A Model for Using Environmental Data-Driven Inquiry and Exploration to Teach Limnology to Undergraduates

Cayelan C. Carey, Rebekka Darner Gougis, Jennifer L. Klug, Catherine M. O'Reilly, David C. Richardson

Limnologists are increasingly using large volumes of data, both from high-frequency sensors as well as long-term studies, to address new research questions. Undergraduate students, i.e., future limnologists and informed citizens, need quantitative reasoning skills and tools to be able to analyze these large datasets. However, most undergraduate curricula typically remains focused on small-scale local studies, potentially contributing to many students' inability to see the applicability of their classroom experiences (Prokop et al. 2007). In response, we have developed undergraduate teaching modules that integrate the use of high-frequency and long-term datasets from many lakes around the world. Here, we describe two modules that are designed to increase conceptual understanding of climate change and lake metabolism while simultaneously improving quantitative reasoning, building data manipulation skills, and highlighting the inherent variability in real data (Fig. 1). These two modules were developed by a team of limnologists and education researchers committed to improving environmental data literacy in undergraduate classrooms as part of the Environmental Data-Driven Inquiry and Exploration Project (Project EDDIE; http://www.projecteddie.org). In addition to describing the modules, we also share both the students' and instructors' experiences during module implementation, and highlight the potential for scaling these modules across different skill levels, both within and across different types of institutions. Our experience suggests that students appreciate the value of high-resolution and long-term data, and that working with large datasets cements the "real world" application of basic freshwater ecology concepts.

The emerging approach of using large and variable datasets to study the environment requires different skill sets than those currently taught in most undergraduate curricula (Brewer and Gross 2003). To date, few educational initiatives have been developed that train undergraduate science students to use high-frequency or long-term datasets (but see Langen et al. 2014). In response to this challenge, we have developed sensor-based and time series data analysis activities that can be integrated into undergraduate classrooms to improve quantitative skills and reasoning and increase student engagement. Each exercise has a modular "A-B-C" structure with three student activities that build from relatively simple to more complex (Fig. 2). The full ABC module allows students to complete a learning cycle involving data exploration, explanation, and extension into a new situation (Bybee et al. 2006). The flexible format of the module enables instructors to choose the activities most appropriate for their classroom, as some activities of the module can be completed in a standard one-hour lecture period whereas the entire module could be taught in a three-hour laboratory session (Fig. 2). All modules include a Microsoft PowerPoint file for instructors to introduce the topic, a student handout that gives an overview of the activities, collated datasets for the students, and an instructor's manual

