ISSN 1938-7024

LESSONS IN Conservation

ISSUE NO. 3 SEPTEMBER 2010

Network of Conservation Educators and Practitioners
Center for Biodiversity and Conservation
AMERICAN MUSEUM OF NATURAL HISTORY

http://ncep.amnh.org/linc

Lessons in Conservation (LinC) is the official journal of the Network of Conservation Educators and Practitioners and is published as issues become available. The Network of Conservation Educators and Practitioners (NCEP) is a collaborative project of the American Museum of Natural History's Center for Biodiversity and Conservation and a number of institutions and individuals around the world. The vision of NCEP is a highly trained network of individuals effectively managing and sustaining the world's biological and cultural diversity. Teaching modules presented here in LinC are available in modifiable form for teachers on the NCEP website (http://ncep.amnh.org). All materials are distributed free of charge. Any opinions, findings and conclusions, or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the American Museum of Natural History or the funders of this project. All components (Syntheses, Exercises, and Case Studies) have been peer-reviewed and approved for publication by NCEP.

Cover photo by: Eleanor J. Sterling

Network of Conservation Educators and Practitioners
Center for Biodiversity and Conservation
AMERICAN MUSEUM OF NATURAL HISTORY

LINC

http://ncep.amnh.org/linc

Lessons in Conservation (LinC)
Developing the capacity to sustain the earth's diversity

Dear Reader,

We are delighted to present the third issue of LinC, Lessons in Conservation, the official journal of the Network of Conservation Educators and Practitioners (NCEP, http://ncep.amnh.org) of the Center for Biodiversity and Conservation (CBC) of the American Museum of Natural History. On these pages, you will find selected NCEP teaching modules, presented in an easy-to-browse PDF format. LinC is designed to introduce NCEP teaching materials to a broad audience. After browsing through LinC, we hope that university faculty members and other teachers and trainers will be inspired to visit and download additional materials from the NCEP site, and to try them in the classroom. We welcome feedback on our modules and we especially welcome those wishing to become further involved in the project!

In this issue of LinC, we present a selection of more advanced topics in biodiversity conservation and related topics, including ecological consequences of extinction, species distribution modeling, and biodiversity conservation and human health. Future issues will be released semi-annually, and will include Case Studies to complement our Syntheses and Exercises.

On our back cover, we are pleased to acknowledge the foundations and individuals that have supported NCEP and LinC. Happy reading, we look forward to your input and comments, and to seeing you again soon on these pages!

Eleanor Sterling	Nora Bynum
Co-Editor	Co-Editor

NCEP workshop participants in Madagascar and Peru, Source: CBC

Table of Contents

SYNTHESES

EXERCISES

Biodiversity Conservation and Human Health

Andrés Gómez [*] and Elizabeth Nichols [†]

[*] The American Museum of Natural History, New York, NY, U.S.A., email agomez@amnh.org

[†] The American Museum of Natural History, New York, NY, U.S.A., email nichols@amnh.org

K. Frey

Table of Contents

Biodiversity Conservation and Human Health

Andrés Gómez and Elizabeth Nichols

Introduction

Current levels of *anthropogenic* environmental disturbance have led to unprecedented loss of biodiversity at a global scale. Human health directly and indirectly depends on the goods and services provided by biodiversity, and thus can be negatively affected by its loss. The World Health Organization (WHO), among others, has highlighted that the linkages between biodiversity and human health have been the focus of much recent attention (WHO, 2006). Because goods and services provided by biodiversity are critical for maintaining human health, health has become a conservation topic.

In this synthesis, we present an overview of the current understanding of links between biodiversity and human health, as well as the health implications of biodiversity loss and conservation actions. Here we use the WHO's definition of health, which includes physical, mental, and social stability (WHO, 1946). We define the term biodiversity as *"the variability among living organisms from all sources including,* inter alia*, terrestrial, marine and other aquatic ecosystems and the ecological complexes of which they are part; this includes diversity within species, between species and of ecosystems"* (Convention on Biological Diversity, Article 2). We divide the linkages between biodiversity and health into two categories (direct and indirect), discuss the evidence that supports them, and touch upon cases that illustrate the potential conflicts between biodiversity conservation and public health.

In this document, we use the WHO's Ecosystems program and its linkages as the backbone of our review. Our intention is to provide a succinct compilation of the links as described by the WHO and supported by the scientific literature, and *not* to present the reader with the notion that *all* of biodiversity at *all* times will have a net positive effect on human health. To this end, we have included questions for critical analysis and that highlight: 1) even when there is evidence for a strong positive association, it may be only supplied by a limited subset of species under certain specific ecological contexts; 2) in some cases, links depend upon the functioning of ecosystems at spatial and/or temporal scales that are not amenable to traditional conservation action; and 3) there are many instances in which biodiversity *unfriendly* practices will result in significant improvements in human health. Finally, we want to make the reader aware of the fact that the explicit consideration of the interface between biodiversity conservation and human health is a relatively new field in rapid development. The linkages between biodiversity and human health *"are not usually easy to unravel, describe, or understand, because of their complexity, not their absence"* (Osofsky et al., 2000). We expect that the future will bring additional evidence for those linkages outlined here, as well as a better understanding of functions and services that are not included in this review.

It is safe to say that without the natural world - without plants, microorganisms, fungi, animals, and other components of biological diversity – humans would cease to exist. Some measure of biological diversity and its interactions are absolutely required to sustain human life – and, therefore, human health. A more pragmatic question to ask, perhaps, is how much? Which species, in populations of what sizes, and/or which interactions are required for humans to not only have access to clean water, food, and shelter, but also to enjoy cultural or spiritual fulfillment, happiness, and security?

"Ecosystem services" is the term coined to represent the benefits people gain from ecosystems (MEA, 2005). As the concept of ecosystem services has evolved into the common practice of valuing them, and incorporating those values into the human economy as a conservation practice, the definition has narrowed slightly, and can be currently considered as "the components of nature, directly enjoyed, consumed or used to yield human well-being" (Boyd and Banzhaf, 2007). The difference can be thought of as the difference in describing an ecosystem good as either a pollinated apple or apple pollination, with the value of the former encompassing the ecosystem functions (including soil formation, water, and atmospheric regulation, as well as bee pollination) of the latter.

This distinction effectively outlines the various ways humans and human health depend on, or relate to, biological diversity, and therefore to biodiversity conservation. We can describe direct linkages between human health and biodiversity as including "ecosystem services" and more indirect linkages as connections between humans and 'ecosystem functions'.

Biodiversity and Human Health: Direct Linkages

Humans depend on several ecosystem services. A short list could include a continuous food supply and good nutrition, pharmaceutical products and medical models, as well as "sentinel" species that act as bellwethers for environmental change.

Food Supply and Nutrition

Humans are dependent on managed, semi-wild, and wild ecosystems for a continuous food supply (Waltner-Toews and Lang, 2000). An adequate provision and diversity of food resources is critical to maintaining the daily caloric and nutrient intake required for basic human health. Reductions in the magnitude and stability of food supplies can lead to malnourishment, a major threat to health and well-being. Childhood and maternal malnutrition alone account for 10% of the global *disease burden*, and an estimated 824 million people are malnourished on a regular basis (Corvalan et al., 2005).

Adequate provision and diversity of food is critical for basic human health. Source: K. Frey

Human health and biodiversity directly relate through food in two broad ways. First and most simply, all of the food (and many key vitamins and minerals) that we consume is derived from a plant, fungus, or animal species. People meet their daily caloric and nutritional needs through some combination of wild and domesticated sources. People use wild sources of food in both developed and underdeveloped areas, though wild edible species often are disproportionately critical to meeting the dietary requirements of the rural poor (WRI et al., 2005). Current levels of environmental change and biodiversity loss, coupled with overexploitation are threatening globally important food sources, such as marine (Orensanz et al., 1998; Baum et al., 2003) and freshwater fisheries (Abramovitz, 1996), and wild mammals (Jerozolimski and Peres, 2003; Marshall et al., 2006). As a result many human communities are now hunting, fishing, and collecting less-preferred food supplies (de Merode et al., 2004). In other cases, the resultant scarcity of a food item (e.g., many types of seafood) has rendered them more valuable and thus subject to greater harvest pressure.

Second, wild species are critical to the human food supply as

Critical Thinking Box

The Millennium Ecosystem Assessment, and countless authors and scientists, argue for the preservation and conservation of natural habitats on the premise that degradation of the ecosystem services they provide will undermine human welfare. Yet such degradation has resulted in enormous gains in human health and well-being (Ghazoul, 2007). For example, consider the benefits of fertilizers and industrial agriculture to the human food supply (even as certain segments of the human society begin to face health concerns from *over* rather than *under* consumption) along with their tremendous environmental impacts (Pollan, 2006). *What are some of the repercussions of human agricultural expansion? Does this represent a gain in human health at the cost of "overall" health? What are some of the spatial and temporal trade-offs involved in these gains and losses?*

a genetic library for the future selection of plants and animals more suitable to ever-changing agricultural ecosystems. Also, increased genetic diversity within agricultural systems often confers a degree of resistance to plant pests and pathogens (Lavelle et al., 2004), which can otherwise affect large areas in which only one susceptible species or genotype is planted (Zhu et al., 2000).

Pharmaceuticals and Other Molecules

Vascular plants and their extracts, but also mosses, fungi, and animal parts, have been the main source of traditional medicine since prehistoric times (Table 1).

The importance of nature as a source of medicinal compounds and other molecules with therapeutic properties has not diminished over time; many are currently used to treat pain, fever, high blood pressure, anxiety, and improve heart function. New drugs are continually being discovered, and it may be that nature holds the key to cures for currently untreatable conditions and *emerging infectious diseases*. Compounds found in nature may provide new protection against diseases such as HIV/AIDS, as well as resistant strains of bacteria and cancer. Bacterial infection remains a particularly serious threat to human health, with years of use (and misuse) of common antibiotics having led to the evolution of resistant bacterial strains. Indeed, several strains are resistant to multiple antibiotics, and at least one strain of bacteria is resistant even to the newest antibiotics on the market. It is important to keep our therapeutic arsenal well-stocked and new antibiotic compounds derived from natural sources, which can potentially fight infections with these resistant strains, have recently been identified (Wang et al., 2006). When a species goes extinct, biodiversity is lost and the potentially useful compounds are lost with it.

Table 1. Illustrative list of commonly used drugs derived from natural sources

Drug	Purpose	Source
Amoxicillin	Antibiotic	Fungus
Captopril	Antihypertensive	Animal
Digitoxin	Cardiotonic	Plant
Morphine	Analgesic	Plant
Penicillin	Antibiotic	Fungus
Quinine	Antimalarial, antipyretic	Plant
Salicin	Analgesic	Plant
Vinblastine	Antitumor	Plant
Exenatide	Antidiabetic	Animal
Ecteinascidin	Antitumor	Animal
Ziconotide	Analgesic	Animal

In addition to new medicines, molecules derived from living beings have other positive impacts on human health. For example, a group of molecules derived from a bacterium found in thermal waters in Yellowstone National Park (USA) became the basis for the polymerase chain reaction, a procedure that allows researchers to make multiple copies of DNA molecules (Chien et al., 1976) and to diagnose infectious diseases and genetic disorders, among other important biomedical applications. Molecules derived from animals (e.g., spiders) and plants (e.g., neem tree) also provide newer and safer insecticides and pesticides, (see section 2.1.).

The therapeutic compounds derived from nature have enormous social value. In fact, it has been estimated that at least 80% of the world's population relies on compounds derived chiefly from plants as their main source of health care (Fabricant and Farnsworth, 2001; Kumar, 2004). The importance of medicines derived from living things is not limited to the developing world. More than half of the most commonly prescribed drugs in the United States come from, are derived from, or are patterned after one or more compounds originally found in a live organism (Grifo et al., 1997).

Medical Models

Species belonging to many different taxa are, and will continue to be, invaluable in biomedical research. Biological diversity contributes to human health by playing a critical role in advancing our understanding of human behavior, anatomy, physiology, and disease. The use of these models in research has provided insights that may have been impossible to obtain otherwise. In some cases, a species may be irreplaceable as a disease model; for example, nine-banded armadillos are unique for the study of human leprosy.

Early research and training tools for medical professionals were based on experimentation with animal models. Even today, the use of animals for teaching basic anatomy and surgical procedures to medical students is a common practice in many countries. Animal species are also used to assess the efficacy of new vaccines, anesthetics, and other chemicals with potential therapeutic effects. Such new drugs must be proven to be both effective and safe in vertebrate models that include mice, rats, dogs, and non-human primates before human trials can be conducted.

Medical models are also used to understand the metabolism of particular systems or of the whole body under specific environmental circumstances. Species as diverse as plants and yeasts are routinely used as models to further our understanding of molecular processes, such as gene expression and mutation, providing insights into human health issues like tumor formation and aging. Finally, animals and their cell and tissue cultures serve to advance our knowledge of the effects of specific diseases on human cells, tissues, and organs.

Mice and rats are routinely used as laboratory animals. However, several less common species are used in biomedical research, including horseshoe crabs (for anatomical research), cone snails (for the study of the physiology of cell receptors, neurotransmitters, and ion channels), and sea squirts (for the study of the formation of kidney stones).

Sentinels

The study of the distribution, abundance, and/or health of certain species can provide valuable information about environmental stressors such as chemical pollution or the presence of pathogens, which can potentially threaten human health (Table 2). These species are the proverbial "canary in the coal mine" and are referred to as sentinels because they can warn of potential human health risks. Species belonging to many different taxa, from mosses to dolphins, can serve as sentinels; in general, non-domesticated species are more frequently used as sentinels of both chemical/physical hazards and infectious agents (Rabinowitz et al., 2005).

The distribution and abundance of certain key species can indicate that a specific environmental stressor is currently acting or has been present in the recent past in a particular ecosystem; for example, the disappearance of certain species of mollusks indicates when water pollution has reached a harmful threshold (Funes et al., 2006). Clues about the quality of the environment can also be gleaned by monitoring health parameters in selected species. Many species accumulate pollutants in their tissues, for example, and can therefore provide an accurate picture of the long-term flows of such pollutants through the atmosphere, water, or the food web (Schintu et al., 2005; Alleva et al., 2006). The exposure of sentinel species

Table 2. Illustrative list of organisms used to monitor specific environmental stressors potentially deleterious to human health

Species	Sample	Environmental Stressor
Marine/freshwater invertebrates	Distribution and abundance	Chemical pollution
Mosses	Tissue	Chemical pollution
Wild birds	Egg shells, feathers, blood, tissue	Chemical pollution
Amphibians	Distribution and abundance, deformities	Multiple/undetermined
Sea turtles	Pathological changes	Multiple/undetermined
Sea otters	Distribution and abundance, blood, tissue	Chemical pollution, ecosystem integrity
Chickens	Blood	Viral exposure
Wild birds and mammals	Blood	Viral exposure
Lichens	Distribution and abundance	Air pollution
Domestic mammals	Blood	Viral exposure, chemical pollution

to specific infectious agents is regularly used to estimate the human infection risk; for example, wild and domestic birds and mammals are commonly employed as sentinels of *arbovirus* activity (Komar, 2001; Komar et al., 2001; Peterson et al., 2004). In other cases, the detection of specific pathological changes in sentinel species may suggest that multiple sources of environmental change are acting simultaneously with deleterious health effects. The emergence of tumors in sea turtles, for example, may be indicative of more than one source of anthropogenic environmental change in the world's oceans (Aguirre and Lutz, 2004).

These examples demonstrate how species diversity may benefit human health by enabling human populations to detect and react to situations where their health would otherwise be compromised. However, further research is needed to identify the appropriate sentinels for specific health risks in specific environments. Also, even when a certain kind of environmental alteration can be detected, quantified, or monitored through the use of sentinel species, a direct link to human health may not always be present (Rabinowitz et al., 2005).

Biodiversity and Human Health: Indirect Linkages

Our dependence on the natural world extends not only to the final goods and products provided by nature (ecosystem services), but also to the ecosystem processes provided by large-scale ecosystem interactions (see the NCEP module: *Why is Biodiversity Important?*). Ecosystem functions, such as pollination, pest control, soil creation and maintenance, nitrogen fixation, and a host of aquatic processes, support all productive ecosystems. Pollination by diverse groups of wild, unmanaged species and domesticated pollinators such as the European honeybee (*Apis mellifera*) (Kremen et al., 2002) enables the production of approximately one third of the average human daily caloric intake (McGregor, 1976; Buchmann and Nabhan, 1997). Natural biological control of plant pests can help maintain crop yields without investment in artificial chemicals that have negative (and often poorly understood) impacts on both human and animal health (Shetty, 2002). Ecosystem "engineering" by corals and oysters in marine and estuarine environments create habitat for a huge diversity of organisms, many of which are key players in marine and coastal food webs that benefit humans. Nutrient cycling is critical for the persistence of both natural and wild ecosystems (John et al., 2007) and mediated by a large and diverse group of bacteria, protozoa, fungi, and invertebrates.

These processes are representative of the many natural processes that can be indirectly linked to human health. The strength of the evidence supporting these linkages, however, is variable and there is often little direct research connecting declines in these services to concomitant declines in human health. This is because ecosystem processes are time and space extensive, and result from the concerted action of many organisms. Hence, extrapolating 1:1 relationships between human health (which is itself complex and multi-factorial) and biodiversity through the lens of ecosystem services is challenging. Additionally, both the study of biodiversity and ecosystem functioning (BEF) and of the explicit relationship between human health and the environment are new areas of research, and thus are undergoing a period of rapid growth and refinement.

Hydrological Control

Clean water, free from biotic and chemical pollutants, is an essential resource for all humans. The capture and slow filtration of water through naturally vegetated watersheds reduces sediment and organic component loads, a process commonly referred to as water purification (Haines et al., 1993). Much of the developed world can afford water treatment facilities, which mimic this ecosystem service at a financial cost. However, for at least two billion people, these services are unavailable. Over one billion people currently lack access to clean water supplies and water-related infectious disease is estimated to cause more than 3.2 million deaths annually (Corvalan et al., 2005). While it is well understood that the preservation of natural vegetation in watersheds is linked to the availability of clean water supplies downstream, the mechanisms for these ecological functions are only grossly resolved.

One form of hydrologic regulation is flood control. Floods are the world's most frequent natural disaster and often the most costly in both economic and human health terms. Intact wetlands, for example, are valued for their ability to reduce the frequency and magnitude of flooding events at local watershed scales (Andreassian, 2004) by securing soil sediment, and increasing or maintaining soil porosity and infiltration capacity (Bronstert et al., 2000; Tollan, 2002).

Study of the linkages between naturally vegetated ecosystems and hydrology is an extremely active avenue of research. Several recent findings run contrary to the conventional wisdom that forested landscapes are unequivocally beneficial for flood control. While successful in limiting the magnitude and frequency of flood events, the *afforestation* of naturally non-forested ecosystems for the purposes of flood control often instigates soil salinization, resulting in diminished soil fertility and tremendous losses in agricultural productivity, as well as below- and above-ground biodiversity (Jobbagy and Jackson, 2004). Naturally forested watersheds typically exhibit in-

Intact ecosystems play a role in flood mitigation and its mpacts on human health, Source: K. Frey creased stream flow, higher evaporative water loss, lower soil moisture, and reduced groundwater recharge relative to deforested watersheds (Jobbagy and Jackson, 2004). In studies of the effects of tree cover *removal* on flood events, the observed effects are most pronounced at small scales and for frequent flood magnitudes (Tollan, 2002). At larger scales, the effects of deforestation on flooding can be negligible (Mudelsee et al., 2003).

Intact ecosystems play a role in mitigating flood events and, therefore, the primary and secondary impacts of floods on human health (e.g., physical destruction, water- and vector-borne disease outbreaks, and water and soil contamination (Ahern et al., 2005). Flood events are associated with an increased risk of vector-borne (e.g., malaria, dengue, West Nile Fever), water-borne (e.g., cholera, leptospirosis), and non-epidemic, water-borne infection (e.g., wound infections, dermatitis). However, the overall risk for disease outbreak is often low unless there is significant population displacement and/or water sources are compromised; even when this happens, the risk of outbreaks can be minimized with rapid disaster response (Gayer and Connolly, 2005). As such, the strongest linkages between human health and flooding tend to be found in the developing world (Conti et al., 1984; Greenough et al., 2001). Only one of the 14 major floods that occurred globally between 1970 and 1994 (Sudan in 1980) led to a major diarrheal disease outbreak (WHO, 2006). Floods may spur an increase in vector-borne diseases through the expansion in the number and extent of vector habitats (Gayer and Connolly, 2005). The major risk factors for outbreaks associated with flooding are water-borne disease following the contamination of drinking-water facilities, and leptospirosis, a zoonotic bacterial disease that is instigated by rodent population booms following heavy rainfall-induced flooding events (Gayer and Connolly, 2005).

Waste Removal and Decomposition

Excrement from livestock, wildlife, and humans (particularly in rural areas with poor sanitation) is removed by a suite of macro-invertebrates, including dung beetles, termites, and earthworms. The global value of fecal waste removal services was estimated at USD 2.277 trillion in 1997 (Costanza,1997). Dung beetles lay their eggs within a dung mass or more commonly relocate dung below the soil surface – actions which serve to significantly reduce the amount of dung remaining on the soil surface (Lindquist, 1933). Potential linkages between human health and dung beetles include the suppression of dung-breeding fly populations (Horgan and Fuentes, 2005), and the reduction in the transmission of endoparasites and protozoa through contact with contaminated dung (Bryan, 1973; Mathison and Ditrich, 1999; Nichols et al., 2008). In contrast, several authors have suggested that rapid and efficient dung removal by dung beetles allows humans to repeatedly visit defecation areas, potentially increasing rather than decreasing the risk of parasite exposure (Miller, 1954). Additional epidemiological research will be required to identify the positive or negative effects of dung beetle activity on human parasites and pathogens.

Critical Thinking Box

Approximately 14 species of dung beetles live in and near New York City, USA. Dung beetles remove large quantities of dung from the forest surface, suppress fly populations, and may reduce the transmission of endoparasites, like nematodes, from livestock and wildlife to humans. However, they require large blocks of forest cover and dung resources from diverse mammal communities to maintain these services. *Are dung beetles important enough to human health to use their services as an impetus for conservation? Even in the face of modern medicine and expensive land? Are their services substitutable with technology, such as waste treatment facilities? Are their services even required, as wildlife in human-dominated areas declines?*

The decomposition of dead organic matter by biological entities is a critical ecological function carried out by many species belonging to different taxa, but primarily by bacteria and fungi. The term "bioremediation" refers to the technology of using biological processes to remove pollutants from the environment. The term encompasses the methods for facilitating the establishment, growth, and reproduction of the organisms involved, as well as the technologies used to improve the efficiency of the removal processes (including genetic engineering) (Kulkarni and Chaudhari, 2007; Padmavathiamma and Li, 2007; Saier, 2007; Zhuang et al., 2007). Environmental managers interested in removing harmful chemicals from water, soil, or even man-made surfaces (e.g., concrete), take advantage of natural metabolic processes to break them down into harmless metabolites or store them in living tissue. Bioremediation is used for the removal of a wide variety of pollutants, including heavy metals, industrial solvents, hydrocarbons, and pesticides, and is considered an efficient, safe, and cost-effective method for cleaning contaminated environments (Kulkarni and Chaudhari, 2007; Saratale et al., 2007; Shi et al., 2007; Urik et al., 2007; Xu et al., 2007)

Biotic Regulation

Biodiversity can act as a buffer for disease by helping to control the populations of *vectors* and *hosts* involved in disease transmission cycles. In general terms, the loss of any species or functional group with a regulatory role in an ecosystem will lead to drastic increases in the abundance of the species it normally regulates. For example, the loss of carnivorous predators can lead to an explosion in prey populations and the diseases for which they are hosts (Packer et al., 2003; Ostfeld and Holt, 2004; Stronen et al., 2007), although recent evidence suggests that the opposite effect can result under certain ecological conditions (Holt and Roy, 2007). In this section, we highlight two ways in which higher diversity leads to lowered disease prevalence.

The dilution effect

Lyme disease, a tick-borne infection, is one of the most common vector-borne diseases in the United States. Larval and nymphal ticks acquire the infection during a blood meal taken from an infected vertebrate. White-footed mice are the most *competent* host for Lyme disease in the United States; other species vary in their *competence* levels.

Empirical and theoretical evidence support the idea that vertebrate diversity can act as a buffer for Lyme disease incidence in humans, i.e., high vertebrate diversity lowers the risk of human exposure to certain infectious diseases. The dilution effect, as stated by Van Burkirk and Ostfeld (1998) and Ostfeld and Keesing (2000b), proposes that, for Lyme disease, increasing species diversity in the host community reduces the incidence of infected vectors by increasing the probability an uninfected tick will feed on hosts other than mice (Ostfeld et al., 2002). In more species-rich communities, the prevalence of infected vectors (nymphal infection prevalence, NIP) is lower (Figure 1). A recent study found that, in the United States, higher richness of small mammals and lizards is correlated with lower Lyme disease incidence in humans (Ostfeld and Keesing, 2000b).

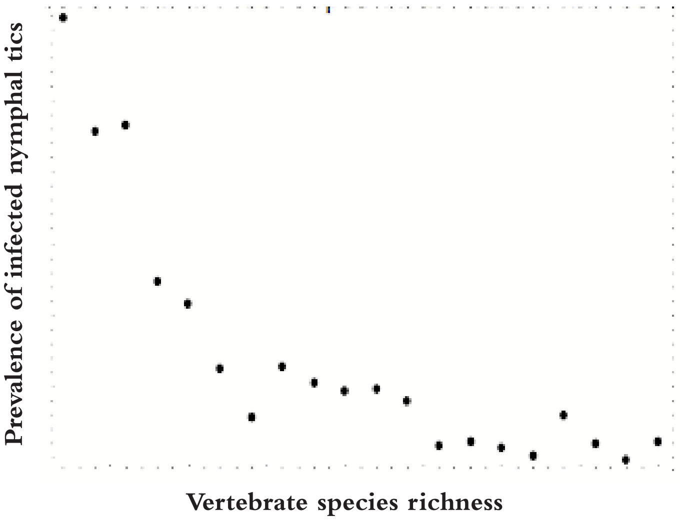


Figure 1. The dilution effect: Prevalence of infected nymphal ticks (a proxy for human Lyme disease risk) as a function of vertebrate species richness. Modified from the model by Schmidt and Ostfeld (2001).

A theoretical exploration of the dilution effect model finds that it hinges upon four basic conditions: 1) the vector must be a generalist; 2) vectors must acquire the infection orally; 3) competence must vary among the species present in the community; and 4) the most competent species must be the most abundant (Ostfeld and Keesing, 2000a). Although it has been suggested that the dilution effect describes a common mechanism by which biodiversity lowers the risk of disease, further evidence is needed to assess its generality. However, the dilution effect provides a powerful argument for biodiversity conservation in areas in which Lyme disease is endemic.

Recent studies have found evidence in favor of the dilution effect in diseases other than Lyme disease. For example, the prevalence of West Nile virus infection in humans and mosquitoes is reduced in areas with higher diversity of non-passerine bird species (Ezenwa et al., 2006), and higher species richness lowers the prevalence of a flea-borne bacterial infection in rodents (Telfer et al., 2005). In addition, mathematical models of disease transmission suggest that high species diversity lowers overall disease prevalence in other disease transmission cycles (Dobson, 2004; Rudolf and Antonovics, 2005). However, the net effects of species diversity over disease incidence could depend on the pathogen's mode of transmission; in these models, species diversity increases disease incidence if pathogen transmission depends on the density of susceptible hosts (Dobson, 2004).

The dilution effect explains a mechanism by which high species richness provides a direct benefit to human health. Addressing the drivers of biodiversity loss in areas where certain diseases are prevalent can serve the dual purposes of conserving species and improving human health.

The buffering effect

Similar to the dilution effect, the buffering effect describes the reduced prevalence of directly transmitted (i.e., not vector-borne) diseases in areas of higher diversity. The buffering effect, first demonstrated for hantavirus infection in Panama, is postulated to be the result of multi-species interactions that lead to regulation of the abundance of pathogen hosts (Suzan, 2005). In this case, experimental species removals resulted in increased hantavirus prevalence. Highly diverse communities have lower densities of hantavirus hosts and thus lower overall disease prevalence. The opposite was found to be true for low diversity ecosystems (Suzan, 2005). Note that while the *dilution effect* modulates the intensity of infection in the disease vectors, the *buffering effect* is concerned with the abundance of hosts with different competence levels.

Atmospheric Regulation

Forest biodiversity can act as a carbon sink, by taking carbon from the atmosphere (carbon dioxide, CO^2, is one of the main greenhouse gases) and converting it into plant biomass (Harper et al., 2007; Kirby and Potvin, 2007) through a process called carbon sequestration. Corals, and many organisms with calciferous body parts, participate in global carbon cycles by metabolizing both organic and inorganic carbon through photosynthesis, respiration (Ridgwell et al., 2003), and the calcification process (Gattuso et al., 1999). Because of these sequestration processes, forest and coral reef conservation are considered important parts of the global strategy to mitigate the negative effects of fossil fuel emissions (Malhi et al., 2002), potentially including the negative effects on human health.

