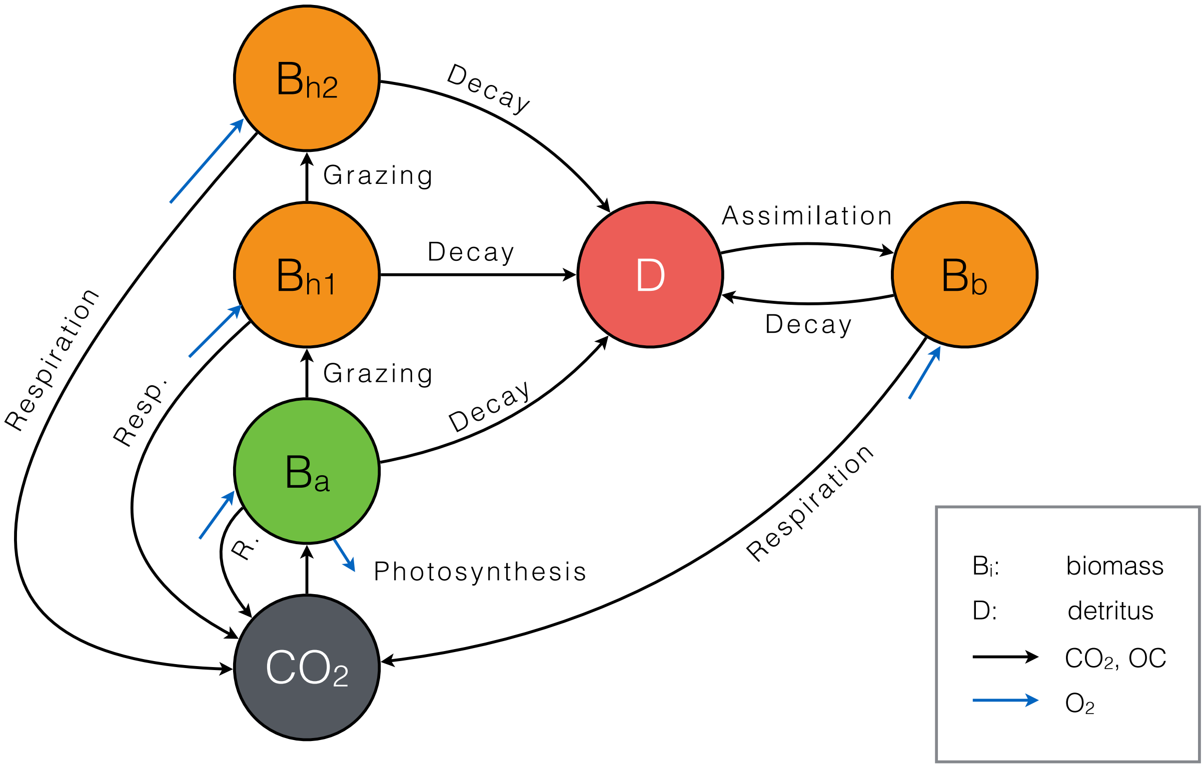
High-frequency monitoring and modeling of shallow lake metabolism

Final Report



NKFIH Grant #: 112934 PD

PI: Mark Honti

25/09/2017

Progress and Summary of Results

Operation of the monitoring station in the Keszthely Basin of Lake Balaton yielded three more years of high-frequency data of meteorological variables, water temperature, turbidity, dissolved oxygen, and delayed fluorescence, elongating the time series database to 17 years (Fig. 1 shows the first 16). This is an exceptionally long high-frequency record in international comparison (*cf.* www.gleon.org). In certain periods additional physical variables such as wave height and flow velocities were recorded, too. Weekly manual control measurements and regular maintenance twice a week contributed to good data quality and coverage.