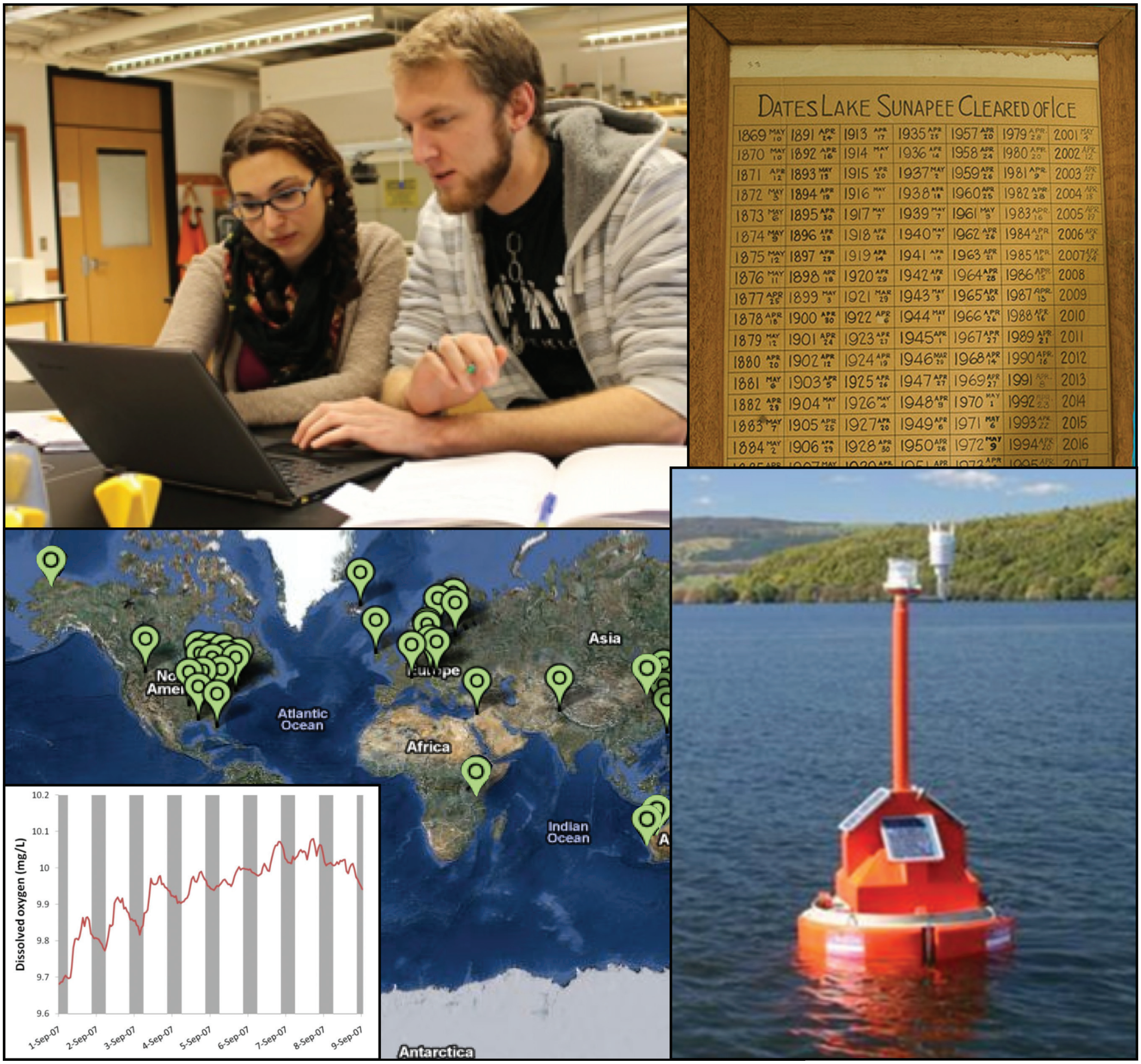

Fig 1. Clockwise from top left: Undergraduates completing the Lake Ice Phenology module (photo credit: Jen Klug); Dataset of ice-off dates observed in Lake Sunapee (New Hampshire), recorded in the Sunapee Town Hall since 1869 (photo credit: Joseph Brophy); Buoy with high-frequency sensors in Lake Tarawera, New Zealand (photo credit: Warrick Powrie); Map of Global Lakes Ecological Observatory Network (GLEON) lake sites (gleon.org); and Dissolved oxygen measurements recorded every 10 min by a high-frequency sonde in Lake Sunapee.

(all module materials are available at www.projecteddie.org).

We piloted two of the EDDIE modules, "Lake Ice Phenology" and "Lake Metabolism" (Fig. 2) in undergraduate Limnology/Freshwater Ecology courses at three different institutions in the United States: one large research university, one medium-sized public university, and one small private university. The Lake Ice Phenology module focuses on analyzing long-term datasets of ice-off date from lakes around the world (Benson and Magnuson 2000), and the Lake Metabolism module uses a comparative approach to explore patterns of gross primary productivity, respiration, and eutrophication across lakes in the Global Lakes Ecological Observatory Network (GLEON; Solomon et al. 2013). To assess student responses to the Lake Metabolism module, we administered surveys at all three institutions after students completed the activities (Table 1).

Survey results indicate that students were challenged by data manipulation and

	EDDIE ABC module conceptual structure	EDDIE Lake Ice Phenology module	EDDIE Lake Metabolism module
A	***Engage* in initial data exploration and skill development using simple analyses**: e.g., students dissect and discuss a prepared figure, graph one response variable, and/or apply basic analysis to one dataset.	**Examine trends in lake ice-off dates for a single lake**: students work in small groups to graph and apply linear regression to ice-off datasets from different lakes, use the entire dataset to predict future ice-off dates, and then compare results and discuss how lake characteristics (e.g., latitude) influenced their findings.	**Examine high-frequency variability in lake parameters (oxygen, light, wind) over a week**: students work in small groups to graph parameters collected on the minute scale for different lakes and time periods, and compare within- and among-day variability in these parameters.
B	***Explore* and *Explain* through more detailed analyses and comparisons**: e.g., students answer open-ended questions that require further exploration of the dataset from A or similar datasets, independently discuss and decide which analyses are appropriate for the data, and/or explain the implications of data variability.	**Apply segmented regression to ice-off dataset**: students explore how the rate of change in ice-off date varies between different periods (e.g., 1900-1970, 1970-present), examine how the length of the dataset influences regression results and outliers, and generate predictions of future ice-off date using the segmented datasets to compare with predictions from A.	**Calculate daily metabolism metrics over a week**: students work in small groups to calculate gross primary production, respiration, and net ecosystem production for their lake and time period using high-frequency dissolved oxygen measurements and light data to delineate day and night.
C	***Expand* on developed ideas to other sites, datasets, and concepts**: e.g., students access data from databases of their choice to explore questions they developed in small groups, manipulate and analyze the datasets independently, and/or explore primary literature for broader context.	**Cross-lake analysis and heat budget calculations**: students download additional ice-off datasets to independently conduct A and B analyses for lakes on size and latitudinal gradients, discuss ecological implications of altered ice cover, and use temperature profiles to calculate annual heat budgets for the same lake in years with early vs. late ice-off dates.	**Comparison of metabolism results across lakes**: students discuss and analyze their results from B to understand how trophic state and dissolved organic carbon concentrations alter lake metabolism and read published metabolism syntheses to put their results in context.

