UC Irvine
UC Irvine Previously Published Works

Title

Complexity in Ecology and Conservation: Mathematical, Statistical, and Computational Challenges

Permalink

https://escholarship.org/uc/item/8gx123f6

Journal

BioScience, 55(6)

ISSN

0006-3568

Authors

GREEN, JESSICA L
HASTINGS, ALAN
ARZBERGER, PETER
et al.

Publication Date

2005

DOI

10.1641/0006-3568(2005)055[0501:cieacm]2.0.co;2

Copyright Information

Peer reviewed

Complexity in Ecology and Conservation: Mathematical, Statistical, and Computational Challenges

JESSICA L. GREEN, ALAN HASTINGS, PETER ARZBERGER, FRANCISCO J. AYALA, KATHRYN L. COTTINGHAM, KIM CUDDINGTON, FRANK DAVIS, JENNIFER A. DUNNE, MARIE-JOSÉE FORTIN, LEAH GERBER, AND MICHAEL NEUBERT

Creative approaches at the interface of ecology, statistics, mathematics, informatics, and computational science are essential for improving our understanding of complex ecological systems. For example, new information technologies, including powerful computers, spatially embedded sensor networks, and Semantic Web tools, are emerging as potentially revolutionary tools for studying ecological phenomena. These technologies can play an important role in developing and testing detailed models that describe real-world systems at multiple scales. Key challenges include choosing the appropriate level of model complexity necessary for understanding biological patterns across space and time, and applying this understanding to solve problems in conservation biology and resource management. Meeting these challenges requires novel statistical and mathematical techniques for distinguishing among alternative ecological theories and hypotheses. Examples from a wide array of research areas in population biology and community ecology highlight the importance of fostering synergistic ties across disciplines for current and future research and application.

Keywords: ecological complexity, quantitative conservation biology, cyberinfrastructure, metadata, Semantic Web

Ecological communities comprise complex networks of individuals interacting with each other and with their environment. Two major challenges for scientists, managers, and policymakers are (1) understanding the forces that organize these heterogeneous systems across space and time (Levin 1992, Levin et al. 1997, Brown et al. 2002) and (2) using this understanding to address environmental problems at local to global scales. These challenges require novel approaches at the interface of ecology, mathematics, statistics, and computer science. Recent developments have prompted dramatic advances in these disciplines. Exponential gains in computing power, data storage capacity, and network connectivity, together with algorithm development, enable researchers to construct increasingly complex models. *In silico* experiments that model biological systems on the computer are often as important as the traditional in vivo and in vitro experiments to which biology was previously restricted (Gentleman 2002, Chow et al. 2004). Advanced networking capabilities allow researchers to share ideas and data rapidly across disciplines and physical locations. Large arrays of spatially embedded sensor networks are revolutionary tools for observing real-world systems at multiple scales. These and other advances can help researchers make fundamental discoveries in basic ecology and can facilitate effective conservation.

Jessica L. Green (e-mail: jgreen@ucmerced.edu) worked at the University of California in Davis when this article was being prepared; she is now at the School of Natural Sciences, PO Box 2039, University of California, Merced, CA 95344. Alan Hastings works in the Department of Environmental Science and Policy, University of California, Davis, CA 95616. Peter Arzberger is with the National Biomedical Computation Resource, University of California, San Diego, La Jolla, CA 92093. Francisco J. Ayala works in the Department of Ecology and Evolutionary Biology, University of California, Irvine, CA 92697. Kathryn L. Cottingham is with the Department of Biological Sciences, Dartmouth College, Hanover, NH 03755. Kim Cuddington works in the Department of Biological Sciences and the Quantitative Biology Institute, Ohio University, Athens, OH 45701. Frank Davis is with the Donald Bren School of Environmental Science and Management, University of California, Santa Barbara, CA 93106. Jennifer A. Dunne works at the Pacific Ecoinformatics and Computation Ecology Lab, 1604 McGee Avenue, Berkeley, CA 94703, and the Santa Fe Institute, 1399 Hyde Park Road, Santa Fe, NM 87501. Marie-Josée Fortin is with the Department of Zoology, University of Toronto, Toronto M5S 3G5, Canada. Leah Gerber works at the School of Life Sciences, Arizona State University, PO Box 871501, Tempe, AZ 85287. Michael Neubert works in the Biology Department, Woods Hole Oceanographic Institution, Woods Hole, MA 02543. © 2005 American Institute of Biological Sciences.

In December 2003, the National Science Foundation sponsored the last of three workshops on quantitative environmental and integrative biology (Hastings et al. 2005). The workshop sought to identify areas of cutting-edge research in ecology and environmental biology that require integration with novel computational, statistical, and informatics tools. This article reports the workshop's overall findings. It is organized into three sections that reflect major topics discussed at the workshop. First, we describe areas of research in which statistical and computational advances provide new ways of incorporating biological complexity into ecological theory. Second, we illustrate how computational advances improve researchers' ability to solve problems in conservation biology and resource management. Finally, we discuss emerging cyberscience and ecoinformatic tools that facilitate advances in ecological theory and enable researchers to apply this knowledge for informed conservation management decisions.

These topics reflect the participants' expertise, and are illustrative rather than comprehensive. This article is not a review. Nevertheless, our examples illustrate that the interface between ecology, mathematics, statistics, and computer science is rich; that the need to foster integration and collaborations among these disciplines is great; and that the potential impact of this interdisciplinary research is unlimited.

Modeling ecological complexity

> The greatest challenge today, not just in cell biology and ecology but in all of science, is the accurate and complete description of complex systems. Scientists have broken down many kinds of systems. They think they know most of the elements and forces. The next task is to reassemble them, at least in mathematical models that capture the key properties of the entire ensembles. (Wilson 1998, p. 85)

As ecology has matured, our conceptual and theoretical models for how the world works have evolved from the very simple to the very complex (Levin et al. 1997). Simple models that ignore individual and environmental variation, species interactions, and transient dynamics try to capture generalities about systems and offer analytical tractability. However, these models are often insufficient for predicting realistic temporal and spatial patterns. Advances in mathematics, statistics, and computation help us to assess more fully the consequences of such simplifications and to incorporate more realism. In many situations, this translates into more complex models.

