μ and the duration of the daytime period (period after sunrise and before sunset). Because NEP is the balance between GPP and R, the rate of GPP (mg L^{-1} d^{-1}) can be calculated as:

$$GPP = NEP_{day} - R_{day}. \quad (5)$$

For complete and regular time series, the expected number of observations in a day is *n*, such that if Δt = 5 minutes then *n* = 288. The rate of NEP (mg L^{-1} d^{-1}) is the average

Table 1. Comparisons of the structure of the 5 different metabolism models included in LakeMetabolizer. Note that other model attributes are accessible for some models by using the R base package function attr (e.g., posterior draws for *metab.bayesian*, fitted parameters for *metab.ols*, *metab.mle*, *metab.bayesian*, and *metab.kalman*).

Model	Underlying statistics	Error structure	Error type	Parameters fit	Photosynthesis-irradiance relationship	Respiration-temperature relationship	Output
metab.bookkeep	Algebra	None	None	None	None	None	GPP, R, NEP
metab.bayesian	Bayesian	Gaussian	Process & observation	$\iota, \rho, K, \tau_w, \tau_v$	Linear	Log-linear	GPP, R, NEP, $\sigma_{GPP}, \sigma_R, \sigma_{NEP}$
metab.kalman	Maximum likelihood & Kalman filter	Gaussian	Process & observation	ι, ρ, Q, H	Linear	Log-linear	GPP, R, NEP
metab.mle	Maximum likelihood	Gaussian	Observation or process	ι, ρ	Linear	Log-linear	GPP, R, NEP
metab.ols	Linear regression	Gaussian	Process	ι, ρ	Linear	Log-linear	GPP, R, NEP

Table 2. Required time series and metadata inputs for each gas flux coefficient model. Some data requirements can be calculated from other commonly observed variables.

Model	Wind speed	Wind sensor height	Atm pressure	Air temp	Downwelling shortwave	Net longwave	Lake latitude	Lake area	Active mix layer depth	Water surface temp	Relative humidity
k.cole	X	X									
k.crusius	X	X									
k.vachon	X	X						X			
k.heiskanen	X	X	X	X	X	X			X	X	X
k.macIntyre	X	X	X	X	X	X			X	X	X
k.read	X	X	X	X	X	X	X	X	X	X	X
k.read.soloviev	X	X	X	X	X	X	X	X	X	X	X

of NEP_t for all values of t, multiplied by *n*. The rate of R (mg L^{-1} d^{-1}) is the product of R_μ and *n*.

Ordinary least squares

The model *metab.ols* estimates metabolism from parameters from a regression model fit using ordinary least squares (OLS). The approach taken in *metab.ols* follows Batt and Carpenter (2012), although McNair et al. (2013) have also employed OLS to estimate metabolism. In *metab.ols*, linear regression is used to predict biologically driven changes in DO from observations of irradiance (I) and the natural logarithm of water temperature ($\log_e T$). The regression equation for *metab.ols* is:

$$O = X\beta + \varepsilon, \tag{6}$$

where *O* is an $n \times 1$ (*n* is the number of time steps, t, in a day) vector of O_2 values calculated from the term $NEP_t \times \Delta t$ (mg L^{-1}) from equation 1; *X* is an $n \times 2$ matrix of predictor variables with I (irradiance in arbitrary light units; e.g., µmol m^{-2} s^{-1}) in the first column and $\log_e T$ ($\log_e$°C) in the second column; β is a 2 × 1 vector of parameters to be estimated (ι [{mg L^{-1}}{µmol m^{-2} s^{-1}}$^{-1}$], ρ [{mg L^{-1}}{$\log_e$°C}$^{-1}$]); and ε is an $n \times 1$ vector of residuals that sum to zero, which are assumed to be normally and identically distributed with a mean of 0 and a variance of σ^2. Thus, after fitting the parameters in β (ι, ρ), NEP at a time step t can be estimated as:

$$NEP_t \times \Delta t = \iota \times I_t + \rho \times \log_e T_t, \tag{7}$$

where $\iota \times I_t = GPP_t \times \Delta t$, and $\rho \times \log_e T_t = R_t \times \Delta t$. This formulation of respiration reflects the dependency of biochemical reaction rates on temperature via the Boltzmann-Arrhenius equation, in particular the relatively greater temperature dependence of respiration relative to photosynthesis as well as the empirical evidence for the importance of temperature to respiration rates across ecosystems, especially in aquatic ecosystems (Yvon-Durocher et al. 2012). Daily rates of metabolism (as O_2 in mg L^{-1} d^{-1}) were calculated by averaging NEP_t, GPP_t, and R_t across all t, then multiplying each average by *N*.

Maximum likelihood estimation

The function *metab.mle* implements both pure observation and process error dynamic linear regression models to estimate metabolism by finding the parameter set that corresponds to the maximum likelihood of the model given the data (Hanson et al. 2008, Solomon et al. 2013). A simplified representation of DO (α) dynamics is:

$$\alpha_{\mathrm{t}} = \alpha_{\mathrm{t-1}} + \iota \times \mathrm{I}_{\mathrm{t-1}} + \rho \times \log_{\mathrm{e}}\mathrm{T}_{\mathrm{t\text{-}1}} + F^{*}_{\mathrm{t\text{-}1}} + \varepsilon_{\mathrm{t}}. \quad (8)$$

In this representation, parameters and notation follow those used in equation 7 for *metab.ols*: ι is a parameter describing GPP per unit of incoming light, I is incoming light (arbitrary light units; e.g., µmol m^{-2} s^{-1}), ρ is a parameter describing average rate of respiration per natural log of water temperature, and $\log_{\mathrm{e}}\mathrm{T}$ is the natural log of water temperature (°C) measured at the same depth of the DO observations. In the process error version of *metab.mle*, $\alpha_{\mathrm{t-1}}$ is set to observed value ($\mathrm{DO}_{\mathrm{t-1}}$), and ε_{t} is the process error ($\varepsilon \sim N(0,\sigma^2)$). In the observation error version, $\alpha_{\mathrm{t\text{-}1}}$ is set to the last modeled value, α_{o} is fitted as an unknown parameter, and ε_{t} then becomes observation error. F^* is the discrete atmospheric gas exchange (O_2; mg L^{-1}), calculated using a reformulation of equation 2, where DO_{t} is replaced by α_{t}:

$$F_{\mathrm{t}}^{*} = \frac{k_{\mathrm{t}} \times \Delta \mathrm{t}}{z_{\mathrm{t}}} \times \left(O_{s,\mathrm{t}} - \alpha_{\mathrm{t}}\right). \quad (9)$$

Unlike *metab.ols*, the regression equation in *metab.mle* does not contain a term for DO observations; the DO_{t} values from equations 1 and 6 are replaced by α_{t}, which is the model-estimated concentration of oxygen. Thus, the current estimate of oxygen is a function of the previous estimate, not of the observation at the previous time step. To yield more accurate metabolism estimates when Δt is large or when F^* is a significant portion of the DO mass balance, the gas exchange term can be solved in continuous time, and the process model (equation 8) re-expressed as:

$$\alpha_{\mathrm{t}} = \alpha_{\mathrm{t}} \times k_{\mathrm{t\text{-}1}} + -\mathrm{e}^{-k_{\mathrm{t\text{-}1}}} \times \alpha_{\mathrm{t}} \times k_{\mathrm{t\text{-}1}} + \mathrm{e}^{-k_{\mathrm{t\text{-}1}}} \times \alpha_{\mathrm{t\text{-}1}} + \varepsilon_{\mathrm{t}} \quad (10)$$

$$\alpha_{\mathrm{t}} = \iota \times I_{\mathrm{t-1}} + \rho \times \log_{\mathrm{e}}\mathrm{T}_{\mathrm{t\text{-}1}} + k_{\mathrm{t\text{-}1}} \times O_{s,\mathrm{t-1}}, \quad (11)$$

where k is the gas exchange coefficient. The negative log likelihood (L) of the model given the data (DO observations) was calculated from:

$$L = \sum_{\mathrm{t}=1}^{N} \frac{1}{2} \log_{\mathrm{e}}(2\pi\sigma^2) + \frac{1}{2\sigma^2}(\mathrm{DO}_{\mathrm{t}} - \alpha_{\mathrm{t}})^2, \quad (12)$$

an algebraic rearrangement of the normal probability density function suited to calculating the likelihood of N discrete observations. The observations are DO_{t}, the variance of the distribution (σ^2) is the variance of the residuals (ε) in equation 8, and the means (α_{t}) are oxygen estimates. To find the parameter estimates (and therefore the values of σ^2 and αt) that minimize L (and maximize the likelihood), LakeMetabolizer uses the *optim* function (from R's stats package; Nelder-Mead routine). Daily rates of GPP and R are calculated as in *metab.ols*, and NEP is calculated as the sum of GPP and R.