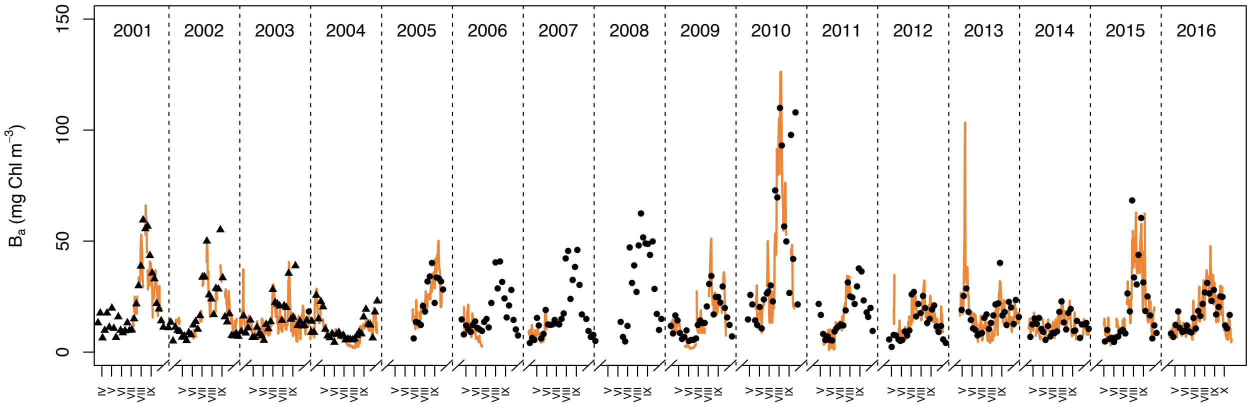


Fig. 1. Algal biomass (B_a) from 2001 to 2016 measured by on-line delayed fluorescence (orange line), and manual sampling (black symbols) in the Keszthely Basin of Lake Balaton.

Together with previous measurements, 8 years of data were available to analyze lake metabolism and other limnological topics. Raw data indicated that the monitoring site is a rather complex physical environment. Besides the expected intense resuspension, temporary stratification during the summer was often observed. Strong gradients of temperature and dissolved oxygen (DO) were typical. These circumstances left their fingerprint on the observed oxygen concentrations: about 20% of days produced intricate DO patterns that were obviously not to be reproduced by any known model of lake metabolism.

Metabolic Modeling

Discrepancies of the observed DO time-series could not be unambiguously bound to any physical boundary condition observed at the monitoring site. Neither temporary stratification nor the existence or lack of strong currents were good predictors of deviations from the expected DO profile. By measuring at a single location, it was impossible to delineate the (dynamically changing) volume, from which the DO signal has originated. This shortcoming is currently addressed in a follow-up research project [NKFIH #120551], where satellite sensor pods were installed around the monitoring site to detect the transfer time and direction of transient phenomena.

As systematic errors could not be reduced by expanding the monitoring program to basic hydrodynamic variables, these errors were treated statistically. A Bayesian representation of systematic deviations in form of an autoregressive random process (Reichert and Schuwirth 2013) was used during model calibrations (Honti et al. 2016, Istvánovics and Honti 2017). However, this alone was not sufficient to calibrate the metabolic model for a longer time-series. Common metabolic models are partially empirical in the sense that they rely on ecosystem parameters that change in time. This means that a single calibrated parameter set cannot represent periods that span over the characteristic changing time of the ecosystem. This problem is usually resolved by calibrating the metabolic model for 1-3 days wide windows (e.g. Hanson et al. 2008, Staehr et al. 2010, McNair et al. 2015), however this does not guarantee that the calibrated parameters follow credible ecological dynamics: productivity or respiration can jump over a wide domain on a day to day basis. To solve this problem, a sequential Bayesian learning algorithm was implemented in the metabolic model. Calibration took place in 3 days wide sliding windows (Fig. 2). The 3-days width was chosen to avoid parameter values that would provide a good fit to a single day of data at the price of longer-term instability and to ensure that fitting took place for a shorter time period than the typical generation times of phytoplankton (5 to 7 days). Posterior parameter distributions of the preceding window were used as priors, which meant that preceding parameter values were kept almost unchanged unless there was strong evidence in the data against them. The outcome of this sequential learning procedure was a good model fit to the observations coupled with credible, mostly gradually changing ecosystem parameters that properly reflected resilience of the ecosystem. This calibration (or 'inverse modelling') procedure was applied for data from Lake Balaton (Honti et al. 2016, Istvánovics and Honti 2017) and other lakes in Denmark Brazil and China (Honti et al. 2016, Staehr et al. 2016).

