Overview of the project and outcomes

Understanding the factors that affect lake water quality and the ecological services provided by them is an urgent global environmental issue. Predicting how lakes will respond to global changes not only requires water quality data, but also information about the ecological context of individual lakes across broad spatial and temporal scales. Because lake water quality is usually sampled in limited geographic regions and time periods, determining the environmental controls of water quality requires scientists to combine existing smaller data sets into an integrated database. Such comprehensive databases have not been available across large regions in the U.S., until now. Our NSF-funded Macrosystems Biology research team created LAGOS-NE-- a database for all 50,000 lakes in an area of 1,800,000 km2 in 17 northeast and midwestern US states. We developed the methods, data, and infrastructure for building other such databases. LAGOS-NE is one of the largest and most comprehensive databases of its type because it includes both in situ measurements and ecological context data.

Our scientific goal was to develop this database to answer fundamental research questions about the controls of lake water quality across broad geographic regions of the country. This database contains the needed information to better understand water quality in thousands of lakes and the complex relationships between climate, land use, lake characteristics, and water quality. And, because ecological context can be used to study a variety of other questions about lakes, streams, and wetlands, LAGOS-NE can be used as the foundation for other studies of freshwaters at broad spatial and ecological scales. Our project has resulted in the following important outcomes:

- 32 peer-reviewed publications, as well as 4 under review and 8 in preparation
- 3 of the above manuscripts described novel computer science methods for studying any continental-scaled environmental data
- 10 data packages in online, public repositories
- 4 software packages in online, public repositories, including novel computer science methods
- Many downloads of our data:
    - The software to access LAGOS-NE - downloaded 980 times
    - LAGOS-NE data - downloaded 513 times
- 1 website providing extensive documentation and information related to these data
- A new blog related to big data and data visualization in environmental research
- 4 PhD dissertations; all four students have already obtained full-time positions
- 4 post-doctoral researchers trained; they are all now either in full-time or other post-doctoral positions

These products and the knowledge gleaned from them have already been incorporated into new interdisciplinary research, as well as evidence-based policy decisions such as the Michigan lake-specific lake criteria.

Significant conclusions

Our project has improved basic knowledge of water quality in thousands of lakes across a broad geographic region (subcontinental). The factors that explain lake water quality, are known to differ when studying a single lake versus studying thousands of lakes. However, until recently, scientists have not had the data to study whole populations of lakes (e.g., thousands of lakes at once). An important knowledge gap that our study filled was to measure and understand what

controls variation in thousands of lakes as compared to the understanding gleaned from decades of research conducted on individual lakes. For example, we confirmed that the factors controlling water quality in an individual lake are different from the factors that control water quality across lakes and that water quality is controlled by different factors in a highly-agricultural region as compared to a mostly-forested region. We also found little evidence that water quality has drastically changed in the last 25 years, although there is good evidence that lake nitrogen is mostly decreasing through time. Finally, we found much larger differences in lake water quality across regions than changes in individual lakes through time. The combination of these results point to the importance of monitoring many lakes across regions of different landscape contexts and including nitrogen as a sample parameter (sample sizes were much smaller for nitrogen than for other variables) to capture broad-scale variation in lake water quality and understand responses to global changes.

Our project was able to answer many complex environmental research questions through innovative research strategies such as: using big-data approaches for database development and analysis; using current team science approaches that leverage the knowledge and skills from multiple disciplines; and, using open-science approaches that make our research products available for future researchers to build off of in a way that saves federal funds. Our project approach can be a model for other multi-institution, interdisciplinary, broad-scale research projects to solve complex problems facing society. Our hope is that our database and the associated support tools and documentation will serve as a powerful resource and a foundation for future research and decision-making by a broad community of scientists, policy-makers, and natural resource managers.