Kalman filter

Like *metab.ols* and *metab.mle*, *metab.kalman* is a metabolism model that includes process error and fits parameters, and, like *metab.mle*, it fits these parameters using maximum likelihood estimation (Batt and Carpenter 2012). However, in addition to error derived from the differences between the true data generating process and the process defined in the metabolism model (process error), error can also result from inaccuracies in the observations of DO (often manifesting as noisy observations). This second type of error is called observation error. Process errors propagate throughout a time series because at each time step they are added to the process generating DO data, and part of that process includes past DO data. By contrast, observation errors do not propagate in such a manner and can be thought of as noise added to the DO data after they are generated. A model that makes this distinction between error types must explicitly consider both the true state of the system (which is unknown but would be the same as the observed state if observations had zero error) and the observed state of the system (which we measure). In such a model, the likelihood of the values of estimated parameters given the data can be computed using a Kalman filter (Kalman 1960, Harvey 1990). The negative log likelihood function in *metab.kalman* involves 2 key sets of equations that describe the process and observation components of the model:

$$y_{\mathrm{t}} = \alpha_{\mathrm{t}} + \eta_{\mathrm{t}};\ \eta \sim N(0,H) \quad (13)$$

$$\alpha_{\mathrm{t|t\text{-}1}} = \alpha_{\mathrm{t\text{-}1}} + \iota \times \mathrm{I}_{\mathrm{t\text{-}1}} + \rho \times \log_{\mathrm{e}}\mathrm{T}_{\mathrm{t\text{-}1}} + F^{*}_{\mathrm{t\text{-}1}} + \varepsilon_{\mathrm{t}};\ \varepsilon \sim N(\varnothing,Q), \quad (14)$$

where y is observed DO; α is the true value of DO; η are observation errors; H is the variance of η; ι, and ρ are parameters to be estimated (as in *metab.mle*); T is water temperature; F^* is discrete atmospheric gas exchange; ε is process error; and Q is the variance of ε. The subscript t|t−1 indicates that the estimates of α are only based on observations of y up to $y_{\mathrm{t-1}}$ and have not been updated to reflect information gained by y_{t}. Written in a form that expands F^* and solves for gas exchange in continuous time, the process equation becomes:

$$\alpha_{\mathrm{t|t\text{-}1}} = \alpha_{\mathrm{t}} \times k_{\mathrm{t\text{-}1}} + -\mathrm{e}^{-k_{\mathrm{t\text{-}1}}} \times \alpha_{\mathrm{t}} \times k_{\mathrm{t\text{-}1}} + e^{-k_{\mathrm{t\text{-}1}}} \times \alpha_{\mathrm{t\text{-}1}} + \varepsilon_{\mathrm{t}} \qquad (15)$$

$$\alpha_{\mathrm{t}} = \iota \times \mathrm{I}_{\mathrm{t\text{-}1}} + \rho \times \log_{\mathrm{e}}\mathrm{T}_{\mathrm{t\text{-}1}} + k_{\mathrm{t\text{-}1}} \times O_{s,\mathrm{t\text{-}1}}. \qquad (16)$$

In addition to making predictions for the system state at each time step, the Kalman filter also makes predictions of the error covariance matrix, P, which is a measure of the accuracy of the estimate of the state in equation 13:

$$P_{\mathrm{t|t-1}} = P_{\mathrm{t-1}}\left(\frac{k_{t-1}}{z_{t-1}}\right)^2. \qquad (17)$$

To incorporate the new information gained from the current observation of DO, y_{t}, the Kalman filter updates the estimates of the predicted values by accounting for the current observation and the relative uncertainty surrounding the predictions and the observations. This process is akin to a weighted average of the prediction and the observation using precision (inverse of variance) as the weights. The updating equations are:

$$E_{\mathrm{t}} = P_{\mathrm{t|t\text{-}1}} + H, \qquad (18)$$

$$\alpha_{\mathrm{t}} = \alpha_{\mathrm{t|t-1}} + \frac{P_{\mathrm{t|t-1}}(y_{\mathrm{t}} - \alpha_{\mathrm{t|t-1}})}{\mathrm{E_t}}, \qquad (19)$$

$$P_{\mathrm{t}} = P_{\mathrm{t|t-1}} - \frac{E_{\mathrm{t}}}{P_{\mathrm{t-1}}^2}, \qquad (20)$$

where E_{t} is the total variance of model error at time t (process and observation error variance). In this implementation of the Kalman filter, *metab.kalman* initiates $P_{\mathrm{t=1}}$ with Q, and $\alpha_{\mathrm{t=1}}$ with $y_{\mathrm{t=1}}$. The parameters to be estimated are Q, H, ι, and ρ. The negative log likelihood, L, of the parameter estimates (model) given the data is:

$$L = \textstyle\sum_{\mathrm{t=1}}^{N} \frac{1}{2}\log_{\mathrm{e}}(2\pi) + \frac{1}{2}\log_{\mathrm{e}}(E_{\mathrm{t}}) + \frac{1}{2E_{\mathrm{t}}}(y_{\mathrm{t}} - \alpha_{\mathrm{t|t-1}})^2. \qquad (21)$$

After fitting parameters, a set of equations similar to those used in the Kalman filter can be used to smooth the DO time series. The Kalman smoother works by using the same observation, prediction, and updating equations as the Kalman filter but includes an additional smoothing step at each iteration:

$$P_{\mathrm{t}}^{*} = P_{\mathrm{t}} \times \frac{k_{\mathrm{t+1}}}{z_{\mathrm{t+1}}} \times P_{\mathrm{t|t+1}}, \qquad (22)$$

$$\alpha_{\mathrm{t}}^{*} = \alpha_{\mathrm{t}} + P_{\mathrm{t}}^{*} \times (\mathrm{\alpha}_{\mathrm{t+1}}^{*} - \alpha_{\mathrm{t|t\text{-}1}}), \qquad (23)$$

where $\alpha^*{}_{\mathrm{t}}$ is the smoothed estimate of DO at time t. This smoothed time series is different from a time series of predicted values, such as may be produced by *metab.ols* or *metab.mle*, because the smoothed values are weighted between the process and the observations. For example, when observation variance, H, is high relative to the state uncertainty, P, then the smoother will yield values that adhere more closely to the model estimates than to the observed data.

Bayesian

The final metabolism model in LakeMetabolizer is *metab.bayesian*, which invokes a Bayesian statistical philosophy to estimate metabolism parameters (Holtgrieve et al. 2010, 2013). The Bayesian model has the same underlying model structure as *metab.kalman* and includes both process and observation error. As was the case in the Kalman filter model, *metab.bayesian* distinguishes among 3 categories of DO values: y represents the observations of DO that come from sensor measurements, α the true but unknown values of DO, and α^* the model's estimates of the true value. The *metab.bayesian* models observations of DO as random deviations from the true value of DO:

$$y_{\mathrm{t}} \sim N(\alpha_{\mathrm{t},}\tau_{v}). \qquad (24)$$

An observation of DO is normally distributed with a mean equal to its corresponding true value, and a precision (precision is the reciprocal of variance and is commonly used in Bayesian formulations of the normal distribution) of τ_v (equation 24). Thus, τ_v is the precision of the observation error (similar to H in equation 13), and the smaller τ_v is, the noisier the observations will be with respect to the true values. Note that all models in LakeMetabolizer (with the exception of *metab.bookkeep*) "fit" parameter values to data by making a comparison between observed and theoretical values of DO. In *metab.bayesian*, this comparison is made in equation 25. To make this comparison, we define a true value of DO, α_{t}, as being normally distributed with a mean α_{t}^{*} and precision τ_w:

