Meetings

Integrating empirical–modeling approaches to improve understanding of terrestrial ecology processes

Strategies to Promote Integrated Data–model Approaches to Terrestrial Ecosystem Study, in Bethesda, Maryland, USA, March 2012

Recent decades have seen tremendous increases in the quantity of empirical ecological data collected by individual investigators, as well as through research networks such as FLUXNET (Baldocchi *et al.*, 2001). At the same time, advances in computer technology have facilitated the development and implementation of large and complex land surface and ecological process models. Separately, each of these information streams provides useful, but imperfect information about ecosystems. To develop the best scientific understanding of ecological processes, and most accurately predict how ecosystems may cope with global change, integration of empirical and modeling approaches is necessary. However, true integration – in which models inform empirical research, which in turn informs models (Fig. 1) – is not yet common in ecological research (Luo *et al.*, 2011).

The goal of this workshop, sponsored by the Department of Energy, Office of Science, Biological and Environmental Research (BER) program, was to bring together members of the empirical and modeling communities to exchange ideas and discuss scientific practices for increasing empirical–model integration, and to explore infrastructure and/or virtual network needs for institutionalizing empirical–model integration (Yiqi Luo, University of Oklahoma, Norman, OK, USA). The workshop included presentations and small group discussions that covered topics ranging from model-assisted experimental design to data driven modeling (e.g. benchmarking and data assimilation) to infrastructure needs for empirical–model integration. Ultimately, three central questions emerged. How can models be used to inform experiments and observations? How can experimental and observational results be used to inform models? What are effective strategies to promote empirical–model integration?

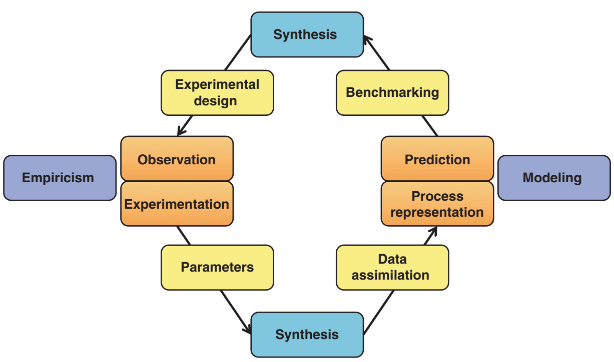

Fig. 1 Conceptual framework for empirical–model integration process. Adapted from original design of Peter Thornton (Oak Ridge National Laboratory, Oak Ridge, TN, USA).

> *'Although empirical–model integration promotes greater understanding of ecological processes, there may be a risk of experiments becoming too influenced by models, hindering scientific discovery and serendipity.'*

Models informing empirical research: designing new experiments while leaving room for discovery

Many presenters and participants suggested that, for successful empirical–model integration, it is crucial for empiricists and modelers to work together from the start of a project, rather than approaching modeling as an afterthought to data collection. Two examples of projects that were designed to facilitate data–model integration are the new Next-Generation Ecosystem Experiments (NGEE Arctic) and Spruce and Peatland Changes Responses Under Climatic and Environmental Change (SPRUCE) projects, both of which are proposing to improve climate prediction through empirical studies in globally important tundra and boreal ecosystems. From the perspective of an experimentalist, Stan Wullschleger (Oak Ridge National Laboratory, Oak Ridge, TN, USA) suggested that key lessons from the development of the NGEE Arctic project included: structuring projects with integration as a priority, knowing how modelers will use empirical data, and asking modelers to identify what they do not know (about a system). Similarly, Paul Hanson (Oak Ridge National Laboratory, Oak Ridge, TN, USA) provided examples of how pre-experiment modeling helped shaped the experimental design of the SPRUCE project, leading to the implementation of a range of experimental temperatures – rather than the two temperature levels originally planned – in order to provide richer information on the shape of ecosystem response functions to warming. Although not explicitly covered by presenters, participants also discussed Bayesian data assimilation (e.g. Clark *et al.*, 2011) as another valuable avenue for models to inform experimental design and data collection.

