Changes in technology and methodology can have a big influence on how we do science. In this essay, I will discuss how new methods for the acquisition and analysis of data have affected biogeography and macroecology.

The underlying data used by macroecologists are geo-referenced specimen collections (GBIF 2008). For many decades, biogeographers explored the globe to collect and cata-

macroecology are just verbal descriptions of mechanisms ("higher productivity in the tropics allows for more biodiversity"). But since multiple explanations can generate the same qualitative patterns ("greater temperature stability in the tropics allows for more biodiversity"), we are not going to easily distinguish these mechanisms through qualitative assessment of correlations alone.

In this regard, I think the most important recent breakthrough in macroecology has been the development of metabolic theory (Allen et al. 2002). This theory, derived from first principles that do not depend in a circular way on existing data, predicts a quantitative relationship between temperature and biodiversity. Instead of just testing a null hypothesis of a slope of zero, we can now test whether observed slopes (with appropriate transformations) deviate from -0.65, the predicted value from the model (Hawkins et al. 2007). Controversy over the empirical support for metabolic theory (Hawkins et al. 2007, Gillooly and Allen 2007) should not obscure its importance: metabolic theory makes quantitative, not just qualitative, predictions and that is what we need right now in macroecology.

Theoreticians should step up to the plate and develop quantitative theories for other hypotheses in macroecology. As recently proposed by O'Brien (2006), the water-energy model may provide an emerging framework that will generate functional forms for water and energy variables derived from first principles of physiology and physical constraints imposed by the energetics of liquid water. For now, however, these models are either entirely verbal (Vetaas 2006), or they are derived from fitted regression functions that are specific to particular taxa, spatial scales, and continents (O'Brien 1998).

In addition to the development of new theory, we need to move beyond analytical methods that simply fit curves to data and test patterns against simple statistical null hypotheses. Some macroecologists are beginning to develop stochastic simulation models that include explicit algorithms for the origin, spread, and extinction of species in a bounded geographic domain (e.g. Storch et al. 2006, Rahbek et al. 2007, Rangel et al. 2007) These mechanistic simulation models (Grimm et al. 2005) have their roots in the mid-domain effect (Colwell and Lees 2000), a pleasingly simple explanation for species richness gradients that emerged from the random placement of contiguous species ranges in a bounded domain. This kind of modeling exercise raises its own challenges: how do we empirically estimate model parameters, and how do we explore the behavior of such a model over a potentially very large parameter space? But this simulation approach may allow macroecology to move beyond statistical correlations, and

the temporal series of spatial coordinates of a swinging pendulum. Their algorithm repeatedly "sampled" the data set from the most critical regions (where the pendulum was changing direction) and iteratively arrived successfully at the correct equations for motion.

Interestingly, the same methods were not so successful when applied to the famous ecological time series of snowshoe hare and Canadian lynx populations (Elton and Nicholson 1942). The algorithm did generate a pair of coupled differential equations (Bongaard and Lipson 2007). However, we know that the hare-lynx cycle is not caused entirely by coupled predator-prey interactions.

The problem, of course, is not the algorithm, but the limited data that it was fed. The time series of pelt records from the Hudson Bay Company does not reveal the critical observations of hare populations on islands in eastern Canada that cycle in the absence of the lynx (Keith 1963). The analysis also did not include time series on the secondary plant compounds in tundra vegetation, which accumulate under intense grazing and may be ultimately responsible for endogenous cycles of the hare (Keith 1983). And the model did not include time-series on snowpack depth or solar sunspot activity, both of which probably contribute to the regional synchrony of hare lynx cycles (Sinclair et al. 1993).

Without such "expert knowledge" it is easy to understand why the model failed. If those data inputs were provided, I think it is very likely the model would reveal the correct functional form of the relationships among hare, lynx, vegetation, and climate. But for now, the use of passive machine-learning algorithms applied to large data sets is an inefficient way to test hypotheses and make progress in macroecology. And given the pressing need to understand how biotas will respond to climate change, I am not sure we have the luxury of waiting for these comprehensive data sets to accumulate.

Nevertheless, the paradigm of machine learning seems to be the direction that much of the bioclimatic niche modeling research is going. If the goal of this research is to understand how biotas will shift in response to climate change, I think it is going to be much more fruitful if we combine it with an experimental approach. Experimental translocation of individuals beyond their current range boundaries (Hellmann et al. 2008) and experimental manipulations of abiotic variables to mimic effects of climate change on populations and communities (Harte and Shaw 1995, Suttle et al. 1997) are very powerful approaches. Experiments can provide realistic parameter estimates for bioclimatic niche models. Even simple models that are supported by experimental data will probably be more trustworthy than sophisticated models that are not.

In sum, the availability of large data bases, the emergence of quantitative predictive theories, and the development of new computational tools and simulation methods make this an exciting time to be studying macroecology. There are pressing applied problems of global climate change that we can address with these new tools and data. And along the way, perhaps we will even answer some unresolved questions in biogeography about species richness gradients.

Acknowledgements

This essay was inspired by the work of the Synthetic Macroecological Models of Species Diversity Working Group supported by the National Center for Ecological Analysis and Synthesis, a Center funded by NSF (Grant #DEB-0553768), the University of California, Santa Barbara, and the State of California.