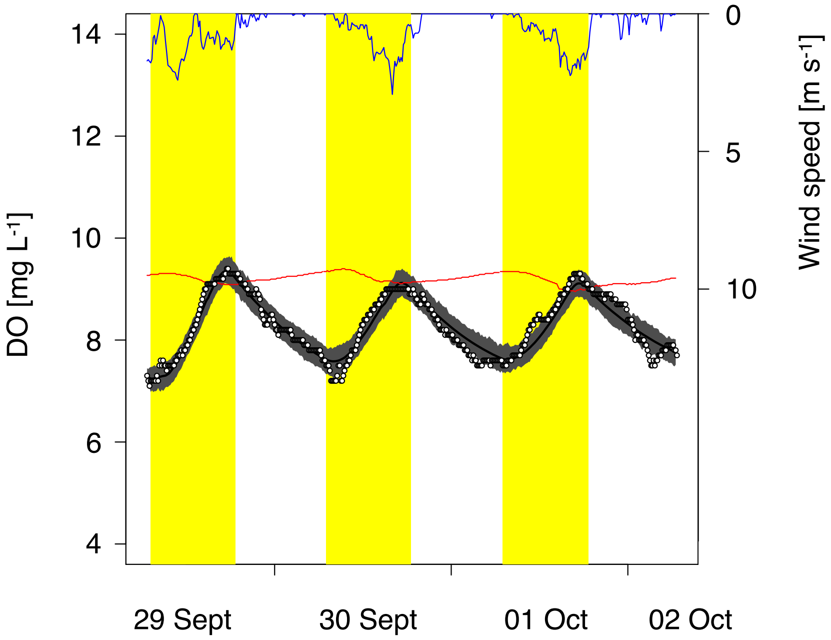


Fig. 2. An example of model fit in a 3-day calibration window. Open circles: observed DO concentration, black line with gray band: model fit with 95% uncertainty interval, red line: saturated DO concentration, blue line: wind speed, yellow bands: photoperiod.

The stability and credibility of the model was enhanced by involving delayed fluorescence data (a proxy of phytoplankton biomass and photosynthesis; Istvánovics et al. 2005, Honti and Istvánovics 2011) in case of Lake Balaton (Honti et al. 2016, Istvánovics and Honti 2017). Bayesian inference supports the assimilation of external data in the form of parameter priors that represent uncertain information about the modelled system. Phytoplankton biomass was a modelled state variable with appropriate feedback links to metabolism. Initial values for the phytoplankton biomass was a parameter with a prior based on the observed values. Modelled daily mean phytoplankton biomass was part of the calibration as well, it was compared to the observations and deviations were penalized by the parameter likelihood function. This additional criterion compromised the good model fit negligibly, indicating that after excluding the unconstrained variation in algal biomass the model still contained enough degrees of freedom.

The operative applicability of Bayesian learning in modeling lake metabolism was tested on a system that is simpler than Lake Balaton. Two months of high-frequency temperature and DO data were obtained from City Park Lake (Városligeti-tó), Budapest, Hungary. An extremely simple metabolic model was developed for the purpose of short-term operative forecasting.