The effects of global climate change are expected to affect the incidence of several infectious and non-infectious human diseases (Rogers and Randolph, 2000; Patz et al., 2005). Since disease vectors are highly dependent on humidity and temperature for reproduction and development, in regions where climate change is expected to lead to higher mean temperatures and increases in rainfall, the incidence of vector-borne diseases is expected to increase (Rogers and Randolph, 2000; Patz, 2001; Epstein, 2002; Patz et al., 2005;). Global warming is also predicted to increase the altitudinal and latitudinal range of vector-borne diseases by extending the total land area that meets the minimum temperatures necessary for vector development. Global climate change places an additional burden on human health through non-infectious effects, such as increased mortality due to extreme heat and cold events (Patz et al., 2005).

Critical Thinking Box

It is sometimes suggested that linking biodiversity to general human well-being considerably broadens the number of links between humans and biodiversity, indeed so broadly that attempts to draw these connections may be fruitless. *Do you think this is the case? Review the linkages described by this synthesis between human health and the natural world. Can you think of others? Do they impact the quality of human life, without fitting into the box of "human health"?*

Psychological Health

It has been argued that biodiversity has beneficial effects on psychological health because:

1*)* Natural settings provide opportunities for relaxation, exercise, and leisure for millions of people; and
2*)* Contact with nature can be associated with increased rates of recovery for patients under treatment.

Grifo and Chivian (1999) suggest the following as evidence of the relationship between biodiversity and psychological well-being:

- Patients offered views of nature showed accelerated recovery from surgery and rehabilitation;
- Inmates offered views of nature sought health care less frequently;
- Nature-related activities were selected most often by recovering cancer patients;
- College students with views of nature have a higher capacity for concentration; and
- Contact with nature may help reduce mental fatigue.

Finally, for many people, cultural survival and community cohesion is tied to specific activities that depend on natural resources, e.g., whale hunting or berry picking. Although, a direct connection to human health is difficult to measure in these cases, the protection of the specific ecosystems in which these activities take place is tied to the preservation of these traditional practices, and the concomitant sense of identity and belonging, which, in turn, form part of the larger concept of human well-being.

Well-being is the state of having the basic materials for a secure, good life – a state that encompasses aspects of a secure and adequate livelihood, good social relations, security and personal safety from natural and human-made disasters, freedom of choice and action, and, importantly, good health (Mooney et al., 2005). We use the concept of well-being to illustrate that wild nature has an impact on people's lives that cannot be confined to the narrow definitions of human "health." The loss of local biodiversity has impacts on human well-being that run the gamut from lost jobs, migration, or

A woman cradles a langur in Cuc Phoung,Vietnam,Source: K. Frey lowered tourism revenue, to the collapse of entire civilizations (Balmford and Bond, 2005).

Conflicts Between Human Health and Biodiversity Conservation

The information presented, thus far, suggests that there are links between biodiversity and human health. However, the net effects of environmental change on human disease risk may depend on specific ecological contexts (Ostfeld et al., 2002; Holt and Roy, 2007). The strength and generality of the links previously outlined are then still in need of research.

Some forms of environmental change, while positive for biodiversity, may have negative consequences for human health. For example, forest regeneration was associated with higher risk of leishmaniasis in Sudan (Gratz, 1999), while in the United States, reforestation of abandoned farmland is associated with higher risk of Lyme disease (Telford III, 2002). An unexpected result of some marine mammal conservation efforts has been the increased incidence of intestinal worm infestations in fish and humans (McCarthy and Moore, 2000; Olson et al., 2004).

In other cases, the most environmentally destructive forms of anthropogenic alteration may be associated with the least risks for human health. Clear-cut logging results in relatively low contact rate between humans and wild animals and, therefore, carries less risk of disease emergence than selective extraction (Wolfe et al., 2005). Oil-palm agriculture replaced natural forests in Sarawak and, in the process, reduced the populations of four species of malaria vectors (Gratz, 1999). Similarly, cattle-ranching and sugar cane cultivation reduced the populations of malaria vectors in Honduras (Reid, 1997).

Emerging infectious diseases in humans may pose a unique challenge to biodiversity conservation. Most are *zoonotic*, many with wild animal *reservoirs* . Humans are at risk of disease from close contact with wildlife species, and eradication of wildlife may be seen as the only option available. Although the efficacy of culling wildlife reservoirs has been questioned, and the prevention of disease transmission events through other means has been advocated, the eradication of wildlife species is still considered a valid strategy for use in such disease control efforts (Donnelly et al., 2006).

Potential conflicts between the objectives of conservation biology and those of public health are also of interest when considering the issue of pathogen conservation. Pathogens are critical players in ecological and evolutionary processes. The interactions between pathogens and their hosts have resulted in the evolution of complex biological systems. Pathogens act as powerful selection agents and are drivers of genetic diversity in their hosts, and, by changing host distribution and abundance, affect the diversity of other species in the ecosystem. Pathogen species are themselves a major component of the planet's diversity and represent unique evolutionary lineages. For these reasons, some conservation biologists have argued for the need for conservation of pathogen species (Gompper and Williams, 1998; Windsor, 1998).

Conclusion

The health of humans, and of all other species on the planet, is ultimately connected through our shared ecological realities. Health is, therefore, a unique lens through which to view and attempt to understand the effects of human activities. We have shown that there are direct and indirect linkages between biodiversity and human health. Human activities that lead to loss of biological diversity can also have deleterious consequences for human health and well-being. However, the impacts of these links on the practice of conservation biology are not always straightforward.

To understand the links between human health and biodiversity conservation further research is needed. In particular, we need to investigate their strength and generality. Additionally, our current understanding of these links strongly suggests that not all biodiversity will have a positive net effect on human health (i.e., biodiversity can also have a neutral effect, and some species and ecological processes can even have a negative net effect on humans). Accurately connecting biodiversity to the provision of goods and services will increase our capacity to properly assess their value relative to human health, and to design and implement adequate conservation strategies. Human health depends on a series of complex interactions among environmental, social, economic, and public health factors, and the relative importance of each may differ in different regions, and also change over time. A recent study found no correlation among global indicators of biodiversity loss and human health (Huynen et al., 2004), suggesting that these relationships are better studied at different scales, or that, currently, improvements in public health policies have disproportionately improved human health in the face of great environmental destruction. The multi-factorial nature of the dynamic balance we call "health" suggests that multi-disciplinary approaches involving the fields of biomedical science, public health, conservation biology, anthropology, sociology, ecology, and earth science are needed to better understand the environmental and social determinants of human health risks.

Anthropogenic environmental alteration can negatively affect human health by increasing the incidence of non-infectious diseases and diminishing ecosystem resilience; for example, deforestation and fossil fuel emissions (due to their role in global climate change) may increase the extent and magnitude of the damage caused by extreme weather events. However, biodiversity loss may not be involved as a causal agent of these increased health risks, but may instead be a concurrent consequence of human activities.

The relationship between environmental alteration, biodiversity loss, and changes in disease risks is complex and the causal links are not always clear. Anthropogenic environmental change is considered an important driver of severely negative impacts on human health. The emergence of infectious diseases of plants, animals, and humans, for example, is often linked to human activities (Table 3).

Even when we understand the direct and indirect links between biodiversity and human health, the goods and services involved may accrue at spatial and temporal scales that are intractable for regular conservation initiatives. To illustrate, the biochemical compounds found in nature are the result of processes taking place in evolutionary time, and the regulatory services mediated by biodiversity, even at the local scale, often depend on interactions among processes happening at distant locations. Those situations in which biodiversity conservation can have negative outcomes for human health should be carefully evaluated in a conservation context. However, the health consequences of biodiversity loss are an important consider-

Table 3. Illustrative list of diseases in which environmental alteration is considered to have played a role in emergence or reemergence

Disease	Alteration factors	Geographical Extent	References
Malaria	Deforestation	Latin America	Walsh et al., 1993
Nipah virus	Encroachment, agricultural intensification	South East Asia	Daszak et al., 2001
Hookworm	Deforestation leading to silting	Haiti	Lilley et al., 1997
Hemorrhagic viruses	Land use changes, encroachment	South America	Enria et al., 1998
Leishmaniasis	Deforestation	Latin America	Patz et al., 2000
Schistosomiasis	Intensive irrigation	Africa	Patz et al., 2000
Filariasis	Irrigation, standing water	Asia	Dzodzomenyo et al., 1999
Arboviral diseases	Deforestation, irrigation, agriculture	Global	Molyneux, 2003
Lyme disease	Habitat fragmentation	USA	Allan et al., 2003

ation for environmental policy and such linkages between biodiversity and health will likely remain a powerful motivator for conservation action.

Terms of Use

Reproduction of this material is authorized by the recipient institution for non-profit/non-commercial educational use and distribution to students enrolled in course work at the institution. Distribution may be made by photocopying or via the institution's intranet restricted to enrolled students. Recipient agrees not to make commercial use, such as, without limitation, in publications distributed by a commercial publisher, without the prior express written consent of AMNH.

All reproduction or distribution must provide full citation of the original work and provide a copyright notice as follows: "A. Gómez and E. Nichols. 2010. Biodiversity Conservation and Human Health. Synthesis. American Museum of Natural History, *Lessons in Conservation*. Available at http://ncep.amnh.org/linc."

"Copyright 2010, by the authors of the material, with license for use granted to the Center for Biodiversity and Conservation of the American Museum of Natural History. All rights reserved."

This material is based on work supported by the National Science Foundation under the Course, Curriculum and Laboratory Improvement program (NSF 0127506).

Any opinions, findings and conclusions, or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the American Museum of Natural History or the National Science Foundation.

Literature Cited

Abramovitz, J. N. 1996. Imperiled waters, impoverished future: the decline of freshwater ecosystems. Page 80. Worldwatch Paper, Washington, D.C.

Aguirre, A. A. and P. A. Lutz. 2004. Marine turtles as sentinels of ecosystem health: is fibropapillomatosis an indicator? EcoHealth 1:275-283.

Ahern, M., R. S. Kovats, P. Wilkinson, R. Few, and F. Matthies. 2005. Global health impacts of floods: epidemiologic evidence. Epidemiologic Reviews 27:36-46.

Allan, B. F., F. Keesing, and R. S. Ostfeld. 2003. Effect of forest fragmentation on Lyme disease risk. Conservation Biology 17:267-272.

Alleva, E., N. Francia, M. Pandolfi, A. M. De Marinis, F. Chiarotti, and D. Santucci. 2006. Organochlorine and heavy-metal contaminants in wild mammals and birds of Urbino-Pesaro province, Italy: An analytic overview for potential bioindicators. Archives of Environmental Contamination and Toxicology 51:123-134.

Andreassian, V. 2004. Waters and forests: from historical controversy to scientific debate Journal of Hydrology 291:1-27.

Balmford, A. and W. Bond. 2005. Trends in the state of nature and their implications for human well-being. Ecology Letters 8:1218-1234.

Baum, J. K., R. A. Myers, D. G. Kehler, B. Worm, S. J. Harley, and P. A. Doherty. 2003. Collapse and conservation of shark populations in the Northwest Atlantic. Science 299:389.

Boyd, J. and S. Banzhaf. 2007. What are ecosystem services? The need for standardized environmental accounting units. Ecological Economics 63:616-626.

Bronstert, A., A. Jaeger, A. Guntner, N. Hauschild, P. Doll, and M. Krol. 2000. Integrated modelling of water availability and water use in the semi-arid Northeast of Brazil. Physics and Chemistry of the Earth Part B: Hydrology Oceans and Atmosphere 25:227-232.

Bryan, R. P. 1973. The effects of dung beetles activity on the numbers of parasitic gastrointestinal helmintic larvae recovered from pasture samples. Australian Journal of Agricultural Research 24:161-168.

Buchmann, S. and G. Nabhan 1997. The Forgotten Pollinators. Island Press, Washington DC.

Chien, A., D. B. Edgar, and J. Trela. 1976. Deoxyribonucleic acid polymerase from the extreme thermophile Thermus aquaticus. Journal of Bacteriology 174:1550-1557.

Conti, J. A., D. J. Forrester, R. K. Frohlich, G. L. Hoff, and W. J. Bigler. 1984. Helminths of urban gray squirrels in Florida. Journal of Parasitology 70:143-144.

Corvalan, C., S. Hales, and A. McMichael. 2005. Ecosystems and human well-being: health synthesis: a report of the Millennium Ecosystem Assessment. World Health Organization. Available from: http://www.millenniumassessment.org/en/Products.Synthesis.aspx

Costanza, R., R. dArge, R. deGroot, S. Farber, M. Grasso, B. Hannon, K. Limburg, S. Naeem, R. V. Oneill, J. Paruelo, R. G. Raskin, P. Sutton, and M. van den Belt. 1997. The value of the world's ecosystem services and natural capital. Nature 387:253-260.

Daszak, P., A. A. Cunningham, and A. D. Hyatt. 2001. Anthropogenic environmental change and the emergence of infectious diseases in wildlife. Acta Tropica 78:103-116.

de Merode, E., K. Homewood, and G. Cowlishaw. 2004. The value of bushmeat and other wild foods to rural households living in extreme poverty in Democratic Republic of Congo. Biological Conservation 118:573-581.

Dobson, A. 2004. Population dynamics of pathogens with multiple host species. American Naturalist 164:S64-S78.

Donnelly, C. A., R. Woodroffe, D. R. Cox, F. J. Bourne, C. L. Cheeseman, R. S. Clifton-Hadley, G. Wei, G. Gettinby, P. Gilks, H. Jenkins, W. T. Johnston, A. M. Le Fevre, J. P. McInerney, and W. I. Morrison. 2006. Positive and negative effects of widespread badger culling on tuberculosis in cattle. Nature 439:843-846.

Dzodzomenyo, M., S. K. Dunyo, C. K. Ahorlu, W. Z. Coker, M. A. Appawu, E. M. Pedersen, and P. E. Simonsen. 1999. Bancroftian filariasis in an irrigation project community in southern Ghana. Tropical Medicine & International Health 4:13-18.

Enria, D. A., A. M. Briggiler, and M. R. Feuillade. 1998. An overview of the epidemiological, ecological and preventive hallmarks of Argentine haemorrhagic fever (Junin virus). Bulletin De L'Institut Pasteur 96:103-114.

Epstein, P. R. 2002. Biodiversity, climate change and emerging infectious diseases. Pages 27-40 in A. A. Aguirre, R. Ostfeld, G. M. Tabor, C. House, and M. C. Pearl, editors. Conservation medicine: ecological health in practice. Oxford University Press, New York.

Ezenwa, V. O., M. S. Godsey, R. J. King, and S. C. Guptill. 2006. Avian diversity and West Nile virus: testing associations between biodiversity and infectious disease risk. Proceedings of the Royal Society B 273:109-117.

Fabricant, D. S., and N. R. Farnsworth. 2001. The value of plants used in traditional medicine for drug discovery. Environmental Health Perspectives 109:69-75.

Funes, V., J. Alhama, J. I. Navas, J. Lopez-Barea, and J. Peinado. 2006. Ecotoxicological effects of metal pollution in two mollusc species from the Spanish South Atlantic littoral. Environmental Pollution 139:214-223.

Gattuso, J., M. Frankignoulle, and S. V. Smith. 1999. Measurement of community metabolism and significance of coral reefs in the CO2 source-sink debate. Proceedings of National Academy of Sciences USA 96:13017-13022.

Gayer, M. and M. Connolly. 2005. Communicable Disease Control After Disasters. Chapter 5 in E. Noji, editor. Public Health Consequences of Disasters. Oxford University Press, Oxford.

Ghazoul, J. 2007. Recognising the complexities of ecosystem management and the ecosystem service concept. Gaia-Ecological Perspectives for Science and Society 16:215-221.

Gompper, M. E. and E. S. Williams. 1998. Parasite conservation and the black-footed ferret recovery program. Conservation Biology 12:730-732.

Gratz, N. G. 1999. Emerging and resurging vector-borne diseases. Annual Review of Entomology 44:51-75.

Greenough, G., M. McGeehin, S. M. Bernard, J. Trtanj, J. Riad, and D. Engelberg. 2001. The potential impacts of climate variability and change on health impacts of extreme weather events in the United States. Environmental Health Perspectives 109:191–198.

Grifo, F. and E. Chivian. 1999. The implications of biodiversity loss for human health. Pages 197-208 in J. Cracraft, and F. Grifo, editors. The Living Planet in Crisis: Biodiversity Science and Policy. Columbia University Press, New York.

Grifo, F., D. Newman, A. S. Fairfield, B. Bhattacharya, and J. T. Grupenhorf. 1997. The Origin of Prescription Drugs.

Pages 131-164 in F. Grifo, and J. Rosenthal, editors. Biodiversity and Human Health. Island Press, Washington, D.C.

Haines, A., P. R. Epstein, A. J. McMichael, M. Hughjones, C. F. Hutchinson, L. S. Kalkstein, S. A. Lloyd, S. S. Morse, N. Nicholls, M. Parry, J. Patz, S. Postel, K. Sherman, and R. Slooff. 1993. Global Health Watch: Monitoring Impacts of Environmental Change. Lancet 342:1464-1469.

Harper, R. J., A. C. Beck, P. Ritson, M. J. Hill, C. D. Mitchell, D. J. Barrett, K. R. J. Smettem, and S. S. Mann. 2007. The potential of greenhouse sinks to underwrite improved land management. Ecological Engineering 29:329-341.

Holt, R. D. and M. Roy. 2007. Predation can increase the prevalence of infectious disease. American Naturalist 169:690-699.

Horgan, F. G. and R. C. Fuentes. 2005. Asymmetrical competition between Neotropical dung beetles and its consequences for assemblage structure. Ecological Entomology 30:182-193.

Huynen, M., P. Martens, and R. S. De Groot. 2004. Linkages between biodiversity loss and human health: a global indicator analysis. International Journal of Environmental Health Research 14:13-30.

Jerozolimski, A., and C. Peres. 2003. Bringing home the biggest bacon: a cross-site analysis of the structure of hunter-kill profiles in Neotropical forests. Biological Conservation 111.

Jobbagy, E. G. and R. B. Jackson. 2004. Groundwater use and salinization with grassland afforestation. Global Change Biology 10:1299-1312.

John, R., J. W. Dalling, K. E. Harms, J. B. Yavitt, R. F. Stallard, M. Mirabello, S. P. Hubbell, R. Valencia, H. Navarrete, M. Vallejo, and R. B. Foster. 2007. Soil nutrients influence spatial distributions of tropical tree species. Proceedings of the National Academy of Sciences 104:864-869

Kirby, K. R. and C. Potvin. 2007. Variation in carbon storage among tree species: Implications for the management of a small-scale carbon sink project. Forest Ecology and Management 246:208-221.

Komar, N. 2001. West Nile virus surveillance using sentinel birds. Pages 58-73 in D. White and D. Morse, editors. West Nile Virus: Detection, Surveillance, and Control. New York Academy of Sciences, USA.

Komar, N., N. A. Panella, and E. Boyce. 2001. Exposure of domestic mammals to West Nile Virus during an outbreak of human encephalitis, New York City, 1999. Emerging Infectious Diseases 7:736-738.

Kremen, C., N. Williams, and R. Thorp. 2002. Crop pollination from native bees at risk from agricultural intensification. Proceedings of National Academy of Sciences USA 99:16812-16816.

Kulkarni, M. and A. Chaudhari. 2007. Microbial remediation of nitro-aromatic compounds: An overview. Journal of Environmental Management 85:496-512.

Kumar, P. 2004. Valuation of medicinal plants for pharmaceutical uses. Current Science 86:930-937.

Lavelle, P., M. Blouin, J. Boyer, P. Cadet, D. Laffray, A. T. Pham-Thi, G. Reversat, W. Settle, and Y. Zuily. 2004. Plant parasite control and soil fauna diversity. Comptes Rendus Biologies 327:629-638.

Lilley, B., P. Lammie, J. Dickerson, and M. Eberhard. 1997. An increase in hookworm infection temporally associated with ecologic change. Emerging Infectious Diseases 3:391-393.

Lindquist, A. 1933. Amounts of dung buried and soil excavated by certain Coprini (Scarabaeidae). Journal of the Kansas Entomological Society 6:109-125.

Malhi, Y., P. Meir, and S. Brown. 2002. Forests, carbon and global climate. Philosophical Transactions of the Royal Society of London Series A: Mathematical Physical and Engineering Sciences 360:1567-1591.

Marshall, A. J., Nardiyono, L. M. Engstrom, B. Pamungkas, J. Palapa, E. Meijaard, and S. A. Stanley. 2006. The blowgun is mightier than the chainsaw in determining population density of Bornean orangutans (Pongo pygmaeus morio) in the forests of East Kalimantan. Biological Conservation 129:566-578.

Mathison, B. and O. Ditrich. 1999. The fate of Cryptosporidium parvum oocysts ingested by dung beetles and their possible role in the dissemination of cryptosporidiosis. Journal of Parasitology 85:687-681.

McCarthy, J. and T. A. Moore. 2000. Emerging helminth zoonoses. International Journal for Parasitology 30:1351-1360.

McGregor, S. E. 1976. Insect Pollination Of Cultivated Crop Plants in C. H. B. R. Center, editor. USDA.

MEA. 2005. Millennium Ecosystem Assessment Synthesis. Pages 1-219 in W. Reid, H. A. Mooney, A. Cropper, D. Capistrano, S. R. Carpenter, K. Chopra, P. Dasgupta, T. Dietz, A. Duraiappah Kumar, R. Hassan, R. Kasperson, Rik Leemans, R. M. May, T. A. J. McMichael, P. Pingali, C. Samper, R. Scholes, R. T. Watson, A. H. Zakri, Z. Shidong, N. J. Ash, E. Bennett, P. Kumar, M. J. Lee, C. Raudsepp-Hearne, Henk Simons, J. Thonell, and M. B. Zurek, editors. World Resources Institute, Washington D.C., USA.

Miller, A. 1954. Dung Beetles (Coleoptera, Scarabaeidae) and other insects in relation to human feces in a hookworm area of southern Georgia. American Journal of Tropical Medicine and Hygiene 3:372-389.

Molyneux, D. H. 2003. Common themes in changing vector-borne disease scenarios. Transactions of the Royal Society of Tropical Medicine and Hygiene 97:129-132.

Mooney, H. A., A. Cropper, D. Capistrano, S. R. Carpenter, K. Chopra, P. Dasgupta, R. Leemans, R. M. May, P. Pingali, R. Hassan, C. Samper, R. Scholes, R. T. Watson, A. H. Zakri, Z. Shidong, T. Dietz, and A. Duraiappah Kumar. 2005. Ecosystems and Human Well-being Synthesis. Pages 1-155 in W. Reid, editor. World Resources Institute, Washington, D.C., USA.

Mudelsee, M., M. Borngen, G. Tetzlaff, and U. Grunewald. 2003. No upward trends in the occurrence of extreme floods in central Europe. Nature 425:166-169.

Nichols, E., S. Spector, J. Louzada, T. Larsen, S. Amezquita, M. Favila, and the Scarabaine Research Network. 2008. Ecological functions and ecosystem services provided by Scarabaeine dung beetles. Biological Conservation.

Olson, R. E., J. R. Pierce, K. C. Jacobson, and E. M. Burreson. 2004. Temporal changes in the prevalence of parasites in two Oregon estuary-dwelling fishes. Journal of Parasitology 90:564-571.

Orensanz, J. M. L., J. Armstrong, D. Armstrong, and R. Hilborn. 1998. Crustacean resources are vulnerable to serial depletion - the multifaceted decline of crab and shrimp fisheries in the Greater Gulf of Alaska. Reviews in Fish Biology and Fisheries 8:117-176.

Osofsky, S. A., W. B. Karesh, and S. L. Deem. 2000. Conservation medicine: a veterinary perspective. Conservation Biology 14:336-337.

Ostfeld, R., and F. Keesing. 2000a. The function of biodiversity in the ecology of vector-borne zoonotic diseases. Canadian Journal of Zoology: Revue Canadienne De Zoologie 78:2061-2078.

Ostfeld, R. S. and F. Keesing. 2000b. Biodiversity and disease risk: The case of Lyme disease. Conservation Biology 14:722-728.

Ostfeld, R. S. and R. D. Holt. 2004. Are predators good for your health? Evaluating evidence for top-down regulation of zoonotic disease reservoirs. Frontiers in Ecology and the Environment 2:13-20.

Ostfeld, R., F. Keesing, E. M. Schauber, and K. A. Schmidt. 2002. Ecological context of Lyme disease: biodiversity, habitat fragmentation, and risk of infection. Pages 207-220 in A. A. Aguirre, R. Ostfeld, G. M. Tabor, C. House, and M. C. Pearl, editors. Conservation medicine: ecological health in practice. Oxford University Press, New York.

Packer, C., R. D. Holt, P. J. Hudson, K. D. Lafferty, and A. P. Dobson. 2003. Keeping the herds healthy and alert: implications of predator control for infectious disease. Ecology Letters 6:797-802.

Padmavathiamma, P. K. and L. Y. Li. 2007. Phytoremediation technology: Hyper-accumulation metals in plants. Water Air and Soil Pollution 184:105-126.

Patz, J. A. 2001. Public health risk assessment linked to climatic and ecological change. Human and Ecological Risk Assessment 7:1317-1327.

Patz, J. A., D. Campbell-Lendrum, T. Holloway, and J. A. Foley. 2005. Impact of regional climate change on human health. Nature 438:310.

Patz, J. A., T. K. Graczyk, N. Geller, and A. Y. Vittor. 2000. Effects of environmental change on emerging parasitic diseases. International Journal for Parasitology 30:1395-1405.

Peterson, A. T., N. Komar, O. Komar, A. Navarro-Siguenza, M. B. Robbins, and E. Martinez-Meyer. 2004. West Nile virus in the New World: potential impacts on bird species. Bird Conservation International 14:215-232.

Pollan, M. 2006. The Omnivore's Dilemma: A Natural History of Four Meals Penguin Press, NY, USA.

Rabinowitz, P. M., Z. Gordon, R. Holmez, B. Taylor, M. Wilcox, D. Chudnov, P. Nadkarni, and F. J. Dein. 2005. Animals as sentinels of human environmental health hazards: An evidence-based analysis. EcoHealth 2:26-37.

Reid, W. V. 1997. Opportunities for Collaboration Between the Biomedical and Conservation Communities. Pages 334-353 in F. Grifo, and J. Rosenthal, editors. Biodiversity and Human Health. Island Press, Washington, DC.

Ridgwell, A. J., A. J. Watson, M. A. Maslin, and J. O. Kaplan. 2003. Implications of coral reef buildup for the controls on atmospheric CO^2 since the Last Glacial Maximum. Paleoceanography 18(4): 1083.

Rogers, D. J. and S. E. Randolph. 2000. The global spread of malaria in a future, warmer world. Science 289:2283-2284.

Rudolf, V. H. W. and J. Antonovics. 2005. Species coexistence and pathogens with frequency-dependent transmission. American Naturalist 166:112-118.

Saier, M. H. 2007. Beneficial bacteria and bioremediation. Water Air and Soil Pollution 184:1-3.

Saratale, G., S. Kalme, S. Bhosale, and S. P. Govindwar. 2007. Biodegradation of kerosene by Aspergillus ochraceus NCIM-1146. Journal of Basic Microbiology 47:400-405.

Schintu, M., A. Cogoni, L. Durante, C. Cantaluppi, and A. Contu. 2005. Moss (Bryum radiculosum) as a bioindicator of trace metal deposition around an industrialised area in Sardinia (Italy). Chemosphere 60:610-618.

Schmidt K. A. Ostfeld R. S. 2001 Biodiversity and the dilution effect in disease ecology. Ecology 82(3): 609-619.

Shetty, P. K. 2002. Ecological implications of pesticide use in agro-ecosystems in India. Page 7. National Pesticide Stewardship Alliance, Bangalore.

Shi, D. Q., Z. Jian, Z. L. Gui, D. Jian, T. L. Wang, V. Murygina, and S. Kalyuzhnyi. 2007. Bioremediation of oil sludge in Shengli oilfield. Water Air and Soil Pollution 185:177-184.

Stronen, A. V., R. K. Brook, P. C. Paquet, and S. McLachlan. 2007. Farmer attitudes toward wolves: Implications for the role of predators in managing disease. Biological Conservation 135:1-10.

Suzan, G. 2005. The responses of hantavirus host communities to habitat fragmentation and biodiversity loss in Panama. Page 98. Department of Biology. University of New Mexico, Albuquerque, U.S.A.

Telfer, S., K. J. Bown, R. Sekules, I. Begon, T. Hayden, and R. Birtles. 2005. Disruption of a host-parasite system following the introduction of an exotic host species. Parasitology 130:661-668.

Telford III, S. R. 2002. Deer tick-transmitted zoonoses in the Eastern United States in A. A. Aguirre, R. Ostfeld, G. M. Tabor, C. House, and M. C. Pearl, editors. Conservation medicine: ecological health in practice. Oxford University Press, New York.

Tollan, A. 2002. Land-use change and floods: what do we need most, research or management? Water Science and Technology 45:183-190.

Urik, M., S. Cernansky, J. Sevc, A. Simonovicova, and P. Littera. 2007. Biovolatilization of arsenic by different fungal strains. Water Air and Soil Pollution 186:337-342.

Van Buskirk, J. and R. S. Ostfeld. 1998. Habitat heterogeneity, dispersal, and local risk of exposure to Lyme disease. Ecological Applications 8:365-378.

Walsh, J. F., D. H. Molyneux, and M. H. Birley. 1993. Deforestation: effects on vector-borne disease. Parasitology 106:S55-S75.

