Coupling Cyberinfrastructure and Geographic Information Systems to Empower Ecological and Environmental Research

Authors: Wang, Shaowen, and Zhu, Xin-Guang

Coupling Cyberinfrastructure and Geographic Information Systems to Empower Ecological and Environmental Research

SHAOWEN WANG AND XIN-GUANG ZHU

With the burgeoning global population, complex climate changes, and the growing demand for natural resources, ecological and environmental researchers must look for the best strategies to ensure a sustainable supply of ecosystem services that provide food, water, air, and energy. As in the field of modern molecular biology, where advanced computational technologies play an important role in managing and analyzing massive quantities of genomic data, cyberinfrastructure-based ecological and environmental sciences will contribute significantly to the quest.

"Cyberinfrastructure" describes integrated information and communication technologies for distributed information processing and coordinated knowledge discovery, which promises to revolutionize the way that science and engineering are done in the 21st century and beyond (Atkins et al. 2003). Similar to the respective roles of telescopes and particle accelerators in astronomy and high-energy physics, cyberinfrastructure acts as a computational test bed for scientific discoveries. Such computational test beds cross disciplines and serve all sciences; they empower researchers by giving them access to several interrelated components, such as high-performance computers, data, information resources, networking, digitally enabled sensors and instruments, virtual organizations, and observatories, along with an interoperable suite of software (i.e., middleware) services and tools (NSF 2007).

In addition to enabling scientific discoveries, cyberinfrastructure itself is an evolving subject of research. Both the individual components of cyberinfrastructure and their interactions are highly complex. To manage this complexity and thus assure the usability of cyberinfrastructure for scientific discovery, science and engineering gateways to cyberinfrastructure are being designed to provide customizable and seamless access to cyberinfrastructure through problem-solving environments tailored to the needs of specific science communities. The National Science Foundation's Tera-Grid (*www.teragrid.org*)—a key element of the US and world cyberinfrastructure—has facilitated the development and operation of more than two dozen science and engineering gateways.

GISolve (*www.gisolve.org*) is one of the TeraGrid science and engineering gateways that focuses on the development and provision of cyberinfrastructure-enhanced GIS (geographic information systems) capabilities. GISolve functions include spatiotemporal database management, spatial analysis and modeling, visualization, and virtual organization support for collaborative problem solving. All GISolve capabilities are accessible through Web interfaces that use a set of GIS-aware middleware to (a) integrate cyberinfrastructure capabilities within GIS functions (such as for spatial analysis and modeling) and (b) hide the complexity of the cyberinfrastructure (figure 1). GISolve has been widely used in geospatial sciences for research and education. In this article, we discuss the great potential of GISolve for ecological and environmental research.

The quality and quantity of spatiotemporal data collected using geospatial technologies such as satellite remote sensing, the global positioning system, and sensor networks have improved dramatically in the past a few decades, and this trend will most likely continue in the foreseeable future. We envision that the data assimilation capabilities and significant computational power accessible within GISolve will enhance the capacity of individual-based models (Levin et al. 1997). Greater computational power from cyberinfrastructure allows models to include detailed physiology and physical processes, as well as a larger number of species and individuals, to achieve high prediction accuracy and resolutions.

GISolve provides cyberinfrastructure-based spatial analysis algorithms (e.g., spatial interpolation and geostatistical modeling) within user-friendly GIS functions for solving large-scale geospatial problems. The methods used in these algorithms to harness cyberinfrastructure power can be applied to integrate multiscale models from different disciplines, such as individual-based models, models of soil geochemical cycles, models of watershed hydrology, models of

Shaowen Wang (e-mail: shaowen@uiuc.edu) is the director of the GISolve project, assistant professor of geography, and senior research scientist at the National Center for Supercomputing Applications at the University of Illinois at Urbana–Champaign (UIUC). Xin-Guang Zhu (e-mail: zhu3@uiuc.edu) is a research scientist at UIUC. *© 2008 American Institute of Biological Sciences.*

vegetation-atmosphere interactions, ecological economics models, and agent-based models of nature and human interactions. Such integration is critical to evaluate the consequences of different ecosystem and environmental management practices.

Ecosystem and environmental management practices related to agriculture, forest, land use, and water often occur at large spatial scales with high economic and environmental stakes, which makes it highly desirable to achieve realistic predictions of consequences of different options before management practices are put into action. Recent interest in biomass-based energy production can be a unique and legitimate driving force for coupling the aforementioned multiscale models. Sustainable production of biomass must be ensured, and to do this, system-level understanding of the interactions among all the ecological and environmental processes and human activities is required. We have started to develop the coupling of several multiscale models supported by sizable spatiotemporal databases within GISolve.