Meteorological variables were restricted to a qualitative light status (dark/cloudy/bright), and water temperature. The latter did not vary too much because to lake is fed by thermal water of nearly constant temperature. As the lake has a significant nutrient load (high oxygen production by photosynthesis), and a thick organic sediment (high oxygen consumption by respiration), diurnal DO fluctuations varied between moderate to large, which facilitated testing the predictive capability of the model under both high and low DO variability. The simplified model was moved in 3-hour steps, it was trained on the preceding 24 hours of observed data and forecasted the forthcoming 24 hours in accordance with the Bayesian learning algorithm. The prediction of DO minima was reliable 6-12 hours in advance: the probability of not foreseeing a low DO event (DO<4 mg L^{-1}) was no more than 3-4%. These results show, that the Bayesian learning technique can be used to compensate for structural model deficiencies by allowing the time-variability of model parameters. The tested operative forecast method could be used in cases when short-term water quality changes can cause economic or ecological harm. Such systems include intensively farmed fishponds or hyper/eutrophic urban lakes, where the 6-12 hours forecast time-horizon is ample to intervene by changing the flow routing or activating aerators.

Error Propagation in Metabolic Modelling

The practice of metabolic modeling based on free-water DO hasn't changed much since the pioneering work of Odum (1956). The metabolic equation – when lateral and vertical transport is neglected – is:

$$\frac{dDO}{dt} = X + GPP - R \tag{1}$$

The observed change of dissolved oxygen is split amongst the unknown quantities of gas exchange with the atmosphere (X, here positive for flux from air into the water), gross primary production (GPP) and community respiration (R). As there are no periods when only one of these processes would be active, there is no way to solve this equation unambiguously. To tackle this problem, a sequential inference is usually performed. First X is estimated based on empirical models of piston velocity and the saturation deficit as the driving gradient of gas exchange. Then the remaining change of DO is attributed to net ecosystem production (NEP = GPP–R). From nighttime NEP the parameters of R are estimated and assumed to describe the daytime respiration as well. Having sorted out all but one of the right hand side components, the rest of DO change is attributed to GPP.

The algorithm of this calculation is obvious when the metabolic components are calculated manually, by e.g. the classical bookkeeping method (Odum 1956, Staehr et al. 2010, McNair 2015). However, the essence is the same when an inverse modelling technique is used: assuming a certain gas-exchange model or fixing certain model parameter values, or telling the model that GPP must be zero during darkness are all helping to reduce the degrees of freedom of the problem. While this technique (reducing degrees of freedom by making assumptions on certain quantities) is mathematically viable, it has some side-effects on the results, which have not been fully recognized and acknowledged in studies of lake metabolism.

Errors are inevitable during the reduction of degrees of freedom. Any estimate, regardless if it is on a parameter value or a mechanism, contains errors. Uncertainty partially stems from insufficient representation of the study system, that is from assumptions on certain mechanisms or fixed parameter values. Besides, uncertainty derives from the imported information: limnological studies tend to report single values or equations without mentioning the corresponding uncertainty. The latter is nicely illustrated by for example of empirical models of piston velocity that are based on meteorological boundary conditions. Epistemic uncertainty (stemming from the lack of knowledge) can be found in the sheer diversity of these models and the range occupied by their predictions for the same conditions (Dugan et al. 2016). While the individual models might be valid for the systems where they were developed, case-specific mechanisms may render them nearly useless in other systems. Besides this, metabolic calculations usually neglect the uncertainty belonging to the specific models. Cole and Caraco (1998) show that popular piston velocity models usually have an uncertainty of 20-30% based on the observations in the pilot system, yet this is not considered anywhere during metabolism modeling. In summary, the estimate of X is likely to contain an error that is comparable to its magnitude.

The estimation errors propagate through the calculations, and as the basic model (eq. (1)) is linear, they produce artificial linear correlations between X and NEP, and GPP and R (Honti and Istvánovics, submitted). Systematic errors of X (e.g. a systematic over- or underestimation) can either reduce or increase the correlation between NEP and X. Reduced correlation occurs when the estimated variance of X is less than the true variance. In contrast to this, random errors always strengthen the correlation. The same phenomenon applies to the stage when GPP is extracted from NEP by estimating R, yet relationships between errors and the correlation are more complicated due to the involvement of estimation errors of both X and R. In summary, the typical error levels of data about X and R suggest that a significant part of the correlation between X and NEP, and GPP and R could be caused by artefacts. Therefore, the limnological and ecological interpretation of such correlations should be preferably avoided unless independent, supporting information is available for decreasing the estimation uncertainties. Most importantly, strong correlation between X and NEP, as well as the near-zero value of NEP means that the CO_2 balance of lakes cannot be properly estimated from the free water dissolved oxygen method alone as has been done recently in a series of climate-change related studies (Cole et al. 2007, Tranvik et al. 2009). A possible solution is to independently measure gas exchange at least sporadically to choose a gas exchange model that matches the conditions of the examined lake relatively well, or to develop a lake-specific gas exchange function.