$$\alpha_{\mathrm{t}} \sim N(\alpha_{\mathrm{t}}^{*},\tau_{w}). \qquad (25)$$

The process precision, τ_w, is analogous to the reciprocal of the process variance in *metab.kalman* (Q in equation 14). The important distinction between process and observation error is that process error at time t affects the state of the system at time t + k (k = Ø, 1, … T − t) because the state evolves dynamically, whereas observation error at time t only affects the state at the same time. Both τ_v and τ_w are given minimally informative priors that follow a gamma distribution with shape and rate parameters of 0.001 (~Γ[0.001, 0.001]). The equation structure for *metab.bayesian* is similar to equations 13–16 in *metab.kalman*; however, uncertainties are handled differently in the 2 approaches:

$$\alpha_t^* = \begin{cases} \alpha_{t-1} + \alpha_t, if\ k_t = \emptyset \\ \frac{\alpha_t}{k_{t-1}} + \frac{-e^{-k_{t-1}\times\alpha_t}}{k_{t-1}} + e^{-k_{t-1}} \times \alpha_{t-1},\ otherwise \end{cases}, \quad (26)$$

$$\alpha_t = X_{t\text{-}1} \times \beta + k_{t\text{-}1} \times O_{s,t-1}, \text{ and} \quad (27)$$

$$k_t = \frac{K_t^* \times \Delta t}{z_t}. \quad (28)$$

Equation 26 shows how α_t^* depends on the value of α at the previous time step (most clearly seen in the case where k = Ø), and thus why variability associated with τ_w propagates through the time series. In equation 28, K^*_t is a stochastic node, with $K_t^* \sim N(K_t, \sigma_\kappa)$, where the mean is the coefficient of gas exchange and the precision is supplied by the user. Although the gas exchange models do not return variance estimates to be supplied as prior information, variance may be drawn from competing models of gas exchange, for example, because competing models sometimes produce widely different estimates of K. Precision of K does not have to be supplied by the user, however, in which case the default is $1/(0.1 \times K_t)$. Just as the gas exchange coefficient is prior information for K, the user can also supply prior information for the fitted GPP and R parameters (ι and ρ), implemented in *metab.bayesian* as normal distributions about the mean, $\iota_t \sim N(\iota_t, \sigma_\iota)$ and $\rho_t \sim N(\rho_t, \sigma_\rho)$. If not specified, the priors for ι and ρ are noninformative with a mean of 0 and variance of 1×10^5. To estimate metabolism as in the other parameterized models, the median of the posterior of the parameters is used to represent the posterior estimate as a scalar. The medians of the posteriors are used to estimate GPP and R as in previous models. The uncertainty of GPP and R is returned to the user as the standard deviation and is expressed as the square root of the variance of the posterior of the respective parameter estimates (ι and ρ), multiplied by the square of the corresponding covariate (irradiance and $\log_e$ [water temperature]). Similarly, NEP is calculated as the sum of GPP and R, and the standard deviation of NEP is the square root of the sum of the variances of GPP and R. The full set of posterior draws is also returned to allow the user to check for model convergence and examine the distribution of posterior draws.

Gas transfer coefficient models

A large number of models have been published that estimate gas flux. An overview of all models currently included with LakeMetabolizer (Table 2) shows the required input data for each. All models of gas exchange return a k_{600} value, a gas exchange normalized to a Schmidt number (Sc) of 600, or CO_2 and O_2 at 20 and 17.5 °C, respectively.

Logarithmic wind speed profile

For consistency across different systems, wind speed used to estimate gas flux is normalized to $U10$, which is the wind speed at a height of 10 m above the water surface. In most situations, it is prohibitively difficult or at least impractical to build a 10 m tall tower on a lake, and as such, many over-lake wind speed measurements are observed at heights between 1 and 3 m. To estimate $U10$ from the lower height measurements, a neutrally stable boundary layer assumption can be used (Arya 1988). Using this assumption, $U10$ is estimated as:

$$U10 = \text{Wind} \times \left(\frac{10}{\text{H}_{\text{wind}}}\right)^{\frac{1}{7}}, \quad (29)$$

where H_{wind} is the height of the wind observation in meters and "Wind" is the observed wind speed to be scaled. This functionality is expressed in the LakeMetabolizer function wind.scale.

Empirical wind-based gas exchange models

Air–water gas exchange estimated using tracer gas budgets have yielded strong relationships between k and over-lake measurements of wind speed. Many of these studies resulted in the development of empirical models for k based on measured wind speed, and these models assume the only dynamic contribution to k is that from wind. Two of the most commonly cited wind-based parameterizations of k are Cole and Caraco (1998) and Crusius and Wanninkhof (2003), whereas Vachon and Prairie (2013) also include lake area effects in their wind-based k estimate.

Cole and Caraco. One of the most widely used models in lakes is based on a power relationship of k_{600} with wind speed (Cole and Caraco 1998). This work was based on the relationship between mean wind speed and k_{600} estimates using sulfur hexafluoride (SF_6) tracers across multiple lakes, formulated as:

$$k_{600} = 2.07 + (0.215 \times U10^{1.7}). \quad (30)$$

This functionality is expressed in LakeMetabolizer as *k.cole.*

Crusius and Wanninkhof. Crusius and Wanninkhof (2003) related k_{600} in lakes to wind speed using an SF_6 tracer study in a small oligotrophic lake. From their observations, they created 4 different mathematical models relating wind speed to k_{600}. Because the coefficients of the simple linear model were not reported in the manuscript, we only included 3 of the 4 forms (bilinear, power, and constant/linear). These different models can be accessed using the method parameter of the *k.crusius* function.

Vachon and Prairie. Although k~wind relationships are usually strong, several reasons explain why an exclusively wind-based gas exchange model might inaccurately estimate k. For example, in small lakes, higher wind sheltering and lower fetch reduce wave height for a given wind speed, thus reducing energy transferred from wind to waves. Seeing that lake ecosystem size may influence k~wind relationships, Vachon and Prairie (2013) developed a wind-based gas exchange model that takes into account the effect of lake ecosystem size. They developed this model empirically with gas transfer data collected from lakes varying in lake size, formulated as:

$$k_{600} = 2.51 + 1.48 \times U10 + 0.39 \times U10 \times \log_{10} \mathrm{LA}, \quad (31)$$

where LA is lake area in km^2. This gas transfer function is expressed in LakeMetabolizer as *k.vachon*.

Surface renewal gas exchange models

Surface renewal gas exchange models take into account a suite of processes that influence gas exchange other than wind. For example, transitioning from day to night causes epilimnetic heat loss, which can be an important source of turbulence in small lakes (MacIntyre et al. 2010, Read et al. 2012). The higher relative importance of convective mixing on gas exchange in small lakes has increased the popularity of surface renewal models in both gas exchange and metabolism research (Raymond et al. 2013, Dugan et al. 2016).

MacIntyre. The surface renewal model was formulated by MacIntyre et al. (2010) as:

$$k_{600} = c_1\,(\varepsilon \upsilon)^{0.25}, \quad (32)$$

where $c_1 = 1.2$ is an empirically derived constant, υ is the kinematic viscosity of water (Mays 2005), and ε is the rate of dissipation of turbulent kinetic energy, calculated following Lombardo and Gregg (1989) as:

$$\varepsilon = c_2\left(c_3 \times \beta + 1.76 \times \frac{{u_{*w}}^3}{\kappa \times z_{\mathrm{aml}}}\right), \quad (33)$$

where c_2 and c_3 are both empirically derived constants 0.84 and 0.58, respectively; u_{*w} is the water-side friction velocity (m s^{-1}), calculated using the rLakeAnalyzer package (Winslow et al. 2016); $\kappa = 0.41$ is the von Karman constant; and z_{aml} is the depth of the actively mixing layer (m). The buoyancy flux, β ($\mathrm{m}^2\ \mathrm{s}^{-3}$), is defined as:

$$\beta = \frac{g\alpha H^*}{C_{\mathrm{pw}}\rho_o}, \quad (34)$$

where $g = 9.81$ is the gravitational acceleration (m s^{-2}), α is the thermal expansion coefficient for water (°C^{-1}), H^* is the effective heat flux (Kim 1976; J $\mathrm{m}^{-2}\ \mathrm{s}^{-1}$), C_{pw} is the specific heat of water at constant pressure (4186 J kg^{-1} °C^{-1}), and ρ_{o} is the density of water (kg m^{-3}). This surface renewal model is implemented in LakeMetabolizer as *k.macIntyre*.

