# NETLAKE toolbox for the analysis of high-frequency data from lakes

## Factsheet #4

## Lake Metabolizer

**R. Iestyn Woolway**

### *Objective*

Metabolism is a fundamental ecological process that occurs at scales ranging from individual organisms to whole ecosystems. Whole ecosystem metabolism represents the balance between carbon fixation (gross primary production; GPP) and biological carbon oxidation (ecosystem respiration; R) in an ecosystem. At an ecosystem scale, metabolism estimates provide insight into the support of food webs through primary productivity, rates of carbon accumulation or loss in an ecosystem, and anticipating changes in ecosystem state. Lake metabolism can be estimated from high frequency free-water dissolved oxygen (DO) concentrations (e.g., Staehr et al. 2010). The value of quantifying lake metabolism and the availability of the necessary data has led to a rapid proliferation of computational methodologies for estimating metabolism. While technological advances in automated sensors and the expansion of cross-site collaborations have increased greatly the accessibility of high frequency DO time series, barriers are presented by the statistics, programming, and multitude of models used to convert sensor observations into estimates of lake metabolism. This analytical barrier may be overcome by the use of a new RPackage called Lake Metabolizer, which is designed to estimate lake metabolism from commonly collected sensor data.

Lake Metabolizer is an Rpackage for estimating lake metabolism and related terms from data collected by high frequency, *in situ* lake monitoring stations with relative ease. The package can be used to calculate lake metabolism using five different methods: bookkeeping, ordinary least squares, maximum likelihood, Kalman filter, and Bayesian (Table 1). For further information of the differences between the metabolism models, see Winslow et al. (*in press*) and Honti (2016). In addition, each of these five methods can be combined with one of seven models for computing the gas transfer coefficient, which influences the rate of gas exchange at the air-water interface. Lake Metabolizer also includes a number of functions that compute conversions and calculations that are commonly applied to raw data prior to estimating lake metabolism (e.g. optical conversion models). This package contains example data, example use-cases, and function documentation.

| Model | Underlying statistics | Error structure | Error type |
|---|---|---|---|
| **Bookkeeping** | Algebra | None | None |
| **Bayesian** | Bayesian | Gaussian | Process and Observation |
| **Kalman filter** | Maximum likelihood and Kalman filter | Gaussian | Process and Observation |
| **Maximum likelihood** | Maximum likelihood | Gaussian | Process and observation |
| **Ordinary least squares** | Linear regression | Gaussian | Observation |

**Table 1.** Table comparing the structure of the five different metabolism models included in LakeMetabolizer.

## *Specific application*

The main application of this program is the calculation of lake metabolism using a number of different approaches published in the scientific literature. An example is the calculation of net ecosystem production (NEP), which is the difference between GPP and R, and is used to delineate heterotrophic systems (negative NEP) from autotrophic systems (positive NEP). Example output calculations for NEP and the gas transfer coefficient (k600, which estimates the amount of gas exchange at the air-water interface) for Sparkling Lake are shown in Figures 1 and 2 below. The example dataset from Sparkling Lake is included in the package and can be accessed in 'R'. In addition, the 'R' code used to generate metabolism estimates and figures for Sparkling Lake is available within the package as a demo (access using *demo(package='LakeMetabolizer')* 'R' function call).

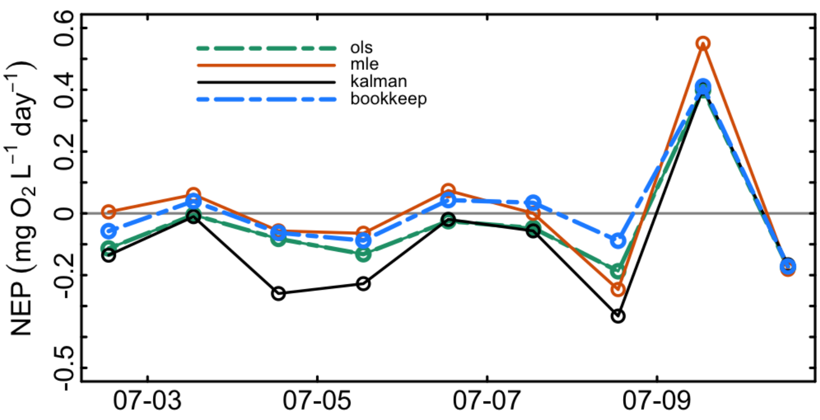


**Figure 1.** Comparison of four different metabolism models (OLS = ordinary least squares; MLE = maximum likelihood; Kamlan = Kalman filter; bookkeep = Bookkeeping) for estimating Net Ecosystem Production (NEP).