The error propagation problem and the linked spurious correlations and linearity are not limited to metabolic studies. Error propagation affects all modeling problems where observed dynamics is decomposed into sub-processes that are weakly or not at all observable. When the model is built on sums or differences of sub-processes, error propagation will change the linear correlation between them. Potentially affected models can be – just from limnology – the growth and mortality of phytoplankton, or the settling and resuspension of sediments. In these cases, similarly to the metabolic decomposition of the DO signal, an overall stock is monitored without having hints of the in- and outward fluxes.

Metabolism and Environmental Boundary Conditions

On the basis of the Bayesian metabolic model (trained by sequential learning) a simpler, empirical multivariate model of GPP and R was set up using data from 2009 to 2015 (Istvánovics and Honti 2017). Water temperature, underwater light intensity, and phytoplankton biomass are the inputs of the nonlinear model. A generic function that, depending on its parameter values, could resemble a saturation mechanism (sigmoid) or an optimality function was used to formulate contribution of input factors to GPP and R. Contributions were multiplied and scaled with a maximal possible magnitude. Surprisingly, such a simple model structure could simulate the complex dynamics of GPP and R with a rather high fidelity (r^2 reached 0.6 for R and 0.8 for GPP, n=891). The success of this tool means that metabolic components can be estimated without the tedious and error-prone DO observations and metabolic modeling. This was validated by testing the empirical model on the modelled data from 2016, which were not used in the original regression. Thus, the complicated Bayesian modeling exercise can be avoided, at least in the case of the Keszthely Basin of Lake Balaton. This means that the best available estimate for GPP and R can be derived from simple observational data on environmental conditions and phytoplankton biomass. The multivariate empirical model revealed general patterns in the relationship of metabolism and the environment. GPP showed less temperature-dependence than R. R showed "temperature-saturation" at higher values than GPP, suggesting that a potential warming during climate change may shift NEP slightly towards the negative domain, e.g. making the lake more heterotrophic. Underwater light had surprisingly low effect on lake metabolism. Despite the general turbidity of the water column, there was mostly enough light, because each species present in such an environment must tolerate low light conditions. Consequently, the linear light-dependence function that worked in many lakes (Hanson et al 2008) would not be applicable to Lake Balaton – and probably many lakes at this latitude.

According to the newly revealed issues with error propagation, it is unknown how the results of the empirical model relate to reality. Successful validation of the multivariate model, however, suggests that the error component is system-specific and stable. If gas exchange could reliably be measured and estimated from meteorology data in the future, it would be possible to correct for the presently unknown error component of metabolic rates.