Although empirical–model integration promotes greater understanding of ecological processes, there may be a risk of experiments becoming too influenced by models, hindering scientific discovery and serendipity. Through a series of manipulative precipitation experiments, Alan Knapp (Colorado State University, Fort Collins, CO, USA) showed that central US grasslands respond to changes in precipitation quantity and frequency differently depending on the natural precipitation regime (Heisler-White *et al.*, 2009). Experimental results were largely inconsistent with what had been predicted *a priori* from other experiments and models, demonstrating the value of 'curiosity-driven' experiments to improve process understanding. Similarly, Joe Berry (Carnegie Institution for Sciences, Stanford, CA, USA) discussed the history of efforts to use atmospheric carbonyl sulfide measurements to constrain the terrestrial carbon cycle, illustrating that new approaches and tools can come from unexpected sources.

Empirical studies informing models: benchmarking, data assimilation, and parameterization

Although empirical data frequently informs ecological models in some capacity, recent work has made use of data in a much more methodical manner, through processes such as benchmarking and data assimilation (e.g. Raupach *et al.*, 2005; Williams *et al.*, 2009; Luo *et al.*, 2012). Benchmarking is the process of systematically comparing model predictions against measured data in order to evaluate model performance and identify processes that may be poorly represented. Jim Randerson (University of California, Irvine, CA, USA) shared recent efforts in benchmarking carbon cycle processes in land surface models as part of the International Land Model Benchmarking (ILAMB) project. This project aims to develop benchmarks for land model performance, apply these benchmarks to global models, support the development of new, open-source, benchmarking software for model intercomparison, and improve linkages between experimental, monitoring, remote sensing, and climate modeling communities. Rich Norby (Oak Ridge National Laboratory, Oak Ridge, TN, USA) discussed benchmarking of ecosystem response models with Free Air CO_2 Enrichment (FACE) data. In this work, a suite of ecosystem process and land surface models were evaluated against data from two long-term FACE experiments to determine how well the models reproduced measured carbon, water, and nitrogen cycle processes.

Data assimilation combines empirical data with models to improve forecasts, by using data to constrain initial conditions and model parameters. The data assimilation process can improve model parameterization, shape model structure, and analyze uncertainty in forecasts (Luo *et al.*, 2011). Daniel Riccuito (Oak Ridge National Laboratory, Oak Ridge, TN, USA) described uncertainty analysis through data assimilation in the Community Land Model (CLM), suggesting that it is necessary to consider uncertainty in empirical observations as well as models, and that uncertainty analysis in large, complex models can be facilitated by the use of reduced model forms and simpler process models. An example of how to decompose carbon cycle models into traceable components, supporting model–model comparisons, data assimilation and benchmarking was presented by Jianyang Xia (University of Oklahoma, Norman, OK, USA). Changhui Peng (University of Quebec at Montreal, Canada) discussed the use of data assimilation to estimate parameters and predict net ecosystem productivity for seven forest AmeriFlux sites in North America. Jinfeng Xiao (University of New Hampshire, Durham, NH, USA) also used AmeriFlux data, to demonstrate how data driven approaches can be used to scale up site data to the regional scale. Together, such efforts are helping to develop a framework for the use of data assimilation in ecological models.

Finally, appropriate parameterization and process understanding is crucial for model forecasting accuracy. However, some processes are not well represented in empirical databases. Charlie Koven (Lawrence Berkley National Laboratory, Berkeley, CA, USA) highlighted the conceptual problems with black box soil carbon models, which frequently fit observed data well, but are limited in their ability to represent processes that cannot easily be measured, such as turnover of slow pools of carbon. Rosie Fisher (National Center for Atmospheric Research, Boulder, CO, USA) discussed vegetation dynamics models and suggested that, although there are a growing number of databases for plant traits, comprehensive data on mortality and abiotic limits to recruitment for use in vegetation dynamics models are still lacking.

Promoting experiment–model integration

A recurring theme throughout the workshop was the need for the development of technological infrastructure for facilitating empirical–model integration. Robert Cook (Oak Ridge National Laboratory, Oak Ridge, TN, USA) presented a vision for cyberinfrastructure that promotes data use and archiving. Such an infrastructure should provide a central clearinghouse in which data is aggregated and available in standard formats, allowing both empiricists and modelers to spend more time on science and less on data management. This infrastructure must be developed through collaboration between producers and users of data products, and have dedicated financial support. Another type of infrastructure development was presented by David Lawrence (National Center for Atmospheric Research, Boulder, CO, USA), who described the institutional support system for the Community Earth System Model (CESM), a model freely available to the climate science community. Factors contributing to the success of the CESM include sustained funding, a simple organizational structure, clearly defined mission, goals, and processes, regular communication among members and users, and comprehensive documentation and robust support.