Heiskanen. The *k.heiskanen* function returns gas flux estimates following methods from Heiskanen et al. (2014), who derived k_{600} using a boundary layer approach that includes wind shear and buoyancy flux. They assert that their equation is a better independent method of estimating k_{600} than surface renewal models because there is no similarity scaling. However, 2 coefficients are used to calibrate the model based on eddy covariance results:

$$k_{600} = \mathrm{Sc}^{-0.5}\sqrt{(C_1 \times U)^2 + (C_2 \times w_*)^2}, \quad (35)$$

where the coefficients were fitted as $C_1 = 0.00015$ and $C_2 = 0.07$, U is the measured wind speed (m s^{-1}) scaled to a height of 10 m, Sc is the Schmidt number, and w_* is the penetrative convective velocity (m s^{-1}) as calculated by Imberger (1985). The parameter w_* is defined similarly to MacIntyre et al. (2010) earlier as:

$$w_* = \sqrt[3]{-\beta \times z_{\mathrm{aml}}}. \quad (36)$$

Heiskanen et al. (2014) also applied the MacIntyre et al. (2010) model with c_1, c_2, and c_3 dimensionless coefficients as 0.50, 0.77, and 0.3, respectively, and z_{aml} as a constant at 0.15 m.

Read. Read et al. (2012) parameterized k_{600} as a function of surface mixed layer turbulence due to the additive effects of wind shear and convection, expressed as *k.read* in LakeMetabolizer. Similar to MacIntyre et al. (2010), the surface renewal model for k_{600} follows equation 32, but the formulation of ε and the value of c_1 (referred to as μ by Read et al. 2012) differ. The coefficient c_1 was set to 0.29 following the lower bounds of estimates by Zappa et al. (2007), and ε was calculated as:

$$\varepsilon = \frac{\left(\frac{\tau_{\mathrm{t}}}{\rho_w}\right)^{3/2}}{\kappa \times \delta_\upsilon} - \beta, \quad (37)$$

where τ_{t} is the tangential shear stress and δ_υ is the thickness of the viscous sublayer (see Soloviev et al. 2007 for additional details).

Read and Soloviev. We expanded the model used in Read et al. (2012) with the addition of a breaking wave component from Soloviev et al. (2007), ε_w, as:

$$\varepsilon_w = A_p^4 \times \alpha_w \left(\frac{3}{B \times S_q}\right)^{0.5} \times \frac{\left(\frac{\mathrm{Ke}}{\mathrm{Kecr}}\right)^{1.5}}{\left(1+\frac{\mathrm{Ke}}{\mathrm{Kecr}}\right)^{1.5}} \times \frac{u_{*w} g v}{0.062\kappa \times c_T (2\pi A_w)^{1.5}} \times \frac{\rho_a}{\rho_w}, \tag{38}$$

where A_p is a weighting coefficient due to turbulence patchiness of breaking-wave generated turbulence; $\alpha_w = 100$, $S_q = 0.2$, $B = 16.6$, and $c_T = 0.6$ and are dimensionless constants defined by Soloviev et al. (2007); *Ke* and *Kecr* are the Keulegan and critical Keulegan numbers, respectively; and A_w is wave age. For further details on the parameters A_p and A_w, see Soloviev et al. (2007). The parameterization for the air–water gas exchange is then represented by the sum of the interfacial (ε) and bubble-mediated (K_b) components as:

$$k_{600} = \varepsilon + K_b, \tag{39}$$

where $\varepsilon = \varepsilon_c + \varepsilon_u + \varepsilon_w$ is the sum of convection ε_c, shear ε_u, and wave ε_w terms; and K_b is parameterized following Woolf (1997) as:

$$K_b = W \times \frac{2450}{\beta_0 \left(1 + \frac{1}{\left(14 \times \beta_0 \times Sc^{-0.5}\right)^{1/1.2}}\right)^{1.2}}, \tag{40}$$

where W is a parameterization of the whitecap fraction due to wave breaking, β_0 is the Ostwald gas solubility and *Sc* is the Schmidt number. This model is expressed in LakeMetabolizer as *k.read.soloviev*.

Converting k_{600} to a gas- and temperature-specific value

The transfer velocity estimated using any of the gas transfer models described earlier must be converted to a gas- and temperature-specific transfer velocity, which is typically O_2 for lake metabolism models. The function *k600.2.kGAS* converts to a gas- and temperature-specific transfer velocity by taking in arguments of k_{600}, water temperature, and type of gas. This function supports the calculation of 8 different gas transfer velocities (He, O_2, CO_2, CH_4, SF_6, N_2O, Ar, N_2) in water temperatures ranging from 4 to 35 °C (Raymond et al. 2012).

Estimating 100% saturation of DO from temperature, pressure, and salinity

The rate at which gas exchanges between the water and atmosphere (F) is dependent both on the piston velocity, modeled using one of the k_{600} methods described earlier and the concentration gradient between observed or modeled DO and 100% saturation of DO at a given temperature, salinity, and barometric pressure (Garcia and Gordon 1992). Deviation from saturation can be caused by physical and biological processes, such as entrainment of oxygen depleted hypolimnetic water and production of oxygen from primary producers. The function *o2.at.sat* uses water temperature measured at the oxygen sensor, barometric pressure (alternatively altitude if barometric pressure not supplied), and optional salinity to calculate O_s. The default fitting method is garcia, based on equations in Garcia and Gordon (1992), with 2 other fitting methods available if specified (benson from Benson and Krause 1984; weiss from Weiss 1970).

External Resources

To avoid duplicating existing functionality, LakeMetabolizer imports basic functions from external sources when applicable.

rLakeAnalyzer

For physically related calculations and a few helper functions, LakeMetabolizer uses functionality from rLakeAnalyzer (Winslow et al. 2016), an R translation of the Matlab-based lake physics tool (Read et al. 2011). Specifically, rLakeAnalyzer is used where water density is calculated from temperature (*ts.water.density*) and the depth of the upper mixed layer is estimated (*ts.meta.depths*). Helper functions for importing a common text-file format for time series limnological data are also recommended for use in several code examples (*load.ts*, *load.all.data*). rLakeAnalyzer is also distributed on the Comprehensive R Archive Network (CRAN), so this dependency will automatically install during the installation of LakeMetabolizer.

JAGS

Estimates of the posterior distribution of the parameters in *metab.bayesian* were sampled using Gibbs sampling implemented in Just Another Gibbs Sampler (JAGS, http://mcmc-jags.sourceforge.net/). JAGS is a free, open-source, multiplatform software external to R (not an R package) and therefore must be downloaded and installed separately. However, interfacing with JAGS in R is simplified through the use of several R packages: R2jags and rjags. Both of these packages are available on CRAN and are installed automatically as default dependencies of LakeMetabolizer.

Table 3. Required input data to calculate lake metabolism using the free-water technique, with corresponding helpful derivation or estimation functions.

Input	Details	Helpful Functions
doobs	DO concentration observations (mg L^{-1})	*load.all.data*, *load.ts* (from rLakeAnalyzer)
do.sat	Equilibrium DO concentration for specific temperature, pressure and salinity (mg L^{-1})	*o2.at.sat*
k.gas	Gas and temperature specific gas transfer coefficient (m^{-1})	*k600.2.kGAS*, k.* models
irr	Photosynthetically active radiation (typically µmol m^{-2} s^{-1})	*sw.to.par*
z.mix	Actively mixed layer depth (m)	*ts.meta.depths* (from rLakeAnalyzer)
wtr	Water temperature used to: calculate k.gas, estimate the *z*.mix, and fit respiration (°C)	*load.all.data*, *load.ts* (from rLakeAnalyzer)

Table 4. Example workflow for calculating lake metabolism estimates using LakeMetabolizer.

Step	Description
1) Data inventory	Determine which types of data and metadata are available (e.g., columns of Table 2). Make sure that data have at least the minimum level of QA/QC.
2) Consider potential methods	Compare list of data available with Table 2 and 3 to determine which model(s) are available for use. Some data can be calculated from other commonly observed data (e.g., shortwave radiation from PAR).
3) Choose models	Choose gas transfer coefficient and metabolism model.
4) Load data	Load necessary time series and metadata. *load.ts*, *load.meta*, *load.bathy*, and *load.all.data* are useful loading functions for this step.
5) Derive time series data	Additional time series data may need to be derived depending on model choice and available data (see Table 2 and 3).
5a) do.sat	Calculate do.sat using the function *o2.at.sat*.
5b) k.gas	Calculate k.gas using one of the gas transfer models and convert from k_{600} to gas-specific k.
5c) z.mix	Calculate z.mix using the imported function *ts.meta.depths* from rLakeAnalyzer.
6) Run metabolism model	Run the metabolism model using the function metab and specify the method as either bayesian, bookkeep, kalman, mle, or ols.