The values of annual NEP showed little excursion from zero during the 8 years covered by observations. There was a seasonal rhythm though: NEP started in negative in the spring, grew to positive during the summer and returned to negative in the autumn. This pattern fits to the ecological expectations. After the collapse of the early spring diatom bloom (which we can't measure because installation of the monitoring site is permitted only after 1 April), heterotrophs are more abundant and keep NEP negative. After this 'clear water phase' autotrophs take over and form the usual summer algal blooms, which convert NEP to positive. With the autumn cooling the blooms fall out and the heterotrophic activity again recycles the accumulated organic matter, which pushes NEP back to negative. On the recommended annual scale NEP did not show any interesting features in Lake Balaton. It was obvious without DO measurements that the lake can't be clearly heterotrophic, because there is no sustained external source of organic matter that could fuel that (Hoellein et al. 2013). Clear autotrophy could be excluded as well, because most available organic matter is rapidly recycled in shallow, well oxygenated Lake Balaton, resulting in low sediment organic carbon content (Máté 1987). This outcome is typical for many lakes around the world (Hoellein et al. 2013; Solomon et al. 2015). Therefore, NEP is not the most informative indicator of lake metabolism. From a methodological point of view, GPP/R ratios are somewhat less sensitive to error propagation than NEP, because the errors of GPP and R partially cancel out, yet the dimensionless nature of GPP/R is more appropriate as it is not tempting to interpret its value as a proper oxygen flux. From an information-oriented perspective, other, more direct metabolic indicators, such as the gross or net growth rate of phytoplankton or the dynamics of nighttime oxygen consumption could be better proxies for lake metabolism.

High-resolution measurements of delayed fluorescence and weekly manual chlorophyll measurements show that phytoplankton blooms still reach 60-100 μg Chl-a L^{-3} in certain years in the Keszthely Basin of Lake Balaton. This indicates that further reduction of external nutrient load might be required; for example, operation of the Kis-Balaton system should not lose its focus on water quality management of Lake Balaton.

Publications of the Research Project

Publications in Impact-Factor Journals

1. Honti, M., Istvánovics V., Staehr, P. A., Brighenti, L.S., Zhu M-Y. and Zhu G-W. (2016) Robust estimation of lake metabolism by coupling high frequency dissolved oxygen and chlorophyll fluorescence data in a Bayesian framework. Inland Waters 6(4): 608-621. doi:10.5268/IW-6.4.877
2. Staehr, P.A., Brighenti, L.S., Honti M., Christensen, J., and Rose, K.C. (2016) Global patterns of light saturation and photoinhibition of lake primary production. Inland Waters 6(4): 593-607. doi:10.5268/IW-6.4.888
3. Istvánovics, V. and Honti, M. (2017), Coupled simulation of high-frequency dynamics of dissolved oxygen and chlorophyll widens the scope of lake metabolism studies. Limnol. Oceanogr.. doi:10.1002/lno.10615
4. Giling, D. P., Staehr, P. A., Grossart, H. P., Andersen, M. R., Boehrer, B., Escot, C., Evrendilek, F., Gómez-Gener, L., Honti, M., Jones, I. D., Karakaya, N., Laas, A., Moreno-Ostos, E., Rinke, K., Scharfenberger, U., Schmidt, S. R., Weber, M., Woolway, R. I., Zwart, J. A. and Obrador, B. 2017. Delving deeper: Metabolic processes in the metalimnion of stratified lakes. Limnology and Oceanography, 62: 1288–1306. doi:10.1002/lno.10504

Publication Submitted to an Impact-Factor Journal

5. Honti M., and Istvánovics V. (submitted) Error propagation leads to spurious correlations and misinterpretation of lake metabolism. Submitted to Water Research

Conference Presentations

1. Honti, M., and Istvánovics V. (2016) Short-term dynamic forecast of dissolved oxygen concentration with a simple metabolic model and Bayesian learning. XXXIII. SIL Congress, 31 July - 5 August 2016, Torino, Italy.
2. Istvánovics V. and Honti, M. (2016) Heat wave effect on lake metabolism. XXXIII. SIL Congress, 31 July - 5 August 2016, Torino, Italy.
3. Istvánovics V. and Honti, M. (2017) Validation of a multivariate model of lake metabolism. Symposium of European Freshwater Sciences, 2-7 July, 2017, Olomouc, Czech Republic.
4. Honti, M., and Istvánovics V. (2017) A system-analysis perspective on models of lake metabolism – is there life beyond NEP? Symposium of European Freshwater Sciences, 2-7 July, 2017, Olomouc, Czech Republic.

References

Cole, J.J., Caraco, N.F., 1998. Atmospheric exchange of carbon dioxide in a low-wind oligotrophic lake measured by the addition of SF_6. Limnology and Oceanography 43, 647–656. doi:10.4319/lo.1998.43.4.0647.