Waltner-Toews, D. and T. Lang. 2000. A new conceptual base for food and agricultural policy: the emerging model of links between agriculture, food, health, environment and society. Global Change & Human Health 1:116-130.

Wang, J., S. M. Soisson, K. Young, W. Shoop, S. Kodali, A. Galgoci, R. Painter, G. Parthasarathy, Y. S. Tang, R. Cummings, S. Ha, K. Dorso, M. Motyl, H. Jayasuriya, J. Ondeyka, K. Herath, C. Zhang, L. Hernandez, J. Allocco, Ã. n. Basilio, J. R. Tormo, O. Genilloud, F. Vicente, F. Pelaez, L. Colwell, S. H. Lee, B. Michael, T. Felcetto, C. Gill, L. L. Silver, J. D. Hermes, K. Bartizal, J. Barrett, D. Schmatz, J. W. Becker, D. Cully, and S. B. Singh. 2006. Platensimycin is a selective FabF inhibitor with potent antibiotic properties. Nature 441:358-361.

WHO, World Health Organization. 1946. Preamble to the Constitution of the World Health Organization as adopted by the International Health Conference, New York.

WHO, World Health Organization. 2006. Ecosystems and health.

Windsor, D. A. 1998. Equal rights for parasites. Bioscience 48:244-244.

Wolfe, N. D., P. Daszak, A. M. Kilpatrick, and D. S. Burke. 2005. Bushmeat hunting deforestation, and prediction of zoonoses emergence. Emerging Infectious Diseases 11:1822-1827.

WRI, U. N. D. Programme, U. N. E. Program, and W. Bank. 2005. The Wealth of the Poor: Managing Ecosystems to Fight Poverty. WRI, Washington DC.

Xu, G. M., Y. Y. Li, W. Zheng, X. Peng, W. Li, and Y. C. Yan. 2007. Mineralization of chlorpyrifos by co-culture of Serratia and Trichosporon spp. Biotechnology Letters 29:1469-1473.

Zhu, Y. Y., H. R. Chen, J. H. Fan, Y. Y. Wang, Y. Li, J. B. Chen, J. X. Fan, S. S. Yang, L. P. Hu, H. Leung, T. W. Mew, P. S. Teng, Z. H. Wang, and C. C. Mundt. 2000. Genetic diversity and disease control in rice. Nature 406:718-722.

Zhuang, P., Q. W. Yang, H. B. Wang, and W. S. Shu. 2007. Phytoextraction of heavy metals by eight plant species in the field. Water Air and Soil Pollution 184:235-242.

Glossary

Afforestation: the process of establishing a forest on land that is not forested, or has not been a forest for some time.

Anthropogenic: derived from or caused by human activities.

Arbovirus: arthropod-borne virus. Refers to viruses with an arthropod vector, such as West Nile and dengue fever viruses.

Competent: see Competence.

Competence: the capacity of a host to pass the infection on to an uninfected vector.

Disease burden: a measure of the amount of disease caused by a specific factor or group of factors.

Emerging infectious disease: a disease that has recently increased in incidence, expanded its geographical or host range, is newly recognized, or has recently evolved.

Host: the individual or species that is infected by a pathogen.

Reservoir: an individual, population, or species that harbors an infection, but is generally not affected by it and can thus act a source for infection for others.

Vector: an organism, most frequently an arthropod, capable of transmitting a disease through bites. For example, mosquitoes are the vectors for the malaria parasite and ticks transmit the bacterium causing Lyme disease.

Zoonotic: a disease that is shared among animals and humans.

Ecological Consequences of Extinction

Madhu Rao[*] and Trond Larsen[†]

[*] Wildlife Conservation Society, New York, NY, U.S.A., email mrao@wcs.org
[†] Princeton University, Princeton, NJ, U.S.A., email tlarsen@princeton.edu

F. Laso

Table of Contents

Ecological Consequences of Extinction

Madhu Rao and Trond Larsen

Introduction

Extinction is a natural process, but it is occurring at an unnaturally rapid rate as a consequence of human activities. Humans have caused the extinction of between 5-20% of the species in many groups of organisms, and current rates of extinction are estimated to be 100-1,000 times greater than pre-human rates (Lawton and May, 1995; Pimm et al., 1995). Overall, accelerated extinctions of species and loss of *biodiversity* are no longer disputed issues in the scientific community. Although much effort has gone into quantifying the rates of biodiversity loss for particular animal and plant groups, the impact of such losses on ecosystems is less clear, especially when many different kinds of plants and animals are simultaneously lost (Raffaelli, 2004). In a review of the ecosystem consequences of bird declines, Şekercioğlu et al. (2004), report that 21% of all bird species are currently threatened or near threatened by extinction, and 6.5% are *functionally or ecologically extinct*. Their projections indicate that by 2100, 6-14% of all bird species will actually be extinct, and 7-25% (28-56% on oceanic islands) will be functionally extinct. These extinctions are likely to disrupt important ecosystem processes such as decomposition, pollination, and seed dispersal.

Evidence from observational and experimental studies suggests that species extinctions are likely to have far-reaching consequences including further cascading extinctions, disruptions of *ecosystem services*, and the spread of disease (Terborgh et al. 2001; Şekercioğlu et al., 2004; Larsen et al. 2005). Ironically, the accelerating effects of human activities on biodiversity can have direct consequences for ecosystem goods and services that support human activities and life (Daily, 1997; Chapin et al., 2000). These services include the maintenance of soil fertility, climate regulation, natural pest control, and the provision of goods such as food, timber and fresh water.

Biodiversity and Ecosystem Function

Some of the first ideas on how biodiversity affects the way ecosystems function are attributable to Darwin and Wallace, who stated that a diverse mixture of plants should be more productive than a monoculture (Darwin and Wallace, 1858; Darwin, 1859). They also suggested the underlying biological mechanism: because coexisting species differ ecologically, loss of a species could result in vacant *niche-space* and consequent disruption of ecosystem processes maintained by that niche. Darwin and Wallace's early hypothesis predicts that intact, diverse communities are generally more stable and function better than less diverse communities. This hypothesis has provided the basis for much of the later ecological research that will be addressed in this synthesis.

Influence of Community Structure on Ecosystem Function

Several components of biodiversity are relevant to *ecosystem function*, including:

1) the number of species present (*species richness*);
2) their relative abundances (*species evenness*);
3) the particular species present (*species composition*);
4) the interactions among species (*non-additive effects*); and
5) the temporal and spatial variation in these properties of community structure (Symstad et al., 2003).

In addition to their effect on ecosystem function, changes in these components of biodiversity influence the *resilience* and *resistance* of ecosystems to environmental change (Chapin et al., 2000).

Species richness is important for ecosystem functioning for several reasons, including higher complementarity of species' differing ecological roles. Several studies have used experimental species assemblages in an attempt to isolate the inherent role of species richness for ecosystem functioning, such as rates of *primary productivity* , nutrient retention, and decomposition of organic matter (Tilman et al., 1996). Many of these studies seek to identify whether species richness is important for function independently of other biodiversity components such as species composition.

Changes in the *relative abundance* of species (or species evenness) are more frequent than species loss and can have significant consequences for ecosystem function long before a species is actually threatened by extinction. The concept of ecological extinction refers to the, "Reduction of a species to such low abundance that, although it is still present in the community, it no longer interacts significantly with other species," (Estes et al., 1989). The implications of reduced species abundance for trophic interactions are discussed in the final section of this synthesis.

Because species differ ecologically, the identity of species present in a community (species composition) can strongly influence ecosystem functions. For example, particular species can have strong effects on ecosystem processes by directly mediating energy and material fluxes or by altering abiotic conditions that regulate the rates of these processes (Hooper and Vitousek, 1997). Keystone species are examples of species with large effects on ecosystem process and function.

Species interactions, including mutualisms, trophic interactions (predation, parasitism, and herbivory) and competition may affect ecosystem processes directly by modifying resource-use efficiency and pathways of energy and material flow (de Ruiter et al., 1995) or indirectly by modifying the relative abundances of species (Power et al., 1996). Thus the disruption of species interactions through the loss of species can lead to many types of ecosystem effects.

The effects from some species interactions can impact entire ecosystems. Pollination, a mutualism between plants and pollinators, is considered an ecosystem service Source: S. Spector

Diversity-Function Relationships and Hypotheses

Although several aspects of biological communities influence ecosystem function, the majority of research has focused on the role of species richness, sometimes also referred to as *diversity*. There are many types of ecosystem functions and processes that can be influenced by species richness, including productivity, decomposition rates, *nutrient cycling*, and resistance and resilience to perturbations, disease, and *species invasions*.

The diversity-stability hypothesis (MacArthur, 1955) introduced the idea that increasing the number of trophically interacting species in an ecological community should increase the collective ability of member populations to maintain their abundances following disturbance. The hypothesis predicts that ecological communities will improve in energetic efficiency (or productivity), and in the ability to recover from disturbance, as the number of species in the system increases. Conversely, removing any species from a food web will enhance the susceptibility of the system to disruption via perturbation. MacArthur based this hypothesis on the premise that energy flow in complex food webs (meaning those featuring the greatest number of interspecific links or 'connectance') will be least disrupted by disturbance because alternative pathways for energy flow are available. Another hypothesis proposes that ecosystem resistance — the ability of a system to absorb changes in abundances of some species without drastically changing ecosystem performance (e.g., biomass production) — can decline as species are lost, even if system performance appears outwardly unaffected, with potentially sudden and drastic consequences as some threshold is passed (Ehrlich and Ehrlich, 1981). This hypothesis likens species in an ecosystem to rivets holding an airplane together — the removal of rivets beyond some threshold number may cause the airplane, or the ecosystem, to suddenly and catastrophically collapse.

Diversity-function relationships are usually plotted with species richness as the independent variable and a measure of ecosystem function as the dependent variable. This relationship can take many forms. Sometimes no relationship or an idiosyncratic (or unpredictable) relationship is observed. This relationship would be expected, for example, in communities featuring *higher-order interactions* (Lawton, 1994). However, the majority of studies have found a positive but saturating relationship between richness and function, such that ecosystem function approaches its maximum level at some intermediate level of species richness (Schwartz et al., 2000). One explanation for this relationship is based on the ecosystem redundancy hypothesis (Walker, 1992), which proposes that there is some degree of functional redundancy in the roles species play in the ecosystem. At the heart of this concept is the idea that species are segregated into functional groups; those within the same group are predicted to be more expendable in terms of ecosystem function if the remaining species can maintain the same functional role. However, from a functional standpoint, many species may be less expendable than they appear. The insurance hypothesis suggests that even in systems, which appear to show saturation of function at low levels of richness, maintaining high levels of species richness provides long-term insurance to buffer against the disruption of function in response to future environmental change and fluctuation (Yachi and Loreau, 1999).

Other studies have found a positive, linear relationship between richness and function, suggesting that all species, even rare ones, are required to maintain healthy levels of ecosystem function (Johnson et al., 1996). More recently, studies looking specifically at the effects of extinctions have found a positive curvilinear relationship where function increases slowly at low levels of richness and more rapidly at the highest levels of species richness and does not saturate at all (Zavaleta and Hulvey, 2004; Larsen et al., 2005). In these cases, initial extinctions lead to large functional loss, as has been observed in several systems, particularly those with large animal consumers (Duffy, 2003). This type of relationship provides a greater challenge because it suggests that to maintain ecosystem function, it is not only important to conserve the full set of species, but also to protect the most sensitive species that are lost first with disturbance.

Assembly Experiments and Diversity-Function Mechanisms

Studies examining the ecological consequences of extinction can be broadly classified as comparative/observational or experimental (Schwartz et al., 2000). Observational research simply measures changes in biodiversity and associated function, and may or may not make comparisons with other observations of other systems. This helps to describe the complex relationships between extinctions or abundance changes and ecosystem function, but cannot usually address mechanisms. Most experimental studies have sought to isolate the relationship between species richness and ecosystem function by constructing randomly assembled communities (known as *assembly experiments*). Consequently, many of these experimental studies are not designed to address the effects of extinctions, but instead often attempt to identify the mechanisms driving the relationship between species richness and function. However, another recently proposed method uses removal experiments to examine the effects of extinction by actively removing species from naturally assembled communities (Diaz et al., 2003).

Although assembly experiments do not usually examine extinctions, they are briefly addressed here because they have increased our understanding of the mechanisms by which species diversity can be important for ecosystem processes. These experiments have demonstrated that at least three distinct mechanisms contribute to the positive effect of species diversity on ecosystem functions (Chapin et al., 1997; Tilman et al., 1997a,b; Loreau, 1998a,b).

The two primary proposed mechanisms are the *sampling effect* model and the resource use complementarity or *niche differentiation* model. According to the sampling effect model, as species richness increases, so does the probability of including a dominant competitor that will contribute disproportionately to ecosystem function (Tilman et al., 1997b). According to the niche differentiation or resource complementarity model, interspecific niche differences lead to complementarity and higher overall efficiency (usually in resource use), causing an increase in ecosystem function. A good example is when several species of plants with different root lengths are able to maximize the use of nutrients available at different soil depths. The relative functional contribution of sampling effects and resource use complementarity have been widely debated (Wardle et al., 1997; Hector, 1998; Loreau, 1998b). Pacala and Tilman (2001) discuss how plant communities in the short-term can usually be characterized by the sampling effect, while over the longer term, interspecific competition and niche differentiation take over, preventing the dominance of fast growing species. A third model predicts that the frequency of *facilitative interactions* between species will increase as the number of species itself increases (Chapin et al., 2000). Facilitative interactions can increase ecosystem function if certain species interact directly or indirectly with other species in a way that increases that species' efficiency within its niche (Jonsson and Malmqvist, 2003). All three mechanisms are supported by the general rule that communities represented by many species contain a greater range of species traits than do species-poor communities.

Because of greater experimental tractability, a large proportion of assembly experiments have focused on grassland plants or laboratory aquatic microbial systems, with less attention given to how changing animal diversity may influence ecosystem processes (Duffy, 2002). Many experimental plant studies have shown that ecosystem functions such as annual rates of primary productivity, community respiration, disease or drought resistance, decomposition, and nutrient and water retention, increase with increasing plant species richness, but often saturate at a relatively low number of species (Naeem et al., 1994, Tilman et al., 1996; Hector et al., 1999). Other communities such as microbes, fungi, and arthropods also exhibit increased ecosystem function at higher levels of species richness (Didham et al., 1998; Chapin et al., 2000; Jonsson and Malmqvist, 2000; Wolters et al., 2000). Despite the generally observed positive trend, the form of the richness-function relationship is variable, and a few studies find no relationship between increased ecosystem diversity and function or stability (Pfisterer and Schmid, 2002; Schaffers, 2002). It is now becoming clear that the specific form of the diversity-function relationship depends on many factors that vary across a large range of ecological functions, scales, and systems, and also depends on other components of biodiversity and extinction order (Huston et al., 2000; Symstad et al., 2003).

Patterns of Non-Random Species Loss

Contrary to the assumptions of many assembly experiments, the sequence of species loss from a community under human pressure is not random but is determined by traits of organisms (Duffy, 2003). General principles of population biology as well as empirical evidence, confirm that extinction risk in both plants and animals is exacerbated by factors such as rarity, small population size, small geographical range size, slow population growth and specialized ecological habits (Didham et al., 1998; Pimm et al., 1988; Purvis et al., 2000; Duffy, 2003; Şekercioğlu et al. 2004). Şekercioğlu et al (2004) show a very strong positive correlation between bird specialization and extinction-proneness (Figure 1b in Şekercioğlu et al. 2004).

In both terrestrial and aquatic systems, large animals and higher trophic levels have been found to be highly vulnerable to extinction (Diamond, 1982; Redford, 1992; Didham et al., 1998; Pauly et al., 1998; Terborgh, 1988; Purvis et al., 2000; Alroy, 2001; Cardillo and Bromham, 2001; Jackson et al., 2001; Myers and Worm, 2003; Şekercioğlu et al., 2004). These studies have shown that two distinct processes tend to make large species especially vulnerable. First, large animals and those high in the food chain tend to be associated with the aforementioned demographic risk factors of small population size and slow population growth, making them more sensitive to disturbances such as habitat destruction (Pimm et al., 1988). Şekercioğlu et al. (2004) show that different avian functional groups have large differences in extinction-proneness, possibly leading to community disassembly. Although most research has focused on *consumers*, large invertebrate species and plant species can also be more sensitive, perhaps due to higher area and resource requirements (Duarte, 2000; Larsen et al., 2005). Second, large vertebrates (both predators and herbivores) are often selectively targeted by human hunting, and commonness does not necessarily confer protection. This is supported by evidence of mass extinctions of formerly abundant Pleistocene megafauna, which closely followed human arrival on continents and islands throughout the world, and by the decline or extinction of the bison, great auk, and passenger pigeon more recently (Diamond, 1982; Alroy, 2001).

A frequently reported threat that ultimately results in the non-random loss of species is human hunting. Intensive hunting has led to substantial reductions in the abundance of certain target species in both terrestrial and aquatic ecosystems (e.g., Redford, 1992; Myers and Worm, 2003). This has also caused changes in the structure and species composition within communities (Greenstreet and Hall, 1996). In general, as would be expected from classic foraging models, hunters in tropical forest regions primarily target large-bodied species (Peres, 1990; Bodmer, 1995). Large-bodied species provide hunters with a greater return on investment in weapons and ammunition. In the sea, relentless fishing pressure throughout the oceans has systematically depleted top predators, and then shifted to the next most valuable (usually the next largest) animals, a phenomenon known as 'fishing down the food web' (Pauly et al., 1998). Evidence for the parallel phenomenon on land is less well-documented, but there are reports of how extirpations of large, preferred species has led to increased hunting pressures on smaller, less-preferred taxa. In general, the responses of predator populations in aquatic and terrestrial systems to human harvesting seem to follow similar patterns, with removal of the largest-sized fauna from the system, ultimately, leading to dominance by smaller-bodied fauna. Sustained and uncontrolled harvesting will lead to a gradual decline in the body size spectrum of the animal population, and an increase in the biomass of small and mid-sized species as a proportion of the overall community (Rice and Gislason, 1996; Peres, 2000).

Consequences of Non-Random Species Loss

The order in which species go extinct can have dramatic consequences for ecological function (Petchey, 2000; Ostfeld and LoGiudice, 2003; Zavaleta and Hulvey, 2004). These consequences will largely depend on two different types of species-specific traits: 1) Response traits that determine sensitivity to disturbance; and 2) Effect traits that determine the functional contribution of a species (Lavorel and Garnier, 2002; Naeem and Wright, 2003). If response and effect traits are independent, the order of species loss from communities will be random with respect to functional importance. However, if these traits are correlated, the relationship between richness and ecosystem function can be strongly modified (Larsen et al., 2005). For example, as described previously, several studies have found that large species, especially vertebrate consumers, are more extinction-prone. Many large species also have disproportionately strong impacts on ecosystem structure and functioning by influencing processes such as predation, selective grazing, seed predation, seed dispersal, nutrient regeneration, disturbance, and bioengineering activities (Owen and Smith, 1987; Redford, 1992; Terborgh et al., 1999; Jackson et al., 2001 and refs. therein; Duffy, 2002;). Consequently, the initial loss of sensitive species can cause a rapid and drastic decline in ecosystem function not predicted by models based on random species assembly. These results have been observed for mammals, birds, insects, and plants, where the most functionally important species (such as keystone species described below) are also the most extinction-prone (Petchey and Gaston, 2002; Zavaleta and Hulvey, 2004).

However, the opposite trend can occur if the most extinction-prone species do not strongly influence ecosystem functioning. This has been observed in some systems, particularly for plants, in which rare species, due to their low abundance, do not interact strongly in the ecosystem and are especially sensitive to disturbance (Wilsey and Polley, 2004). Although rarity is often associated with extinction-proneness, rare species can also be functionally important; other studies show that extinction-prone rare plants can be disproportionately important for maintaining invasion resistance (Lyons and Schwartz, 2001; Zavaleta and Hulvey, 2004).

In addition to species traits, the functional consequences of extinctions can be strongly influenced by how the remaining biological community responds following species loss. For example, remaining species may maintain the same level of function if they are functionally redundant with the species that were lost (Walker, 1992). Remaining species can also alter their behavior to compensate for the roles played by the lost species. *Density compensation* can occur, frequently as a result of relaxation of competition in a less diverse community, whereby remaining species increase in abundance following species loss. In some cases, density compensation may act as a buffering mechanism to maintain ecosystem function (Lawton and Brown, 1993; Tilman and Downing, 1994; Ruesink and Srivastava, 2001).

Keystone Species

The keystone species concept has been the focus of scientific interest since its introduction by Robert T. Paine (Mills et al., 1993; Paine, 1966; Power et al., 1996; Navarrette and Menge, 1996; Kotliar, 2000). A keystone species is defined as one whose impact on its community or ecosystem is not only large, but disproportionately large relative to its abundance (Power et al., 1996). By definition, keystone species differ from dominant species in that their effects are significantly greater than would be predicted from their abundance. Given their importance in the community, loss of keystone species is expected to have major consequences for ecosystem structure and function. Identifying keystone species becomes essential to understanding how their loss will affect ecosystems. Ambiguity in the use of the term keystone, and the lack of an operational definition, led to initial criticism of its continued application in research and policy contexts (Mills et al., 1993) and was later followed by clarification of the concept (Power et al., 1996).

Several case studies of keystone species reviewed by Bond (1993), Mills et al., (1993) and Menge et al., (1994) make the following generalizations. First, keystone species occur in all of the world's major ecosystems. Second, keystone species are often, but not always of high trophic status (e.g. predators). For example, certain plant species may be keystone resources for pollinators or dispersers if they flower or fruit in times of scarcity (e.g. Terborgh, 1986). Third, keystone species influence communities through mechanisms that include consumption, competition, mutualism, dispersal, pollination, disease, and by modifying habitats and abiotic factors (as "keystone modifiers") (Bond, 1993; Mills et al., 1993). Keystone modifiers include ecosystem engineers such as beavers, which inundate forests and meadows (Naiman et al., 1988; Lawton and Jones, 1993; Pollock et al., 1995); and gophers and leaf cutter ants, whose tunnels pipe water through hillslopes (Elmes, 1991).

Knowledge of keystone species is clearly important, as conserving them is necessary for maintaining intact communities and ecosystems. The following secitons outline some of the challenges in identifying keystone species (Power et al., 1996).

Keystone Species Versus Keystone Guilds

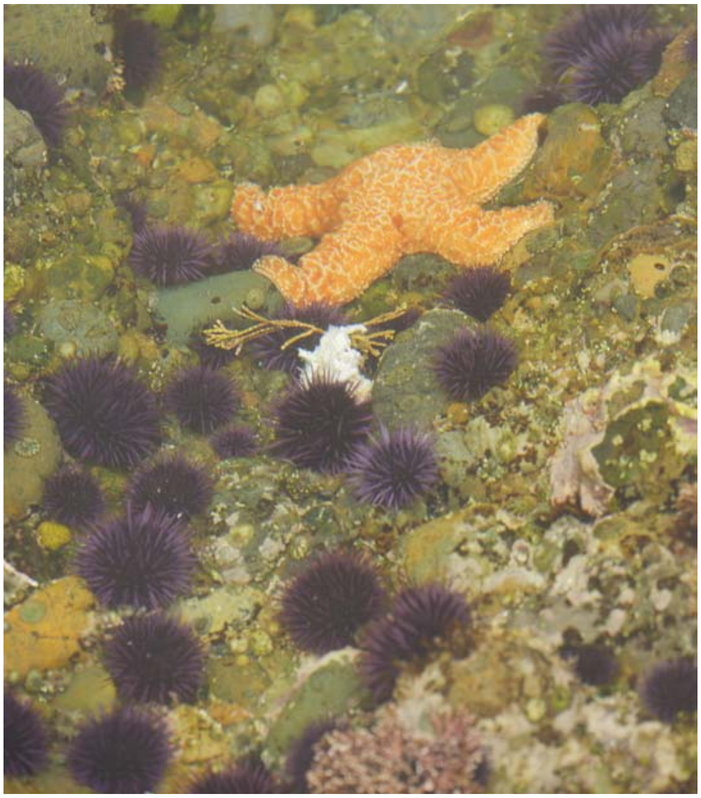
Although the clearest application of the keystone concept is to single species, identifying "*keystone guilds*" may be relevant to scientific understanding and management (Brown and Heske, 1990; Power, 1990). This term refers to groups of species that are known to have impacts that are disproportionately large relative to their collective biomass.

Time Scale

It is often a challenge to assess the impacts of species loss or abundance changes because the consequences can manifest themselves at different time scales. For example, the full impact of top predator removal from tropical forest ecosystems takes decades to centuries to become apparent, and considerably longer to ripple through different elements of the community (Terborgh, 1986; Dirzo and Miranda, 1991; Jackson et al., 2001).

Context Dependence

An increasing body of evidence suggests that keystone status is context-dependent. That is, keystone species are not necessarily dominant controlling agents in all parts of their range or at all times. Some taxa play keystone roles only under certain conditions. For example, the keystone starfish species *Pisaster ochraceus* occupies an unambiguous keystone role on wave-exposed rocky headlands (Menge et al., 1994). In more wave-sheltered habitats, however, the impact of *Pisaster* predation may be weak or nonexistent.



Some starfish act as keystone species. Source: K. Frey

Overall, the keystone species concept shows how the loss of a species with low abundance may have surprisingly dramatic effects. In addition, the concept implies that focusing conservation concerns on a single species may be problematic since other species (keystone) may be unknowingly influencing its abundance. For further discussion on the utility of the keystone species concept as outlined in Power et al (1996), see Hurlbert (1997) and Kotliar (2000).

Table 1. Potential impacts of the loss of species at different trophic levels

Trophic Interaction	References
Predation: Loss of predators leads to an increase in herbivore densities and generally reduces primary productivity	Estes and Palmisano, 1978; Leigh et al., 1993; McLaren and Peterson, 1994; Terborgh and Wright, 1994; Estes and Duggins, 1995; Terborgh et al., 2001
Predation: Loss of predators leads to *mesopredator* release and increased levels of predation by mesopredators on their prey	Palomares et al., 1995; Ostfeld et al., 1996; Crooks and Soulé, 1999; Terborgh et al., 1999
Frugivory: Loss of frugivores frequently decreases plant regeneration through reduced seed dispersal, increased seed predation, and reduced seedling recruitment	Dirzo and Miranda, 1991; Redford, 1992; Hamann and Curio, 1999; Andresen, 1999; Ganzhorn et al., 1999; Wright et al., 2000
Herbivory: Loss of herbivores increases the density of seedlings and consequently plant regeneration	Dirzo and Miranda, 1991; Redford, 1992; Hamann and Curio, 1999; Andresen, 1999; Ganzhorn et al., 1999; Wright et al., 2000
Pollination: Loss of pollinators reduces seed and fruit set; causes erratic pollination service (fewer visits, mixed pollen loads, etc.); lowers crop yields; leads to reproduction by clonal growth; negative consequences for fig tree reproduction and further cascading effects on frugivorous birds and bats.	Janzen, 1974; Cox, 1983; Terborgh, 1986; Cropper and Calder, 1990; Cox et al., 1991; Thomson et al., 1991; Johnson and Bond, 1992; O'Toole, 1993 Steiner, 1993; Compton et al., 1994; Kearns and Inouye, 1997; Allen-Wardell et al., 1998; Kremen et al., 2002

Trophic Interactions

Loss of species results in the disruption of trophic interactions, which has consequences for ecosystem structure and function. A brief summary of the potential impacts of the loss of species at different trophic levels is provided in the following section, with the main points summarized in Table 1.

Loss of Predators

Ecological consequences of the loss of predators can be broadly classified into two categories: i) Increased herbivory due to higher densities of herbivorous prey following the loss of their predators; and ii) Increased densities of smaller predators known as mesopredators.

i. Increased herbivory

There is considerable evidence to suggest that the removal of top predators results in increased herbivory, and ultimately affects primary productivity. McLaren and Peterson (1994) investigated tree growth in Isle Royale National Park in Michigan, U.S.A. They found that plant growth rates were regulated by cycles in animal density, and trees increased in primary productivity only when released from herbivory due to predation by wolves on herbivores. It has also been shown that sea otters can have a profound effect on the structure of marine communities by controlling densities of herbivorous sea urchin populations, which feed on kelp beds. Absence of sea otters resulted in high densities of sea urchins, increased herbivory, and depletion of kelp beds (Estes and Palmisano, 1978; Estes and Duggins, 1995; Estes et al., 1998).

ii. Mesopredator release

Loss of predators may also lead to increased densities of smaller predators, a phenomenon known as 'mesopredator release' (Soulé et al., 1988). This release has been implicated in the decline and extinction of prey species (Soulé et al., 1988; Palomares et al., 1995; Sovada et al., 1995; Rogers et al., 1998; Crooks and Soulé, 1999). Crooks and Soulé (1999) describe a study of urban habitat fragments in coastal southern California. They test the hypothesis that the decline of the most common large predator (coyote) would result in the ecological release of mesopredators, both native (striped skunk, raccoon, grey fox) and exotic (domestic cat, opossum), and that increased predation by these mesopredators would result in higher mortality and local extinction rates of scrub-breeding birds. They found that bird species diversity decreased with total mesopredator abundance and was higher in fragments where coyotes were either present or more abundant. Other examples of direct and indirect effects of predator loss are given in Table 2.