A challenge in modeling any system is the choice of level of detail. The challenge resides in identifying which details at one level of organization are driving phenomena at other levels, and which details can be ignored. In many cases, developing a suite of complementary models operating at different scales and levels of complexity will help elucidate the mechanisms underlying observed macroscopic patterns. However,

building more detailed and complex models is not always better. Complexity typically demands additional data and computation time, and makes model results difficult to analyze. Researchers need tools for identifying the situations in which building detailed models will increase our ability to understand and predict the structure and dynamics of ecological systems. For example, recent statistical advances based on model selection (box 1) show substantial promise for distinguishing among alternative ecological hypotheses and theories. In general, situations that call for more detailed models will either require mathematical approximations of added complexity or advances in computer science that allow more efficient computation. In the following sections, we describe three areas in which advances in computational science may improve ecological theory by providing ways to incorporate increased biological complexity.

Transient dynamics. Ecological theory has traditionally focused on long-term or asymptotic behavior as a way to understand natural systems, with stability analysis as the primary tool (May 1974). Even models that incorporate nonequilibrium dynamics, such as limit cycles or chaos, primarily look at long-term behavior. However, it is widely recognized that theoretical studies of short-term dynamics are also needed (Levin 1992) in order to understand and interpret ecological experiments, most of which occur on time scales of less than 1 year (Kareiva and Andersen 1988). Adaptive management and restoration practices require understanding both short- and long-term effects of field manipulations. Transient dynamics, which characterize the behavior of a dynamical system before its terminal behavior, are garnering more attention in the ecological literature (Hastings 2004).

Recent investigations of transient dynamics have changed researchers' view of ecological systems. It is now understood that traditional analyses of ecosystem stability and resilience may give a misleading picture of how ecological systems respond to environmental perturbation. Resilience, which measures how rapidly a stable system returns to its original state after a perturbation, is an asymptotic property giving the rate of decay of perturbations after a very long time. Novel measures of transient response, including reactivity (the maximal instantaneous rate at which perturbations can be amplified) (Neubert and Caswell 1997), have shown that perturbations can grow for a time before decaying, causing dramatic and long-lasting changes that are entirely overlooked by studies of asymptotic behavior. Spatially structured models also suggest that after a major perturbation, population dynamics may become unpredictable for a long time without ever attaining simple asymptotic behavior (Hastings and Higgins 1994). Such complex transient behavior may explain sudden outbreaks in populations (e.g., insects) for which no recent change in environmental conditions has been detected. Recently, the recognition that transient dynamics can be an important aspect of species coexistence has received much attention. This transient coexistence may

Box 1. Model selection.

Basic and applied ecology require high standards of model testing and evaluation. The traditional approach of evaluating models in ecology is to generate a null hypothesis and ask whether the hypothesis can be rejected at some specified level of statistical significance. Yet a model's validity, reliability, and accuracy can be most meaningfully judged by explicit comparison against alternative models (Kirchner et al. 1996). The task of using data to evaluate different competing hypotheses has increasingly occupied the attention of statisticians and ecologists. The underlying ideas of maximizing the likelihood of the data given the parameters of a model (by varying the parameters), and of comparing across models by choosing the one that has the highest information per parameter, provide a powerful approach for connecting theory to data (Burnham and Anderson 2002). The implementation of these ideas of model selection has been facilitated by advances in computational power and statistical developments.

The model-selection approach has begun to be adopted for a diverse range of applications in conservation biology, such as estimating survival rates, modeling the effects of habitat fragmentation on populations, and modeling the effect of fire on community organization (Johnson and Omland 2004). Applications of this approach are still few, and there is substantial opportunity and need for novel uses and concomitant advances in the underlying statistical theory. In particular, these approaches provide some of the best opportunities for making ecological predictions, with uncertainties explicitly included in a rigorous way, which is essential for decisionmaking in the public policy arena.

elucidate mechanistic explanations for patterns in the distribution and abundance of species (Hastings 2004).

Ecologists are just beginning to explore the importance of transient dynamics. Advances in mathematics, statistics, and computing enable more sophisticated analyses of complex dynamical systems and, hence, provide a deeper understanding of how transient dynamics can affect the persistence and structure of ecological communities.

Environmental variability. Many modelers of population, community, and ecosystem dynamics seek to incorporate the effects of temporal environmental variation. Environmental stochasticity is important when attempting to develop predictions for the management of endangered species, invasive populations, harvested populations, or whole reserve areas. Harsh environmental conditions in a single year, or repeatedly bad conditions over a series of years, may decimate a population.

Variable environmental conditions are frequently simulated using a "white-noise" model. Underlying this model is the assumption that environmental fluctuations are temporally

uncorrelated. However, many environmental signals (e.g., temperature, rainfall, river height) are positively autocorrelated, or have a "reddened" noise signal, with continually increasing variance in time (Steele 1985, Vasseur and Yodzis 2004). A run of bad conditions is more likely than swiftly alternating conditions. There is, however, little work demonstrating the impact of positively autocorrelated environmental signals on commonly used management methods, such as population viability analysis (PVA). Studies that do incorporate realistic reddened noise signals usually rely on simulations to achieve generalization (Johst and Wissel 1997, Cuddington and Yodzis 2002).

Because researchers continue to use white-noise models, even in the face of contradictory environmental data, research that focuses on providing a solid theoretical framework for the analysis of reddened environmental variation is sorely needed. Marion and colleagues (2000) suggest some simple models of colored environmental noise, and they have made progress in analyzing the population effects of such variation by applying analytical approximations such as local linearization of stochastic differential formulations and moment-closure techniques. Further research that focuses on the development of analytically tractable methods for incorporating environmental stochasticity will be of great import, especially where such methods yield techniques that can be applied to the protection of endangered species (extinction risk analysis, below) or the management of harvested populations (box 2).