Example workflow

All data must be quality assured/quality controlled (QA/QC) before use in metabolism models. At a minimum, data should be cleaned of extreme errors in an automated fashion (e.g., negative DO values, data logger error codes). In addition, we strongly suggest manual inspection of all the data to identify data gaps, anomalies not detected by the automated process, sensor drift during deployment, non-real repeat values, or periods of sensor calibration. More advanced QA/QC includes gap filling techniques that attempt to recreate data not observed or poorly observed.

Estimating whole-lake metabolism

At a minimum, high-frequency DO (recommended 5 min frequency), irradiance (typically photosynthetically active radiation [PAR]), wind speed, and water temperature at the depth of the DO sensor are required for estimating metabolism using the free-water oxygen technique (Table 3). The datasets imported with LakeMetabolizer include the minimum data requirements described above as well as high-frequency temperature profiles, relative humidity, and lake metadata.

A series of steps are necessary for calculating estimates of lake metabolism (Table 4). The R code used to generate metabolism estimates and figures for Sparkling Lake is available within the package as a demo [view available demos using *demo*(package='LakeMetabolizer') R function call, and run examples using demo('fig_metab', package='Lake Metabolizer')]. From the high-frequency data, several derived data need to be calculated before running the metabolism model, including the gas exchange coefficient (k.gas), mixed layer depth (z.mix), and DO at 100% saturation (do.sat). In our metabolism example, we use the wind-based Cole and Caraco (1998) method discussed

DOI: 10.5268/IW-6.4.883

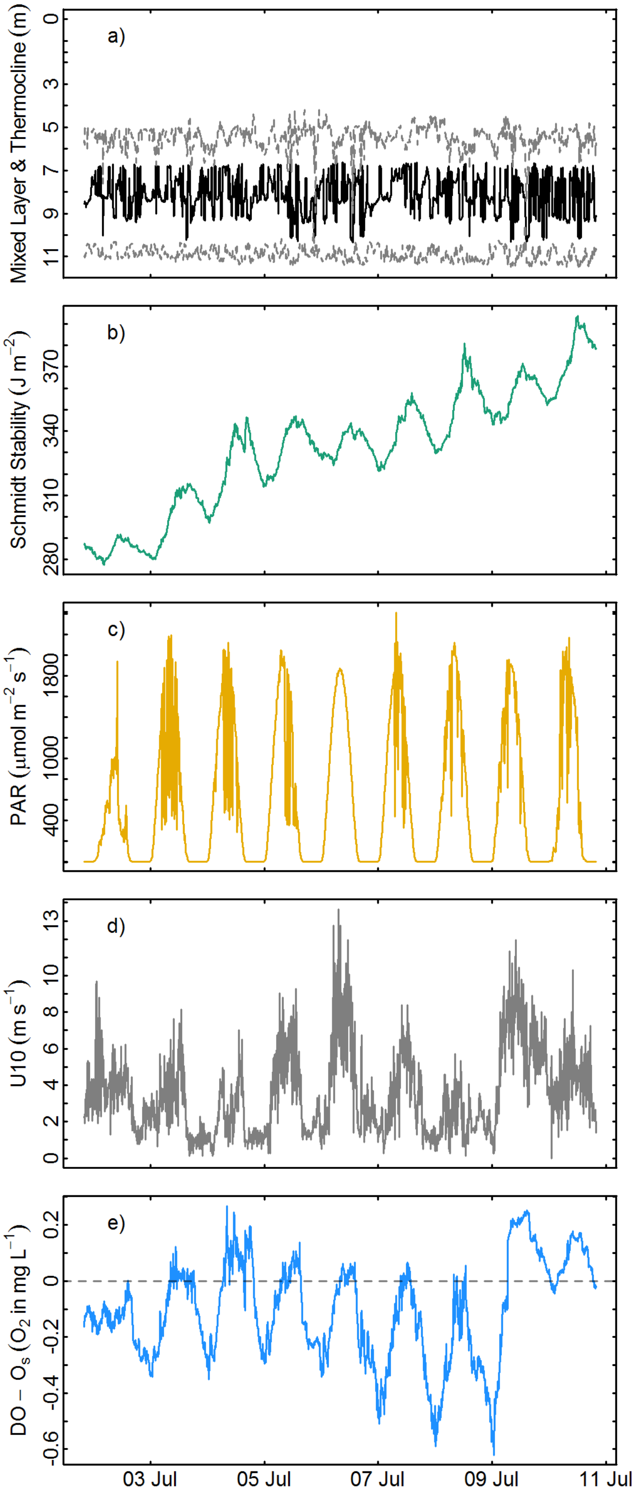

Fig. 1. Time series data for the Sparkling Lake dataset included in the LakeMetabolizer package showing variations in (a) thermocline depth (black solid line) and metalimnion top and bottom (grey dashed lines), (b) Schmidt stability, (c) photosynthetically active radiation (PAR), (d) wind speed, and (e) DO deviations from saturation.

earlier for calculating k.gas (Table 4, step 5b). The supplied wind speed data must first be scaled to 10 m from recorded wind height before using any gas exchange model, and the gas exchange coefficient must be converted from k_{600} to gas-specific k, oxygen in our case (Table 4, step 5b). Coefficient z.mix is calculated using *ts.meta.depths*, a function imported from rLakeAnalyzer (Table 4, step 5c), and do.sat is calculated using the temperature at the DO sensor depth and altitude (Table 4, step 5a). A more accurate calculation of do.sat uses barometric pressure instead of altitude; however, our example dataset does not contain barometric pressure.

After calculating these derived data, a metabolism model can be applied to each day of the time series data using the metab function, which recognizes necessary time series data based on column headings in the time series data frame, including datetime, do.obs, do.sat, k.gas, z.mix, irr, and wtr. Because all metabolism models can be run using the same input, we ran all 5 models against the example dataset.

Dataset description

An example dataset from Sparkling Lake (46.01°N, 89.70°W) is included in the package to show the necessary steps to estimate whole-lake metabolism. The example dataset is provided by the North Temperate Lakes Long Term Ecological Research Program (NTL LTER; https://lter.limnology.wisc.edu/; Magnuson and Bowser 1990). Sparkling Lake is a 64 ha, clear (mean light extinction coefficient = 0.35), deep (mean depth = 11 m), oligotrophic (mean Chl-*a* = 2.2 µg L^{-1}), seepage lake (Krabbenhoft et al. 1990).

The 9-day dataset was taken 2–10 July 2009 when Sparkling Lake was strongly stratified, with Schmidt stability steadily increasing over the time series and varying between 278 and 394 J m^{-2} (calculated using *ts.schmidt.stability* from rLakeAnalyzer). The seasonal thermocline varied between 6.6 and 10.4 m (calculated using *ts.thermo.depth* from rLakeAnalyzer; Fig. 1), and the metalimnion thickness varied between 1.5 and 7.1 m with a mean metalimnion thickness of 5.4 m (calculated using *ts.meta.depths* from rLakeAnalzyer). Wind speed exhibited a diel cycle over most days, and DO deviations from saturation displayed diel patterns consistent with physical and metabolic expectations (e.g., primary production during day increased observed DO, respiration at night decreased observed DO at night, wind-driven gas transfer equilibrated observed DO toward O_s; Fig. 1).