Loss of Herbivores and Frugivores

There is a large body of research documenting the important role played by mammals and large birds in herbivory, seed dispersal, and seed predation (Redford, 1992). In Mexico, Dirzo and Miranda (1991) compared two tropical forests, one with a full complement of large mammals (peccaries, deer and tapir) and another in which these species had been extirpated by hunters. A striking differences between the two forests exists; the hunted forest was typified by seedling carpets, piles of uneaten rotting fruits and seeds, and herbs and seedlings undamaged by mammalian herbivores — phenomena much less evident in the non-hunted forest. Similarly, in central Panama, a study by Wright et al. (2000) showed that poachers reduce the abundance of herbivorous mammals, which in turn alters seed dispersal, seed predation, and seedling recruitment for two palms (*Attalea butyraceae* and *Astrocaryum standleyanum*).

Hamann and Curio (1999) assessed the potential impact of frugivore extirpations on forest regeneration in a wet tropical rainforest ecosystem in the Philippine islands. They found that dispersers of late-successional tree species were mostly endangered species including hornbills and fruit pigeons. Late-*successional* tree species were most specialized with respect to dispersers and could therefore be susceptible to extirpation following the loss of their dispersers due to overhunting.

In a comprehensive review, Şekercioğlu et al. (2004) present a general framework for characterizing the ecological and societal consequences of biodiversity loss and apply it to the global avifauna. The table in Appendix 1 describes the ecological and economical contributions of avian functional groups and the consequences of their loss. The unique study is perhaps the first to provide a rigorous and comprehensive understanding of the ecosystem consequences of bird declines. The major findings suggested that 21% of bird species are currently extinction-prone and 6.5% are functionally extinct contributing negligibly to ecosystem processes. Their projections indicate that by 2100, 6-14% of all bird species will be extinct, and 7-25% will be functionally extinct.

A number of studies have highlighted the role of primates in forest regeneration. A study in a rainforest in southeastern Peru by Andresen (1999) showed a complex web of interactions among seed dispersers, seed predators, and secondary dispersers that influence the fate of seeds. The study documented the significance of primates for seed dispersal by showing that spider monkeys (*Ateles paniscus*) and howler monkeys (*Alouatta seniculus*) dispersed the seeds of 71 and 14 plant species respectively. In a dry deciduous forest in Madagascar, Chapman and Onderdonk (1999) assess the potential importance of primates as seed dispersers in tropical forests and evaluate the possible consequences of hunting primates for recruitment in tropical tree communities. They use a case study in the Kibale National Park, Uganda to show that disrupting the complex interactions among primates and fruiting trees can have negative and possibly cascading effects on ecosystem processes.

Loss of Pollinators

Many authors have documented exclusive mutualisms between plant species and their pollinators, and highlighted the potential consequences of disruptions of mutualisms for plant regeneration and food crop yields (Cox et al., 1991; Compton et al., 1994; Allen-Wardell et al., 1998). Pollinator loss can affect plants in several ways, including loss of, or reduced, seed set (Kearns and Inouye, 1997). In addition, a scarcity of pollinators may affect a plant's mating system, resulting in the production of less vigorous offspring. This is because in the absence of pollinators, a higher percentage of seeds may be set through *self-pollination*, decreasing *heterozygosity* and increasing the expression of deleterious traits associated with *inbreeding*. On a broader scale, loss of pollinators or disruption of pollination systems may cause reduced seed and fruit production and ultimately, plant extinction. Any of these events will affect the organisms that consume seeds, fruits, or plants, or that use plants for nest construction. The plants most at risk from the loss of a pollinator are those that are *dioecious* and self-incompatible, those that have a single pollinator, and those that propagate only by seeds.

Bond (1994) developed a vulnerability index to rank the threat of extinction of a plant species due to the loss of pollinator or disperser mutualisms. The index assesses the vulnerability of a species by considering the following variables: the number of pollinator and disperser species needed, the level of *vegetative propagation*, whether the species is self-compatible, and whether the species depends on seedling recruitment. Analysis of case studies suggests that plants often compensate for high risk in one of the three categories by low risk in another. For example, self-incompatible plants with rare specialist pollinators often propagate vegetatively. Many of the species that appear vulnerable have compensatory mechanisms that buffer them in part from pollinator failure: the bird-pollinated species can be pollinated by beetles, and the dioecious ones can be wind pollinated (Bond, 1994). Some systems, including elements of the Cape flora in South Africa and lowland tropical rain forests, lack compensatory traits and the risk of plant extinction from failed mutualisms is high.

There is widespread concern regarding the potential consequences of pollinator declines on the conservation of biodiversity and stability of food crop yields (Allen-Wardell et al., 1998; Kremen et al., 2002). Approximately 30% of human food is derived from bee-pollinated crops (O'Toole, 1993) and thousands of wild plants depend on the services of bees for seed and fruit formation. Managed and feral European honey bees, throughout the United States and some European countries, are experiencing major population declines due to introduced parasitic mites, pesticide misuse, bad weather, or threats from Africanized honey bees (Matheson et al., 1996). Worldwide, nearly 200 species of wild vertebrate pollinators may be on the verge of extinction along with an untold number of invertebrate pollinators (Matheson et al., 1996). These declines are expected to have consequences that are both ecologically and economically significant, with serious implications for natural and agricultural systems.

Flying foxes (*Chiroptera: Pteropodidae*) are known to be of extraordinary ecological and economic importance throughout the forests of the old World tropics, but are seriously threatened by overexploitation. Many species appear to be in severe decline and several species are already extinct. A study by Fujita and Tuttle (1991) showed that at least 289 plant species rely to varying degrees on large populations of flying foxes for propagation. These plants, in addition to their many ecological contributions, produce some 448 economically valuable products. Additional evidence for the significance of flying foxes as pollinators is found in Cox et al. (1991) and Elmqvist et al. (1992).

There are several examples of extirpations of animal partners in reproductive mutualisms, but very few of subsequent plant extinctions. For example, the oil-collecting bee pollinator of a rare fynbos shrub *Ixianthes*, has become locally extirpated but the plant still sprouts and is not immediately threatened with extinction (Steiner, 1993). In Hawaii, extinction of native bird pollinators resulted in a change of pollinators for the *Freycinetia arborea*, but not extinction (Cox, 1983). Memmott et al. (2004) explored probable patterns of extinction in two large networks of plants and flower visitors by simulating the removal of pollinators and consequent loss of the plants that depend upon them for reproduction. Plant species diversity declined most rapidly with preferential removal of the most-linked or most generalized pollinators as compared to the least-linked or most specialized pollinators. However, both pollination networks were relatively tolerant to loss of component species. Overall, the evidence for functional declines associated with the loss of pollinators is mixed. Systems with *redundancy* in which pollinator species can interact with a single plant may tolerate declines, while other plant species in non-redundant systems may go extinct and crop yields may decline.

Table 2. Examples of documented trophic cascades (adapted from Pace et al., 1999)

Ecosystem	Cascade	Effect	References
Marine			
Open ocean	Salmon-zooplankton-phytoplankton	Twofold higher phytoplankton when salmon are abundant	Shiomoto, et al., 1997
Coastal	Whales-otter-urchins-kelp	Increased predation by whales on otters leads to increased urchin grazing and up to ten times fewer kelp	Estes et al., 1998
Freshwater			
Streams	Fish-invertebrates-periphyton	Production of periphyton affected (six-fold) by predation of invertebrate populations	Huryn, 1998
Shallow lake	Fish-zooplankton-phytoplankton	Significant changes in fish populations due to mortality lead to shifts in zooplankton size structure and corresponding strong effects on phytoplankton	Jeppesen et al., 1998
Terrestrial			
Tropical forest	Beetles-ants-insects-Piper plants	Beetles prey upon ants that remove herbivorous insects that consume plants; more foliage consumed in the presence of beetles	Letourneau and Dyer,1998b
Boreal forest	Wolves-moose-balsam fir	Wolf predation controlled moose densities which in turn influenced primary productivity	McLaren and Peterson, 1994

Weak Interactors

‘Strong interactors’ are similar to keystone species, and usually have disproportionately large effects on the rest of the ecosystem. As described in the previous section, the loss of strong interactors can cause dramatic changes in communities (Paine, 1969; Estes and Palmisano, 1978; Menge et al., 1994; Power et al., 1996). Experiments indicate that in many communities only a few species will have such strong effects, whereas most will have weak effects owing to small per capita effects and/or low abundance (Power et. al., 1996, Paine, 1992; McGrady-Steed et al., 1997, Berlow, 1999). Alarmingly, some studies have shown that even the extinction of these ‘weak’ interactors could significantly alter natural communities because they play important stabilizing or ‘noise-dampening’ roles (Navarrete and Menge, 1996; Bengtsson et al., 1997; McGrady-Steed et al., 1997; McCann et al., 1998). Using a simple rocky-intertidal food web as a model system, Berlow (1999) showed that weak interactors play an important, but unappreciated role in maintaining landscape-scale diversity if their effects on species abundances are strongly context-dependent, or highly variable over space and time. Overall, these studies indicate that even though strong interactors clearly play important functional roles, it is important not to overlook the non-obvious stabilizing effects of some weak interactors.

Trophic Cascades

All of the above changes in communities that follow species loss, especially changes in trophic interactions, can lead to extended cascading effects throughout ecosystems due to species interactions. Trophic cascades result in inverse patterns in abundance or biomass across more than one trophic link in a food web (Carpenter and Kitchell, 1998). For a three-level food chain, abundant top predators result in lower abundances of mid-level herbivores and higher abundance of basal producers. In this case, removing a top predator would result in a greater abundance of consumers and fewer pro-

Abundance changes of organisms on one trophic level has repercussions on other trophic levels. Source: S. Spector

ducers. Global extinction of a species or local extirpation of a population can result in disruptions of trophic cascades leading to dramatic shifts in community composition, structure, and function.

The 'trophic cascade' concept arose from the observations and experiments of field ecologists who observed the powerful effects of predators in the marine intertidal zone (Paine, 1980) and in lakes (Shapiro et al., 1975). Estes et al., (1989) provide a classic example of a trophic cascade in the sea otter-urchin-kelp interaction of coastal North America. Otters stabilize a system of abundant kelp forests by reducing urchin grazing. Removal of otters shifts the system to urchin dominance with substantial reductions in kelp coverage and productivity. Studies have documented trophic cascades in diverse ecosystems such as streams, lakes, marine intertidal zone, terrestrial and marine ecosystems including fields, soils, forests, and the open ocean. Table 2 provides a non-exhaustive summary of documented trophic cascades.

There may be a number of mechanisms suppressing cascades but recent studies have emphasized the importance of omnivores. Omnivory by top predators and mid-level consumers can exert strong regulation of other trophic levels in ways not predicted by cascading trophic interactions. In Costa Rican lowland streams, electric enclosures were used to limit access by fish and shrimp to benthic communities. Increases in the number of insects in these enclosures did not lead to a significant reduction in algae, contrary to expectations based on cascading interactions (Pringle and Hamazaki, 1998). Similar results were observed when fish were excluded from areas of Venezuelan streams (Flecker, 1996). In these cases, the top predators are omnivores that consume both insects and algae thereby precluding the potential for cascades. Recent work has shown that higher predator diversity can dampen the magnitude of trophic cascades (Finke and Denno, 2004).

Overall, there appears to be much evidence for cascading effects of species loss from diverse ecosystems. Given accelerated human alteration of ecosystems, increased management of species and ecosystems may become necessary to either prevent cascading effects or remedy the disruption of cascades responsible for maintaining ecosystem structure and function.

Species Co-extinctions

The term "co-extinction" has been used to describe the process of the loss of parasitic insects with the loss of their hosts (Stork and Lyal, 1993). The concept has been expanded to describe the demise of a broader array of interacting species including predators with their prey and specialist herbivores with their host plants (Koh et al., 2004a). Koh et al. (2004b) define co-extinction as the loss of a species (the affiliate) upon the loss of another (the host).

Larval host plant specificity was an important ecological determinant of butterfly extirpations in Singapore (Koh et al., 2004a). The authors indicate that the number of locally extinct butterfly species is expected to increase exponentially with that of extinct host plants. Further, butterflies may go extinct sooner than their host plants when the declines, and not necessarily extinctions, of certain host plants (e.g., long-lived tree species) reduce butterflies to below their *minimum viable population*s (see section on ecological extinctions below). Although the loss of butterflies may result in declines of flowering plants that need these butterflies for pollination, the reverse is more likely to be true, whereby the loss of host plants drives butterfly extinctions. This is because caterpillars are often more dependent on plants for food than plants are on adult butterflies as generic pollinators (Ehrlich and Raven, 1964, Corlett, 2004).

Koh et al. (2004b) use a model to examine the relationship between co-extinction levels (proportion of species extinct) of affiliates and their hosts across a wide range of co-evolved interspecific systems: pollinating *Ficus* wasps and *Ficus*, parasites and their hosts, butterflies and their larval host plants, and ant butterflies and their host ants. They estimate that 6,300 affiliate species are "co-endangered" with host species currently listed as endangered, thus calling for a need to increase current extinction estimates by taking species co-extinctions into account. Species co-extinction is a manifestation of the interconnectedness of organisms in complex ecosystems. The loss of species through co-extinction represents the loss of irreplaceable evolutionary and co-evolutionary history (Nee and May, 1997; Purvis et al., 2000).

Ecological Extinctions

Extinction is irreversible. However, human activities influence the relative abundances of species more frequently than they cause extinction. Species abundance distributions warrant increased attention, because they usually respond more rapidly to human activities than do changes in species richness, and because they can have important consequences for ecosystems long before a species is threatened by extinction (Chapin et al., 2000; Wilsey and Potvin, 2000; Smith and Knapp, 2003). Estes et al. (1989) defined ecological extinction as the reduction of a species to such low abundance that, although it is still present in the community, it no longer interacts significantly with other species. Very little is currently known about the prevalence of ecological extinctions, which can often be difficult to measure empirically.

It has been suggested that vertebrate populations which meet demographic and genetic criteria for viability (i.e., they occur at minimum viable population (MVP) size) could still be so sparse that they no longer interact as intensely as they previously did when they occurred at higher densities (Conner, 1988; Redford and Feinsinger, 2001). For example, if populations of seed-dispersing primates were maintained only at MVP levels, total numbers of seeds dispersed might decline abruptly and vegetation dynamics would be affected by a reduction (not extinction) of the primate population (Redford and Feinsinger, 2001). In other words, even if a species is not locally or globally extinct, a reduction in population size could initiate cascading effects. According to Redford and Feinsinger (2001): if the population of species A no longer maintains interactions with species B to Z due to a reduction in the population size of species A, the cascading effects that result will be indistinguishable from those where species A has gone completely extinct.

Redford and Feinsinger (2001) examine the impacts of population reduction of a target species due to harvesting and present two models of species interaction in forests with demographically viable but ecologically extinct animal populations. In the uniform model, a reduction in consumer population either by harvesting or another phenomenon reduces the population by 50%. All else being equal, this should simply lead to a 50% reduction in use, on average, across the resources used by the population, i.e. the response is uniform. The population displays the same pattern of relative selectivity as before, and even the least preferred class of resource still has its share of users. Reduction in population density has only quantitative but not qualitative effects. However, available ecological data do not support the assumption that individuals within a population will respond in a homogeneous fashion irrespective of that population's density. Evidence suggests that changes in animal population density appear to have qualitative as well as quantitative effects through either one or both of two mechanisms: 1) Individuals that make up the consumer population are not identical. Every animal's choices may differ consistently from those of others. 2) An individual may change its choices as a direct or indirect result of the density of competitors of the same species.

According to the differential model, substantial reductions in animal population density will rarely, if ever, result in uniform reduction in interactions across the various classes of foods. Instead, depending on which particular individuals disappear and take their unique diet-related traits with them, some kinds of food may be consumed nearly as frequently as previously (when the population is at a higher density) while others will scarcely be consumed at all. For example, less preferred resources may be passed over by surviving foragers. If the different classes of resources are different species, then population dynamics of species less preferred by consumers might change dramatically, with resulting cascading effects.

The concept of ecological extinctions has been applied to terrestrial as well as marine settings. A study by Şekercioğlu et al. (2004) found 7% of birds being ecologically extinct whereas 1% were actually extinct (Figure 1a in Şekercioğlu et al., 2004). A study by Novaro et al. (2000) concludes that native large-bodied prey species (guanacos, *Lama guanicoe*, and rheas, *Pterocnemia pennata*) are ecologically extinct as prey and a source of carrion for native carnivores in northwestern Patagonia and likely throughout Argentine Patagonia. The geographical ranges and densities of native prey species have been greatly reduced due to hunting, habitat degradation, and competition with introduced livestock and wild exotic species. This is likely to lead to reductions in abundance and/or extinctions of native carnivores. In two different systems, Larsen et al. (2005) found that human landscape alteration strongly reduced bee and dung beetle abundance. Reduced abundance of bees was associated with lower pollination estimates for crops and reduced abundance of dung beetles was associated with disruption of estimated dung burial and associated functions such as seed dispersal.

Summary

Human domination of Earth's ecosystems is accelerating the extinction of species and is significantly changing the structure and dynamics of biological communities worldwide. Within this context, a relevant and pragmatic question that arises is the extent to which this loss of biodiversity matters and whether stability, productivity, and other aspects of the functioning of both managed and natural ecosystems are dependent on biodiversity. This synthesis attempts to provide a brief overview of the consequences of biodiversity loss to ecosystem functions and processes, focusing on evidence from field experiments, mechanistic theory, and quantitative field observations. Overall, increasing research and understanding of the ecological consequences of species loss has led to the emergence of generalities relevant for conservation planning and management.

Terms of Use

"M. Rao and T. Larsen. 2009. Ecological Consequences of Extinction. Synthesis. American Museum of Natural History, *Lessons in Conservation*. Available at http://ncep.amnh.org/linc."

This material is based on work supported by the National Science Foundation under the Course, Curriculum and Laboratory Improvement program (NSF 0127506), and the United States Fish and Wildlife Service (Grant Agreement No. 98210-1-G017).

Literature Cited

Allen-Wardell, G. , P. Bernhardt, R. Bitner, A. Burquez, S. Buchmann, J. Cane, P.A. Cox, V. Dalton, P. Feinsinger, M. Ingram, D. Inouye, C.E. Jones, K. Kennedy, P. Kevan, H. Koopowitz, R. Medellin, S. Medellin-Morales, and G.P. Nabhan. 1998. The potential consequences of pollinator declines on the conservation of biodiversity and stability of food crop yields. Conservation Biology 12: 8-17.

Alroy, J. 2001. A multispecies overkill simulation of the end-Pleistocene megafaunal extinction. Science 292: 1893-1896.

Andresen, E. 1999. Seed dispersal by monkeys and the fate of dispersed seeds in a Peruvian rainforest. Biotropica 31: 145-158.

Bengtsson, J., J. Jones, and H. Setala. 1997. The value of biodiversity. Trends in Ecology and Evolution 12:334-336.

Berlow, E.L. 1999. Strong effects of weak interactions in ecological communities. Nature 398: 330-334.

Bodmer, R.E. 1995. Managing Amazonian wildlife: biological correlates of game choice by detribalized hunters. Ecological Applications 5: 872-877.

Bond, W. J. 1993. Keystone species. Pages 237-253 in D. Schulze and H.A. Mooney, editors. Ecosystem Function and Biodiversity. Springer-Verlag, Berlin, Germany.

Bond, W. J. 1994. Do mutualisms matter? Assessing the impact of pollinator and disperser disruption on plant extinction. Philosophical Transactions of the Royal Society, London B: Biological Sciences 344: 83-90.

Brown, J. H. and E.J. Heske. 1990. Control of a desert-grassland transition by a keystone rodent guild. Science 250: 1705-1707.

Cardillo, M. and L. Bromham. 2001. Body size and risk of extinction in Australian mammals. Conservation Biology 15: 1435-1440.

Carpenter, S. R., and J.F. Kitchell. 1988. Consumer control of lake productivity. Bioscience 38: 764-769.

Carpenter, S. R., J.F. Kitchell, and J.R. Hodgson. 1998. Cascading trophic interactions and lake productivity. Bioscience 35(10): 634-639.

Case, T.J. and E.A. Bender. 1981. Testing for higher order interactions. American Naturalist 118: 920-929.

Chapin III, F.S., B.H. Walker, R.J. Hobbs, D.U. Hooper, J.H. Lawton, O.E. Sala and D. Tilman. 1997. Biotic controls over the functioning of ecosystems. Science 277: 500-504.

Chapin III, F.S., E. S. Zavaleta, V.T. Eviner, R.L. Naylor, P.M. Vitousek, H.L. Reynolds, D.U. Hooper, S. Lavorel, O.E. Sala, S.E. Hobbie, M.C. Mack and S. Diaz. 2000. Consequences of changing biodiversity. Nature 405: 234-242.

Chapin III, F.S., O. E. Sala, I.C. Burke, J. P. Grime, D. U. Hooper, W. K. Lauenroth, A. Lombard, H. A. Mooney, A. R. Mosier, S. Naeem, S.W. Pacala, J. Roy, W. L. Steffen, and D. Tilman. 1998. Ecosystem consequences of changing biodiversity. Experimental evidence and a research agenda for the future. Bioscience 48(1): 45-52.

Chapman, C. A. and D.A. Onderdonk.1998. Forests without primates: primate/plant codependency. American Journal of Primatology 45: 127-241.

Compton, S. G., S.J. Ross, and I.W.B. Thornton. 1994. Pollinator limitation of fig tree reproduction on the Island of Anak Krakatau (Indonesia). Biotropica 26(2): 180-186.

Conner, R. N.1988. Wildlife populations: minimally viable or ecologically functional. Wildlife Society Bulletin 16: 80-84.

Corlett, R. T. 2004. Flower visitors and pollination in the oriental (Indo Malayan) region. Biological Reviews 79(3): 497-532.

Cox, P.A. 1983. Extinction of the Hawaiian avifauna resulted in a change of pollinators for the ieie, *Freycinetia arborea*. Oikos 41: 195-199.

Cox, P. A., T. Elmqvist, E.D. Pierson, and W.E. Rainey. 1991. Flying foxes as strong interactors in South Pacific island ecosystems: a conservation hypothesis. Conservation Biology 5(4): 1-7.

Crooks, K. R. and M.E. Soulé. 1999. Mesopredator release and avifaunal extinctions in a fragmented system. Nature 400: 563-566.

Cropper, S. C., and D.M. Calder. 1990. The floral biology of *Thelymitr epipactoides* (Orchidadaeae) and the implications of pollination by deceit on the survival of this rare orchid. Plant Systematics and Evolution 170: 11-27.

de Ruiter, P.C., A. Neutel, and J.C. Moore. 1995. Energetics, patterns of interactions, strengths, and stability in real ecosystems. Science 269: 1257-1260.

Daily, G. C. 1997. Nature's Services: Societal Dependence on Natural Ecosystems. Island Press, Washington D.C.

Darwin, C. 1859. On the Origin of Species by Means of Natural Selection Murray, London.

Darwin, C. and A. Wallace. 1858. On the tendency of species to form varieties; and on the perpetuation of varieties and species by natural means of selection. Zoological Journal of the Linnaean Society of London 3, 45-62.

Diamond, J. M. 1982. Man the exterminator. Nature 298: 787-789.

Diaz, S., A. J. Symstad, F.S. Chapin III, D.A. Wardle and L.F. Huennekes. 2003. Functional diversity revealed by removal experiments. Trends in Ecology and Evolution 18(3): 140-146.

Didham, R. K., J. H. Lawton, P.M. Hammond, and P. Eggleton. 1998. Trophic structure stability and extinction dynamics of beetles (Coleoptera) in tropical forest fragments. Philosophical Transactions of the Royal Society, London B: Biological Sciences 353: 437-451.

Dirzo, R. and A. Miranda. 1991. Altered patterns of herbivory and diversity in the forest understory: a case study of the possible consequences of contemporary defaunation. Pages 273-287 in P. W. Price, P.W. Lewinshon, G.W. Fernandes, and W.W Benson, editors. Plant-Animal Interactions: Evolutionary Ecology in Tropical and Temperate Regions. John Wiley and Sons, New York, NY, USA.

Duarte, C.M. 2000. Marine biodiversity and ecosystem services: an elusive link. Journal of Experimental Marine Biology and Ecology 250:117-131.

Duffy, E. 2002. Biodiversity and ecosystem function: the consumer connection. Oikos 99: 201-219.

Duffy, E. 2003. Biodiversity loss, trophic skew and ecosystem functioning. Ecology Letters 6: 680-687.

Ehrlich, P. R. and A.H. Ehrlich. 1981. Extinction: The Causes and Consequences of the Disappearance of Species. Random House, New York, NY, USA.

Ehrlich, P. R., and P.H. Raven. 1964. Butterflies and plants: a study in coevolution. Evolution 18: 586-608.

Elmes, G. W. 1991. Ant colonies and environmental disturbance. Symposium of the Zoological Society of London 63: 15-32.

Elmqvist, T., P.A. Cox, W.E. Rainey, and E.D. Pierson. 1992. Restricted pollination on oceanic islands: pollination of Ceiba Pentandra by Flying Foxes in Samoa. Biotropica 24(1): 15-23.

Elton, C.S. 1927. Animal Ecology. Sidgwick and Jackson, London.

Estes, J. A. and D.O. Duggins. 1995. Sea otters and kelp forests in Alaska: generality and variation in a community ecological paradigm. Ecological Monographs 65: 75-100.

Estes, J. A. and J.F. Palmisano. 1978. Sea otters: their role in structuring nearshore communities. Science 185: 1058-1060.

Estes, J. A., D.O. Duggins, and G.B. Rathbun. 1989. The ecology of extinctions in kelp forest communities. Conservation Biology 3: 251-264.

Estes, J. A., M.T. Tinker, T.M. Williams, and D.F. Doak. 1998. Killer whale predation on sea otters linking oceanic and nearshore ecosystems. Science 282: 473-476.

Finke, D.L. and R.F. Denno. 2004. Predator diversity dampens trophic cascades. Nature 429: 407-410.

Flecker, A. S. 1996. Ecosystem engineering by a dominant detritivore in a diverse tropical stream. Ecology 77: 1845-1854.

Fujita, M. S., and M.D. Tuttle. 1991. Flying foxes (Chiroptera: Pteropodidae): threatened animals of key ecological and economic importance. Conservation Biology 5(4): 455-463.

Ganzhorn, J. U., J. Fietz, E.Rakotovao, D. Schawb, and D. Zinner.1999. Lemurs and the regeneration of dry deciduous forest in Madagascar. Conservation Biology 13(4): 794-804.

Greenstreet, S. P. and S.J. Hall. 1996. Fishing and ground-fish assemblage structure in the north-western North Sea: an analysis of long-term and spatial trends. Journal of Animal Ecology 65: 577-598.

Grinnell, J. 1917. The Niche relationship of the California Thrasher. The Auk 34: 427-433.

Hamann, A. and E. Curio. 1999. Interactions among frugivores and fleshy fruit trees in a Philippine Submontane rainforest. Conservation Biology 13(4): 766-773.

Hector, A. 1998. The effect of diversity on productivity: detecting the role of species complementarity. Oikos 82: 597-599.

Hector, A., B. Schmid, C. Beierkuhnlein, M.C. Caldeira, M. Diermer, P.G. Dimitrakopoulos, J.A. Finn, H. Freitas, P.S. Giller, J. Good, R. Harris, P. Hogberg, K. Huss-Danell, J. Joshi, A. Jumpponen, C. Korner, P.W. Leadley, M. Loreau, A. Minns, C.P.H. Mulder, G. O'Donovan, S.J. Otway, J.S. Pereira, A. Prinz, D.J. Read, M. Scherer-Lorenzen, E.-D. Schulze, A.-S. D. Siamantziouras, E.M. Spehn, A.C. Terry, A.Y. Troumbis, F.I. Woodward, S. Yachi, and J.H. Lawton. 1999. Plant diversity and productivity experiments in European grasslands. Science 286: 1123-1127.

Holling, C.S. 1973. Resilience and stability of ecological systems. Annual Reviews of Ecology and Systematics 4: 1-23

Hooper, D.U. and P.M. Vitousek. 1997. The effects of plant composition and diversity on ecosystem processes. Science 277: 1302-1305.

Hurlbert, S.H. 1997. Functional importance vs. keystoneness: reformulating some questions in theoretical biocenology. Australian Journal of Ecology: 22-369-382.

Huryn, A. D. 1998. Ecosystem-level evidence for top-down and bottom-up control of production in a grassland stream system. Oecologia 115: 173-183.

Huston, M.A., L.W. Aarsen, M.P. Austin, B.S. Cade, J.D. Fridley, and E. Garnier. 2000. No consistent effect of plant diversity on productivity. Science 289:1255a.

Hutchinson, G. E. 1957. Concluding remarks, Cold Spring Harbor Symposium. Quantitative Biology 22: 415-427.

Jackson, J.B.C., M. X. Kirby, W.H. Berger, K. A. Bjorndal, L. W. Botsford, B. J. Bourque, R. H. Bradbury, R. Cooke, J. Erlandson, J. A. Estes, T. P. Hughes, S. Kidwell, C. B. Lange, H. S. Lenihan, J. M. Pandolfi, C. H. Peterson, R. S. Steneck, M. J. Tegner, and R. R. Warner. 2001. Historical overfishing and the recent collapse of coastal ecosystems. Science 293(5530): 629-638.