Complex ecological networks. A challenge in the study of complex systems is integrating recent research on network structures with advances in modeling the dynamics of large nonlinear systems (Strogatz 2001). Networks of many interacting species are widely observed in nature, but few models have successfully simulated persistent dynamics of complex ecosystems. Since the 1970s, mathematical approaches have been used to describe general aspects of the network structure, dynamics, and stability of food webs, but much of the early work inspired by May (1974) was based on simple, analytically tractable models. More detailed ecological data and new approaches to network analysis have supported advances in the description of general properties of complex food web structure (figure 1; Dunne et al. 2002, Garlaschelli et al. 2003). Researchers have used biologically realistic, nonlinear mathematical models to explore trophic dynamics (reviewed in McCann 2000), but they have focused on relatively small systems with fewer than 10 species. A few studies have explored ways to integrate complex structure and dynamics in more diverse empirical (Yodzis 2000) and model ecosystems (Kondoh 2003), but such studies often include questionable assumptions about structure and dynamics (Brose et al. 2003a).

Research on complex ecological networks is computationally intensive and was effectively impossible a decade ago. The increase in personal computer power, as well as the availability of local, low-cost supercomputing power, has made such research widely feasible. The study of complex eco-
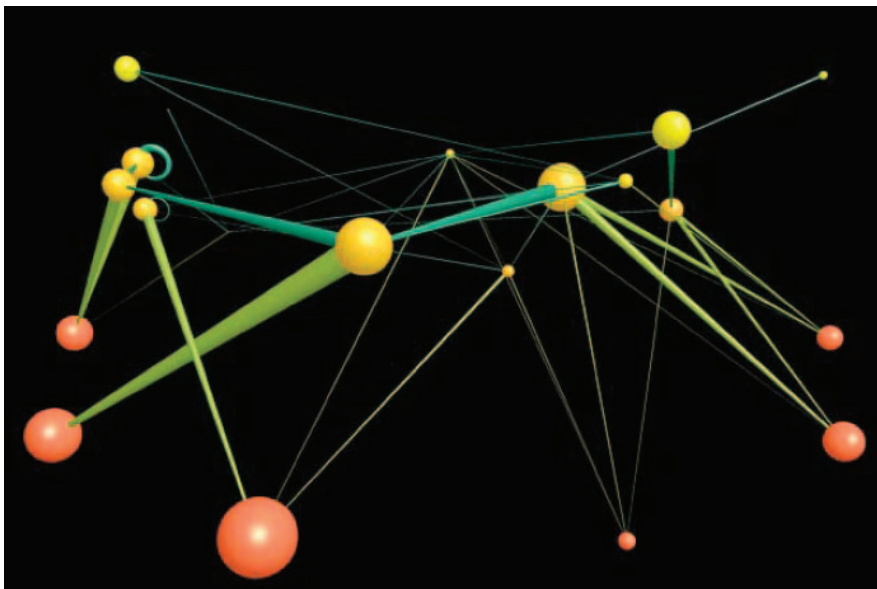
*Figure 1. Three-dimensional snapshot of a 20-species, chaotically persisting food web whose structure is generated by the empirically successful niche model and whose dynamics are simulated by extending a nonlinear bioenergetic approach (Yodzis and Innes 1992) to n species (Brose et al. 2003a). Spheres represent species (plant species at bottom); the volume of each sphere is proportional to the species' biomass. Cones represent feeding links between species (cannibalism represented by circular cones); the cross-sectional area of each cone is proportional to the species' feeding rates. Graphic courtesy of the Webs on the Web project (*www.foodwebs.org*).*

Biological diversity is being lost at a record pace as a result of habitat loss and fragmentation, climate change, pollution, introduction of exotic species, and overharvesting (Vitousek et al. 1997). Effective policies for preserving global biodiversity depend on accurate predictions of species' temporal and spatial distributions. As anthropogenic stresses escalate, the need for reliable quantitative approaches in environmental problem solving is hard to overstate.

In the previous section, we discussed the need to develop new mathematical and computational techniques for understanding complex biological systems. We now highlight some of the challenges of applying these techniques to specific issues in conservation and resource management. Conservation biology currently relies on quantitative methods, but there are many hurdles to solving complex problems, including estimating past and predicting future population dynamics, and optimizing the spatial design of reserves in a changing environment. We discuss some of these challenges in the areas of extinction risk analysis, landscape connectivity analysis, and biodiversity estimation.

logical networks encompasses three major challenges that will drive, and take advantage of, emerging quantitative and computational methods. First, recent insights into the complex structure of food web networks need to be integrated with modeling the transient, long-term, and evolutionary dynamics of diverse nonequilibrium, nonlinear ecosystems. Second, ecological network research needs to more effectively encompass other interactions such as parasitism, pollination, competition, mutualism, and trait-mediated indirect effects (Bascompte et al. 2003). Third, approaches for exploring and constraining the large parameter spaces generated by high-dimensional models need to be developed. In general, the synthetic nature of ecological network analysis and model development will be facilitated by advances in ecoinformatics, such as the Semantic Web (see below), that will increase the ability of researchers to access and integrate diverse types of data.

Applications in conservation biology and resource management

> The rates, scales, kinds, and combinations of [global] changes occuring now are fundamentally different from those at any other time in history; we are changing Earth more rapidly than we are understanding it.
> (Vitousek et al. 1997, p. 498)

Extinction risk analysis. In the United States and other countries, the development of a PVA is a legal requirement of any survival plan for threatened and endangered species. Typical objectives of PVA include assessing the risk of reaching some threshold, such as extinction, and projecting population growth, either under current conditions or those predicted by proposed management plans (Morris and Doak 2002). There is growing concern over the use of PVA models for making conservation decisions, in part because census data for threatened species are often sparse and error prone, causing substantial difficulties in estimating population trends (Gerber et al. 1999). A primary challenge in PVA is characterizing and accounting for uncertainties that result from process noise (e.g., random variations among individuals or in environmental conditions) and observation error (e.g., the difficulty of accurate measurements under field conditions).