While full empirical–model integration may not yet be common in ecology, several presenters shared their experiences and observations with empirical–model integration. Belinda Medlyn (Macquarie University, North Ryde, NSW, Australia) presented a series of case studies that demonstrated successful empirical–model integration efforts, ranging from leaf to ecosystem level processes. Based on these experiences, Medlyn suggested that modelers should interact with empiricists to suggest useful measurements, analyze empirical data and make models transparent, while experimentalists should interact with modelers in order to understand model assumptions, verify models and share and preserve data. Markus Reichstein (Max-Planck Institute for Biogeochemistry, Jena, Germany) pointed out that while model evaluation can be model or data driven, it is critical that empirical–model dialogues target the same scale or level of process.

Conclusions and future work

Overall, workshop participants identified several key community needs in order to enhance empirical–model integration in the future. First, there is a clear requirement to store, aggregate, and share ecological data. Additionally, these databases should facilitate not just data exchange, but also dialogues in which modelers could ask for help to find values for particular parameters, or identify parameters for which data is lacking in order to motivate future collection of that data. Second, there was a call for greater transparency of models. Ideally this would mean web-based versions of varying complexity, including some simple enough for students or others with little modeling experience to be able to manipulate. Finally, there was a great deal of support for programs such as INTERFACE (http://www.bio.purdue.edu/INTERFACE/index.php), which sponsors graduate student 'exchanges' in which students who primarily work with models can spend up to 1 month working in an empirical setting, or students who primarily work on empirical research can spend up to 1 month working in a modeling environment. These kinds of programs can help close knowledge and communication gaps between empirical and modeling communities.

Heather R. McCarthy[1]*, Yiqi Luo[1] and Stan D. Wullschleger[2]

[1]Department of Botany & Microbiology, University of Oklahoma, Norman, OK, USA; [2]Environmental Sciences Division, Oak Ridge National Laboratory, Oak Ridge, TN, USA (*Author for correspondence: tel +1 405 325 7793; email heather.mccarthy@ou.edu)

References

Baldocchi D, Falge E, Gu L, Olson R, Hollinger D, Running S, Anthoni P, Bernhofer C, Davis K, Evans R *et al.* 2001. FLUXNET: a new tool to study the temporal and spatial variability of ecosystem-scale carbon dioxide, water vapor, and energy flux densities. *Bulletin of the American Meteorological Society* **82**: 2415–2434.

Clark JS, Agarwal P, Bell DM, Flikkema PG, Gelfand A, Nguyen X, Ward E, Yang J. 2011. Inferential ecosystem models, from network data to prediction. *Ecological Applications* **21**: 1523–1536.

Heisler-White JL, Blair JM, Kelly EF, Harmoney K, Knapp AK. 2009. Contingent productivity responses to more extreme rainfall regimes across a grassland biome. *Global Change Biology* **15**: 2894–2904.

Luo Y, Ogle K, Tucker C, Fei S, Gao C, LaDeau S, Clark JS, Schimel DS. 2011. Ecological forecasting and data assimilation in a data-rich era. *Ecological Applications* **21**: 1429–1442.

Luo Y, Randerson J, Abramowitz G, Bacour C, Blyth E, Carvalhais N, Ciais P, Dalmonech D, Fisher J, Fisher R *et al.* 2012. A framework of benchmarking land models. *Biogeosciences Discussions* **9**: 1899–1944.

Raupach MR, Rayner PJ, Barrett DJ, Defries RS, Heimann M, Ojima DS, Quegan S, Schmullius CC. 2005. Model–data synthesis in terrestrial carbon observation: methods, data requirements and data uncertainty specifications. *Global Change Biology* **11**: 378–397.

Williams M, Richardson AD, Reichstein M, Stoy PC, Peylin P, Verbeeck H, Carvalhais N, Jung M, Hollinger DY, Kattge J *et al.* 2009. Improving land surface models with FLUXNET data. *Biogesciences* **6**: 1341–1359.

Key words: benchmarking, data assimilation, data–model integration, ecological forecasting, ecosystem processes, global change.