Janzen, D. H. 1974. The deflowering of Central America. Natural History 83: 49-53.

Jennings, S., S.P.R. Greenstreet, and J.D. Reynolds. 1999. Structural changes in an exploited fish community: a consequence of differential fishing effects on species with contrasting life histories. Journal of Animal Ecology 68: 617-627.

Jeppesen, E., M. Søndergaard, J. P. Jensen, E. Mortensen, A. M. Hansen, and T. Jørgensen. 1998. Cascading trophic interactions from fish to bacteria and nutrients after reduced sewage loading: an 18-year study of a shallow hypertrophic Lake. Ecosystems 1: 250-267.

Johnson, S. D. and W.J. Bond. 1992. Habitat dependent pollination success in a Cape orchid. Oecologia 91: 455-456.

Johnson, K. H., K.A. Vogt, H.J. Clark, O.J. Schmitz, and D.J. Vogt. 1996. Biodiversity and the productivity and stability of ecosystems. Trends in Ecology and Evolution 11(9): 372-377.

Jonsson, M. and B. Malmqvist. 2000. Ecosystem process rate increases with animal species richness: evidence from leaf-eating, aquatic insects. Oikos 89: 519-523.

Kearns, C. A. and D.W. Inouye. 1997. Pollinators, flowering plants and conservation biology. Conservation Biology 47: 297-307.

Koh, L. P., N.S. Sodhi, and B.W. Brook. 2004a. Co-extinctions of tropical butterflies and their hostplants. Biotropica 36: 272-274.

Koh, L.P, R.R. Dunn, N.S. Sodhi, R.K. Colwell, H.C. Proctor, and V.S. Smith. 2004b. Species coextinctions and the biodiversity crisis. Science 305: 1632-1634.

Kotliar, N. B. 2000. Application of the new keystone-species concept to Prairie dogs: how well does it work. Conservation Biology 14(6): 1715-1721.

Kremen, C., N. M. Williams, and R.W. Thorp. 2002. Crop pollination from native bees at risk from agricultural intensification. Proceedings of the National Academy of Sciences of the United States of America 99(26): 16812-16816.

Larsen, T. H., N. M. Williams and C. Kremen. 2005. Extinction order and altered community structure rapidly disrupt ecosystem functioning. Ecology Letters 8: 538-547.

Lavorel, S. and E. Garnier. 2002. Predicting changes in community composition and ecosystem functioning from plant traits: revisiting the Holy Grail. Functional Ecology 16: 545–556.

Lawton, J. 1994. What do species do in ecosystems? Oikos 71: 367-374.

Lawton, J. H., and V.K. Brown. 1993. Redundancy in ecosystems. Pages 255- 270 in E. D. Schulze and H.A. Mooney, editors. Biodiversity and Ecosystem Function. Springer-Verlag, Berlin, Germany.

Lawton, J. H., and C.G. Jones. 1993. Linking species and ecosystems: organisms as ecosystem engineers. Pages 141-150 in C.G. Jones and J.H. Lawton, editors. Linking species and ecosystems. Chapman and Hall, New York, New York, USA.

Lawton, J.H. and R.M. May, editors. 1995. Extinction Rates. Oxford University Press, Oxford, UK.

Leigh, E.G., Jr., S.J. Wright, F.E. Putz, and E.A. Herre. 1993. The decline of tree diversity on newly isolated tropical islands: a test of a null hypothesis and some implications. Evolutionary Ecology 7: 76-102.

Letourneau, D.K. and L.A. Dyer. 1998b. Density patterns of *Piper* ant-plants and associated arthropods: top-predator trophic cascades in a terrestrial system. Biotropica 30: 162-169.

Loreau, M. 1998a. Biodiversity and ecosystem functioning: a mechanistic model. Proceedings of the National Academy of Science 95: 5632–5636.

Loreau, M. 1998b. Separating sampling and other effects in biodiversity experiments. Oikos 82: 600-602.

Matheson, A., S.L. Buchmann, C. O'Toole, P. Westrich, and I.H. Williams. 1996. The Conservation of Bees. Academic Press, Harcourt Brace, London, U.K.

McArthur, R. 1955. Fluctuations of animal populations and a measure of community stability. Ecology 36: 533-536.

McCann, K., A. Hastings, and G.R. Huxel. 1998. Weak trophic interactions and the balance of nature. Nature 395: 794-798.

McGrady-Steed, J., P.M. Harris, and P.J. Morin. 1997. Biodiversity regulates ecosystem predictability. Nature 390:162-165.

McGrady-Steed, J. and Morin, P. J. 2000. Biodiversity, density compensation, and the dynamics of populations and functional groups. Ecology 81: 361–373.

McLaren, B. E., and R.O. Peterson. 1994. Wolves, moose, and tree rings on Isle Royale. Science 266: 1555-1558.

Meffe, G.K. and Carroll, C.R. 1994. Principles of Conservation Biology. Sinauer Associates, Inc., Sunderland, USA.

Memmott, J., N.M. Waser and M.V. Price. 2004. Tolerance of pollination networks to species extinctions. Proceedings of the Royal Society of London B: Biological Sciences 271: 2605-2611.

Menge, B.A., E.L. Berlow, C. Blanchette, S.A. Navarrete, and S.B. Yamada. 1994. The keystone species concept: variation in interaction strength in a rocky intertidal habitat. Ecological Monographs 64: 249-286.

Mills, L. S., M.E. Soulé, and D.F. Doak. 1993. The keystone species concept in ecology and conservation. Bioscience 43: 219-224.

Myers, N. 1996. Two key challenges for biodiversity: discontinuities and synergisms. Biodiversity and Conservation 5: 1025-1034.

Myers, N. and B. Worm. 2003. Rapid worldwide depletion of predatory fish communities. Nature 423: 280-283.

Naeem, S. and J. P. Wright. 2003. Disentangling biodiversity effects on ecosystem functioning: deriving solutions to a seemingly insurmountable problem. Ecology Letters 6: 567-579.

Naeem, S., L.J. Thompson, S.P. Lawler, J.H. Lawton and R.M. Woodfin. 1994. Declining biodiversity can alter the performance of ecosystems. Nature 368: 734-737.

Naiman, R.J., J.M Melillo, and J.E. Hobbie. 1988. Alteration of North American streams by beaver (*Castor canadensis*.).

Ecology 67: 1254-1289.

Navarrete, S.A., and B.A. Menge. 1996. Keystone predation and interaction strength: interactive effects of predators on their main prey. Ecological Monographs 66(4): 409-429.

Nee, S. and R.M. May. 1997. Extinction and the loss of evolutionary history. Science 278: 692-694.

Novaro, A. J., M.C. Funes, and R.S. Walker. 2000. Ecological extinction of native prey of a carnivore assemblage in Argentine Patagonia. Biological Conservation 92: 25-34.

Ostfeld, R.S. and K. LoGiudice. 2003. Community disassembly, biodiversity loss, and the erosion of an ecosystem service. Ecology 84: 1421-1427.

Ostfeld, R. S., C.G. Jones, and J.O. Wolff. 1996. Of mice and mast: ecological connections in eastern deciduous forests. Bioscience 46(5): 323-330.

O'Toole, C. 1993. Diversity of native bees and agroecosystems. Pages 169-196 in J. LaSalle and I.D. Gauld, editors. Hymenoptera and biodiversity. Center for Agriculture and Biosciences (CAB) International, Wallingford, U.K.

Owen-Smith, N. 1987. Pleistocene extinctions: the pivotal role of megaherbivores. Paleobiology 13: 351-362.

Pacala, S. and D. Tilman. 2001. From sampling to complementarity. Pages 151-166 in A. Kinzig, S. Pacala and D. Tilman, Editors. The Functional Consequences of Biodiversity. Princeton University Press, Princeton, New Jersey, USA.

Pace, M. L., J.J. Cole, S.R.Carpenter, and J.F. Kitchell. 1999. Trophic cascades revealed in diverse ecosystems. Trends in Ecology and Evolution 14(12): 483-489.

Paine, R.T. 1966. Food web complexity and species diversity. American Naturalist 100: 65-75.

Paine, R. T. 1969. A note on trophic complexity and community stability. American Naturalist 103: 91-93.

Paine, R. T. 1980. Food webs: linkage, interaction strength and community infrastructure. Journal of Animal Ecology 49: 667-685.

Paine, R.T. 1992. Food-web analysis through field measurement of per capita interaction strength. Nature 255: 73-75.

Palomares, F., P. Gaona, P. Ferreras, and M. Delibes. 1995. Positive effects on game species of top predators by controlling smaller predator populations: an example with lynx, mongooses, and rabbits. Conservation Biology 9: 295-305.

Parsons, T. R. 1993. The impact of industrial fisheries on the trophic structure of marine ecosystems. Pages 352-357 in G.A. Polis and K.O. Winemiller, editors. Food Webs: Integration of Patterns and Dynamics. Chapman and Hall, New York.

Pauly, D., V. Christensen, J. Dalsgaard, R. Froese, and F. Torres, Jr. 1998. Fishing down marine food webs. Science 279: 860-863.

Peres, C.A. 1990. Effects of hunting on western Amazonian primate communities. Biological Conservation 54: 47-59.

Peres, C.A. 2000. Effects of subsistence hunting on vertebrate community structure in Amazonian forests. Conservation Biology 14: 240-253.

Petchey, O. L. 2000. Species diversity, species extinction, and ecosystem function. American Naturalist 155: 696-702.

Petchey, O.L. and K.J. Gaston. 2002. Functional diversity, species richness and community composition. Ecology Letters 5: 402-411.

Petchey, O. L., P.T. McPhearson, T.M. Casey, and P.J. Morin. 1999. Environmental warming alters food-web structure and ecosystem function. Nature 402: 69-72.

Pfisterer, A. B. and B. Schmid. 2002. Diversity-dependent production can decrease the stability of ecosystem functioning. Nature 416: 84-86.

Pimm, S.L. 1991. The Balance of Nature. Ecological issues in the conservation of species and communities. The University of Chicago Press, Chicago. USA.

Pimm, S. L., H.L. Jones, and J. Diamond. 1988. On the risk of extinction. American Naturalist 132: 757-785.

Pimm, S.L., G. Russell, J.L. Gittleman and T.M. Brooks. 1995. The future of biodiversity. Science 269: 347-350.

Pollock, M. M., R.J. Naiman, H.E. Erickson, C.A. Johnston, J. Pastor, and G. Pinay. 1995. Beaver as engineers: influences on biotic and abiotic characteristics of drainage basins. Pages 117-126 in C.G. Jones and J.H. Lawton, editors. Linking Species and Ecosystems. Chapman and Hall, New York, New York, USA.

Power, M. E. 1990. Effects of fish in river food webs. Science 250: 411-415.

Power, M. E., D. Tilman, J.A. Estes, B.A. Menge, W.J. Bond, L.S. Mills, G.C. Daily, J.C. Castilla, J. .Lubchenco and R.T. Paine. 1996. Challenges in the quest for keystones. Bioscience 46(8): 609-620.

Pringle, C.M. and T. Hamazaki. 1998. The role of omnivory in a neotropical stream: separating diurnal and nocturnal effects. Ecology 79, 269–280.

Purvis, A. and A. Hector. 2000. Getting the measure of biodiversity. Nature 405: 220-227.

Purvis, A., P.M. Agapow, J.L. Gittleman, and G.M. Mace. 2000. Nonrandom extinction and the loss of evolutionary history. Science 288: 328-330.

Raffaelli, D. 2004. Ecology: how extinction patterns affect ecosystems. Science 306(5699): 1141-1142.

Redford, K. H.1992.The Empty Forest. BioScience 42(6): 412-422.

Redford, K. H., and P. Feinsinger. 2001. Sustainable use and the ecology of interactions. Pages 370-399 in Conservation of Exploited Species. J. D. Reynolds, G.M. Mace, K.H. Redford, and J.G., Robinson, editors. Conservation of Exploited Species. Cambridge University Press, Conservation Biology Series 6.

Rice, J., and H. Gislason. 1996. Patterns of change in the size spectra of numbers and diversity of the North Sea fish assemblage, as reflected in surveys and models. ICES Journal of Marine Science 53: 1199-1213.

Rogers, C. M. and M.J. Caro. 1998. Song sparrows, top carnivores, and nest predation: a test of the mesopredator release hypothesis. Oecologia 116: 227-233.

Ruesink, J.L. and D.S. Srivastava. 2001. Numerical and per capita responses to species loss: mechanisms maintaining ecosystem function in a community of stream insect detritivores. Oikos 93: 221–234.

Schaffers, A. P. 2002. Soil, biomass, and management of semi-natural vegetation. Part II. Factors controlling species diversity. Plant Ecology 158: 247-268.

Schulze, E.-D. and H.A. Mooney, editors 1993. Biodiversity and ecosystem function. Springer-Verlag.

Schwartz, M.W., C.A. Brigham, J.D. Hoeksema, K.G. Lyons, M.H. Mills, and P. J. van Mantgem. 2000. Linking biodiversity to ecosystem function: implications for conservation ecology. Oecologia 122: 297-305.

Şekercioğlu, C. H., G.C. Daily, and P.R. Ehrlich. 2004. Ecosystem consequences of bird declines. Proceedings of the National Academy of Sciences 101(52): 18042-18047.

Shapiro, J., V. Lammara, and M. Lynch. 1975. Biomanipulation: an ecosystem approach to lake restoration. Pages 85-96 in P. L. Brezonik and J.L. Fox, editors. Water Quality Management Through Biological Control. Department of Environmental Engineering Science, FL, USA.

Smith, M.D. and A.K. Knapp. 2003. Dominant species maintain ecosystem function with non-random species loss. Ecology Letters 6: 509-517.

Soulé, M.E., D.T. Bolger, A.C. Alberts, J. Wrights, M. Sorice and S. Hill. 1988. Reconstructed dynamics of rapid extinctions of chaparral-requiring birds in urban habitat islands. Conservation Biology 2: 75-92.

Sovada, M. A., A.B. Sargeant and J.W. Grier. 1995. Differential effects of coyotes and red foxes on duck nest success. Journal of Wildlife Management 59: 1-9.

Steiner, K. E. 1993. Has Ixianthes (Scrophulariaceae) lost its special bee? Plant Systematics and Evolution 185: 7-16.

Stork, N. E. and C.H.C. Lyal. 1993. Extinction or co-extinction rates. Nature 366: 307.

Symstad, A. J., D. Tilman, J. Willson and J.M.H. Knops.1998. Species loss and ecosystem functioning: effects of species identity and community composition. Oikos 81: 389–397.

Symstad, A. J., F. S. Chapin III, D.H. Wall, K. L. Gross, L.F. Huenneke, G.G. Mittelbach, D.P.C. Peters, and D. Tilman. 2003. Long-term and large-scale perspectives on the relationship between biodiversity and ecosystem functioning. Bioscience 53(1): 89-98.

Terborgh, J. 1986. Keystone plant resources in the tropical forest. Pages 330-344 in M. E. Soulé, editor. Conservation Biology. Sinauer, Sunderland, MA, USA.

Terborgh, J. 1988. The big things that run the world: a sequel to E.O. Wilson. Conservation Biology 2: 402-403.

Terborgh, J. and S. J. Wright. 1994. Effects of mammalian herbivores on plant recruitment in two Neotropical forests. Ecology 75: 1829-1833.

Terborgh, J., J.A. Estes, P. Paquet, K. Ralls, D. Boyd-Heger, B.

J. Miller, and R.F. Noss. 1999. The role of top carnivores in regulating terrestrial ecosystems. Pages 39-64 in Continental Conservation. M. E. Soulé and J. Terborgh, editors. Continental Conservation. Island Press, Washington D.C., U.S.A.

Terborgh, J., L. Lopez, P. Nunez, M. Rao, G. Shahabuddin, G. Orihuela, M. Riveros, R. Ascanio, G.H. Adler, T.D. Lambert, and L. Balbas. 2001. Ecological meltdown in predator-free forest fragments. Science 294(5548): 1923-1926.

Thomson, J. D., E.A. Herre, and J.L. Hamrick.1991. Genetic mosaics in strangler fig trees: implications for tropical conservation. Science 254: 1214-1216.

Tilman, D. and J.A. Downing. 1994. Biodiversity and stability in grasslands. Nature 367: 363-365.

Tilman, D., C.L. Lehman and K.T. Thomson. 1997a. Plant diversity and ecosystem productivity: theoretical considerations. Proceedings of the National Academy of Science, USA 94: 1857-1861.

Tilman, D., D. Wedin, and J. Knops. 1996. Productivity and sustainability influenced by biodiversity in grassland ecosystems. Nature 379: 718-720.

Tilman, D., D. Wedin, and J. Knops. 1988. Productivity and sustainability influenced by biodiversity in grassland ecosystems. Nature 396: 69-72.

Tilman, D., J. Knops, D. Wedin, P. Reich, M. Ritchie, and E. Siemann. 1997b. The influence of functional diversity and composition on ecosystem processes. Science 277: 1300-1302.

Walker, B. 1992. Biodiversity and Ecological redundancy. Conservation Biology 6:18-23.

Wardle, D.A., K.I. Bonner and K.S. Nicholson. 1997. Biodiversity and plant litter: experimental evidence which does not support the view that enhanced species richness improves ecosystem function. Oikos 79: 247-258.

Wilsey, B.J. and H.W. Polley. 2004. Realistically low species evenness does not alter grassland species-richness-productivity relationships. Ecology 85: 2693-2700.

Wilsey, B.J. and C. Potvin. 2000. Biodiversity and ecosystem functioning: importance of species evenness in an old field. Ecology 81: 887-892.

Wolters, V., W.L. Silver, D.E. Bignell, D.C. Coleman, P. Lavelle, P., D.P. W.H. Van, R.P. De, J. Rusek, D.H. Wall, D.A. Wardle, L. Brussard, J.M. Dangerfield, V.K. Brown, K.E. Giller, D.U. Hooper, O. Sala, J. Tiedje, and V.J.A. Van. 2000. Effects of global changes on above- and below- ground biodiversity in terrestrial ecosystems. Implications for ecosystem functioning. Bioscience 50: 1089-1098.

Wootton, J.T. 1993. Indirect effects and habitat use in an intertidal community: interaction chains and interaction modifications. American Naturalist 141: 71-89

Wright, S. J., H. Zeballos, I. Dominguez, M.M. Gallardo, M.C, Moreno, and R. Ibanez. , R. 2000. Poachers alter mammal abundance, seed dispersal, and seed predation in a Neotropical Forest. Conservation Biology 14(1): 227-239.

Yachi, S. and M. Loreau. 1999. Biodiversity and ecosystem productivity in a fluctuating environment: the insurance hypothesis. Proceedings of the National Academy of Sciences 96: 1463-1468.

Zavaleta, E. S. and K. B. Hulvey. 2004. Realistic species losses disproportionately reduce grassland resistance to biological invaders. Science 306: 1175-1177.

Glossary

Assembly experiments: experiments that are used to examine the functional relationship between species richness and ecosystem function using artificially assembled communities within controlled environments.

Biodiversity: the variety of life on Earth at all its levels, from genes to ecosystems, and the ecological and evolutionary processes that sustain it.

Consumers: organisms that ingest other organisms or organic matter in a food chain.

Density compensation: an inverse relation between population density and species richness; differences in species richness result in compensatory changes in the abundance of populations. For example, loss of one species could result in an increase in the population density of remaining species.

Dioecious: a dioecious plant has unisexual flowers or functionally equivalent structures of both sexes occurring on different individuals.

Diversity: use of the term diversity has led to some confusion, since diversity sometimes refers to the number of species, but more often refers to a measure combining species richness and evenness (the degree to which species in a community are equally abundant).

Ecosystem function: the term "ecosystem functioning" refers to the biogeochemical processes particular to a system, such as photosynthesis, decomposition, and nitrogen fixation

Ecosystem services: defined as the processes and conditions of natural ecosystems that support human activity and sustain human life. Such services include the maintenance of soil fertility, climate regulation and natural pest control, and provide flows of ecosystem goods such as food, timber and fresh water. They also provide intangible benefits such as aesthetic and cultural values.

Extinction: the global extinction of a species refers to the irreversible loss of all populations (and thus individuals) of the species across its entire range. At a smaller spatial scale, a species is considered locally extinct if there are no populations within a particular geographical area or site, a phenomenon that is also known as extirpation. Unlike global extinction, local extinction or extirpation is theoretically considered to be a reversible phenomenon. The reversibility of local extinctions is influenced by many factors including whether the population of individuals is open (individuals from extant populations in other locations could potentially recolonize) or closed (the population is isolated such that it is not possible for new individuals to recolonize). Ecological extinctions are a separate phenomena, whereby a population is reduced to a low enough level at which the species can no longer maintain its functional role in the ecosystem.

Facilitative interactions: interactions in which certain species interact directly or indirectly with other species in a way that increases species' efficiency within its niche.

Functional or ecological extinction: the concept of ecological extinction refers to the, "Reduction of a species to such low abundance that, although it is still present in the community, it no longer interacts significantly with other species," (Estes et al. 1989).

Heterozygosity: a measure of the genetic diversity in a population, as measured by the number of heterozygous loci across individuals.

Higher-order interactions: refers to any non-additivity of the per capita interaction terms describing different species effects on the per capita growth rate of a focus species (Case and Bender 1981); has also been used to describe a functional change in the interaction of two species caused by a third species (Wootton 1993). These functional changes cannot be extrapolated from the dynamics of single species or species pairs in isolation.

Inbreeding: the mating of individuals who are more closely related than by chance alone.

Keystone guilds: a guild is defined as a group of species that exploit the same class of environmental resources in a similar way. This term groups together species that overlap significantly in their niche requirements without regard to taxonomic position. For example, in Brown and Heske's study (1990), seeds constituted the primary food for three groups of granivores-rodents (*Dipodomys* sp., *Perognathus* sp., and *Peromyscus* sp.), harvester ants (*Pogonomyrmex* sp.) and various species of birds. The three groups of granivores together constitute a guild.

Mesopredator release: the process by which smaller predators known as mesopredators increase in abundance following the loss of top predators that control their densities; the process of mesopredator release leads to a decrease in the population density of small prey species.

Minimum viable populations: the smallest isolated population size that has a specified percent chance of remaining extant for a specified period of time in the face of foreseeable demographic, genetic, and environmental stochasticities, plus natural catastrophes. (Meffe and Carroll 1994)

Niche space: definitions of niche emphasize either an organism's individual characteristics or its relationships within a community. An alternative definition of niche involves arbitrary subdivisions grouping similar species, sometimes called 'habitat' or 'trophic' niches. Grinnell (1917) defined it as all the sites where organisms of a species can live (where conditions are suitable for life). Elton (1927) described the niche as the function performed by the species in the community of which it is a member. Hutchinson (1957) defined a niche as a region (n-dimensional hypervolume) in a multi-dimensional space of environmental factors that affect the welfare of a species.

Nutrient cycling: the processes by which elements are extracted from their mineral, aquatic, or atmospheric sources or recycled from their organic forms, converting them to the ionic form in which biotic uptake occurs and ultimately returning them to the atmosphere, water, or soil (taken from the Millenium Ecosystem Assessment Glossary).

Primary productivity: rate at which new plant biomass is formed by photosynthesis.

Redundancy: refers to a state of being redundant i.e. exceeding what is necessary.

Resilience: ecosystem resilience measured as a rate of change refers to the amount of time taken for an ecosystem that has been displaced from equilibrium returns to it. It is also defined as how fast a variable that has been displaced from equilibrium returns to it (Pimm 1991). Holling (1973) defines resilience to be how large a range of conditions will lead to a system returning to equilibrium. In his definition, highly resilient systems will almost always return to equilibrium, whatever happens to them; systems that are not resilient will often be fundamentally changed after a perturbation, perhaps by losing species or by moving to a new equilibrium involving very different species' densities.

Resistance: measures the consequences when a variable is permanently changed: how much do other variables change as a consequence? If the consequent changes are small, the system is relatively resistant. Resistance is measured as a ratio of a variable before and after the change. (Pimm 1991).

Self-pollination: transfer of pollen from the anther to the stigma of the same flower.

Species invasions: the introduction of species beyond their native ranges. There are many examples of disastrous invasions by such species that has resulted in losses of native species, changes in community structure and function, and even alterations of the physical structure of the system. Not all species invasions result in disaster (Meffe and Carroll 1994).

Succession: the natural, sequential change of species composition of a community in a given area.

Vegetative propagation: a process of asexual reproduction by which new plant individuals arise without the production of seeds or spores.

Appendix I

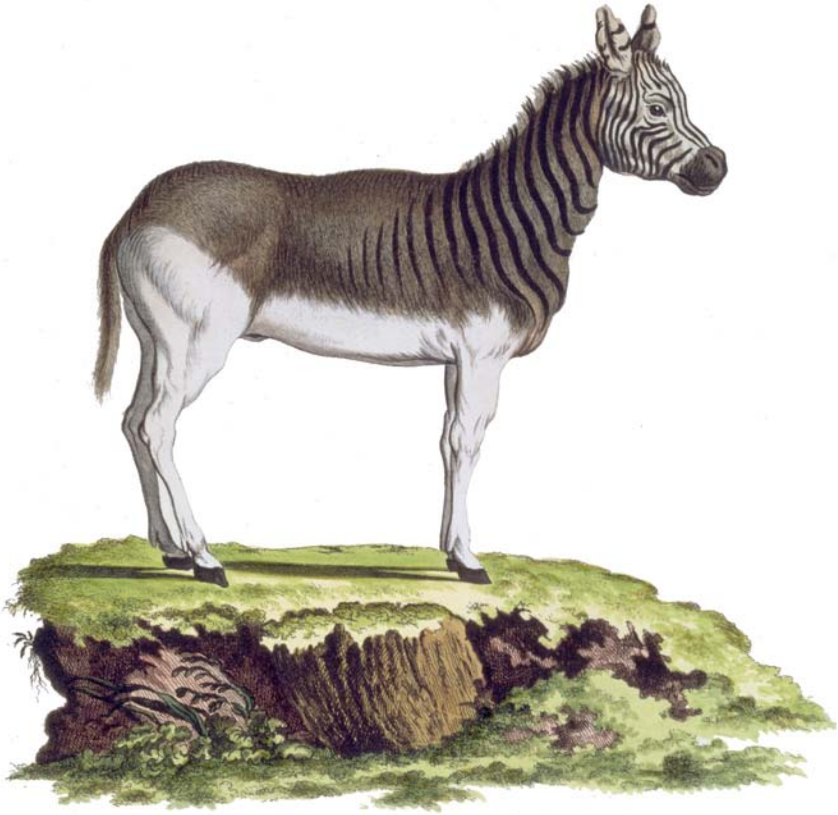
Ecological and Economical Contributions of Avian Functional Groups (From Şekercioğlu et al. 2004. Proceedings of the National Academy of Sciences 101 [52]: 10842–18047. Copyright 2004 National Academy of Sciences, U.S.A).