Fig 2. Flow chart of the conceptual structure and scaffolding of an EDDIE module, modified from the 5E learning cycle (Bybee et al. 2006), with a summary description of the ABC activities of the EDDIE Lake Ice Phenology and Lake Metabolism modules. See http://www.projecteddie.org for more information and teaching materials.

Table 1. Categorized student responses on surveys administered after engagement with piloted Project EDDIE activities across three classrooms ($n = 41$ responses).

Survey question	Student responses	Frequency
What were the most difficult aspects of working with a large dataset?*	Inexperience with Excel	5(12%)
	Managing, organizing, sorting through all the data	31(76%)
	Doing calculations/using formulas in Microsoft Excel	5(12%)
	Graphing in Microsoft Excel	3(7.3%)
	Visualizing results when there is so much data	5(12%)
	Nothing was difficult	1(2.4%)
What did you learn that you might not have learned if you had a smaller dataset (such as one point per day or data points every other day)?†	More data = better or more accurate data	13(32%)
	Allows for better visualization of trends	5(12%)
	Allows us to see changes within one 24-h period	21(51%)
How much more information can you gain when you work with high-frequency data?‡	Much more/more realistic/more is better/more accuracy	23(56%)
	Able to see changes/fluctuations	14(34%)
	More predictive power	1(2.4%)
	More chance of finding statistical significance	2(4.8%)

*One student did not answer this question.
†Three students did not answer this question.
‡Two students did not answer this question.

analysis activities, but still recognized the value of working with high-frequency data (Table 1). Across institutions, the majority of students (76%) found that sorting and managing the vast amount of data were the most difficult aspects of working with large datasets, whereas others cited their inexperience with using a spreadsheet program (e.g., Microsoft Excel), graphing, and visualizing results. When asked what they learned from the larger datasets that they might not have learned with smaller datasets, most students stated that the high-frequency metabolism data allowed them to see fluctuations in the data that they did not know existed at coarser temporal resolution. Interestingly, several students stated that the long-term and high-frequency data allowed them to better visualize trends, and one student acknowledged greater predictive power when working with high-frequency data. The majority of students (59%) acknowledged that they will need quantitative, data management, or database skills for their future careers.

After teaching the modules, the three instructors recorded reflections of their experience in the classroom. One of the biggest challenges at all three institutions was coordinating the interactions among students when there were disparities in data manipulation and analysis skills. For example, the instructors asked the students to complete the modules in pairs to encourage discussion about the data. However, a situation routinely emerged in which the student that was most comfortable with spreadsheet and graphing programs led the typing of the module exercises on a computer and was inherently more engaged in the activity, while the other, less experienced student was relegated to a peripheral observer role. Conversely, if both students had access to computers, they failed to engage with each other and tended to work almost completely in isolation. Instructors may be able to overcome these challenges by providing explicit stopping points in the module to pose discussion questions, and prompt partners to discuss and make decisions about future analyses. Lack of experience with Microsoft Excel was a pervasive problem across the three institutions: several students with less experience manipulating data expected the software to generate a graph that did not need modification, and struggled with unit conversions and formulas. Despite these challenges, however, students valued working with "real" environmental data that came from actual lakes and were used in published papers.