As all methods can be run using the same input files, Lake Metabolizer allows comparisons between methods. For example, in Figure 1 we can see that each of the methods can return different estimates, where even the sign of NEP can vary between the different methods. Furthermore, using the example dataset provided we see that the different gas transfer coefficient models can return very different estimates of k600 (Fig. 2); see Dugan et al. (*in* *press*), with averages ranging from a minimum of approximately 0.5 m day$^{-1}$ to a maximum of approximately 3 m day$^{-1}$. Lake Metabolizer provides a means of estimating lake metabolism and related terms using a consistent method, thereby facilitating global comparisons of high frequency data from lake buoys.

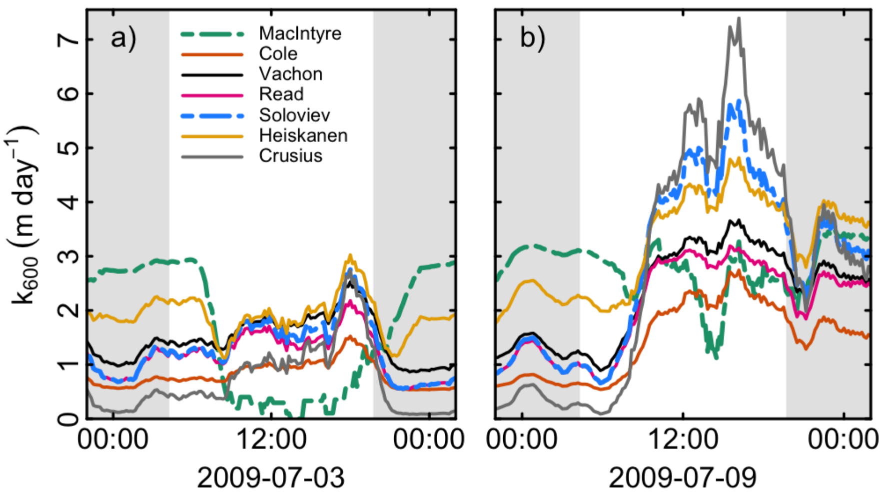


**Figure 2.** Comparison of the seven different gas transfer coefficient models included in the Lake Metabolizer package. Grey regions illustrate night-time, which can also be estimated by the Lake Metabolizer package (i.e. sun rise and sun set times).

## *Background*

The package requires some experience of using 'R'. However, the user manual (see link below) does provide a number of examples for using the different functions.

Having these methods in an 'R' environment allows them to be calculated with relative ease. However, while the tool can be used without having any prior knowledge of lake metabolism, interpretation of the results does require some understanding of the principles behind aquatic metabolism.

## *Type of data and requirements*

At a minimum, high frequency DO (at least hourly observations), 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 - see Staehr et al. 2010). However, to use all of the available gas transfer coefficient models, the user will need additional data. The data required for each gas transfer coefficient model are shown in Table 2.

| | k.cole (Cole and Caraco 1998) | k.crusius (Crusius and Wanninkhof 2003) | k.vachon (Vachon and Prairie 2013) | k.heiskanen (Heiskanen et al. 2014) | k.macIntyre (MacIntyre et al. 2010) | k.read (Read et al. 2012) | k.read.soloviev (Read et al. 2012; Soloviev et al. 2007) |
|---|---|---|---|---|---|---|---|
| **Wind speed** | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ |
| **Air temperature** | | | | | ✓ | ✓ | ✓ |
| **Relative humidity** | | | | ✓ | ✓ | ✓ | ✓ |
| **Short-wave radiation** | | | | ✓ | ✓ | ✓ | ✓ |
| **Water temperature profile** | | | | ✓ | ✓ | ✓ | ✓ |
| **Wind height** | | | | | ✓ | ✓ | ✓ |
| **Atmospheric pressure** | | | | ✓ | ✓ | ✓ | ✓ |
| **Net Long-wave radiation** | | | | ✓ | ✓ | ✓ | ✓ |
| **Latitude** | | | | | | ✓ | ✓ |
| **Area** | | | ✓ | | | ✓ | ✓ |
| **Wind height** | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ |

**Table 2.** Data required for each gas transfer coefficient model included in Lake Metabolizer. References for the most relevant publication associated with each gas transfer coefficient is provided in brackets. k.read.soloviev is a new gas transfer coefficient model used by Dugan et al (*in press*) where the model of Read et al. 2012 is modified to include the influence of breaking waves, from Soloviev et al. 2007, on the gas transfer coefficient.