Functional group	Ecological process	Ecosystem service and economical benefits	Negative consequences of loss of functional group
Frugivores	Seed dispersal (1–4)	Removal of seeds from parent tree (5–8); escape from seed predators (9,10); improved germination (11,12); increased economical yield (13–16); increased gene flow (17–19); recolonization and restoration of disturbed ecosystems (20–24)	Disruption of dispersal mutualisms (25–27); reduced seed removal (28); clumping of seeds under parent tree (29); increased seed predation (10); reduced recruitment (28,30); reduced gene flow (31,32) and germination (12,33,34); reduction (35,36) or extinction (37–40) of dependent species
Nectarivores	Pollination (3,41,42)	Outbreeding of dependent (42–44) and/or economically important species (14,45)	Pollinator limitation (45,46); inbreeding and reduced fruit yield (47–52); evolutionary consequences (41,45,53); extinction (37,54)
Scavengers	Consumption of carrion (55)	Removal of carcasses (56,57); leading other scavengers to carcasses (55); nutrient recycling; sanitation (56,57)	Slower decomposition (55); increases in carcasses (56,57); increases in undesirable species (56,57); disease outbreaks (56,57); changes in cultural practices (56,58)
Insectivores	Predation on invertebrates	Control of insect populations (59–65); reduced plant damage (62,66,67); alternative to pesticides (68–70)	Loss of natural pest control (68,69); pest outbreaks (59,61,71); crop losses (62); trophic cascades (72)
Piscivores	Predation on fishes and invertebrates Production of guano	Controlling unwanted species (73); nutrient deposition around rookeries (74–78); soil formation in polar environments (79); indicators of fish stocks (80); environmental monitors (81)	Loss of guano and associated nutrients (82); impoverishment of associated communities (83); loss of socioeconomic resources (84) and environmental monitors (81); trophic cascades (73,85)
Raptors	Predation on vertebrates	Regulation of rodent populations (86,87) secondary dispersal (88)	Rodent pest outbreaks (89); trophic cascades (72,90,91); indirect effects (92)
All species	Miscellaneous	Environmental monitoring (93,94); indirect effects (88,95–101); birdwatching tourism (102–104); reduction of agricultural residue (105); cultural and economic uses (106)	Losses of socioeconomic resources (102,107) and environmental monitors (108); unpredictable consequences (96)

1. Snow, D. W. (1981) Biotropica 13, 1–14.
2. Howe, H. F. and Smallwood, J. (1982) Annu. Rev. Ecol. Syst. 13, 201–228.
3. Stiles, E. W. (1985) in Conservation of Tropical Forest Birds, eds. Diamond, A. W. and Lovejoy, T. E. (Int. Council for Bird Pres., Cambridge, U.K.), pp. 49–59.
4. Howe, H. F. and Miriti, M. N. (2000) Trends Ecol. Evol. 15, 434–436.
5. Sun, C., Ives, A. R., Kraeuter, H. J. and Moermond, T. C. (1997) Oecologia 112, 94–103.
6. Greenberg, R., Foster, M. S. and Marquez-Valdelamar, L. (1995) J. Trop. Ecol. 11, 619–639.
7. Avila-H, M. L., Hernandez-O, V. H. and Velarde, E. (1996) Biotropica 28, 720–727.
8. Wenny, D. G. and Levey, D. J. (1998) Proc. Natl. Acad. Sci. USA 95, 6204–6207.
9. Janzen, D. (1972) Ecology 53, 350–361.
10. Howe, H. F. (1993) Vegetatio 108, 149–162.
11. Murphy, S. R., Reid, N., Yan, Z. and Venables, W. N. (1993) Oecologia 93, 171–176.
12. Meyer, G. A. and Witmer, M. C. (1998) Am. Midl. Nat. 140, 129–139.
13. Hutchins, H. E., Hutchins, S. A. and Liu, B. W. (1996) Oecologia 107, 120–130.
14. Narang, M. L., Rana, R. S. and Prabhakar, M. (2000) J. Bombay Nat. Hist. Soc. 97, 215–222.
15. Hammond, D. S., Gourlet-Fleury, S., Van Der Hout, P., Ter Steege, H. and Brown, V. K. (1996) For. Ecol. Manage. 83, 99–116.
16. Yumoto, T. (2000) Am. J. Bot. 87, 1181–1188.
17. Howe, H. F., Schupp, E. W. and Westley, L. C. (1985) Ecology 66, 781–791.
18. Hamrick, J. L., Murawski, D. A. and Nason, J. D. (1993) Vegetatio 108, 281–297.
19. Gibson, J. P. and Wheelwright, N. T. (1995) Oecologia 103, 49–54.
20. Robinson, G. R. and Handel, S. N. (1993) Conserv. Biol. 7, 271–278.
21. Tucker, N. I. J. and Murphy, T. M. (1997) For. Ecol. Manage. 99, 133–152.
22. Hjerpe, J., Hedenas, H. and Elmqvist, T. (2001) Biotropica 33, 249–259.
23. Galindo-Gonzalez, J., Guevara, S. and Sosa, V. J. (2000) Conserv. Biol. 14, 1693–1703.
24. Wilkinson, D. M. (1997) J. Biogeogr. 24, 61–65.
25. Stocker, G. C. and Irvine, A. K. (1983) Biotropica 15, 170–176.
26. Clark, C. J., Poulsen, J. R. and Parker, V. T. (2001) Biotropica 33, 606–620.
27. Meehan, H. J., McConkey, K. R. and Drake, D. R. (2002) J. Biogeogr. 29, 695–712.
28. Cordeiro, N. J. and Howe, H. F. (2003) Proc. Natl. Acad. Sci. USA 100, 14052–14056.
29. Bleher, B. and Bohning-Gaese, K. (2001) Oecologia 129, 385–394.
30. Cordeiro, N. J. and Howe, H. F. (2001) Conserv. Biol. 15, 1733–1741.
31. Pacheco, L. F. and Simonetti, J. A. (2000) Conserv. Biol. 14, 1766–1775.
32. Shapcott, A. (1999) Biotropica 31, 579–590.
33. Compton, S. G., Craig, A. J. F. K. and Waters, I. W. R. (1996) J. Biogeogr. 23, 553–563.
34. Peres, C. A. and Van Roosmalen, M. G. M. (1996) Oikos 75, 249–258.
35. Santos, T. and Telleria, J. (1994) Biol. Conserv. 70, 129–134.
36. Santos, T., Telleria, J. L. and Virgos, E. (1999) Ecography 22, 193–204.
37. Bond, W. J. (1994) Philos. Trans. R. Soc. London B 344, 83–90.
38. Hamann, A. and Curio, E. (1999) Conserv. Biol. 13, 766–773.
39. Loiselle, B. A. and Blake, J. G. (1999) Ecology 80, 330–336.
40. Cardoso da Silva, J. M. and Tabarelli, M. (2000) Nature 404, 72–74.
41. Stiles, F. G. (1978) Am. Zool. 18, 715–727.
42. Proctor, M., Yeo, P. and Lack, A. (1996) The Natural History of Pollination (Timber Press, Portland, OR).
43. Keighery, G. J. (1980) Entwicklungsgeschichte Und Systematik Der Pflanzen 135, 171–176.
44. Ford, H. A. (1985) Proc. Ecol. Soc. Aust. 14, 153–158.
45. Nabhan, G. P. and Buchmann, S. L. (1997) in Nature's Services, ed. Daily, G. (Island Press, Washington D.C.), pp. 133–150.
46. Murphy, D. J. and Kelly, D. (2001) N. Z. J. Ecol. 25, 69–81.
47. Montgomery, B. R., Kelly, D. and Ladley, J. J. (2001) N. Z. J. Bot. 39, 559–565.
48. Rathcke, B. J. (2000) Ecology 81, 1951–1958.
49. Feinsinger, P., Wolfe, J. A. and Swarm, L. A. (1982) Ecology 63, 494–506.

50. Paton, D. C. (2000) Conserv. Biol. 14, 1232–1234.
51. Robertson, A. W., Kelly, D., Ladley, J. J. andand Sparrow, A. D. (1999) Conserv. Biol. 13, 499–508.
52. Cox, P. A. and Elmqvist, T. (2000) Conserv. Biol. 14, 1237–1239.
53. Thompson, J. N. (1996) Trends Ecol. Evol. 11, 300–303.
54. Sakai, A. K., Wagner, W. L. and Mehrhoff, L. A. (2002) Syst. Biol. 51, 276–302.
55. Houston, D. C. (1994) in Handbook of the Birds of the World: New World Vultures to Guineafowl, eds. del Hoyo, J., Elliott, A. andand Sargatal, J. (Lynx Edicions, Barcelona), pp. 24–41.
56. Pain, D. J., Cunningham, A. A., Donald, P. F., Duckworth, J. W., Houston, D. C., Katzner, T., Parry-Jones, J., Poole, C., Prakash, V., Round, P., et al., (2003) Conserv. Biol. 17, 661–671.
57. Prakash, V., Pain, D. J., Cunningham, A. A., Donald, P. F., Prakash, N., Verma, A., Gargi, R., Sivakumar, S., and Rahmani, A. R. (2003) Biol. Conserv. 109, 381–390.
58. Parry-Jones, J. (2001) in 4th Eurasian Congress on Raptors, eds. Katzner, T. and Parry-Jones, J. (Raptor Research Foundation, Seville, Spain), pp. 17–18.
59. Crawford, H. S. and Jennings, D. T. (1989) Ecology 70, 152–163.
60. Marquis, R. J. and Whelan, C. J. (1994) Ecology 75, 2007–2014.
61. Kirk, D. A., Evenden, M. D. and Mineau, P. (1996) Current Ornithol. 13, 175–269.
62. Greenberg, R., Bichier, P, Angon, A. C., MacVean, C., Perez, R., Cano, E. (2000) Ecology 81, 1750–1755.
63. Jantti, A., Aho, T., Hakkarainen, H., Kuitunen, M. and Suhonen, J. (2001) Oecologia 128, 488–491.
64. Mols, C. M. M. and Visser, M. E. (2002) J. Appl. Ecol. 39, 888–899.
65. Van Bael, S. A., Brawn, J. D. and Robinson, S. K. (2003) Proc. Natl. Acad. Sci. USA 100, 8304–8307.
66. Sipura, M. (1999) Oecologia 121, 537–545.
67. Sanz, J. J. (2001) Ecol. Res. 16, 387–394.
68. Dolbeer, R. A. (1990) Ibis (Lond. 1859) 132, 309–322.
69. Naylor, R. and Ehrlich, P. R. (1997) in Nature's Services, ed. Daily, G. C. (Island Press, Washington, D.C.), pp. 151–174.
70. Mourato, S., Özdemiroğu, E. and Foster, V. (2000) Environ. Sci. Technol. 34, 1456–1461.
71. Quammen, D. (1997) The Song of the Dodo (Touchstone Books, New York).
72. Terborgh, J., Lopez, L., Nunez, P., Rao, Madhu, Shahabuddin, G., Orihuela, G., Riveros, M., Ascanio, R., Adler, G. H., Lambert, T. D., et al.,... (2001) Science 294, 1923–1926.
73. Wootton, J. T. (1995) Ecoscience 2, 321–328.
74. Powell, G.V. N., Fourqurean, J. W., Kenworthy, W. J. andand Zieman, J. C. (1991) Estuar. Coast. Shelf Sci. 32, 567–579.
75. Anderson, W. B. and Polis, G. A. (1999) Oecologia 118, 324–332.
76. Sanchez-Pinero, F. and Polis, G. A. (2000) Ecology 81, 3117–3132.
77. Palomo, G., Iribarne, O. and Martinez, M. M. (1999) Bull. Mar. Sci. 65, 119–128.
78. Hawke, D. J., Holdaway, R. N., Causer, J. E. and Ogden, S. (1999) Aust. J. Soil Res. 37, 103–113.
79. Heine, J. C. and Speir, T. W. (1989) Polar Biol. 10, 89–100.
80. Crawford, R. J. M. and Shelton, P. A. (1978) Biol. Conserv. 14, 85–109.
81. Gilbertson, M., Elliott, J. E. and Peakall, D. B. (1987) in The Value of Birds, eds. Diamond, A. W. andand Filion, F. L. (Int. Council for Bird Pres., Norfolk, VA), pp. 231–248.
82. Oliver, J. D. and Legovic, T. (1988) Ecol. Model. 43, 195–223.
83. Oliver, J. D. and Schoenberg, S. A. (1989) Oikos 55, 175–182.
84. Haynes-Sutton, A. M. (1987) in The Value of Birds, eds. Diamond, A. W. and Filion, F. L. (Int. Council for Bird Pres., Norfolk, VA), pp. 77–81.
85. Williams, R. J., Berlow, E. L., Dunne, J. A., Barabasi, A.-L. and Martinez, N. D. (2002) Proc. Natl. Acad. Sci. USA 99, 12913–12916.
86. Korpimaki, E. and Norrdahl, K. (1991) Oikos 62, 195–208.
87. Ims, R. A. and Andreassen, H. P. (2000) Nature 408, 194–196.
88. Nogales, M., Quilis, V., Medina, F. M., Mora, J. L. and Trigo, L. S. (2002) Biol. J. Linn. Soc. 75, 345–352.
89. Korpimaki, E. and Norrdahl, K. (1998) Ecology 79, 2448–2455.
90. Crooks, K. R. and Soule, M. E. (1999) Nature 400, 563–566.
91. Dunne, J. A., Williams, R. J. and Martinez, N. D. (2002) Ecology Lett. 5, 558–567.
92. Sih, A., Crowley, P., McPeek, M., Petranka, J. and Strohmeier, K. (1985) Annu. Rev. Ecol. Syst. 16, 269–311.
93. Eriksson, M. O. G. (1987) in The Value of Birds, eds. Diamond, A. W. and Filion, F. L. (Int. Council for Bird Pres.,

Norfolk, VA), pp. 183–190.
94. Bryce, S. A., Hughes, R. M. and Kaufmann, P. R. (2002) Environ. Manage. 30, 294–310.
95. Paine, R. T., Wootton, J. T. and Boersma, P. D. (1990) Auk 107, 1–9.
96. Wootton, J. T. (1994) Annu. Rev. Ecol. Syst. 25, 443–466.
97. Milton, S. J., Dean, W. R. J., Kerley, G. I. H., Hoffman, H. T. and Whitford, W. G. (1998) Southwest. Nat. 43, 449–452.
98. Murakami, M. and Nakano, S. (2002) Ecol. Lett. 5, 333–337.
99. Loiselle, B. A. (1990) Oecologia 82, 494–500.
100. Dean, W. R. J., Milton, S. J. and Siegfried, W. R. (1990) Ecology 71, 1299–1306.
101. Izhaki, I. and Safriel, U. N. (1989) Oikos 54, 23–32.
102. Şekercioğlu, Ç. H. (2002) Environ. Cons. 29, 282–289.
103. Bouton, S. N. and Frederick, P. C. (2003) Conserv. Biol. 17, 297–306.
104. Jacquemot, A. and Filion, F. L. (1987) in The Value of Birds, eds. Diamond, A. W. and Filion, F. L. (Int. Council for Bird Pres., Norfolk, VA), pp. 15–21.
105. Bird, J. A., Pettygrove, G. S. and Eadie, J. M. (2000) J. Appl. Ecol. 37, 728–741.
106. Diamond, A. W. (1987) in The Value of Birds, eds. Diamond, A. W. and Filion, F. L. (Int. Council for Bird Pres., Norfolk, VA), pp. 99–109.
107. Filion, F. L. (1987) in The Value of Birds, eds. Diamond, A. W. and Filion, F. L. (Int. Council for Bird Pres., Norfolk, VA), pp. 7–14.
108. Peakall, D. B. and Boyd, H. (1987) in The Value of Birds, eds. Diamond, A. W. and Filion, F. L. (Int. Council for Bird Pres., Norfolk, VA), pp. 113–118.



Species' Distribution Modeling for Conservation Educators and Practitioners

Richard G. Pearson*

* The American Museum of Natural History, New York, NY, U.S.A., email pearson@amnh.org

K. Frey

Table of Contents

Species' Distribution Modeling for Conservation Educators and Practitioners

Richard G. Pearson

Abstract

Models that predict distributions of species by combining known occurrence records with digital layers of environmental variables have much potential for application in conservation. Through using this module, teachers will enable students: to develop species' distribution models; to apply the models across a series of analyses; and to interpret predictions accurately. Part A of the synthesis introduces the modeling approach, outlines key concepts and terminology, and describes questions that may be addressed using the approach. A theoretical framework that is fundamental to ensuring that students understand the uses and limitations of the models is then described. Part B of the synthesis details the main steps in building and testing a distribution model. Types of data that can be used are described and some potential sources of species' occurrence records and environmental layers are listed. The variety of alternative algorithms for developing distribution models is discussed, and software programs available to implement the models are listed. Techniques for assessing the predictive ability of a distribution model are then discussed, and commonly used statistical tests are described. Part C of the synthesis describes three case studies that illustrate applications of the models: 1) Predicting distributions of known and unknown species in Madagascar; 2) Predicting global invasions by plants of South African origin; and 3) Modeling the potential impacts of climate change on species' distributions in Britain and Ireland.

This synthesis document is part of an NCEP (Network of Conservation Educators and Practitioners; http://ncep.amnh.org/) module that also includes a presentation and a practical exercise:

- An introduction to species' distribution modeling: theory and practice (presentation by Richard Pearson)
- Species' distribution modeling using Maxent (practical by Steven Phillips)

This module is targeted at a level suitable for teaching graduate students and conservation professionals.

Learning objectives

Through use of this synthesis, teachers will enable students to:

1. Understand the theoretical underpinnings of species' distribution models
2. Run a distribution model using appropriate data and methods
3. Test the predictive performance of a distribution model
4. Apply distribution models to address a range of conservation questions

Introduction and Theory

Introduction

Predicting species' distributions has become an important component of conservation planning in recent years, and a wide variety of modeling techniques have been developed for this purpose (Guisan and Thuiller, 2005). These models commonly utilize associations between environmental variables and known species' occurrence records to identify environmental conditions within which populations can be maintained. The spatial distribution of environments that are suitable for the species can then be estimated across a study region. This approach has proven valuable for generating biogeographical information that can be applied across a broad range of fields, including conservation biology, ecology, and evolutionary biology. The focus of this synthesis is on conservation-oriented applications, but the methods and theory discussed are also applicable in other fields (see Table 1 for a list of some uses of species' distribution models in conservation biology and other disciplines).

This synthesis aims to provide an overview of the theory and practice of species' distribution modeling. Through use of the synthesis, teachers will enable students to understand the theoretical basis of distribution models, to run models using a variety of approaches, to test the predictive ability of models, and to apply the models to address a range of questions. Part A of the synthesis introduces the modeling approach and describes the usefulness of the models in addressing conservation questions. Part B details the main steps in building a distribution model, including selecting and obtaining suitable data, choosing a modeling algorithm, and statistically assessing predictive performance. Part C of the synthesis provides three case studies that demonstrate uses of species' distribution models.

What is a species' distribution model?

The most common strategy for estimating the actual or potential geographic distribution of a species is to characterize the environmental conditions that are suitable for the species, and to then identify where suitable environments are distributed in space. For example, if we are interested in modeling the distribution of a plant that is known to thrive in wet clay soils, then simply identifying locations with clay soils and high precipitation can generate an estimate of the species' distribution. There are a number of reasons why the species may not actually occupy all suitable sites (e.g. geographic barriers that limit dispersal, competition from other species), which we will discuss later in this synthesis. However, this is the fundamental strategy common to most distribution models.

The environmental conditions that are suitable for a species may be characterized using either a *mechanistic* or a *correlative* approach. Mechanistic models aim to incorporate physiologically limiting mechanisms in a species' tolerance to environmental conditions. For example, Chuine and Beaubien (2001) modeled distributions of North American tree species by estimating responses to environmental variables (including mean daily temperature, daily precipitation, and night length) using mechanistic models of factors including frost injury, phenology, and reproductive success. Such mechanistic models require detailed understanding of the physiological response of species to environmental factors and are therefore difficult to develop for all but the most well understood species.

Correlative models aim to estimate the environmental conditions that are suitable for a species by associating known species' occurrence records with suites of environmental variables that can reasonably be expected to affect the species' physiology and probability of persistence. The central premise of this approach is that the observed distribution of a species provides useful information as to the environmental requirements of that species. For example, we may assume that our plant species of interest favors wet clay soils because it has been observed growing in these soils. The limitations of this approach are discussed later in the synthesis, but it has been demonstrated that this method can yield valuable biogeographical information (e.g., Raxworthy et al., 2003; Bourg et al., 2005). Since spatially explicit occurrence records are available for a large number of species, the vast majority of species' distribution models are correlative. The correlative approach to distribution modeling is the focus of this synthesis.

The principal steps required to build and validate a correlative species' distribution model are outlined in Figure 1. Two types of model input data are needed: 1) known species' occurrence records; and 2) a suite of environmental variables. 'Raw' environmental variables, such as daily precipitation records collected from weather stations, are often processed to generate model inputs that are thought to have a direct physiological role in limiting the ability of the species to survive. For example, Pearson et al. (2002) used a suite of seven climate variables and five soil variables to generate five model input variables, including maximum annual temperature, minimum temperature over a 20-year period, and soil moisture availability. At this stage, care should be taken to ensure that data are checked for errors. For example, simply plotting the spe-

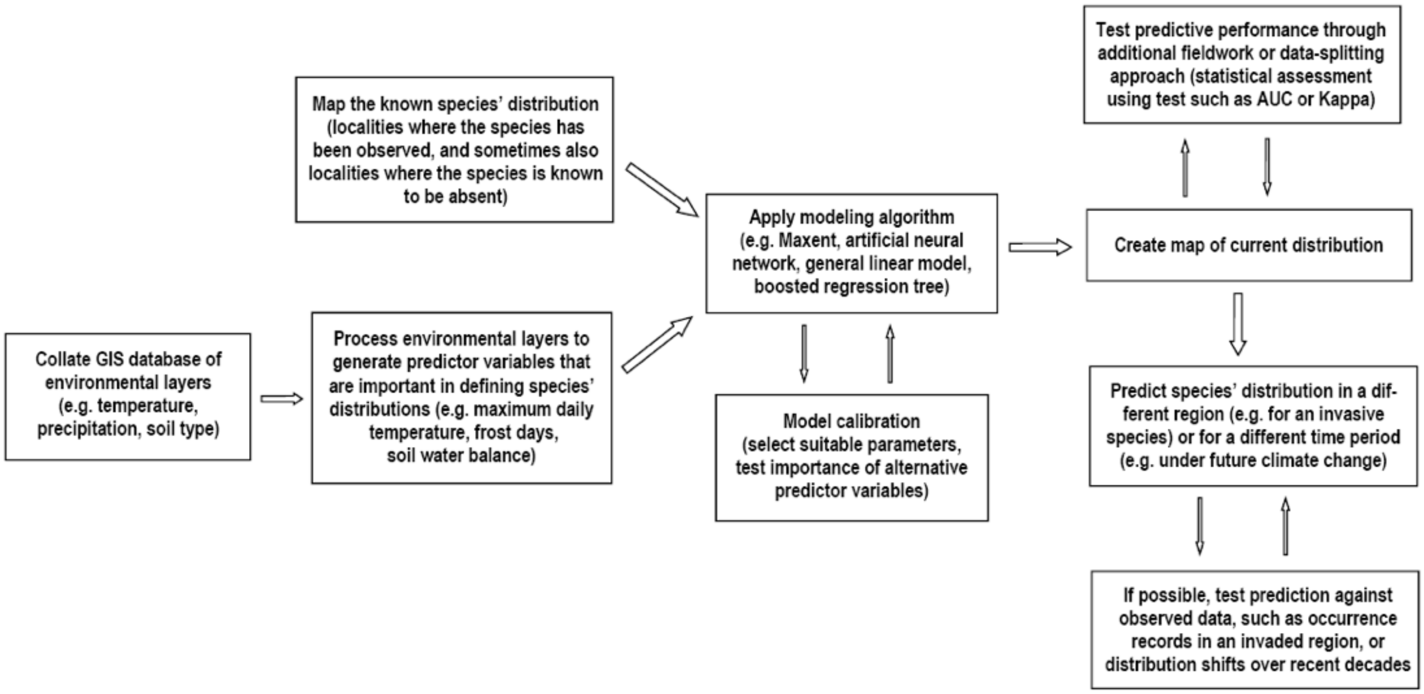


Figure 1. Flow diagram detailing the main steps required for building and validating a correlative species distribution model.

cies' occurrence records in a GIS can help identify records that are distant from other occupied sites and should therefore be checked for accuracy. Data types and sources are discussed in detail in the Section 3: 'Data types and sources'.

The species occurrence records and environmental variables are entered into an algorithm that aims to identify environmental conditions that are associated with species occurrence. If just one or two environmental variables were used, then this task would be relatively straightforward. For example, we may readily discover that our plant species has only been recorded at localities where mean monthly precipitation is above 60mm and soil clay content is above 40%. In practice, we usually seek algorithms that are able to integrate more than two environmental variables, since species are in reality likely to respond to multiple factors. Algorithms that can incorporate interactions among variables are also preferable (Elith et al., 2006). For example, a more accurate description of our plant's requirements may be that it can occur at localities with mean monthly precipitation between 60mm and 70mm if soil clay content is above 60%, and in wetter areas (>70mm) if clay content is as low as 40%.

A number of modeling algorithms that have been applied to this task are reviewed in Section 4. Depending on the method used, various decisions and tests will need to be made at this stage to ensure the algorithm gives optimal results. For example, a suitable 'regularization' parameter will need to be selected if applying the Maxent method (see Phillips et al., 2006 and Box 3), or the degrees of freedom must be selected if running a generalized additive model (see Guisan et al., 2002). The relative importance of alternative environmental predictor variables may also be assessed at this stage so as to select which variables are used in the final model.

Having run the modeling algorithm, a map can be drawn showing the predicted species' distribution. The ability of the model to predict the known species' distribution should be tested at this stage. A set of species occurrence records that have not previously been used in the modeling should be used as independent test data. The ability of the model to predict the independent data is assessed using a suitable test statistic. Different approaches to generating test datasets and alternative statistical tests are discussed in the section: 'Assessing predictive performance'. Since a number of modeling algorithms predict a continuous distribution of environmental suitability (i.e. a prediction between 0 and 1, as opposed to a binary prediction of 'suitable' or 'unsuitable'), it is sometimes useful to convert model output into a prediction of suitable (1) or unsuitable (0). This is a necessary step before applying many test statistics; thus, methods for setting a threshold probability, above which the species is predicted as present, are also outlined in the section: 'Assessing predictive performance'.

Once these steps have been completed, and if model validation is successful, the model can be used to predict species' occurrence in areas where the distribution is unknown. Thus, a set of environmental variables for the area of interest is input into the model and the suitability of conditions at a given locality is predicted. In many cases the model is used to 'fill the gaps' around known occurrences (e.g., Anderson et al., 2002a; Ferrier et al., 2002). In other cases, the model may be used to predict species' distributions in new regions (e.g. to study invasion potential, for review see Peterson, 2003) or for a different time period (e.g. to estimate the potential impacts of future climate change, for review see Pearson and Dawson, 2003). Three examples of the use of predictions from species distribution models are presented in Part C. Ideally, model predictions in different regions or for different time periods should be tested against observed data; for example, Thuiller et al. (2005; see Case Study 2) tested predictions of invasion potential using occurrence records from the invaded distribution, while Araújo et al. (2005a) tested predictions of distribution shifts under climate change using observed records from different decades.

This modeling approach has been variously termed 'species distribution', 'ecological niche', 'environmental niche', 'habitat suitability' and 'bioclimate envelope' modeling. Use of the term 'species distribution modeling' is widespread but it should be noted that the term is somewhat misleading since it is actually the distribution of suitable environments that is being modeled, rather than the species' distribution *per se*. Regardless of the name used, the basic modeling process is essentially the same (see Part B) and the theoretical underpinnings of the models are similar. It is essential that these theoretical underpinnings are properly understood in order to interpret model outputs accurately. The following section describes this theoretical framework.

Section 2. Theoretical Framework

This section outlines some of the fundamental concepts that are crucial for understanding how species' distribution models work, what types of questions they are suitable for addressing, and how model output should be interpreted.

Geographical versus environmental space

We are used to thinking about the occurrence of species in geographical space; that is, the species' distribution as plotted on a map. To understand species' distribution models it is important to also think about species occurring in *environmental space*, which is a conceptual space defined by the environmental variables to which the species responds. The concept of environmental space has its foundations in ecological niche theory. The term 'niche' has a long and varied history of use in ecology (Chase and Leibold, 2003), but the definition proposed by Hutchinson (1957) is most useful in the current context. Hutchinson defined the *fundamental niche* of a species as the set of environmental conditions within which a species can survive and persist. The fundamental niche may be thought of as an '*n*-dimensional hypervolume', every point in which corresponds to a state of the environment that would permit the species to exist indefinitely (Hutchinson, 1957). It is the axes of this n-dimensional hypervolume that define environmental space.

Visualizing a species' distribution in both geographical and environmental space helps us to define some basic concepts that are crucial for species' distribution modeling (Fig. 2). Notice that the observed localities constitute all that is known about the species' *actual distribution*; the species is likely to occur in other areas in which it has not yet been detected (e.g., Fig. 2, area *A*). If the actual distribution is plotted in envi-

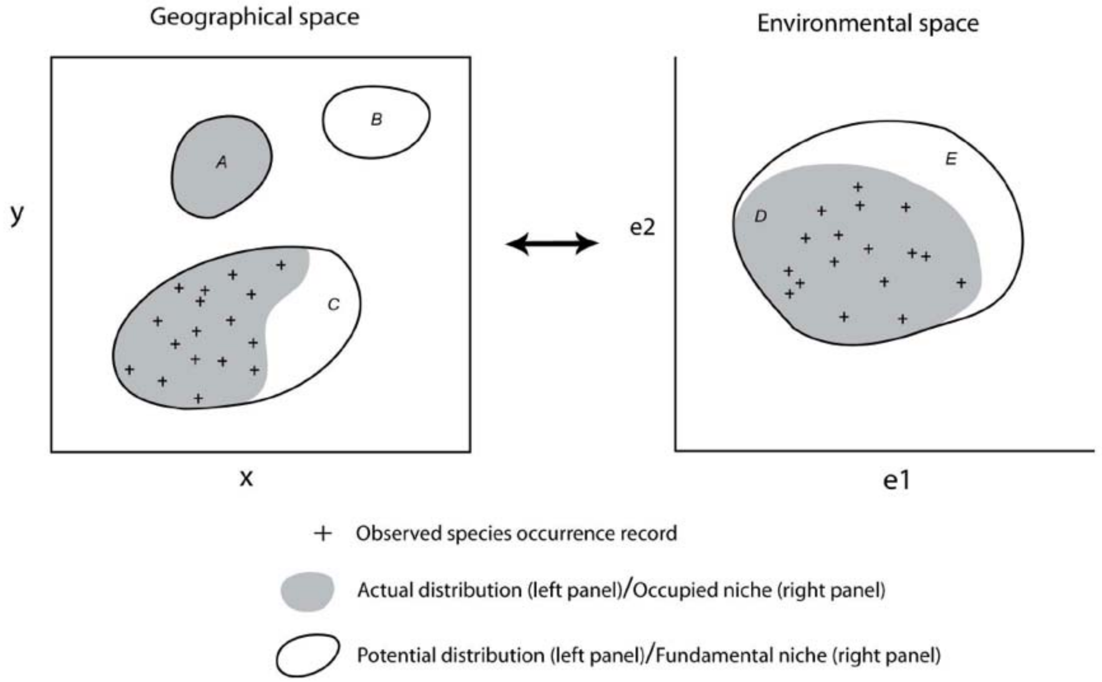


Figure 2. Illustration of the relationship between a hypothetical species' distribution in geographical space and environmental space. Geographical space refers to spatial location as commonly referenced using x and y coordinates. Environmental space refers to Hutchinson's n-dimensional niche, illustrated here for simplicity in only two dimensions (defined by two environmental factors, e_1 and e_2). Crosses represent observed species occurrence records. Grey shading in geographical space represents the species' actual distribution (i.e. those areas that are truly occupied by the species). Notice that some areas of actual distribution may be unknown (e.g. area A is occupied but the species has not been detected there). The grey area in environmental space represents that part of the niche that is occupied by the species: the occupied niche. Again, notice that the observed occurrence records may not identify the full extent of the occupied niche (e.g. the shaded area immediately around label D does not include any known localities). The solid line in environmental space depicts the species' fundamental niche, which represents the full range of abiotic conditions within which the species is viable. In geographical space, the solid lines depict areas with abiotic conditions that fall within the fundamental niche; this is the species' potential distribution. Some regions of the potential distribution may not be inhabited by the species due to biotic interactions or dispersal limitations. For example, area B is environmentally suitable for the species, but is not part of the actual distribution, perhaps because the species has been unable to disperse across unsuitable environments to reach this area. Similarly, the non-shaded area around label C is within the species' potential distribution, but is not inhabited, perhaps due to competition from another species. Thus, the non-shaded area around label *E* identifies those parts of the fundamental niche that are unoccupied, for example due to biotic interactions or geographical constraints on species dispersal.

ronmental space then we identify that part of environmental space that is occupied by the species, which we can define as the *occupied niche*.