Including two types of error in time-series models for fluctuating populations is particularly challenging. First, there are rarely good estimates for both kinds of error, requiring errors to be estimated along with parameters. Second, analytical approaches incorporating both kinds of error in fits of nonlinear time-series models to observed data are scarce (De Valpine and Hastings 2002). Newer methods using maximum-likelihood approaches offer a promising avenue for making PVA predictions with process and observation

uncertainties clearly specified and described. This approach still needs to be extended to more complex models, particularly those that include spatial structure (Reed et al. 2002), individual variation (Kendall and Fox 2002), and autocorrelated environmental variation (as discussed in "Environmental variability," above). Finally, current computational approaches to implement these methods can be prohibitively complex. Further work is needed to develop algorithms that are as efficient as possible (Miller 1998).

Spatiotemporal landscape connectivity analysis. Protecting wildlife populations requires quantifying how changes in landscape spatial composition, and the arrangement of habitats of differing quality, affect animal movement in fragmented landscapes (figure 2; Bélisle et al. 2001). Species are affected differently by landscape fragmentation because of their specific range size, dispersal ability, habitat and food requirements, and behavior. Moreover, species' abilities to move across a landscape vary depending on the spatial configuration of habitats, the distance separating habitats, and the intervening cover types (D'Eon et al. 2002). To implement

conservation goals and maintain populations in fragmented landscapes, structural connectivity among habitats needs to be preserved through time. Researchers need to develop quantitative measures of spatiotemporal landscape connectivity that characterize the degree to which the landscape impedes or facilitates the movement of organisms.

There are several statistical and modeling challenges to overcome before this can be achieved, especially within a spatially explicit modeling environment. For instance, connectivity of habitats is dynamic, and fluctuates as a result of succession and disturbances that modify habitat quality and resource availability. Graph theory offers considerable promise in the analysis of landscape connectivity at multiple spatial and temporal scales (Urban and Keitt 2001). Approaches developed in other disciplines, such as circuitry and network optimization in computer science, can provide a quantitative framework for modeling the flux of populations between habitat patches in different landscape mosaics. Although existing applications of graph theory to landscape ecology account for patch size in quantifying landscape connectivity, an important avenue for future research is to consider the shapes of patches and their dynamic properties. Environmental perturbations such overharvesting (box 2), wildfire (box 3), urban sprawl, and global climate change operate across a variety of temporal scales, and thus effective

---

Box 2. Fisheries management and reserve design.

The closure of much of the groundfish fishery off the West Coast of the United States in 2004 (*www.nwr.noaa.gov/1sustfsh/gdfsh01.htm*) is an indication that new and revolutionary methods are needed to manage fisheries. Traditional management is based on extensions (incorporating much more sophisticated ideas) of the basic Schaefer model (Schaefer 1957),

$$dn/dt = rn \,(1 - n/K) - hn,$$

of logistically growing stock (n) harvested at the rate h. This model is used to determine the maximum sustainable yield ($Y = Kr/4$), which is achieved when the harvest is set at half the intrinsic rate of growth of the stock (r), which in turn reduces the stock to half its size at equilibrium in the absence of harvesting (K). As the groundfish example shows, a combination of scientific and political issues makes management along these lines problematic.

One potential solution is to use marine reserves that allow similar yields for fisheries (Hastings and Botsford 1999) but may have the advantages of sustainability. This has begun to be implemented in areas such as the Channel Islands, off the coast of southern California (*www.cinms.nos. noaa.gov/marineres/main.html*). The design of successful marine reserves requires substantial theoretical developments, including models that incorporate better descriptions of physical influences, biological interactions, monitoring, and the economics of fishing.
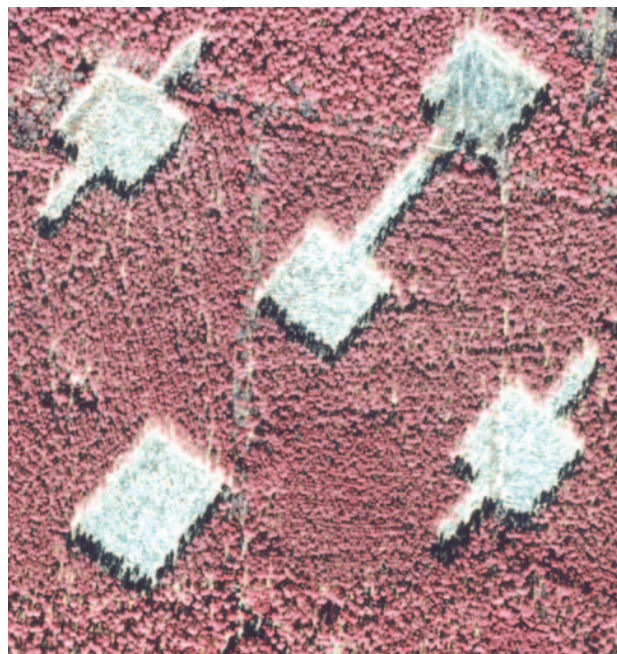

Figure 2. Infrared aerial photograph from a habitat fragmentation and corridor experiment at the Savannah River Site, South Carolina. Experimental patches are clearings, and are surrounded by pine plantation forest (shown in red). The length of the corridor is 150 meters. The experiment has been used to test for corridor and fragmentation effects on plant and animal movement, population demography, and diversity (Tewksbury et al. 2002). Photograph courtesy of the USDA Forest Service.

conservation practices rely on dynamical models that incorporate the rate of habitat loss into landscape connectivity analyses.

Biodiversity extrapolation techniques. Patterns in the spatial distribution of species are a central concern in ecology, providing information about the forces that regulate biodiversity, the design of nature reserves, and the likelihood of species extinction following climate change or habitat loss. For most habitats and taxonomic groups, detailed species distribution maps are unavailable, and researchers have invested considerable effort in developing methods for estimating the total number of species in particular localities, regions, and biomes using sparse sample data. Current extrapolation approaches have many shortcomings, and new computational and statistical techniques for estimating biodiversity are critically needed.