All instructors were encouraged by the discussions brought on by the modules. In an upper-level classroom, the students engaged in an informative discussion about outliers, prompted by the instructor telling them that there might be "bad data" in the dataset. The students initiated a class discussion and asked: What are "bad data?" What is an outlier? When can data points be excluded? When is it inappropriate to exclude data points? This instructor concluded that if the goal of this module was to highlight that real datasets are large and variable, then the module was a success. For the students in lower-level classes, who tended to have less Excel experience, their motivational challenges seemed to be lessened when each student pair (analyzing data from different lakes) presented results to the rest of the class, enabling a discussion on the differences in lake metabolism among eutrophic, oligotrophic, and dystrophic systems. Concepts that students often do not understand (i.e., gross vs. net production) were more easily developed by having students calculate these metrics, rather than teaching the topic in a standard lecture format. Engaging students in these modules also allowed the instructors to assess different strengths and weaknesses of the students. For example, some of the strongest students by traditional measures (e.g., exam scores) were the ones that exhibited the most obvious cognitive dissonance with the more independent and open approach required by the module activities.

In conclusion, data from this pilot study indicate that the modules were successful in improving quantitative literacy and increasing appreciation for large datasets. All students stated that they could attain better or more information using high-frequency datasets than they would with fewer data. Project EDDIE activities also provided the opportunity for some students to practice sophisticated cognitive tasks, such as data visualization and realizing the predictive power of a dataset. These two cognitive tasks—prediction and data visualization—seem simple but are notoriously difficult to teach. Prediction, as Pace (2001) explains, is not simply a goal in itself, but a means through which scientists come to greater understanding of natural phenomena and judge the adequacy of tentatively accepted knowledge. From a psychological perspective, making predictions leads to greater cognitive understanding because it calls on the predictor, either the student or the instructor, to make explicit assumptions, expectations, and causal reasoning. Once explicit, the predictor must then reflect on their reasoning in the light of evidence, which in turn leads to integration of evidence-based theory into one's cognitive understanding (Runnel et al. 2013). Consequently, we contend that integrating data manipulation, visualization, and analysis activities into undergraduate classrooms, such as through teaching the modules we describe here, will greatly advance the training of the next generation of quantitatively literate citizens and limnology researchers.

Acknowledgments

We thank the entire Project EDDIE team, collaborators, and our students for their assistance in developing and providing feedback on the Lake Ice Phenology and Lake Metabolism modules. We acknowledge the Global Lakes Ecological Observatory Network (GLEON) and the data providers that generously shared datasets for the modules: R. Adrian, B. Benson, C.-Y. Chiu, J. Fichter, E. E. Gaiser, S. Hendricks, V. Istvànovics, J. J. Magnuson, and K. C. Weathers. Project EDDIE is supported by the U.S. National Science Foundation (DUE-1351823 and DUE-1245707). This study was approved for exempt status by Institutional Review Boards (IRBs) at all institutions.

References

Benson, B., and J. J. Magnuson. 2000. Global lake and river ice phenology database. National Snow and Ice Data Center. doi: 10.7265/N5W66HP8 (see: http://nsidc.org/data/G01377).

Brewer, C. A., and L. J. Gross. 2003. Training ecologists to think with uncertainty in mind. Ecology 84: 1412–1414. doi:10.1890/0012-9658(2003)084[1412:TETTWU]2.0.CO;2

Bybee, R. W., and others. 2006. BSCS 5E instructional model: Origins and effectiveness. A report prepared for the Office of Science Education, National Institutes of Health. BSCS.

Langen, T. A., and others. 2014. Using large public datasets in the undergraduate ecology classroom. Front. Ecol. Environ. 12: 362–363. doi:10.1890/1540-9295-12.6.362

Pace, M. L. 2001. Prediction and the aquatic sciences. Can. J. Fish. Aquat. Sci. 58: 63–72. doi:10.1139/cjfas-58-1-63

Prokop, P., M. Prokop, and S. D. Tunnicliffe. 2007. Is biology boring? Student attitudes toward biology. J. Biol. Educ. 42: 36–39. doi: 10.1080/00219266.2007.9656105

Runnel, M. I., M. Pedaste, and A. Leijen. 2013. Model for guiding reflection in the context of inquiry-based science education. J. Balt. Sci. Educ. 12: 107–118.

Solomon, C. T., and others. 2013. Ecosystem respiration: Drivers of daily variability and background respiration in lakes around the globe. Limnol. Oceanogr. 58: 849–866. doi:10.4319/lo.2013.58.3.0849

Cayelan C. Carey, Department of Biological Sciences, Virginia Tech, Blacksburg, Virginia

Rebekka Darner Gougis, School of Biological Sciences, Illinois State University, Normal, Illinois

Jennifer L. Klug, Department of Biology, Fairfield University, Fairfield, Connecticut

Catherine M. O'Reilly, Department of Geography-Geology, Illinois State University, Normal, Illinois

David C. Richardson, Department of Biology, SUNY New Paltz, New Paltz, New York