Table 5. Estimated average metabolism based on the 9-day Sparkling Lake example dataset. Values in parentheses are means with impossible GPP and R values first removed

Model	Mean GPP (as O_2 mg L^{-1} d^{-1})	Mean R (as O_2 mg L^{-1} d^{-1})	Mean NEP (as O_2 mg L^{-1} d^{-1})
metab.ols	0.210 (0.262)	−0.239 (−0.336)	−0.029
metab.mle	0.235 (0.284)	−0.275 (−0.275)	−0.040
metab.kalman	0.198 (0.235)	−0.291 (−0.351)	−0.093
metab.bayesian	0.210 (0.292)	−0.241 (−0.275)	−0.031
metab.bookkeep	0.185 (0.255)	−0.215 (−0.244)	−0.028

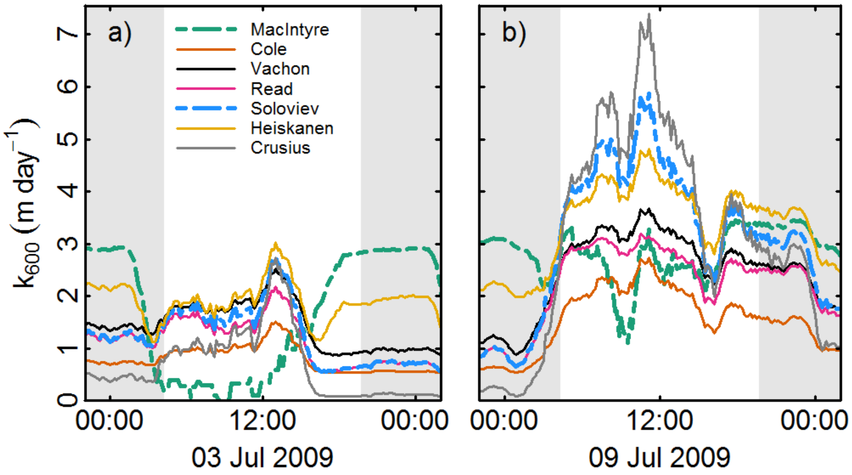

Fig. 2. Time series of estimated k_{600} values showing the results of each model for 2 days: (a) a low wind day, (b) a high wind day. A 3-hour moving average was applied to each time series to remove noise.

Results

Different gas transfer velocity models returned different estimates of k_{600} (Fig. 2), with averages ranging from a minimum of 0.47 m^{-1} for *k.vachon* to a max of 2.37 m^{-1} for *k.heiskanen*. The models based solely on wind (*k.cole*, *k.crusius*) have correlating diel patterns with different magnitudes. The other models differed in both magnitude and diel patterns, depending on the underlying model, and had values dependent on lake-specific hydrologic, morphologic, and atmospheric conditions. Although not explored here, choice of gas transfer velocity model can have a significant influence on metabolism estimates; see Dugan et al. (2016) for a more in-depth discussion of this topic.

Metabolism estimates varied among models (Table 5, Fig. 3). For both R and GPP, *metab.bookkeep* had the lowest estimates, whereas *metab.mle* and *metab.kalman* had the highest GPP and R estimates, respectively. For this 9-day dataset, NEP was not consistently positive or negative, indicating the uncertainty of such estimates, especially on short timescales. The *metab.kalma*n and *metab.bookkeep* models returned the smallest and largest NEP values, respectively.

Discussion

Code availability and distribution

To maximize usefulness and availability, we built LakeMetabolizer as a native R package and submitted it to CRAN under the name LakeMetabolizer. Through CRAN, the package can be installed using the command install.packages('LakeMetabolizer'). The code for LakeMetabolizer has been released under the GPL version 2 open-source license and is available both as a package on CRAN and under the version management repository used for development (https://github.com/GLEON/LakeMetabolizer). In the full spirit of open-source software and open science, we welcome and encourage contributions that improve or expand the package functionality.

Performance

Different components of LakeMetabolizer have different runtimes typically related to the complexity of metabolism and gas flux models used. For example, the gas flux model *k.cole* is a rapid, simple algebraic operation applied to wind speed, and as such the runtime is ~50 ms for our

9-day Sparkling Lake example (~1000 observations). In contrast, *k.read* is a complex model that must first calculate a number of subparameters before combining them to estimate a gas flux coefficient. This lengthens *k.read* runtimes to ~1.5 s for the 9-day example. Metabolism calculations have an even larger runtime range; the simplest model, *metab.bookkeep*, has a runtime of ~10 ms, and the more computationally intensive *metab.bayesian* takes closer to 45 s for the 9-day example.

For certain functions, we worked to improve performance by using external programs (like JAGS for *metab.bayesian*) or by translating key components from the interpreted R language to the compiled C language. The components in the package written in C are substantially faster than those written in R. By translating the central model loops of *metab.kalman* and *metab.mle*, the performance of these models was improved by several orders of magnitude. Although other components may benefit from the translation of pieces from R to C, we currently do not plan to add such complexity until improved performance is clearly needed.

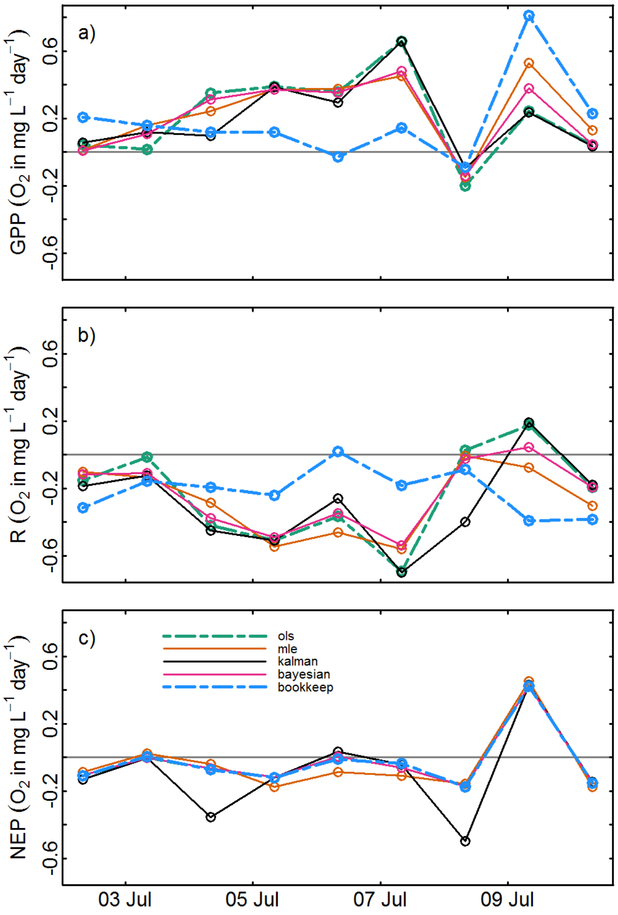

Fig. 3. Results of estimating metabolism for the Sparkling Lake example using all available models. Each line shows metabolism estimated with a different model.

Program limitations and future questions

The package estimates metabolism with the most widely used modeling techniques in the field. Although published implementations differ in a number of areas, no clear community consensus points to a single model strategy.

Dealing with unrealistic estimates

As defined, negative GPP and positive R are ecologically impossible, but unfortunately, unconstrained metabolism estimates using free-water oxygen often return negative GPP and positive R. Two general strategies exist to handle these model outputs. First, the model can be run unconstrained and the impossible estimates can be classified as nonsensical and removed from subsequent analysis. These impossible results are often from days when physical processes (e.g., wind mixing) dominate the DO signal and therefore are days when the biological signal is overwhelmed by other sources of DO variability (Rose et al. 2014). Second, the model can be written to constrain the parameters and force the estimation of positive GPP and negative R using *a priori* information about the possible values of GPP and R.

It is unclear which technique should be chosen to address impossible metabolism parameter estimates. Forcing the parameters may simply be masking a data-quality problem, returning an estimate with the correct sign while not improving accuracy. Alternatively, impossible values could arise from rough likelihood surfaces where multiple parameter sets are similarly likely but only one is possible; thus, if we incorporated this prior information about the sign of these values, we may achieve better results. Evaluating this difference is itself challenging because most alternative methods of estimating metabolism are time consuming and suffer from their own biases. Without a clear path, we chose to produce only unconstrained parameter estimates from the metabolism models. Future work is required to determine an optimal strategy.

Even if constrained or unconstrained metabolism models produce estimates with the correct sign, the model fit can be poor. In these situations, estimates of uncertainty can help guide the investigator, but there is not a consensus on an optimal strategy. Strategies include, but are not limited to, keeping all metabolism estimates except for extremely uncertain values (Solomon et al. 2013), setting more rigorous thresholds of uncertainty (Cremona et al. 2014), down-weighting poorly fit metabolism days based on uncertainty estimates (Rose et al. 2014), or fitting metabolism parameters over multiple days (Van de Bogert et al.