Cole JJ, Prairie YT, Caraco NF, McDowell WH, Tranvik LJ, Striegl RG, et al. 2007. Plumbing the global carbon cycle: Integrating inland waters into the terrestrial carbon budget. Ecosystems 10(1):172–85.

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., Hetherington, A.L., Oliver, S.K., Read, J.S., Winters, K.M., Hanson, P.C., Read, E.K., Winslow, L.A., Weathers, K.C., 2016. Consequences of gas flux model choice on the interpretation of metabolic balance across 15 lakes. Inland Waters 6, 581–592. doi:10.5268/IW-6.4.836.

Hanson, P.C., Carpenter, S.R., Kimura, N., Wu, C., Cornelius, S.P., Kratz, T.K., 2008. Evaluation of metabolism models for free-water dissolved oxygen methods in lakes. Limnology and Oceanography: Methods 6, 454–465. doi:10.4319/lom.2008.6.454.

Hoellein, T.J., Bruesewitz, D.A., Richardson, D.C., 2013. Revisiting Odum (1956): A synthesis of aquatic ecosystem metabolism. Limnology and Oceanography 58, 2089–2100. doi:10.4319/lo.2013.58.6.2089.

Honti M. and Istvánovics V. 2011. On-line monitoring of phytoplankton light response curves using a novel delayed fluorescence device. Lakes & Reservoirs - Research and Management 16(2): 153-158.

Istvánovics V., Honti M., Osztoics A., H. M. Shafik, Padisák J., Y. Yacobi and W. Eckert. 2005. On-line delayed fluorescence excitation spectroscopy, as a tool for continuous monitoring of phytoplankton dynamics and its application in shallow Lake Balaton (Hungary). Freshwater Biology 50: 1950-1970.

Máté F. 1987. Mapping of recent sediments in Lake Balaton (In Hungarian). In: A Magyar Állami Földtani Intézet évi jelentése az 1985. évről, pp. 367–379.

McNair, J.N., Sesselmann, M.R., Kendall, S.T., Gereaux, L.C., Weinke, A.D., Biddanda, B.A., 2015. Alternative approaches for estimating components of lake metabolism using the free- water dissolved-oxygen (FWDO) method. Fundamental and Applied Limnology / Archiv für Hydrobiologie 186, 21–44. doi:10.1127/fal/2015/0626.

Odum, H.T., 1956. Primary production in flowing waters. Limnology and Oceanography 1, 102–117. doi:10.4319/lo.1956.1.2.0102.

Reichert, P., and Schuwirth, N. 2012. Linking statistical bias description to multiobjective model calibration, Water Resour. Res., 48, W09543, doi:10.1029/2011WR011391.

Solomon, C.T., Bruesewitz, D.A., Richardson, D.C., Rose, K.C., de Bogert, M.C.V., Hanson, P.C., Kratz, T.K., Larget, B., Adrian, R., Babin, B.L., Chiu, C.Y., Hamilton, D.P., Gaiser, E.E., Hendricks, S., Istvánovics, V., Laas, A., O'Donnell, D.M., Pace, M.L., Ryder, E., Staehr, P.A., Torgersen, T., Vanni, M.J., Weathers, K.C., Zhu, G., 2013. Ecosystem respiration: Drivers of daily variability and background respiration in lakes around the globe. Limnology and Oceanography 58, 849–866. doi:10.4319/lo.2013.58.3.0849.

Staehr, P.A., Bade, D., de Bogert, M.C.V., 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. doi:10.4319/lom.2010.8.0628.

Tranvik L.J., Downing J.A., Cotner J.B., Loiselle S.A., Striegl R.G., Ballatore T.J., et al. 2009. Lakes and reservoirs as regulators of carbon cycling and climate. Limnology and Oceanography 54(1): 2298-2314.

Title page graphics: scheme of the artificial ecosystem used to study error propagation in metabolic modeling.