The distinction between the occupied niche and the fundamental niche is similar, but not identical, to Hutchinson's (1957) distinction between the realized niche and the fundamental niche. With reference to the case of two species utilizing a common resource, Hutchinson described the realized niche as comprising that portion of the fundamental niche from which a species is not excluded due to biotic competition. The definition of the occupied niche used in this synthesis broadens this concept to include geographical and historical constraints resulting from a species' limited ability to reach or re-occupy all suitable areas, along with biotic interactions of all forms (competition, predation, symbiosis, and parasitism). Thus, the occupied niche reflects all constraints imposed on the actual distribution, including spatial constraints due to limited dispersal ability, and multiple interactions with other organisms.

If the environmental conditions encapsulated within the fundamental niche are plotted in geographical space then we have the *potential distribution*. Notice that some regions of the potential distribution may not be inhabited by the species (Fig. 2, areas B and C), either because the species is: excluded from the area by biotic interactions (e.g., presence of a competitor or absence of a food source); the species has not dispersed into the area (e.g., there is a geographic barrier to dispersal, such as a mountain range, or there has been insufficient time for dispersal); or the species has been extirpated from the area (e.g. due to human modification of the landscape).

Before we go on to discuss how these concepts are used in distribution modeling, it is important to appreciate that the environmental variables used in a distribution model are unlikely to define all possible dimensions of environmental space. Hutchinson (1957) originally proposed that all variables, "both physical and biological" are required to define the fundamental niche. However, the variables available for modeling are likely to represent only a subset of possible environmental factors that influence the distribution of the species. Variables used in modeling most commonly describe the physical environment (e.g. temperature, precipitation, soil type), though aspects of the biological environment are sometimes incorporated (e.g. Araújo and Luoto 2007, Heikkinen et al. 2007). However, the distinction between biotic and abiotic variables is often problematic; for example, land cover type is likely to incorporate both abiotic (e.g. urban) and biotic (e.g. deciduous forest) classes.

Another important factor that we must be aware of is source-sink dynamics, which may cause a species to be observed in unsuitable environments. 'Source-sink' refers to the situation whereby an area (the 'sink') may not provide the necessary environmental conditions to support a viable population, yet may be frequently visited by individuals that have dispersed from a nearby area that does support a viable population (the 'source'). In this situation, species occurrence may be recorded in sink areas that do not represent suitable habitat, meaning that the species is present outside its fundamental niche (Pulliam, 2000). We can logically expect this situation to occur most frequently in species with high dispersal ability, such as birds. In such cases, it is useful to only utilize records for modeling that are known to be from breeding distributions, rather than migrating individuals. Because correlative species distribution models utilize observed species occurrence records to identify suitable habitat, inclusion of occurrence localities from sink populations is problematic. However, it is often assumed that observations from source areas will be much more frequent than observations from sink areas, so this source of potential error is commonly overlooked.

One more thing to be aware of before we move on is that some studies explicitly aim to only investigate one part of the fundamental niche, by using a limited set of predictor variables. For example, it is common when investigating the potential impacts of future climate change to focus only on how climate variables impact species' distributions. A species' niche defined only in terms of climate variables may be termed the


Humpback whales have been recently sighted in unlikely waters around Hong Kong and New York Harbor (Source: Chad King)

climatic niche (Pearson and Dawson, 2003), which represents the climatic conditions that are suitable for species existence. An approximation of the climatic niche may then be mapped in geographical space, giving what is commonly termed the *bioclimate envelope* (Huntley et al., 1995; Pearson and Dawson, 2003).

Estimating niches and distributions

Let us now consider the extent to which species' distribution models can be used to estimate the niche and distribution of a species. We will assume in this section that the model algorithm is excellent at defining the relationship between observed occurrence localities and environmental variables; this will enable us to focus on understanding the ecological assumptions underlying distribution models. The ability of different modeling algorithms to identify the relationship between occurrence localities and environmental variables is discussed in the section: 'Modeling Algorithms'.

Let us first ask what the aim of the modeling is: what element of a species' distribution are we trying to estimate? There are many potential uses of the approach (Table 1) and these require modeling either the actual distribution or the potential distribution. For example, if a model is being used with the purpose of selecting sites that should be given high conservation priority, then modeling the actual distribution will be the aim (since there would be less priority given to conserving sites where the environment is suitable for the species, but the species is not present). In contrast, if the purpose is to identify sites that may be suitable for the reintroduction of an endangered species, then modeling the potential distribution is an appropriate aim. We will now consider the degree to which alternative aims are achievable using the species' distribution modeling approach.

Correlative species' distribution models rely on observed occurrence records for providing information on the niche and distribution of a species. Two key factors are important when considering the degree to which observed species occurrence records can be used to estimate the niche and distribution of a species:

Table 1. Some published uses of species' distribution models in conservation biology (based in part on Guisan and Thuiller, 2005).

Type of use	Example reference(s)
Guiding field surveys to find populations of known species	Bourg et al., 2005; Guisan et al., 2006
Guiding field surveys to accelerate the discovery of unknown species	Raxworthy et al., 2003
Projecting potential impacts of climate change	Iverson and Prasad, 1998; Berry et al., 2002; Hannah et al., 2005; for review see Pearson and Dawson, 2003
Predicting species' invasion	Higgins et al., 1999; Thuiller et al., 2005; for review see Peterson, 2003
Exploring speciation mechanisms	Kozak and Wiens, 2006; Graham et al., 2004b
Supporting conservation prioritization and reserve selection	Araújo and Williams, 2000; Ferrier et al., 2002; Leathwick et al., 2005
Species delimitation	Raxworthy et al., 2007
Assessing the impacts of land cover change on species' distributions	Pearson et al., 2004
Testing ecological theory	Graham et al., 2006; Anderson et al., 2002b
Comparing paleodistributions and phylogeography	Hugall et al., 2002
Guiding reintroduction of endangered species	Pearce and Lindenmayer, 1998
Assessing disease risk	Peterson et al., 2006, 2007

1) *The degree to which the species is at 'equilibrium' with current environmental conditions.*

   A species is said to be at equilibrium with the physical environment if it occurs in all suitable areas, while being absent from all unsuitable areas. The degree of equilibrium depends both on biotic interactions (for example, competitive exclusion from an area) and dispersal ability (organisms with higher dispersal ability are expected to be closer to equilibrium than organisms with lower dispersal ability) (Araújo and Pearson, 2005). When using the concept of 'equilibrium,' we should remember that species distributions change over time, so the term should not be used to imply stasis. However, the concept is useful for us here to help understand that some species are more likely than others to occupy areas that are abiotically suitable.

2) *The extent to which observed occurrence records provide a sample of the environmental space occupied by the species.*
In cases where very few occurrence records are available, perhaps due to limited survey effort (Anderson and Martinez-Meyer, 2004) or low probability of detection (Pearson et al., 2007), the available records are unlikely to provide a sufficient sample to enable the full range of environmental conditions occupied by the species to be identified. In other cases, surveys may provide extensive occurrence records that provide an accurate picture as to the environments inhabited by a species in a particular region (for example, breeding bird distributions in the United Kingdom and Ireland are well known) (Gibbons et al., 1993). It should be noted that there is not necessarily a direct relationship between sampling in geographical space and in environmental space. It is quite possible that poor sampling in geographical space could still result in good sampling in environmental space.

Each of these factors should be carefully considered to ensure appropriate use of a species' distribution model (see Box 1). In reality, species are unlikely to be at equilibrium (as illustrated by area B in Fig. 2, which is environmentally suitable but is not part of the actual distribution) and occurrence records will not completely reflect the range of environments occupied by the species (illustrated by that part of the occupied niche that has not been sampled around label D in Fig. 2). Fig. 3 illustrates how a species' distribution model may be fit under these circumstances. Notice that the model is calibrated (i.e. built) in environmental space and then projected into geographical space. In environmental space, the model identifies neither the occupied niche nor the fundamental niche; instead, the model fits only to that portion of the niche that is represented by the observed records. Similarly, the model identifies only some parts of the actual and potential distributions when projected back into geographical space. Therefore, it should not be expected that species' distribution models are able to predict the full extent of either the actual distribution or the potential distribution.

This observation may be regarded as a failure of the modeling approach (Woodward and Beerling, 1997; Lawton, 2000; Hampe, 2004). However, we can identify three types of model predictions that yield important biogeographical information: species' distribution models may identify 1) the area around the observed occurrence records that is expected to be occupied (Fig. 3, area 1); 2) a part of the actual distribution that is currently unknown (Fig. 3, area 2); and/or 3) part of the potential distribution that is not occupied (Fig. 3, area 3). Prediction types 2 and 3 can prove very useful in a range of applications, as we will see in the following section.

Uses of species' distribution models

Consider modeled area 2 in Fig. 3, which identifies part of the actual distribution for which no occurrence records have been collected. Although the model does not predict the full extent of the actual distribution, additional sampling in the area identified may yield new occurrence records. A number of studies have demonstrated the utility of species' distribution modeling for guiding field surveys toward regions where there is an increased probability of finding new populations of a known species (Fleishman et al., 2002; Bourg et al., 2005; Guisan et al., 2006; also see Case Study 1). Accelerating the discovery of new populations in this way may prove extremely useful for conservation planning, especially in poorly known and highly threatened landscapes.

Consider now predicted area 3 in Fig. 3. Here, the model identifies an area of potential distribution that is environmentally similar to where the species is known to occur, but which is not inhabited. The full extent of the potential distribution is not predicted, but the model can be useful for identifying sites that may be suitable for reintroduction of a species (Pearce and Lindenmayer, 1998) or sites where a species is most likely to become invasive (if it overcomes dispersal barriers and if biotic competition does not prevent establishment) (Peterson,

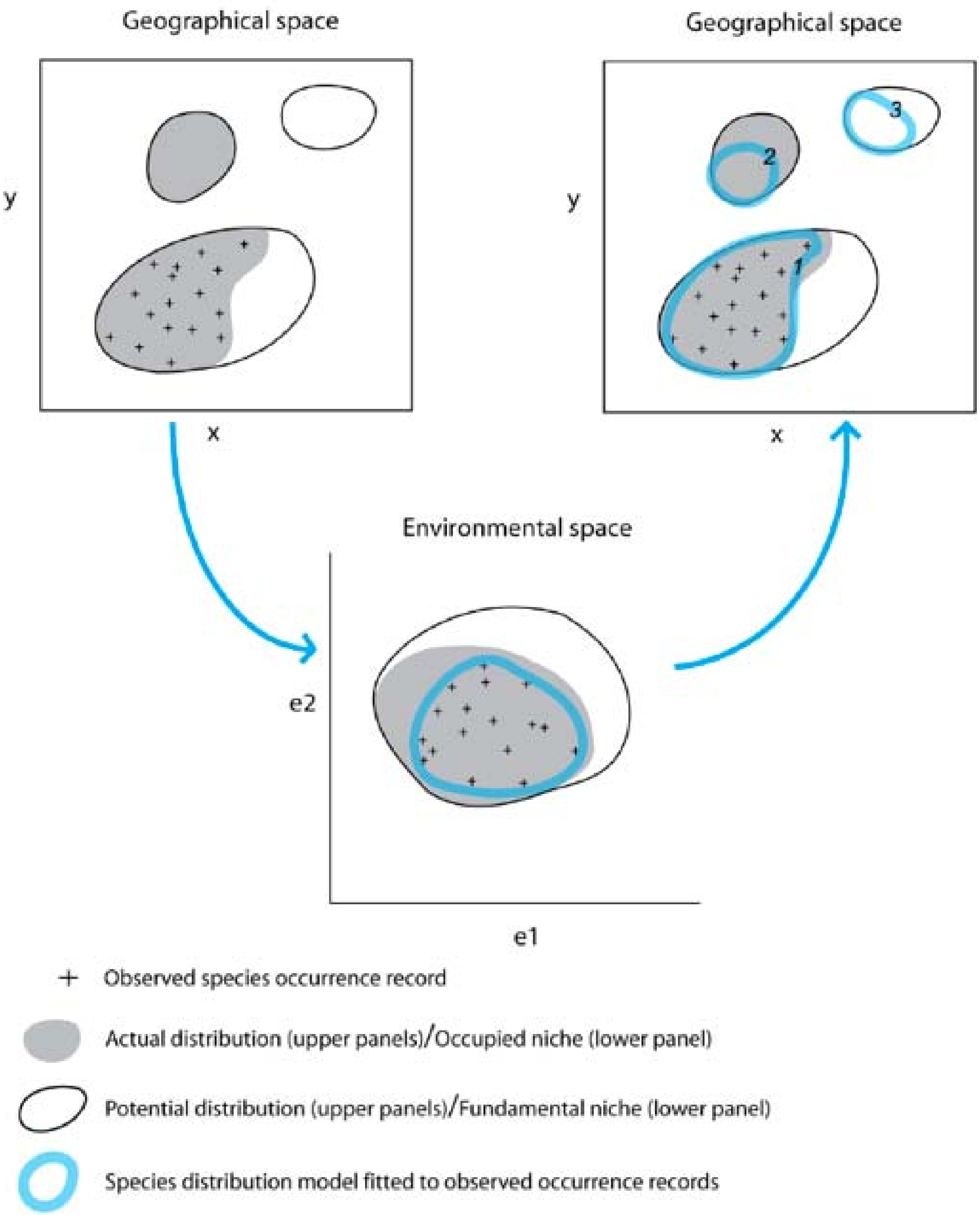


Figure 3. Diagram illustrating how a hypothetical species' distribution model may be fitted to observed species occurrence records (using the same hypothetical case as in Fig. 2). A modelling technique (e.g. GARP, Maxent) is used to characterize the species' niche in environmental space by relating observed occurrence localities to a suite of environmental variables. Notice that, in environmental space, the model may not identify either the species' occupied niche or fundamental niche; rather, the model identifies only that part of the niche defined by the observed records. When projected back into geographical space, the model will identify parts of the actual distribution and potential distribution. For example, the model projection labeled *1* identifies the known distributional area. Projected area *2* identifies part of the actual distribution that is currently unknown; however, a portion of the actual distribution is not predicted because the observed occurrence records do not identify the full extent of the occupied niche (i.e. there is incomplete sampling; see area *D* in Fig. *2*). Similarly, modeled area *3* identifies an area of potential distribution that is not inhabited (the full extent of the potential distribution is not identified because the observed occurrence records do not identify the full extent of the fundamental niche due to, for example, incomplete sampling, biotic interactions, or constraints on species dispersal; see areas *D* and *E* in Fig. 2).

2003). Model predictions of this type also have the potential to accelerate the discovery of previously unknown species that are closely related to the modeled species and that occupy similar environmental space but different geographical space (Raxworthy et al., 2003; see Case Study 1).

Model predictions as illustrated in Fig. 3 therefore have the potential to yield useful information, even though species are not expected to inhabit all suitable locations and sampling may be poor. Additional uses of species' distribution modeling include identifying potential areas for disease outbreaks (Peterson et al., 2006), examining niche evolution (Peterson et al., 1999; Kozak and Wiens, 2006) and informing taxonomy (Raxworthy et al., 2007). However, some potential applications require an estimation of the actual distribution of a species. For example, if a model is being used with the purpose of selecting priority sites for conservation, then an estimate of the actual species' distribution is desired since it would be inefficient to conserve sites where the species is not present (Loiselle et al., 2003). In such cases, it should be remembered that modeled distributions represent environmentally suitable regions but do not necessarily correspond closely with the actual distribution. Additional processing of model output may be required to improve predictions of the actual distribution. For example, predicted areas that are isolated from observed occurrence records by a dispersal barrier may be removed (Peterson et al., 2002) and the influence of competing species may be incorporated (Anderson et al., 2002b).

It is useful to note that mechanistic distribution models (e.g., Chuine and Beaubien, 2001) are subject to the same basic caveat as correlative approaches: the models aim to identify areas with suitable environmental conditions, but do not inform us which areas are actually occupied. Mechanistic models are ideally suited to identifying a species' fundamental niche, and hence its potential distribution. This is because mechanistic approaches model physiological limitations in a species' environmental tolerance, without relying on known occurrence records to define suitable environments. However, the detailed understanding of species' physiology that is required to build mechanistic models prohibits their use in many instances.

The discussion in this section should help clarify the theoretical basis of the species' distribution modeling approach. It is crucial that any application of these models has a sound theoretical basis and that model outputs are interpreted in the context of this framework (see Box 1). It should now be apparent why the terminology used to describe these models is so varied throughout the literature. The terms 'ecological niche model', 'environmental niche model', 'bioclimate envelope model' and 'environmental suitability model' usually refer to attempts to estimate the potential distribution of a species. Use of the term 'species distribution model' implies that the aim is to simulate the actual distribution of the species. Nevertheless, each of these terms refers to the same basic approach, which can be summarized as follows: 1) the study area is modeled as a raster map composed of grid cells at a specified resolution; 2) the dependent variable is the known species' distribution; 3) a suite of environmental variables are collated to characterize each cell; and 4) a function of the environmental variables is generated so as to classify the degree to which each cell is suitable for the species (Hirzel et al., 2002). Part B of this synthesis details the principal steps required to build a distribution model, including selecting and obtaining suitable data, choosing a modeling algorithm, and statistically assessing predictive performance.

Developing a Species' Distribution Model

Data Types and Sources

Correlative species' distribution models require two types of data input: *biological data*, describing the known species' distribution, and *environmental data*, describing the landscape in which the species is found. This section discusses the types of data that are suitable for distribution modeling, and reviews some possible sources of data (see Table 2).

Data used for distribution modeling are usually stored in a Geographical Information System (GIS; see Box 2). The data may be stored either as point localities (termed 'point vector' data; e.g. sites where a species has been observed, or locations of weather stations), as polygons defining an area (termed

Box 1. Caution! On the use and misuse of models

Garbage in, garbage out: This old adage is as relevant to distribution modeling as it is to other fields. Put simply, a model is only as good as the data it contains. Thus, if the occurrence records used to build a correlative species' distribution model do not provide useful information as to the environmental requirements of the species, then the model cannot provide useful output. If you put garbage into the model, you will get garbage out.

Model extrapolation: 'Extrapolation' refers to the use of a model to make predictions for areas with environmental values that are beyond the range of the data used to calibrate (i.e. develop) the model. For example, suppose a distribution model was calibrated using occurrence records that spanned a temperature range of 10–20°C. If the model is used to predict the species' distribution in a different region (or perhaps under a future climate scenario) where the temperature reaches 25°C, then the model is extrapolating. In this case, because the model has no prior information regarding the probability of the species' occurrence at 25°C, the prediction may be extremely uncertain (see Pearson et al., 2006). Model extrapolation should be treated with a great deal of caution.

The lure of complicated technology: Many approaches to modeling species' distributions utilize complex computational technology (e.g. machine learning tools such as artificial neural networks and genetic algorithms) along with huge GIS databases of digital environmental layers. In some cases, these approaches can yield highly successful predictions. However, there is a risk that model users will be swayed by the apparent complexity of the technology: "it is so complicated, it must be correct!" Always remember that a model can only be useful if the theoretical underpinnings on which it is based are sound.

For additional discussion of the limitations of ecological models, see the NCEP module: *Applications of Remote Sensing to Ecological Modeling.*

'polygon vector' data; e.g. areas with different soil types) or as a grid of cells (termed 'raster' data; e.g. land cover types derived from remote sensing [see NCEP module: *Remote Sensing for Conservation Biology*]). For use in a distribution model, it is usual to reformat all environmental data to a raster grid. For example, temperature records from weather stations may be interpolated to give continuous data over a grid (Hijmans et al., 2005). Formatting all data to the same raster grid ensures that environmental data are available for every cell in which biological data have been recorded. These cells, containing both biological data and environmental data, are used to build the species' distribution model. After the model is constructed, its fit to test occurrence records is evaluated (see section: 'Assessing Predictive Performance') and, if the fit is judged to be acceptable, the occurrence of species in cells for which only environmental data are available can be predicted. Note that in some applications the model may be used to predict the species' distribution in a region from which data were not used to build the model (e.g. to predict the spread of an invasive species) or under a future climate scenario. In these cases, environmental data for the new region or climate scenario must also be collated.

A consideration when collating data is the spatial scale at which the model will operate. Spatial scale has two components: *extent* and *resolution* (see NCEP module *Applications of remote sensing to ecological modeling*). Extent refers to the size of the region over which the model is run (e.g. New York state or the whole of North America) whilst resolution refers to the size of grid cells (e.g. 1 km^2 or 10 km^2). Note that it is common for datasets with large extent to have coarse resolution (e.g. data for North America at 10 km^2) and datasets with small extent to have fine resolution (e.g. New York state at 1 km^2). Spatial scale can play an important role in the application of a species' distribution model. In particular, ideally the data resolution should be relevant to the species under consideration: the appropriate data resolution for studying ants is likely to be very different from that for studying elephants.

Biological data

Data describing the known distribution of a species may be obtained in a variety of ways:

1) *Personal collection*: occurrence localities can be obtained during field surveys by an individual or small group of researchers. For example, Fleishman et al. (2001) built models using butterfly occurrence records collected by the researchers during surveys in Nevada, USA.
2) *Large surveys*: distribution information may be available from surveys undertaken by a large number of people. For example, Araújo et al. (2005a) built distribution models using data from *The New Atlas of Breeding Birds in Britain and Ireland*: *1988-1991* (Gibbons et al., 1993), which represents the sampling effort of hundreds of volunteers.
3) *Museum collections:* occurrence localities can be obtained from collections in natural history museums. For example, Raxworthy et al. (2003) utilized occurrence records of chameleons in Madagascar that are held in museum collections.
4) *Online resources:* distributional data from a variety of sources are increasing being made available over the internet (see Table 2). For example, the Global Biodiversity Information Facility (www.gbif.org) is collating available datasets from a diversity of sources and making the information available online via a searchable web portal.

Species distribution data may be either *presence-only* (i.e. records of localities where the species has been observed) or *presence/absence* (i.e. records of presence and absence of the species at sampled localities). Different modeling approaches have been developed to deal with each of these cases (see section on: 'Modeling Algorithms'). In some instances, the inclusion of absence records has been shown to improve model performance (Brotons et al., 2004). However, absence records are often not available and may be unreliable in some cases. In particular, absences may be recorded when the species was not detected even though the environment was suitable. These cases are often referred to as 'false absences' because the model will interpret the record as denoting unsuitable environmental conditions, even though this is not the case. False absences can occur when a species could not be detected although it was present, or when the species was absent but the environment was in fact suitable (e.g. due to dispersal limitation or metapopulation dynamics). Inclusion of false absence records may seriously bias analyses, so absence data should be used with care (Hirzel et al., 2002).

There are a number of additional potential sources of bias and error that should be carefully considered when collating spe-

Box 2. The Role of Geographic Information Systems (GIS)

GIS is a vital tool in species' distribution modeling. The large datasets of biological and environmental data that are used in distribution modeling are ideally suited to being stored, viewed, and formatted in a GIS. For example, useful GIS operations include changing geographic reference systems (it is essential that all data are referenced to a common coordinate system, so occurrence records can be matched with the environmental conditions at the site), reformatting spatial resolution, and interpolating point locality data to a grid. GIS is also crucial for visualizing model results and carrying out additional processing of model output, such as removing predicted areas that are isolated from observed species records by a dispersal barrier (Peterson et al., 2002). However, the distribution modeling itself is usually undertaken outside the GIS framework. With few exceptions (e.g. Ferrier et al., 2002), the distribution model does not 'see' geographical coordinates; instead, the model operates in environmental space (see section: 'Theoretical Framework'). Some GIS platforms now incorporate distribution modeling tools (e.g. DIVA GIS: www.diva-gis.org/, IDRISI: http://www.clarklabs.org/), or have add-in scripts that enable distribution models to be run (e.g. BIOCLIM script for ArcView: http://arcscripts.esri.com/details.asp?dbid=13745), but running the model within a GIS is not necessary.

Table 2. Some example sources of biological and environmental data for use in species' distribution modeling

Type of data	Source
Species' distributions	
– Data for a wide range of organisms in many regions of the world	Global Biodiversity Information Facility (GBIF): www.gbif.org
– Data for a range of organisms, mostly rare or endangered, and primarily in North America	NatureServe: www.NatureServe.org
Climate	
– Interpolated climate surfaces for the globe at 1km resolution	WorldClim:http://www.worldclim.org/
– Scenarios of future climate change for the globe	Intergovernmental Panel on Climate Change (IPCC): http://www.ipcc-data.org/
– Reconstructed palaeoclimates	NOAA: http://www.ncdc.noaa.gov/paleo/paleo.html
Topography	
– Elevation and related variables for the globe at 1km resolution	USGS: http://edc.usgs.gov/products/elevation/gtopo30/hydro/index.html
Remote sensing (satellite)	
– Various land cover datasets	Global Landcover Facility: http://glcf.umiacs.umd.edu/data/
– Various atmospheric and land products from the MODIS instrument	NASA:http://modis.gsfc.nasa.gov/data/
Soils	
– Global soil types	UNEP: http://www.grid.unep.ch/data/data.php?category=lithosphere
Marine	
– Various datasets describing the world's oceans	NOAA: www.nodc.noaa.gov

cies' distribution data. Errors may arise through the incorrect identification of species, or inaccurate spatial referencing of samples. Biases can also be introduced because collectors tend to sample in easily accessible locations, such as along roads and rivers and near towns or biological stations (Graham et al., 2004a). In some cases, biased sampling in geographical space may lead to non-representative sampling of the available environmental conditions, although this is not necessarily the case. When utilizing records from museum collections, it should be remembered that these data were not generally collected with the purpose of determining the distributional limits of a species; rather, sampling for museum collections tends to be biased toward rare and previously unknown species.

Environmental data

A wide range of environmental input variables have been employed in species' distribution modeling. Most common are variables relating to climate (e.g. temperature, precipitation), topography (e.g., elevation, aspect), soil type and land cover type (see Table 2).Variables tend to describe primarily the abiotic environment, although there is potential to include biotic interactions within the modeling. For example, Heikkinen et al. (2007) used the distribution of woodpecker species to predict owl distributions in Finland since woodpeckers excavate cavities in trees that provide nesting sites for owls.

As noted in the previous section: 'Introduction' variables are often processed to generate new variables that are thought to have a direct physiological or behavioral role in determining species' distributions. In general, it is advisable to avoid predictor variables that have an indirect influence on species' distributions, since indirect associations may cause erroneous predictions when models are used to predict the species' distribution in new regions or under alternative climate scenarios (Guisan and Thuiller, 2005). For example, species do not respond directly to elevation, but rather to changes in temperature and air pressure that are affected by elevation. Thus, a species characterized as living at high elevation in a low latitude region may, in fact, be associated with lower elevations in areas with higher latitude, since the regional climate is cooler.

Environmental variables may comprise either *continuous data* (data that can take any value within a certain range, such as temperature or precipitation) or *categorical data* (data that are split into discrete categories, such as land cover type or soil type). Categorical data cannot be used with a number of common modeling algorithms (see section: 'Modeling Algorithms'). In these cases, it may be possible to generate a continuous variable from the categorical data. For example, Pearson et al. (2002) estimated soil water holding capacity from categorical soil data, and used these values within a water balance model to generate continuous predictions of soil moisture surplus and deficit.

Modern technologies, including remote sensing (see NCEP module *Applications of remote sensing to ecological modeling*), the internet, and GIS have greatly facilitated the collection and dissemination of environmental datasets (see Table 2). In addition, global climate models have been used to generate scenarios of future climates and to simulate climatic conditions since the end of the last glacial period (see Table 2). Predicted future climate scenarios can be used to estimate the potential impacts of climate change on biodiversity (e.g. Thomas et al. 2004; see case study 3), whilst simulations of past climates can be used to test the predictive ability of models (e.g. Martinez-Meyer et al., 2004). Given the vast amounts of data that are available, it is especially important to remain critical as to which variables are suitable for inclusion in the model. Some studies have demonstrated good predictive ability using only three variables (e.g., Huntley et al., 1995), whilst other studies have applied methodologies that can incorporate many more variables (e.g., Phillips et al., [2006] utilized 14 environmental variables, although some of these variables are likely to have been rejected by the algorithm applied because they did not provide useful information beyond that which was included in other variables).