Parametric methods, species accumulation curves, and nonparametric estimators are three tools commonly used to estimate species richness from samples (Magurran 2004). Parametric methods estimate the number of species in a community by fitting sample data to distributional models of relative abundance. It is difficult to know a priori which distribution is appropriate for the region and taxonomic group of interest. Parametric approaches implicitly assume that individuals are randomly sampled in space or, equivalently, that the spatial distribution of individuals across a landscape is random. However, most organisms are spatially aggregated, and parametric extrapolation methods should account for this heterogeneity. The performance of species accumulation curves and nonparametric methods, on the other hand, is not substantially affected by species' spatial distributions (Brose et al. 2003b). Species accumulation curves use an assumed model (e.g., the Michaelis–Menten equation) to extrapolate an asymptote of total species richness from data on richness and sample size. Nonparametric estimators, adapted from mark–release–recapture statistics for estimating the size of animal populations, assume models for how singletons and doubletons are distributed in the sample community. Both methods significantly underestimate biodiversity for low levels of sampling intensity (Condit et al. 1998, Brose et al. 2003b) and thus provide only a lower bound on diversity for highly abundant and diverse taxa, such as invertebrates and microorganisms, that are difficult or impossible to sample extensively.

Answering questions such as "How many species are on Earth?" and "What is the rate of global biodiversity loss?" will require the development of creative, robust statistical techniques. Approaches that utilize patterns of beta diversity (how community composition changes across a landscape) have already been developed and applied to extrapolate microbial diversity from the scale of a soil sample, where richness is quantifiable, to regional scales (Green et al. 2004, Horner-Devine et al. 2004). Future efforts should account for uncertainties in community dominance, in species' spatial distributions, and in sampling intensity. Novel techniques

> ### Box 3. Landscape ecology and wildfire management.
>
> The 2003 wildfires in southern California are a sharp reminder of the importance and complexity of managing wildfire in fire-prone ecosystems. The fires burned more than 290,000 hectares of intermixed wildlands and residential areas and consumed more than 2000 homes. Managing wildfire for human safety, while at the same time managing for the persistence of fire-dependent native species and ecosystems, forest resources, and watershed processes, makes heavy demands on the theory, quantitative methods, and informatics tools of landscape ecology and geography (Brooks et al. 2004). The behavior of individual fire events and the evolution of fire regimes over time involve strong interactions among (a) regional climate, biota, weather, terrain, and fuel conditions; (b) patterns of roads and human settlement; and (c) emergency response capabilities.
>
> Wildfire information systems are being implemented at national (e.g., the USDA Forest Service Wildland Fire Assessment System; *www.fs.fed.us/land/wfas/*) and regional (e.g., the Southern California Wildfire Hazard Center; *www.icess.ucsb.edu/resac/*) levels. Integrating physical, biological, and social models and linking them to geospatial data is a challenging exercise, both scientifically and in terms of informatics theory and practice. Improved capabilities need to be developed for real-time assimilation of remotely sensed data and data gathered on the ground. Such capabilities are being developed at the San Diego Supercomputer Center's National Partnership for Advanced Computational Infrastructure, or NPACI, in the Earth System Science thrust area.

for estimating biodiversity, combined with emerging cyber-science technologies that enhance access to species distribution data, will facilitate our understanding of local and global biodiversity.

Cyberinfrastructure and ecoinformatics

Exciting things are happening in the life sciences. The big challenges such as cancer, AIDS, and drug discovery for new viruses require the interplay of vast amounts of data from many fields that overlap: genomics, proteomics, epidemiology, and so on. Some of this data is public, some very proprietary to drug companies, and some very private to a patient. The Semantic Web challenge of getting interoperability across these fields is great but has huge potential benefits. (Sir Tim Berners-Lee, quoted in Frauenfelder 2004)

Ecological research and its application to conservation management require that researchers acquire existing data from geographically, technologically, and intellectually disparate resources; integrate that information; model and analyze the information; and recommend policies such as establishing ecological reserves (box 2), incorporating wildfire dynamics into urban planning (figure 3, box 3), and managing invasive species. For example, understanding the potential impacts of an invasive species on an ecosystem requires access to diverse information, including the basic taxonomy and population dynamics of invasive and extant organisms, the structure and dynamics of food webs, environmental conditions, and the outcomes of experimental and observational studies from related systems and organisms. Developing ecoinformatics and supporting cyberinfrastructure could greatly enhance researchers' ability to intelligently retrieve information from diverse sources on the Internet, to integrate that information into models that predict the spread of the species under various management options, to store these results in databases for other researchers and managers, and to monitor the impact of the decisions made.

We now present trends in technology that are likely to transform the scenarios we have just described into reality, and discuss issues that the ecological community needs to address to bring this vision to fruition. Pieces of these technologies are in place already. The emerging tools, technologies, and infrastructure can advance current approaches to research and management, and can alter how ecologists look at our science, opening new windows of opportunity for research and application.

Cyberinfrastructure. Technological trends in global information and computing infrastructure are transforming the shape of computing and the conduct of science. The ability to place sensors in the environment allows researchers to conduct remote sampling at new frequencies and larger spatial and temporal scales. Partly as a result of these sensors, scientists are witnessing an avalanche of physical, chemical, and biological data similar to the avalanche of DNA, RNA, and protein sequences faced by molecular biologists. Interpretation of the large data sets being generated by sensor nets and associated technologies requires drastically more powerful and efficient computational capabilities. Fortunately, exponential gains in networking, storage, and computing capabilities enable scientists to compile, manipulate, and analyze data at unprecedented rates. For example, fast, cheap, local cluster computing is now possible through off-the-shelf computational nodes and software, allowing easy construction and maintenance of supercomputers (box 4). These clusters facilitate intensive computational activities such as data mining and nonlinear dynamical modeling and simulation.