2012). Currently, only *metab.bayesian* returns estimates of uncertainty, but future versions of LakeMetabolizer may include bootstrapping functions as an estimate of parameter uncertainty for the other metabolism models.

Photosynthesis–irradiance relationship

All metabolism models except for the bookkeeping method estimate GPP using a linear light dependency of primary production. Although this approach may be adequate for many lakes (Hanson et al. 2008), evidence indicates that light saturation or even inhibition may more accurately model metabolism in some lakes (e.g., Brighenti et al. 2015). Integration of nonlinear primary production relationships with light may be included in later versions of the package.

Metabolism estimates using multiple sensor locations

Currently, LakeMetabolizer supports estimates of metabolism from a DO sensor at a single location. Although single-depth DO data are most common, the usefulness of vertical and horizontal integration of whole-lake metabolism calculated from concurrently deployed sensors is recognized (Coloso et al. 2008, Staehr et al. 2012a, Van de Bogert et al. 2012, Obrador et al. 2014). Future versions of the package may include calculation of whole-lake metabolism across multiple DO sensors, incorporating advective and diffusive exchange across lake layers for vertically spaced sensors, and kriging methods for interpolating horizontally spaced sensors.

Which model to use

Although metabolism estimates were similar for Sparkling Lake, some variation among the models was noted (Table 5, Fig. 3), so which model is correct? The goal of this manuscript was not to suggest one model is more correct than another, a perplexing task because validation of estimates would require a separate method of metabolism estimation, which would have its own methodological biases. Although further work is needed to establish a reliable way to assess model accuracy, we do provide some guidance to metabolism model choice because a certain model may be more appropriate over another in some situations.

The *metab.bookkeep* model may be most appropriate to use when the user wants quick computation time or is limited by time series data, the ecosystem has irregular photosynthetic–irradiance relationships, or in nontraditional diel O_2 sampling situations (e.g., Godwin et al. 2014). Because *metab.bookkeep* does not incorporate light-dependent primary production, it would be the model of choice if the user did not have light data. Additionally, *metab.bookkeep* has the fewest statistical assumptions of any of the metabolism models and likely has the simplest and most transparent interpretation. However, *metab.bookkeep* subsumes any process and observation error into the parameter estimates, which can greatly influence metabolism estimates if error is large (Batt and Carpenter 2012). Accounting for observational error helps overcome these biased estimates, as formulated in *metab.ols*, *metab.mle*, *metab.kalman*, and *metab.bayesian*. Both *metab.kalman* and *metab.bayesian* incorporate process error, and *metab.kalman* has been shown to be most useful when the DO data are noisy, for example, when estimating metabolism in the metalimnion (Batt and Carpenter 2012). The *metab.bayesian* model is unique because it also estimates the gas exchange coefficient and may be most useful in systems where gas exchange is not well known (e.g., wind speed recorded far away from the lake). Many other scenarios likely exist where using one model is advantageous over another, and we recommend the user first assess the quality of time series data (e.g., noisy DO, light-dependent photosynthesis) and the type of question to be addressed (e.g., uncertainty in estimates) before choosing a metabolism model.

We acknowledge that additional limitations of the package models not listed here are likely; however, we release this package under an open source license to maximize utility for the community of users. We strongly encourage suggestions and contributions that will improve the package.

Conclusions

We created a collection of numerical modeling tools for estimating metabolic parameters in lakes using free-water DO observations. Included is a collection of additional models for estimating input parameters that are difficult or rarely directly observed (e.g., gas flux coefficient) based on the most recently available published methodologies. LakeMetabolizer was developed for the free and open-source R scientific computing platform, making it accessible for anyone to use and available on all common platforms (Windows, OS X, Linux). All source code is open and freely available, contributing to reproducible research and scientific transparency. As with previous scientific tools, LakeMetabolizer will play a role in increasing the recognition and impact of open-source scientific software. We hope that with this package, ecologists can focus not on recreating complex metabolism models, but instead on analyzing data and knowledge creation, increasing our understanding of the metabolic distribution of lakes worldwide.

Acknowledgements

For bringing together the original group and being the impetus for the initial model writing, we thank the Global Lakes Ecological Observatory Network (GLEON) Fellowship Program (National Science Foundation DEB-1137353 and DEB-1137327). Through the GLEON fellowship program, Emily Read and Kathleen Weathers helped in supporting the group's goals and ideas. The North Temperate Lakes Long Term Ecological Research program provided open, freely available data for our examples here. We thank Gordon Holtgrieve for his review of the manuscript, assistance with the Bayesian metabolism model, and solving for continuous gas exchange. We also thank Edward Stets and an anonymous reviewer for their feedback on the manuscript and code. Any use of trade, product, or firm names is for descriptive purposes only and does not imply endorsement by the US Government. Each collaborator's contributions were supported by various sources, including National Science Foundation grants DEB-0941510, DGE-1313583, DEB-0917858, DEB-1144683, Rutgers Institute of Marine and Coastal Sciences, grant #A101058 from the Cooperative Institute of the North Atlantic Region.

References

Arya S. 1988. Introduction to micrometerology. Academic Press.

Ask J, Karlsson J, Persson L, Ask P, Byström P, Jansson M. 2009. Terrestrial organic matter and light penetration: effects on bacterial and primary production in lakes. Limnol Oceanogr. 54:2034–2040.

Batt RD, Carpenter SR. 2012. Free-water lake metabolism: addressing noisy time series with a Kalman filter. Limnol Oceanogr-Meth. 10:20–30.

Batt RD, Carpenter SR, Cole JJ, Pace ML, Johnson RA. 2013. Changes in ecosystem resilience detected in automated measures of ecosystem metabolism during a whole-lake manipulation. P Natl Acad Sci USA. 110:17398–17403.

Benson BB, Krause D. 1984. The concentration and isotopic fractionation of oxygen dissolved in freshwater and seawater in equilibrium with the atmosphere. Limnol Oceanogr. 29:620–632.

Brighenti LS, Staehr PA, Gagliardi LM, Brandão LPM, Elias EC, de Mello NAST, Barbosa FAR, Bezerra-Neto JF. 2015. Seasonal changes in metabolic rates of two tropical lakes in the Atlantic forest of Brazil. Ecosystems. 18:589–604.

Brown JH, Gillooly JF, Allen AP, Savage VM, West GB. 2004. Toward a metabolic theory of ecology. Ecology. 85:1771–1789.

Carpenter SR, Kitchell JF, Hodgson JR, Cochran PA, Elser JJ, Elser MM, Lodge DM, Kretchmer D, He X, von Ende CN. 1987. Regulation of lake primary productivity by food web structure. Ecology. 68:1863–1876.

Cremona F, Laas A, Nõges P, Nõges T. 2014. High-frequency data within a modeling framework: on the benefit of assessing uncertainties of lake metabolism. Ecol Model. 294:27–35.

Cole JJ, Caraco NF. 1998. Atmospheric exchange of carbon dioxide in a low-wind oligotrophic lake measured by the addition of SF6. Limnol Oceanogr. 43:647–656.

Cole JJ, Pace ML, Carpenter SR, Kitchell JF. 2000. Persistence of net heterotrophy in lakes during nutrient addition and food web manipulations. Limnol Oceanogr. 45:1718–1730.

Cole JJ, Prairie YT, Caraco NF, McDowell WH, Tranvik LJ, Striegl RG, Duarte CM, Kortelainen P, Downing JA, Middelburg JJ, et al. 2007. Plumbing the global carbon cycle: integrating inland waters into the terrestrial carbon budget. Ecosystems. 10:172–185.

Coloso JJ, Cole JJ, Hanson PC, Pace ML. 2008. Depth-integrated, continuous estimates of metabolism in a clear-water lake. Can J Fish Aquat Sci. 65:712–722.

Crusius J, Wanninkhof R. 2003. Gas transfer velocities measured at low wind speed over a lake. Limnol Oceanogr. 48:1010–1017.

Dugan HA, Woolway RI, Santoso AB, Corman JR, Jaimes A, Nodie ER, Patil VP, Zwart JA, Bentrup JA, Hetherington AL, et al. 2016. Consequences of gas flux model choice on the interpretation of metabolic balance across 15 lakes. Inland Waters. 6:581–592**.**

Field CB, Behrenfeld MJ, Randerson JT, Falkowski P. 1998. Primary production of the biosphere: integrating terrestrial and oceanic components. Science. 281:237–240.