Formatting of the input files is detailed in the user manual. Note that the formatting of the input files is important, as the functions used by the package to load the data assumes that the user has followed the examples provided in the user manual. For example, DO data must be formatted as a tab-delimited text file as:

```
dateTime            doobs_0.5
2009-07-23 00:00          13.550
2009-07-23 00:01          13.493
2009-07-23 00:02          13.458
```

This file format is the same as that required by Lake Heat Flux Analyzer (Woolway et al. 2015, see Jones 2016) and Lake Analyzer (Read et al. 2011), thus allowing them to be used by a number of programs to provide specific details of the lake.

## *Basic procedures*

The procedure to follow is detailed in the user manual of the 'R' package for Lake Metabolizer (see link below), and differs depending on the chosen model. Only a brief synopsis is given here:

1. Collect and clean high frequency data (see de Eyto and Pierson 2016).
2. Determine which types of data and metadata are available (e.g. wind speed, air temperature, short-wave radiation, lake latitude, lake area, etc.).

3. Compare list of data available to determine which model(s) are available for use (see user manual).
4. Choose gas transfer coefficient and metabolism methods for estimating metabolism and related variables.
5. Load necessary time series and metadata in 'R' using the helper functions provided (see user manual)
6. Run metabolism model using the helper function for that particular model.

## *Pitfalls and tips*

- The package estimates metabolism with the most widely used modelling techniques. However, there are a number of areas where implementation differs and it is unclear if there is community consensus that point to a single model strategy.
- As defined, negative Gross Primary Production (GPP) and positive Respiration (R) are ecologically impossible. Unfortunately, unconstrained metabolism estimates using free-water oxygen can return negative GPP and positive R. There are generally two strategies for handling such model output, (i) the model can be run unconstrained and the impossible estimates can be removed, and (ii) the model can be written to constrain the parameters and force the estimation of positive GPP and negative R.
- All methods, except for the bookkeeping method, estimate GPP using a linear light dependency of primary production. Although this approach may be adequate for many lakes, there is evidence that light saturation or even inhibition may more accurately model metabolism in some lakes. Integration of non-linear primary production relationships with light may be included in later versions of the package.
- Currently, LakeMetabolizer supports estimates of metabolism from a surface DO sensor at a single location. Future versions of the package may include calculation of whole-lake metabolism across multiple DO sensors (Obrador et al. 2014, see Obrador et al. 2016).

## *Further reading*

**Key References:**

The reference for the paper describing the code and its uses is:

Winslow, L. A., Zwart, J. A., Batt, R. D., Dugan, H. A., Woolway, R. I., Corman, J. R., Hanson, P. C., Read, J. S. LakeMetabolizer: An R package for estimating lake metabolism from free-water oxygen using diverse statistical models. *Inland Waters* (in press)

**Lake Metabolizer Manual:**

http://cran.r-project.org/web/packages/LakeMetabolizer/LakeMetabolizer.pdf

**Other references:**

For an example of lake metabolism being calculated and used see:

Dugan, H.A., Woolway, R.I., Santoso, A.B., Corman, J.R., Jaimes, A., Nodine, E.R., Patil, V.P., Zwart, J.A., Brentrup, J.A., Heatherington, A.L., Oliver, S.K., Read, J.S., Winters, K.M., Hanson, P.C., Read, E.K., Winslow, L.A., Weathers, K.C. Consequences of gas flux model choice on the interpretation of metabolic balance across 15 lakes. *Inland Waters* (in press)

Other useful references for lake metabolism and the gas transfer coefficients are:

Batt, R.D., Carpenter, S.R. 2012. Free-water lake metabolism: addressing noisy time series with a Kalman filter. *Limnology and Oceanography Methods* 10: 20-30.

Cole, J.J., Caraco, N.F. 1998. Atmospheric exchange of carbon dioxide in a low-wind oligotrophic lake measured by the addition of SF6*. Limnology and Oceanography* 43, 647-656.

Crusius, J., Wanninkhof, R. 2003. Gas transfer velocities measured at low wind speed over a lake. *Limnology and Oceanography* 48: 1010-1017.

de Eyto, E., Pierson, D. 2016. Data handling: cleaning and quality control. In Obrador, B., Jones, I.D. and Jennings, E. (Eds.) *NETLAKE toolbox for the analysis of high-frequency data from lakes* (Factsheet 1). Technical report. NETLAKE COST Action ES1201. pp. 2-6. http://eprints.dkit.ie/id/eprint/532.

Heiskanen, J.J., Mammarellam 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: Chemical and Physical Meteorology* 66: 1-16.