Advances in data generation and computing technologies have increased scientists' understanding of issues including lake metabolism, the movement of elusive species, landscape-scale patterns and processes, and the impact of natural disasters such as hurricanes (Estrin et al. 2003). Building on

Figure 3. Satellite image of fires that spread across southern California, 26 October 2003. This image was captured by the Moderate Resolution Imaging Spectroradiometer (MODIS) on the Terra satellite. The fires, which are marked in bright red, send streamers of gray-blue smoke into the air. The Terra and Aqua satellites view the entire Earth's surface every 1 to 2 days, acquiring data in 36 spectral bands. Imagery and data from MODIS can be rapidly provided to a number of users, including the USDA Forest Service, to facilitate wildfire management. In general, these types of informatics tools will improve understanding of global dynamics and processes occurring on the land, in the oceans, and in the lower atmosphere. Image courtesy of MODIS Rapid Response Project at the National Aeronautics and Space Administration, Goddard Space Flight Center.

those insights and developing a better understanding of how to sustain ecosystems at multiple scales requires an integrated, global-scale research cyberinfrastructure consisting of distributed computer, information, observation, and communication technologies (Atkins et al. 2003). Examples of cyberinfrastructure span many disciplines and countries (box 4; Wooley 2004). Within environmental biology, the National Ecological Observatory Network (NEON), a continental-scale network of observatories with state-of-the-art instrumentation for studying the environment, has been developed in reports by the American

Institute of Biological Sciences (*http://ibrcs.aibs.org/reports*) and the National Academies (NRC 2003). Projects such as NEON will enable integrative research on the nature and pace of biological change at local, regional, continental, and global scales.

Metadata standards. Managing large volumes of complex data is becoming one of science's key challenges (Maurer et al. 2000). Metadata, or "data about data," describe the content, quality, structure, and other characteristics of a data set. Federal programs and university researchers are laboring to produce metadata standards, tools, and online resources to promote multidisciplinary data sharing (box 4). Scientific communities in fields such as geography and genomics adopted metadata standards relatively early on, and automated metadata tools for these communities quickly followed. For example, commercial GIS (geographic information system) packages now include software to simplify preparation of metadata in accordance with national standards (ArcCatalogue software is included in the ESRI ArcGIS software package).

Most ecologists remain uninformed about and uninterested in metadata standards. Historically, ecological data have been collected by individuals or small research groups at small temporal and spatial scales. Because ecologists do not typically publish or share research data, good metadata practices are not the cultural norm (Gross et al. 1995, Michener et al. 1997). Recently, metadata standards have been developed for ecological data (Ecological Metadata Language, or EML). Organizations such as the National Center for Ecological Analysis and Synthesis are developing a network-enabled database framework that allows individual scientists to customize metadata to meet their needs, while also promoting the use of standards such as the US Ecological Metadata Standard and the National Biological Information Infrastructure's Biological Data Profile (Jones et al. 2001). A broad and concerted community effort will ensure long-term acceptance of metadata standards by research ecologists, thereby facilitating the sharing, archiving, and integration of data from different disciplines to help scientists understand broad environmental processes.

Box 4. Examples of cyberinfrastructure, metadata standards, and Semantic Web development that span many countries and disciplines.

Cyberinfrastructure
Asia Pacific Grid: *www.apgrid.org*
Beowulf Cluster Computing: *www.beowulf.org*
Biomedical Informatics Research Network: *www.nbirn.net*
Geosciences Network: *www.geongrid.org*
Grid Physics Network: *www.griphyn.org*
NASA Information Power Grid: *www.ipg.nasa.gov*
National Ecological Observatory Network: *www.neoninc.org*
National Virtual Observatory: *www.us-vo.org*
Network for Earthquake Engineering Simulation: *www.nees.org*
Pacific Rim Applications and Grid Middleware Assembly: *www.pragma-grid.net*
Research Councils UK, e-Science Programme: *www.research-councils.ac.uk/escience/*
Rocks Cluster Distribution: *www.rocksclusters.org/Rocks/*
TeraGrid: *www.teragrid.org*

Metadata standards
Ecological Metadata Language: *http://knb.ecoinformatics.org/software/eml/*
Marine Environmental Data Inventory: *http://ioc.unesco.org/medi/*
National Biological Information Infrastructure: *www.nbii.gov/datainfo/metadata/*
National Center for Supercomputing Applications: *http://metadata.ncsa.uiuc.edu/*
National Oceanographic and Atmospheric Administration: *www.csc.noaa.gov/metadata/*
US Geological Survey: *http://geology.usgs.gov/tools/metadata/*

The Semantic Web
DARPA Agent Markup Language: *www.daml.org*
Gene Ontology Next Generation Consortium: *http://gong.man.ac.uk/*
Knowledge Network for Biocomplexity: *http://knb.ecoinformatics.org*
NASA Semantic Web for Earth and Environmental Terminology: *http://sweet.jpl.nasa.gov*
National Biological Information Infrastructure: *www.nbii.org*
Science Environment for Ecological Knowledge: *http://seek.ecoinformatics.org*
Semantic Prototype in Research Ecoinformatics: *http://spire.umbc.edu*
US National Cancer Institute Metathesaurus Project: *http://ncimeta.nci.nih.gov*
Web Ontology Language: *www.w3.org/TR/owl-features/*
Webs on the Web: *http://foodwebs.org*
World Wide Web Consortium: *www.w3.org*

The Semantic Web. An increasing amount of information is available on the World Wide Web (WWW) in dispersed databases, Web pages, publications, images, and other formats. Scientists already rely on the WWW to support their research endeavors in ways that include searching for and distributing publications, accessing databases, and ordering supplies (Hendler 2003). Even though an incredible array of potentially scientifically useful information resides in countless WWW pages, the conventions of HTML (hypertext markup language) and current search engines are limited in their ability to take advantage of such information. Because of the huge variety of formats and the inability of current search tools to do more than look for text with no semantic content indicating the contextual meaning of the word, efforts are under way to develop languages and tools for a Semantic Web that will allow for more sophisticated, content-based access to data and information on the WWW (Berners-Lee et al. 2001, Hendler 2003, Frauenfelder 2004). Semantic annotation of WWW content and other Semantic Web technologies will facilitate the development and deployment of new WWW-based databases, ontologies, and intelligent search agents with greatly enhanced powers to strategically gather and retrieve information that is diverse in terms of

both content and format, yet appropriate and useful for basic and applied tasks. Based on ontologies that generally specify the types of information within certain domains, the Semantic Web can assist scientific research goals by providing a platform for swift, integrated, global access to scientific information, facilitating previously time-consuming or currently intractable disciplinary and interdisciplinary syntheses and analyses.