Garcia H, Gordon L. 1992. Oxygen solubility in seawater: better fitting equations. Limnol Oceanogr. 37:1307–1312.

Godwin S, Kang A, Gulino LM, Manefield M, Gutierrez-Zamora ML, Kienzle M, Ouwerkerk D, Dawson K, Klieve AV. 2014 Investigation of the microbial metabolism of carbon dioxide and hydrogen in the kangaroo foregut by stable isotopeprobing. ISME J. 8:1855–1865.

Hanson PC. 2007. A grassroots approach to sensor and science networks. Front Ecol Environ. 5:343.

Hanson P, Carpenter SR, Kimura N, Wu C, Cornelius SP, Kratz TK. 2008. Evaluation of metabolism models for free-water dissolved oxygen methods in lakes. Limnol Oceanogr-Meth. 6:454–465.

Harvey AC. 1990. Forecasting, structural time series models and the Kalman filter. Cambridge University Press.

Heiskanen JJ, Mammarella I, Haapanala S, Pumpanen J, Vesala T, Macintyre S, Ojala A. 2014. Effects of cooling and internal wave motions on gas transfer coefficients in a boreal lake. Tellus B. 66:1–16.

Hoellein TJ, Bruesewitz DA, Richardson DC. 2013. Revisiting Odum (1956): a synthesis of aquatic ecosystem metabolism. Limnol Oceanogr. 58:2089–2100.

Holtgrieve GW, Arias, ME, Irvine KN, Lamberts D, Ward EJ, Kummu M, Koponen J, Sarkkula J, Richey JE. 2013. Patterns of ecosystem metabolism in the Tonle Sap Lake, Cambodia with links to capture fisheries. PLoS One. 8:e71395.

Holtgrieve GW, Schindler DE, Branch TA, A'mar ZT. 2010. Simultaneous quantification of aquatic ecosystem metabolism and reaeration using a Bayesian statistical model of oxygen dynamics. Limnol Oceanogr. 55:1047–1062.

Imberger J. 1985. The diurnal mixed layer. Limnol Oceanogr. 30:737–770.

Jansson M, Karlsson J, Blomqvist P. 2003. Allochthonous organic carbon decreases pelagic energy mobilization in lakes. Limnol Oceanogr. 48:1711–1716.

Kalman RE. 1960. A new approach to linear filtering and prediction problems. J Basic Eng. 82:35–45.

Krabbenhoft DP, Bowser CJ, Anderson MP, Valley JW. 1990. Estimating groundwater exchange with lakes 1. The stable isotope mass balance method. Water Resour Res. 26:2445–2453.

Lombardo CP, Gregg MC. 1989. Similarity scaling of viscous and thermal dissipation in a convecting surface boundary layer. J Geophys Res. 94:6273–6284.

Lovett GM, Cole JJ, Pace ML. 2006. Is net ecosystem production equal to ecosystem carbon accumulation? Ecosystems. 9:152–155.

MacIntyre S, Jonsson A, Jansson M, Aberg J, Turney DE, Miller SD. 2010. Buoyancy flux, turbulence, and the gas transfer coefficient in a stratified lake. Geophys Res Lett. 37:L24604.

Magnuson JJ, Bowser CJ. 1990. A network for long-term ecological research in the United States. Freshwater Biol. 23:137–143.

McNair JN, Gereaux LC, Weinke AD, Sesselmann MR, Kendall ST, Biddanda BA. 2013. New methods for estimating components of lake metabolism based on free-water dissolved-oxygen dynamics. Ecol Modell. 263:251–263.

Obrador B, Staehr PA, Christensen JPC. 2014. Vertical patterns of metabolism in three contrasting stratified lakes. Limnol Oceanogr. 59:1228–1240.

Odum HT. 1956. Primary production in flowing waters. Limnol Oceanogr. 1:102–117.

Porter JH, Hanson PC, Lin CC. 2012. Staying afloat in the sensor data deluge. Trends Ecol Evol. 27:121–129.

Raymond PA, Hartmann J, Lauerwald R, Sobek S, McDonald C, Hoover M, Butman D, Striegl R, Mayorga E, Humborg C, et al. 2013. Global carbon dioxide emissions from inland waters. Nature. 503:355–359.

Raymond PA, Zappa CJ, Butman D, Bott TL, Potter J, Mulholland P, Laursen AE, McDowell WH, Newbold D. 2012. Scaling the gas transfer velocity and hydraulic geometry in streams and small rivers. Limnol Oceanogr-Fluids Environ. 2:41–53.

Read JS, Hamilton DP, Jones ID, Muraoka K, Winslow LA, Kroiss R, Wu CH, Gaiser E. 2011. Derivation of lake mixing and stratification indices from high-resolution lake buoy data. Environ Model Softw. 26:1325–1339.

Read JS, Hamilton DP, Desai AR, Rose KC, MacIntyre S, Lenters JD, Smyth RL, Hanson PC, Cole JJ, Staehr PA, et al. 2012. Lake-size dependency of wind shear and convection as controls on gas exchange. Geophys Res Lett. 39:L09405.

Rose KC, Winslow LA, Read JS, Read EK, Solomon CT, Adrian R, Hanson PC. 2014. Improving the precision of lake ecosystem metabolism estimates by identifying predictors of model uncertainty. Limnol Oceanogr-Meth. 12:303–312.

Solomon C, Bruesewitz D, Richardson DC, Rose KC, Van de Bogert MC, Hanson PC, KRatz TK, Larget B, Adrian R, Babin BL, et al. 2013. Ecosystem respiration: drivers of daily variability and background respiration in lakes around the globe. Limnol Oceanogr. 58:849–866.

Soloviev A, Donelan M, Graber H, Haus B, Schlüssel P. 2007. An approach to estimation of near-surface turbulence and CO_2 transfer velocity from remote sensing data. J Mar Syst. 66:182–194.

Staehr PA, Bade D, Van de Bogert MC, Koch GR, Williamson C, Hanson P, Cole JJ, Kratz T. 2010. Lake metabolism and the diel oxygen technique: state of the science. Limnol Oceanogr-Meth. 8:628–644.

Staehr PA, Christensen JPA, Batt R, Read J. 2012a. Ecosystem metabolism in a stratified lake. Limnol Oceanogr. 57:1317–1330.

Staehr PA, Testa JM, Kemp WM, Cole JJ, Sand-Jensen K, Smith SV. 2012b. The metabolism of aquatic ecosystems: history, applications, and future challenges. Aquat Sci. 74:15–29.

Vachon D, Prairie Y. 2013. The ecosystem size and shape dependence of gas transfer velocity versus wind speed relationships in lakes. Can J Fish Aquat Sci. 70:1757–1764.

Van de Bogert MC, Bade DL, Carpenter SR, Cole JJ, Pace ML, Hanson PC, Langman OC. 2012. Spatial heterogeneity strongly affects estimates of ecosystem metabolism in two north temperate lakes. Limnol Oceanogr. 57:1689–1700.

Weiss R. 1970. The solubility of nitrogen, oxygen and argon in water and seawater. Deep Sea Res Pt I. 17:721–735.

Winslow LA, Read J, Woolway R, Brentrup J, Leach T, Zwart J. 2016. rLakeAnalyzer:1.8.3 Standardized methods for calculating common important derived physical features of lakes. doi:10.5281/zenodo.58411

Woolf D. 1997. Bubbles and their role in gas exchange. In: Liss PS, Duce RA, editors. The sea surface and global change. Cambridge University Press. p. 173–206.

Yvon-Durocher G, Caffrey JM, Cescatti A, Dossena M, Del Giorgio P, Gasol JM, Montoya JM, Pumpanen J, Staehr PA, Trimmer M, et al. 2012. Reconciling the temperature dependence of respiration across timescales and ecosystem types. Nature. 487:472–476.

Zappa CJ, McGillis WR, Raymond PA, Edson JB, Hintsa EJ, Zemmelink HJ, Dacey JWH, Ho DT. 2007. Environmental turbulent mixing controls on air-water gas exchange in marine and aquatic systems. Geophys Res. Lett. 34:1–6.

DOI: 10.5268/IW-6.4.883