Holtgrieve, G.W., Schindler, D.E., Branch, T.A., A'Mar, T. 2010. Simultaneous quantification of aquatic ecosystem metabolism and reparation using a Bayesian statistical model of oxygen dynamics. *Limnology and Oceanography* 55: 1047-1063.

Honti, M. 2016. Bayesian calibration of mechanistic models of lake metabolism. In Obrador, B., Jones, I.D. and Jennings, E. (Eds.) *NETLAKE toolbox for the analysis of high-frequency data from lakes* (Factsheet 8). Technical report. NETLAKE COST Action ES1201. pp. 40-46. http://eprints.dkit.ie/id/eprint/539.

Jones, I.D. 2016. Lake Heat Flux Analyzer (LHFA). In Obrador, B., Jones, I.D. and Jennings, E. (Eds.) *NETLAKE toolbox for the analysis of high-frequency data from lakes* (Factsheet 3). Technical report. NETLAKE COST Action ES1201. pp. 11-15. http://eprints.dkit.ie/id/eprint/534.

MacIntyre, S., Jonsson, A., Jansson, M., Aberg, J., Turney, D.E., Miller, S.D. 2010. Buoyancy flux, turbulence, and the gas transfer coefficient in a stratified lake. *Geophysical Research Letters* 37 doi: 10.1029/2010GL044164.

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

Obrador, B., Staehr, P.A., Christensen, J. 2014. Vertical patterns of metabolism in three contrasting stratified lakes. *Limnology and Oceanography* 59: 1228-1240.

Obrador, B., Christensen, J., Staehr, P.A. 2016. Determination of whole-column metabolism from profiling data. In Obrador, B., Jones, I.D. and Jennings, E. (Eds.) *NETLAKE toolbox for the analysis of high-frequency data from* lakes (Factsheet 9). Technical report. NETLAKE COST Action ES1201. pp. 47-51. http://eprints.dkit.ie/id/eprint/540.

Read J.S., Hamilton D.P., Jones I.D., Muraoka K., Kroiss R., Wu C.H., Gaiser E. 2011. Derivation of lake mixing and stratification indices from high-resolution lake buoy data. *Environmental*

*Modelling and Software* 26: 1325–1336.

Read, J.S., Hamilton, D.P., Desai, A.R., Rose, K.C., MacIntyre, S., Lenters, J.D., Smyth, R.L., Hanson, P.C., Cole, J.J., Staehr, P.A., Rusak, J., Pierson, D., Brookes, J., Laas, A., Wu, C. 2012. Lake-size dependency of wind-shear and convection as controls on gas exchange. *Geophysical Research Letters* 39: L09405.

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. *Journal of Marine Systems* 66: 182–194.

Staehr, P.A., Bade, D., van de Bogert, M.C., Koch, G.R., Williamson, C., Hanson, P., Cole, J.J., Kratz, T. 2010. Lake metabolism and the diel oxygen technique: State of the science. *Limnology and Oceanography Methods* 8: 628–644

Vachon, D., Prairie, Y. 2013. The ecosystem size and shape dependence of gas transfer velocity versus wind speed relationship in lakes. *Canadian Journal of Fisheries and Aquatic Sciences* 70: 1757-1764.

Woolway, R.I., Jones, I.D., Hamilton, D.P., Maberly, S.C., Muraoka, K., Read, J.S., Smyth, R.L., Winslow, L.A. 2015. Automated calculation of surface energy fluxes with high-frequency lake buoy data. *Environmental Modelling and Software* 70: 191–198.

## Code

The code for LakeMetabolizer has been released under the GPL version 2 open-source license. It is available both as an 'R' package on CRAN, using the command *install.packages('LakeMetabolizer')* and under the version management repository used for development (https://github.com/GLEON/LakeMetabolizer).

## Contact details

R. Iestyn Woolway. University of Reading, UK.
riwoolway@gmail.com

Luke A. Winslow, United States Geological Survey, WI, USA.
Jake A. Zwart, University of Notre Dame, IN, USA.

## Suggested citation

Woolway, R.I. 2016. Lake Metabolizer. In Obrador, B., Jones, I.D. and Jennings, E. (Eds.) *NETLAKE toolbox for the analysis of high-frequency data from lakes* (Factsheet 4). Technical report. NETLAKE COST Action ES1201. pp. 16-22. http://eprints.dkit.ie/id/eprint/535.

## Acknowledgement

This factsheet is based upon work from the NETLAKE COST Action ES1201, supported by COST (European Cooperation in Science and Technology).