Numerous ontologies and information-technology tools are being developed for Semantic Web use within a scientific research context (box 4). Within environmental and integrative biology, there are two major consortia in which natural and computer scientists are collaborating to develop and test the Semantic Web and related technologies. The Science Environment for Ecological Knowledge, or SEEK, project is developing software to assist how researchers (a) gain global access to ecological data and information, (b) rapidly locate and utilize distributed computational services, and (c) use powerful new methods for capturing, reproducing, and extending the analysis itself. The Semantic Prototypes in Research Ecoinformatics, or SPiRE, project is testing Semantic Web capability for biodiversity and food web research using the Web portal of the National Biological Information Infrastructure project in conjunction with the Webs on the Web, or WoW, project on ecological networks. The emerging information technologies of the Semantic Web, if properly developed and widely implemented, have the potential to transform and augment the scope, effectiveness, and efficiency of biological research at all levels of organization.

Facilitating interdisciplinary biosciences research

The examples given show how imaginative approaches at the interface of ecology, statistics, mathematics, informatics, and computational science can improve scientists' understanding of complex ecological systems and our approach to biological conservation and resource management. We have made significant progress, but further advances will demand shifting the way that we approach research and education. Many academic institutions are facilitating interdisciplinary research and teaching programs to accelerate knowledge in the biosciences (e.g., Stanford University's Bio-X program). Individuals, research institutions, and funding agencies must invest more resources in developing and sustaining cross-disciplinary research collaborations to generate more generally applicable research. Educational institutions need to invest in programs that provide biologists with robust quantitative and informational skills, and that provide computer scientists, mathematicians, and statisticians with biological expertise. The combination of mathematical and computational advances, sophisticated informatics technologies, and synergistic ties across disciplines may well lead to this century's most fundamental advances in ecology and environmental biology.

References cited
Atkins DE, Droegemeier KK, Feldman SI, Garcia-Molina H, Klein ML, Messina P, Messerschmitt DG, Ostriker JP, Wright MH. 2003. Revolutionizing Science and Engineering through Cyberinfrastructure: Report of the National Science Foundation Blue Ribbon Advisory Panel on Cyberinfrastructure. Arlington (VA): National Science Foundation. (1 April 2005; *www.cise.nsf.gov/b_ribbon/*)

Bascompte J, Jordano P, Melian CJ, Olesen JM. 2003. The nested assembly of plant–animal mutualistic networks. Proceedings of the National Academy of Sciences 100: 9383–9387.

Bélisle M, Descrochers A, Fortin M-J. 2001. Influence of forest cover on the movements of forest birds: A homing experiment. Ecology 82: 1893–1904.

Berners-Lee T, Hendler J, Lassila O. 2001. The Semantic Web. Scientific American 284: 35–43.

Brooks ML, D'Antonio CM, Richardson DM, Grace JB, Keeley JE, DiTomaso JM, Hobbs RJ, Pellant M, Pyke D. 2004. Effects of invasive alien plants on fire regimes. BioScience 54: 677–688.

Brose U, Williams RJ, Martinez ND. 2003a. Comment on "Foraging adaptation and the relationship between food-web complexity and stability." Science 301: 918b.

Brose U, Martinez ND, Williams RJ. 2003b. Estimating species richness: Sensitivity to sample coverage and insensitivity to spatial patterns. Ecology 84: 2364–2377.

Brown JH, Gupta VJ, Li B, Milne BT, Restrepo C, West GB. 2002. The fractal nature of nature: Power-laws, ecological complexity and biodiversity. Philosophical Transactions: Biological Sciences 357: 619–626.

Burnham KP, Anderson DR. 2002. Model Selection and Multi-model Inference: A Practical Information-Theoretic Approach. New York: Springer-Verlag.

Chow SS, Wilke CO, Ofria C, Lenski RE, Adami C. 2004. Adaptive radiation from resource competition in digital organisms. Science 305: 84–86.

Condit R, Foster RB, Hubbell SP, Sukumar R, Leigh EG, Manokaran N, Loo de Lao S. 1998. Assessing forest diversity on small plots: Calibration using species–individual curves from 50 ha plots. Pages 247–268 in Dallmeier F, Comiskey JA, eds. Forest Biodiversity Diversity Research, Monitoring, and Modeling. Paris: UNESCO, Parthenon Publishing Group.

Cuddington K, Yodzis P. 2002. Predator–prey dynamics and movement in fractal environments. American Naturalist 160: 119–134.

D'Eon RG, Glenn SM, Parfitt I, Fortin M-J. 2002. Landscape connectivity as a function of scale and organism vagility in a real forested landscape. Conservation Ecology 6: 10.

De Valpine P, Hastings A. 2002. Fitting population models incorporating process noise and observation error. Ecological Monographs 72: 57–76.

Dunne JA, Williams RJ, Martinez ND. 2002. Food-web structure and network theory: The role of connectance and size. Proceedings of the National Academy of Sciences 99: 12917–12922.

Estrin D, Michener W, Bonito G. 2003. Environmental Cyberinfrastructure Needs for Distributed Sensor Networks: A Report from a National Science Foundation Sponsored Workshop. (1 April 2005; *www.lternet.edu/sensor_report*)

Frauenfelder M. 2004. Sir Tim Berners-Lee: He created the Web. Now he's working on Internet 2.0. Technology Review (October): 40–45.

Garlaschelli DG, Caldarelli G, Pietronero L. 2003. Universal scaling relations in food webs. Nature 423: 165–168.