Here, we use a concrete example to illustrate how GISolve dramatically increased the capacity to analyze massive spatiotemporal data. To assess the impact of climate and management practices on crop yields in the United States, we have compiled long-term, historical (since 1900) geospatial data for detailed climatic variables and different crops. One of our initial challenges was to effectively organize and manipulate these massive (greater than a half terabyte; a terabyte is approximately one trillion bytes) geospatial data that cover a wide spectrum of spatiotemporal resolutions. In this case, even basic geospatial data manipulations such as map projections and spatial interpolation consume significant computational resources—the equivalent of several weeks of uninterrupted computation using a state-of-the-art personal computer. Using TeraGrid, GISolve can shorten such manipulations to several hours. More advanced analysis and modeling would not even be possible without access to GISolve and the underlying TeraGrid capabilities.
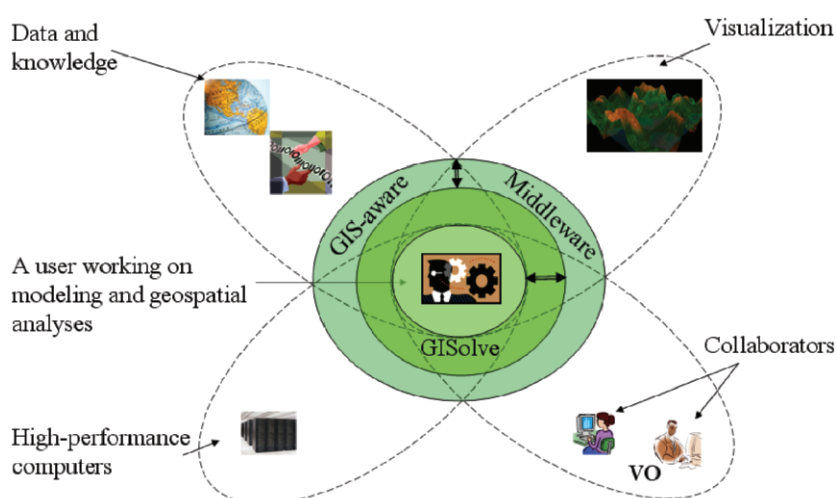

Figure 1. Relationships between high-performance computers, middleware, VO (virtual organizations), data management and knowledge discovery, visualization services, GIS (geographic information systems), and GISolve.

The advanced analysis and modeling empowered by coupling cyberinfrastructure and GIS require effective integration of processes operating at different spatiotemporal scales. For example, changes in soil status are extremely slow, while the growth and development of plantation change relatively faster. In addition, new algorithms to effectively integrate multiscale modeling and GIS are valuable. Cyberinfrastructure capabilities, accessed through GISolve, make the integration of GIS and multiscale modeling feasible for solving large and complex problems. Our experiences so far suggest that the active participation of domain scientists such as biologists and geographers in the evolution of cyberinfrastructure, especially in the development of science and engineering gateways, is critical for making cyberinfrastructure truly relevant to those scientists and for realizing the enormous impact of cyberinfrastructure on scientific discovery.

In summary, scientific discovery can be significantly empowered by introducing computational thinking to individual sciences (Wing 2006). Coupling cyberinfrastructure and GIS can facilitate computational thinking to analyze massive quantities of spatiotemporal data rapidly and economically. Computational modeling and simulation based on cyber-infrastructure-enhanced GIS allows researchers to tackle large and complex ecological and environmental problems that cannot be replicated in laboratories. Biomass-based energy production provides a unique opportunity to develop comprehensive, multiscale models of complex and large systems. We believe that GISolve, designed to shield cyberinfrastructure complexity and integrate cyberinfrastructure into GIS capabilities, can and will empower unprecedented scientific discovery for sustainable biomass-based energy.

References cited

Atkins DE, et al. 2003. Revolutionizing Science and Engineering through Cyberinfrastructure: Report of the National Science Foundation Blue-Ribbon Advisory Panel on Cyberinfrastructure. (9 January 2008; *www.community technology.org/nsf_ci_report/*)

Levin SA, Grenfell B, Hastings A, Perelson AS. 1997. Mathematical and computational challenges in population biology and ecosystems science. Science 275: 334–343.

[NSF] National Science Foundation. 2007. Cyberinfrastructure Vision for 21st Century Discovery. (9 January 2008; *www.nsf.gov/pubs/ 2007/nsf0728/index.jsp*)

Wing JM. 2006. Computational thinking. Communications of the ACM 49: 33–35.