Gentleman W. 2002. A chronology of plankton dynamics *in silico*: How computer models have been used to study marine ecosystems. Hydrobiologica 48: 69–85.

Gerber LR, DeMaster DP, Kareiva PM. 1999. Gray whales illustrate the value of monitoring data in implementing the Endangered Species Act. Conservation Biology 13: 1215–1219.

Green JL, Holmes AJ, Westoby M, Oliver I, Briscoe D, Dangerfield M, Gillings M, Beattie AJ. 2004. Spatial scaling of microbial eukaryote diversity. Nature 432: 747–750.

Gross KL, et al. 1995. Final Report of the Ecological Society of America Committee on the Future of Long-term Ecological Data (FLED). (1 April 2005; *http://intranet.ltrnet.edu/archives/documents/other/fled.vol1.pdf* and *http://intranet.ltrnet.edu/archives/documents/other/fled.vol2.pdf*)

Hastings A. 2004. Transients: The key to long-term ecological understanding? Trends in Ecology and Evolution 19: 39–45.

Hastings A, Botsford LW. 1999. Equivalence of yield from marine reserves and traditional fisheries management. Science 284: 1537–1538.

Hastings A, Higgins K. 1994. Persistence of transients in spatially structured models. Science 263: 1133–1136.

Hastings A, Arzberger P, Bolker B, Collins S, Ives AR, Johnson NA, Palmer MA. 2005. Quantitative bioscience for the 21st century. BioScience 55: 511–517.

Hendler J. 2003. Science and the Semantic Web. Science 299: 520–521.

Horner-Devine M, Lage M, Hughes J, Bohannan B. 2004. A taxa–area relationship for bacteria. Nature 432: 750–753.

Johnson JB, Omland KS. 2004. Model selection in ecology and evolution. Trends in Ecology and Evolution 19: 101–108.

Johst K, Wissel C. 1997. Extinction risk in a temporally correlated fluctuating environment. Theoretical Population Biology 52: 91–100.

Jones MB, Berkley J, Bojilova J, Schildhauer M. 2001. Managing scientific metadata. IEEE Internet Computing 5: 59–68.

Kareiva P, Andersen M. 1988. Spatial aspects of species interactions: The wedding of models and experiments. Pages 38–54 in Hastings A, ed. Community Ecology. New York: Springer Verlag.

Kendall BE, Fox GA. 2002. Variation among individuals and reduced demographic stochasticity. Conservation Biology 16: 109–116.

Kirchner JW, Hooper RP, Kendall C, Neal C, Leavelsley G. 1996. Testing and validating environmental models. Science of the Total Environment 183: 33–47.

Kondoh M. 2003. Foraging adaptation and the relationship between food-web complexity and stability. Science 299: 1388–1391.

Levin S. 1992. The problem of pattern and scale in ecology. Ecology 73: 1943–1967.

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

Magurran AE. 2004. Measuring Biological Diversity. Malden (MA): Blackwell Science.

Marion G, Renshaw E, Gibson G. 2000. Stochastic modelling of environmental variation for biological populations. Theoretical Population Biology 57: 197–217.

Maurer SM, Firestone RB, Scriver CR. 2000. Science's neglected legacy. Nature 405: 117–120.

May RM. 1974. Stability and complexity in model ecosystems. Princeton (NJ): Princeton University Press.

McCann K. 2000. The diversity–stability debate. Nature 405: 228–233.

Michener WK, Brunt JM, Helly JJ, Kirchner TB, Stafford SG. 1997. Non-geospatial metadata for the ecological sciences. Ecological Applications 71: 330–342.

Miller JH. 1998. Active nonlinear tests (ANTs) of complex simulation models. Management Science 44: 820–830.

Morris WF, Doak DF, eds. 2002. Quantitative Conservation Biology: Theory and Practice of Population Viability Analysis. Sunderland (MA): Sinauer.

Neubert MG, Caswell H. 1997. Alternatives to resilience for measuring the responses of ecological systems to perturbations. Ecology 78: 653–665.

[NRC] National Research Council. 2003. NEON: Address the Nation's Environmental Challenges. Washington (DC): National Academies Press.

Reed JM, Mills LS, Dunning JB, Menges ES, McKelvey KS, Frye R, Beissinger SR, Anstett M, Miller P. 2002. Emerging issues in population viability analysis. Conservation Biology 16: 7–19.

Schaefer MB. 1957. Some aspects of the dynamics of populations important to the management of commercial marine fisheries. Journal of the Fisheries Research Board of Canada 14: 669–681.

Steele JH. 1985. A comparison of terrestrial and marine ecological systems. Nature 313: 335–358.

Strogatz SR. 2001. Exploring complex networks. Nature 410: 268–276.

Tewksbury JJ, Levey DJ, Haddad NM, Sargent S, Orrock JL, Weldon A, Danielson BJ, Brinkerhoff J, Damschen EI, Townsend P. 2002. Corridors affect plants, animals, and their interactions in fragmented landscapes. Proceedings of the National Academy of Sciences 99: 12923–12926.

Urban D, Keitt TH. 2001. Landscape connectivity: A graph-theoretic perspective. Ecology 82: 1205–1218.

Vasseur DA, Yodzis P. 2004. The color of environmental noise. Ecology 1146–1152.

Vitousek PM, Mooney HA, Lubchenco J, Melillo JM. 1997. Human domination of Earth's ecosystems. Science 277: 494–499.

Wilson EO. 1998. Consilience: The Unity of Knowledge. New York: Knopf.

Wooley JC. 2004. Building a Cyberinfrastructure for the Biological Sciences—2005 and Beyond: A Roadmap for Consolidation and Exponentiation. Interim Report of an NSF-Sponsored Workshop. (14 April 2005; *http://research.calit2.net/cibio/report*)

Yodzis P. 2000. Diffuse effects in food webs. Ecology 81: 261–266.

Yodzis P, Innes S. 1992. Body-size and consumer-resource dynamics. American Naturalist 139: 1151–1173.