Modeling Algorithms

A number of alternative modeling algorithms have been applied to classify the probability of species' presence (and absence) as a function of a set of environmental variables. The task is to identify potentially complex non-linear relationships in multi-dimensional environmental space. In the section: 'Theoretical Frameworks', we assumed that the modeling algorithm is excellent, enabling us to identify three expected types of model prediction, illustrated by modeled areas *1, 2* and *3* in Fig. 3. If we now admit some degree of error in the algorithm's ability to fit the observed records, then a fourth type of prediction will occur: the model will predict as suitable areas that are part of neither the actual nor the potential distribution. The most useful algorithms will limit these erroneous 'type 4' predictions.

Table 3 lists some commonly used approaches for species' distribution modeling. Some methods that have been applied are statistical (e.g. generalized linear models [GLMs] and generalized additive models [GAMs]), whilst other approaches are based on machine-learning techniques (e.g., maximum entropy [Maxent] and artificial neural networks [ANNs]). Published studies have often applied one or more of these algorithms and have given the resulting model a name or acronym (e.g., 'Maxent' refers to an implementation of the maximum entropy method, while 'BIOMOD' is the acronym given to a model that implements a number of methods, including GLMs and GAMs). Often these models have been implemented in user-friendly software that is free and easy to obtain (Table 3).

There are some important differences between among model algorithms that should be carefully considered when selecting which method(s) to apply. One key factor is whether the algorithm requires data on observed species absence (see 'Data Types and Sources'). Some algorithms operate by contrasting sites where the species has been detected with sites where the species has been recorded as absent (e.g., GLMs, GAMs, ANNs). However, reliable absence data often are not available (see: 'Data Types and Sources'), so other methods have been applied that do not require absence data. We can distinguish three types of presence-only methods:

1) Methods that rely solely on presence records (e.g. BIOCLIM, DOMAIN). These methods are truly 'presence-only' since the prediction is made without any reference

Table 3. Some published methods for species' distribution modeling

Method(s)[1]	Model/software name[2]	Species data type	Key reference/URL
Gower Metric	DOMAIN⋆	presence-only	Carpenter et al., 1993 http://www.cifor.cgiar.org/docs/_ref/research_tools/domain/ http://diva-gis.org
Ecological Niche Factor Analysis (ENFA)	BIOMAPPER⋆	presence and background	Hirzel et al., 2002 http://www2.unil.ch/biomapper/
Maximum Entropy	MAXENT⋆	presence and background	Phillips et al., 2006 http://www.cs.princeton.edu/~schapire/maxent/
Genetic algorithm (GA)	GARP[3]⋆	pseudo-absence[4]	Stockwell and Peters, 1999 http://www.lifemapper.org/desktopgarp/
Artificial Neural Network (ANN)	SPECIES	presence and absence (or pseudo-absence)	Pearson et al., 2002
Regression: generalized linear model (GLM), generalized additive model (GAM), boosted regression trees (BRT), multivariate adaptive regression splines (MARS)	Implemented in R[5]	presence and absence (or pseudo-absence)	Lehman et al., 2002 Elith et al., 2006 Leathwick et al., 2006 Elith et al., 2007
Multiple methods	BIOMOD	presence and absence (or pseudo-absence)	Thuiller, 2003
Multiple methods	OpenModeller	depends on method implemented	http://openmodeller.sourceforge.net/

[1] 'Method' refers to a statistical or machine-learning technique. [2]'Model/software name' refers to a name (or acronym) given to a published model that implements the method(s) stated. Software to implement the method for species' distribution modeling is readily available at no cost for those models marked with an asterisk (⋆); other models are available at the discretion of the author(s). [3]The genetic algorithm for rule-set prediction (GARP) includes within its processing multiple methods, including GLM. [4]Note that Pseudo-absence here refers to the sampling approach implemented in the GARP software; in principle, any presence-absence method can be implemented using pseudo absences. [5]R is a freely available (at no cost) software environment for statistical computing and graphics (http://www.r-project.org/). Based in part on Elith et al. (2006) and Guisan and Thuiller (2005).

to other samples from the study area.

2) Methods that use 'background' environmental data for the entire study area (e.g. Maxent, ENFA). These methods focus on how the environment where the species is known to occur relates to the environment across the rest of the study area (the 'background'). An important point is that the occurrence localities are also included as part of the background.
3) Methods that sample 'pseudo-absences' from the study area. In principle, any presence/absence algorithm can be implemented using pseudo-absences. The aim here is to assess differences between the occurrence localities and a set of localities chosen from the study area that are used in place of real absence data. The set of 'pseudo-absences' may be selected randomly (e.g., Stockwell and Peters, 1999) or according to a set of weighting criteria (e.g., Engler et al., 2004; Zaniewski et al., 2002). An important difference between the pseudo-absence approach and the background approach is that pseudo-absence models do not include occurrence localities within the set of pseudo-absences.

Another key difference among model algorithms is their ability to incorporate categorical environmental variables (see the section: Data Types and Sources'). Methods also differ in the form of their output, which is most commonly a continuous prediction (e.g. a probability value ranging from 0 to 1) but may be a binary prediction (with '0' as a prediction of unsuitable environmental conditions or species absence, and '1' a prediction of highly suitable environmental conditions or species presence). To generate a binary prediction from a model that gives continuous output, it is necessary to set a threshold value above which the prediction is classified as 'highly suitable' or 'present' (see the section: 'Assessing Predictive Performance').

A further consideration when selecting a modeling algorithm is whether it is important to determine the relative influence of different input variables on the model's fit or predictive capacity. Some models may have excellent predictive power but do not enable us to easily understand how the algorithm is operating; such models are often termed 'black box' since the model takes input and produces output but the internal workings are somewhat opaque. For example, artificial neural networks have shown good predictive ability (e.g., Pearson et al., 2002; Thuiller, 2003; Segurado and Araújo, 2004), but identifying the relative contribution of each input variable to the prediction is difficult (sensitivity analysis may be used, but this requires additional analyses). In contrast, a GLM builds a regression equation from which the relative contributions of different variables are immediately apparent (Guisan et al., 2002).

It is not possible within the scope of this synthesis to describe

Box 3. Maximum Entropy (Maxent) Modeling of Species Distributions (based on Phillips et al., 2006)

Maxent is a general-purpose method for characterizing probability distributions from incomplete information. In estimating the probability distribution defining a species' distribution across a study area, Maxent formalizes the principle that the estimated distribution must agree with everything that is known (or inferred from the environmental conditions where the species has been observed) but should avoid making any assumptions that are not supported by the data. The approach is thus to find the probability distribution of maximum entropy (the distribution that is most spread-out, or closest to uniform) subject to constraints imposed by the information available regarding the observed distribution of the species and environmental conditions across the study area.

The Maxent method does not require absence data for the species being modeled; instead it uses background environmental data for the entire study area. The method can utilize both continuous and categorical variables

Box 3. Maximum Entropy (Maxent) Modeling of Species Distributions (Continued)

and the output is a continuous prediction (either a raw probability or, more commonly, a cumulative probability ranging from 0 to 100 that indicates relative suitability). Maxent has been shown to perform well in comparison with alternative methods (Elith et al., 2006; Phillips et al., 2006; Pearson et al., 2007). One drawback of the Maxent approach is that it uses an exponential model that can predict high suitability for environmental conditions that are outside the range present in the study area (i.e. extrapolation, see Box 1). To alleviate this problem, when predicting for variable values that are outside the range found in the study area, these values are reset (or 'clamped') to match the upper or lower values found in the study area.

For a concise mathematic definition of Maxent and for more detailed discussion of its application to species distribution modeling see Phillips et al. (2004, 2006). These authors have developed software with a user-friendly interface to implement the Maxent method for modeling species distributions (for free download see web link in Table 3). The software also calculates a number of alternative thresholds (see Section 5), computes model validation statistics (see section: 'Assessing Predictive Performance'), and enables the user to run a jackknife procedure to determine which environmental variables contribute most to the model prediction (see the practical exercise by Steven Phillips that accompanies this synthesis).

the theory, advantages and disadvantages of a large number of modeling algorithms; the reader is referred to the literature cited in Table 3. However, see Box 3 and the practical exercise by Steven Phillips that accompanies this Synthesis for a more detailed description of one method, Maxent.

The model algorithm is in some ways the 'core' of the distribution model, but it should be remembered that the algorithm is just one part of the broader modeling process; other factors, including selection of environmental variables (see section: 'Data Types and Sources') and application of a decision threshold (see section: 'Assessing Predictive Performance'), are key elements of the modeling process that affect model results and may be varied regardless of the model algorithm being used. Nevertheless, studies comparing different modeling algorithms have demonstrated substantial differences between predictions from alternative methods. The importance of selecting an appropriate algorithm is discussed below.

Differences between methods and selection of 'best' models

Given the variety of possible modeling methods (Table 3), it is important to consider the degree to which different methods yield different results. Furthermore, if model predictions differ substantially, how should we choose which method to apply? This is an active area of research, and unfortunately there are no simple answers.

A number of studies have demonstrated that different modeling approaches have the potential to yield substantially different predictions (e.g., Loiselle et al., 2003; Thuiller, 2003; Brotons et al., 2004; Segurado and Araújo, 2004; Thuiller et al., 2004; Elith et al., 2006; Pearson et al., 2006). Pearson et al. (2006) found especially large differences among predictions of changes in range size under future climate change scenarios based on nine alternative modeling methods. Predicted changes in range size differed in both magnitude and direction (e.g. from 92% range reduction to 322% range increase for a single species). In another study, Loiselle et al. (2003) demonstrated markedly different results when alternative distribution models were used alongside a reserve selection algorithm for identifying priority sites for conservation.

The most comprehensive model comparison to date was provided by Elith et al. (2006). The authors compared sixteen modeling methods using 226 species across six regions of the world. All of the models included in the study were implemented using presence-only data for calibration (some methods required the use of pseudo-absence data), but model performance was assessed using data on both presence and absence. These analyses found differences between predictions from alternative methods, but also found that some methods consistently outperformed others. In general, models classified as 'best' were those that were able to identify complex relationships that existed in the data, including interactions among environmental variables.

Several additional factors that lead to differences among predictions from alternative algorithms have been identified. These include (1) whether the model uses presence-absence or presence-only data (Brotons et al., 2004; Pearson et al., 2006), (2) if the model does not use absence data, whether the model uses solely presence records, 'background' data, or 'pseudo-absences' (Elith et al., 2006), (3) whether the algorithm is parametric or non-parametric (Segurado and Araújo, 2004), and (4) how the model 'extrapolates' beyond the range of data used for its calibration (Pearson et al., 2006; and see Box 1).

In view of these differences among models, selection of an appropriate algorithm is both difficult and crucial. Identifying models that are generically 'best' is problematic since the approach used to assess predictive performance depends on the aim of the modeling. For example, Elith et al. (2006) assessed the ability of models to simulate *actual* distributions by using statistical tests that reward models for correctly classifying both presences and absences (see Section 5). In contrast, Pearson et al. (2007) assessed predictive performance based only on the model's ability to predict observed presences, arguing that the purpose of the modeling was to identify *potential* distributions (in which case use of absence data in assessing performance is invalid since a site classified as absent still may be environmentally suitable). The relative merits of aiming to predict actual versus potential distributions were discussed in Section 2. We will also return to the question of how to identify 'best' models when describing various statistical approaches for assessing predictive performance in Section 5. However, the important point is that it is not straightforward to identify which methods are best, and it is therefore not possible to recommend use of one method over another.

In practice, model selection will be influenced by factors including whether observed absence data are available, whether data on some of the environmental variables are categorical, and whether it is important to evaluate the influence of different variables on the model prediction. We recommend that modeling efforts apply and examine predictions from a range of methods in order to quantify uncertainty arising from the choice of method and to identify when different models are in agreement. Perhaps most importantly, it is vital that the assumptions and behavior of any model are properly understood (e.g. how does the model deal with environmental conditions that are beyond the range of the data used to calibrate the model?) so that model output can be accurately interpreted.

Assessing Predictive Performance

Assessing the accuracy of a model's predictions is commonly termed 'validation' or 'evaluation', and is a vital step in model development. Application of the model will have little merit if we have not assessed the accuracy of its predictions. Validation thus enables us to determine the suitability of a model for a specific application and to compare different modeling methods (Pearce and Ferrier, 2000). This section discusses different approaches for assessing predictive performance, including strategies for obtaining data against which the predictions can be compared, methods for selecting thresholds of occurrence, and various test statistics. As in previous sections, there is no single approach that can be recommended for use in all modeling exercises; rather, the choice of validation strategy will be influenced by the aim of the modeling effort, the types of data available, and the modeling method used.

Strategies for obtaining test data

In order to test predictive performance it is necessary to have data against which the model predictions can be compared. We can refer to these as *test data* (sometimes called *evaluation data*) to distinguish them from the *calibration data* (sometimes called *training data*) that are used to build the model. It is fairly common for studies to assess predictive performance by simply testing the ability of the model to predict the calibration data (i.e. calibration and test datasets are identical). However, this approach makes it difficult to identify models that have over-fit the calibration data (meaning the model is able to accurately classify the calibration data, but the model performs poorly when predicting test data), making it impossible for users to judge how well the model may perform when making predictions (Araújo et al., 2005a). It is therefore preferable to use test data that are different from the calibration data.

Ideally, test data would be collected independently from the data used to calibrate the model. For example, Fleishman et al. (2002) modeled the occurrence of butterfly species in Nevada, USA, using species inventory data collected during the period 1996-1999, and then tested the models using data collected from new sites during 2000-2001 (see Case Study 1 for a comparable study). Other researchers have undertaken validation using independent data from different regions (e.g. Beerling et al., 1995; Peterson, 2003), data at different spatial resolution (e.g., Pearson et al., 2004; Araújo et al., 2005b), data from different time periods (e.g., Araújo et al., 2005a), and data from surveys conducted by other researchers (Elith et al., 2006).

However, in practice it may not be possible to obtain independent test data and it is, therefore, common to partition the available data into calibration and test datasets. Several strategies are available for partitioning data, the simplest being a one-time split in which the available data are assigned to calibration and test datasets either randomly (e.g., Pearson et al., 2002) or by dividing the data spatially (e.g., Peterson and Shaw, 2003). The relative proportions of data included in each data set are somewhat arbitrary, and dependent on the total number of locality points available (though using 70% for calibration and 30% for testing is common, following guidelines provided by Huberty (1994)). An alternative to a one-time split is 'bootstrapping', whereby the data are split multiple times. Bootstrapping methods sample the original set of data randomly with replacement (i.e. the same occurrence record could be included in the test data more than once). Multiple models are thus built, and in each case predictive performance is assessed against the corresponding test data. Validation statistics can then be reported as the mean and range from the set of bootstrap samples (e.g., Verbyla and Litaitis, 1989; Buckland and Elston, 1993). An approach similar to bootstrapping, but sampling without replacement (i.e. the same occurrence record cannot be included in the test data more than once), can also be applied and may be termed 'randomization' (Fielding and Bell, 1997).

Another useful data partitioning method is k-fold partitioning. Here, data are split into k parts of roughly equal size ($k > 2$) and each part is used as a test set with the other k-1 sets used for model calibration. Thus, if we select $k = 5$ then five models will be calibrated and each model tested against the excluded test data. Validation statistics are then reported as the mean and range from the set of k tests (Fielding and Bell, 1997). An extreme form of k-fold partitioning, with k equal to the number of occurrence localities, is recommended for use with very low sample sizes (e.g., < 20; Pearson et al., 2007). This method is termed 'jackknifing' or 'leave-one-out' since each occurrence locality is excluded from model calibration during one partition.

The following sub-sections describe validation statistics that can be calculated after test data have been obtained using one of the above approaches.

The presence/absence confusion matrix

If a model is used to predict a set of test data, predictive performance can be summarized in a *confusion matrix*. Note that binary model predictions (i.e. predictions of suitable and unsuitable, rather than probabilities; see section 4) are required in order to complete the confusion matrix. Later subsections describe methods for converting continuous model outputs into binary predictions (see *Selecting thresholds of occurrence*) and for assessing predictive performance using continuous predictions (see *Threshold-independent assessment*). However, in order to understand these later sections it is important to first look at the confusion matrix.

The confusion matrix is rather more straightforward than its name suggests, and is alternatively termed an 'error matrix' or a 'contingency table'. The confusion matrix records the fre-

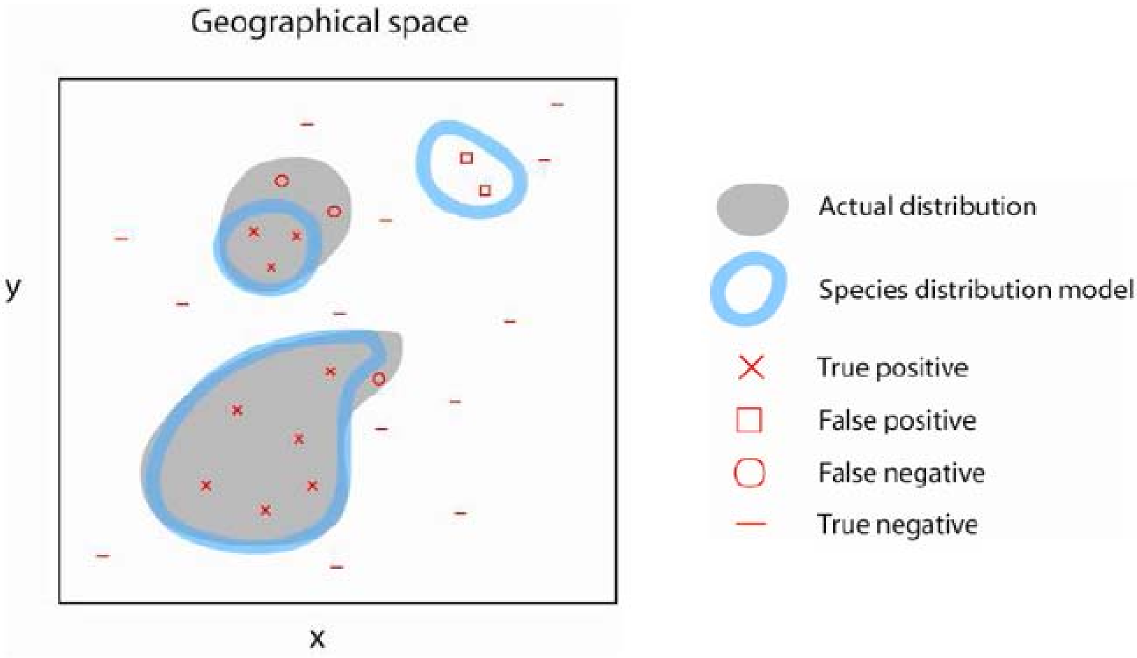


Figure 4. Diagram illustrating the four types of outcomes that are possible when assessing the predictive performance of a species distribution model: true positive, false positive, false negative and true negative. The diagram uses the same hypothetical actual and modeled distributions as in Figure 3. Each instance of a symbol (x, □, o, –) on the map depicts a site that has been surveyed and presence or absence of the species recorded (it is assumed here that if a site falls within the actual distribution then the species will be detected). These survey records constitute the test data. Frequencies of each type of outcome are commonly entered into a confusion matrix (see main text).

quencies of each of the four possible types of prediction from analysis of test data: (a) true positive (the model predicts that the species is present and test data confirms this to be true), (b) false positive (the model predicts presence but test data show absence), (c) false negative (the model predicts absence but test data show presence), (d) true negative (the model predicts and the test data show absence). Frequencies are commonly recorded in a confusion matrix with the following form:

	recorded present	recorded absent
predicted present	*a (true positive)*	*b (false positive)*
predicted absent	*c (false negative)*	*d (true negative)*

Each element of the confusion matrix can be visualized in geographical space as illustrated in Figure 4. In the example depicted, 27 test localities have been sampled and presence or absence of the species recorded at each site. The use of test data comprising only presence localities (i.e. species occurrence records) is discussed below, but for completion of the confusion matrix we require both presence and absence records. Thus, the hypothetical case shown in Figure 4 would yield the following confusion matrix:

	recorded present	recorded absent
predicted present	*9*	*2*
predicted absent	*3*	*13*

The frequencies in the confusion matrix form the basis for a variety of different statistical tests that can be used to assess model performance. Most of the commonly used tests are described below. Terminology related to model performance often varies from study to study and is sometimes not intuitive. In particular, false negative predictions are commonly termed errors of 'omission', whilst false positive predictions are termed errors of 'commission'.

Test statistics derived from the confusion matrix

A simple measure of predictive performance that can be derived from the confusion matrix is the proportion (or percentage) of test localities that are correctly predicted, calculated as

$$(a+d)/(a+b+c+d)$$

This measure may be termed 'accuracy' or 'correct classification rate'. The concept of accuracy is simple and logical, but it is possible to obtain high accuracy using a poor model when a species' prevalence (the proportion of sampled sites in which the species is recorded present) is relatively high or low. For example, if prevalence is 5% then 95% of test localities can be correctly classified simply by predicting all sites as 'absent'. To circumvent this problem, Cohen (1960) introduced a measure of accuracy that is adjusted to account for chance agreement between predicted and observed values. The statistic, Kappa (k), is similar to accuracy but the proportion of correct predictions expected by chance is taken into account (for full derivation see Monserud and Leemans, 1992). Kappa is calculated as:

$$\frac{[(a+d)-(((a+c)(a+b)+(b+d)(c+d))/n)]}{[n-(((a+c)(a+b)+(b+d)(c+d)/n]}$$

Accuracy and Kappa statistics use all values in the confusion matrix and therefore require both presence and absence data. However, absence data are often unavailable (e.g. when using specimens from museum collections) and are inappropriate for use when the aim is to estimate the *potential distribution* (since the environment may be suitable even though the species is absent).

When only presence records are used, the proportion of observed occurrences correctly predicted can be calculated as:

$$a/(a+c)$$

This measure is sometimes termed 'sensitivity' or 'true positive fraction'. Alternatively, we may calculate:

$$c/(a+c)$$

which is often termed 'omission rate' or 'false positive fraction'. Note that these two measures – sensitivity and omission rate – sum to 1. Thus, high sensitivity means low omission, and low sensitivity means high omission. Although sensitivity and omission rate avoid the use of absence records, a serious disadvantage of these tests is that it is possible to achieve very high sensitivity (and low omission) simply by predicting that the species is present at an excessively large proportion of the study area. In short, it is possible to cheat: if the model predicts the entire study area to be suitable, then sensitivity will equal 1, and omission rate will be zero. To avoid this problem, it is necessary to test the statistical significance of a sensitivity or omission rate score.

To test for statistical significance, we ask whether the accuracy of our predictions is greater than would be expected by chance. Imagine, for example, that we are blindfolded and asked to throw darts at a map of our study area. The sites identified by our random throws are then used as species' occurrence localities for model testing. The probability of landing darts in the area predicted by the model to be suitable for the species is equal to the proportion of the study area that is predicted as suitable. Thus, if the model predicts that the species will be present in 40% of the study area, then our probability of successfully landing a dart in the area predicted as suitable is 0.4.

We can apply the same logic to assess the statistical significance of a sensitivity (or omission rate) score. In this case, we use an exact one-tailed binomial test (or for larger sample sizes a chi-square test; for description of binomial and chi-square tests see Zar, 1996) to calculate the probability of obtaining a sensitivity result by chance alone (Anderson et al., 2002a). For example, suppose that the model in Figure 4 predicts that 30% of the study area is suitable for the species. The probability of success by chance alone for each test locality is therefore

0.3. We can calculate from the confusion matrix that the sensitivity = 9/(9+3) = 0.75, and we can use an exact one-tailed binomial test to calculate that the probability of making nine or more successful predictions of presence by chance alone is 0.0017. We may therefore conclude that our result is statistically significant ($p < 0.01$).

A similar assessment of predictive performance can be conducted when only very few occurrence localities are available and test data have been generated using a jackknifing approach (see subsection *Strategies for obtaining test data*). In this case the number of successful predictions from a set of jackknife trials can be calculated (e.g., 9 successes in 12 jackknife trials) and a *p*-value can be calculated using the method presented in Pearson et al. (2007).

In practice, binomial and chi-square tests can be performed in most standard statistical packages, whilst the jackknife *p*-value can be calculated using software provided as Supplementary Material to the Pearson et al. (2007) paper. Various test statistics and thresholds (including those discussed in the remainder of this section) are sometimes calculated automatically by software designed for species distribution modeling (see Table 3), and test statistics may also be calculated using more general applications such as DIVA-GIS (for free download see http://www.diva-gis.org/).

Another statistic that can be derived from the confusion matrix is the proportion of observed absences that are correctly predicted, calculated as:

$$d/(b+d)$$

This statistic is commonly termed 'specificity' or 'true negative fraction'. Specificity is rarely used as a test statistic on its own, since specificity focuses solely on observed absence records. However, specificity is an important measure used in setting decision thresholds and in ROC analysis, which are described in the following two subsections.

Selecting thresholds of occurrence

Binary predictions of 'present' or 'absent' are necessary to test model performance using statistics derived from the confusion matrix. It is therefore often useful to convert continuous model output into binary predictions by setting a threshold probability value above which the species is predicted to be present. Although an alternative test statistic that does not require a threshold is available (Area Under the ROC curve; see the following subsection), this approach is not suitable in many circumstances (notably when absence records are not available, Boyce et al., 2002, but see Phillips et al., 2006). Furthermore, it is essential to learn the techniques described in this subsection to understand how the AUC test operates.

A number of different methods have been employed for selecting thresholds of occurrence (Table 4). Perhaps the simplest approach is to use an arbitrary value, but this method is subjective and lacks ecological reasoning (Liu et al., 2005). Other methods use criteria that are based on the data used to calibrate the model. One approach is to use the lowest predicted value of environmental suitability, or probability of presence, across the set of sites at which a species has been detected. This method assumes that species presence is restricted to locations equally or more suitable than those at which the species has been observed. The approach therefore identifies the minimum area in which the species occurs whilst ensuring that no localities at which the species has been observed are omitted (i.e. omission rate = 0, and sensitivity = 1). An alternative approach is to set the threshold to allow a certain amount of omission (e.g. 5%), which is analogous to setting a fixed sensitivity (e.g. 0.95). This method is less sensitive than the lowest predicted value method to 'outliers' (i.e. locations in which the species is detected despite a low predicted probability of occurrence or suitability), but errors of omission are imposed (i.e. some observed localities will be omitted from the prediction).

Many methods for setting thresholds can be implemented by calculating statistics derived from the confusion matrix across the range of possible thresholds. For example, sensitivity and specificity may be calculated at thresholds increasing in incre-

Table 4. Some published methods for setting thresholds of occurrence

Method	Definition	Species data type[1]	Reference(s)
Fixed value	An arbitrary fixed value (e.g. probability = 0.5)	presence-only	Manel et al., 1999 Robertson et al., 2001
Lowest predicted value	The lowest predicted value corresponding with an observed occurrence record	presence-only	Pearson et al., 2006 Phillips et al., 2006
Fixed sensitivity	The threshold at which an arbitrary fixed sensitivity is reached (e.g. 0.95, meaning that 95% of observed localities will be included in the prediction)	presence-only	Pearson et al., 2004
Sensitivity-specificity equality	The threshold at which sensitivity and specificity are equal	presence and absence	Pearson et al., 2004
Sensitivity-specificity sum maximization	The sum of sensitivity and specificity is maximized	presence and absence	Manel et al., 2001
Maximize Kappa	The threshold at which Cohen's Kappa statistic is maximized	presence and absence	Huntley et al., 1995 Elith et al., 2006
Average probability/suitability	The mean value across model output	presence-only	Cramer, 2003
Equal prevalence	Species' prevalence (the proportion of presences relative to the number of sites) is maintained the same in the prediction as in the calibration data.	presence and absence	Cramer, 2003

[1] Species occurrence records required to set the threshold. Based in part on Liu et al. (2005).

ments of 0.01 from 0 to 1 (i.e. 0, 0.01, 0.02, 0.03…0.99, 1). As the threshold increases, the proportion of the study area predicted to be suitable for the species, or in which the species is predicted to be 'present,' will decrease. Consequently, the proportion of observed presences that are correctly predicted decreases (i.e. decreasing sensitivity) and the proportion of observed absences that are correctly predicted increases (i.e. increasing specificity) (Figure 5A). From these data we can select the threshold at which sensitivity and specificity are equal (labeled *a* in Figure 5A) or at which their sum is maximized. Similarly, it is common to calculate Kappa across the range of possible thresholds and to select the threshold at which the statistic is maximized (Figure 5B).

Choice of an appropriate decision threshold is dependent on the type of data that are available and the question that is being addressed. Some methods require presence and absence records, while others require presence-only records (Table 4). When using both presence and absence records, the general approach is to balance the number of observed presences and absences that are correctly predicted; in effect, to maximize agreement between observed and predicted distributions. Thus, we must be willing to increase the omission rate (i.e. decrease sensitivity) in order to increase the proportion of observed absences that are correctly predicted (i.e. increase specificity). Liu et al. (2005) tested twelve methods for setting thresholds using presence and absence data for two European plant species. Based on four assessments of predictive perfor-

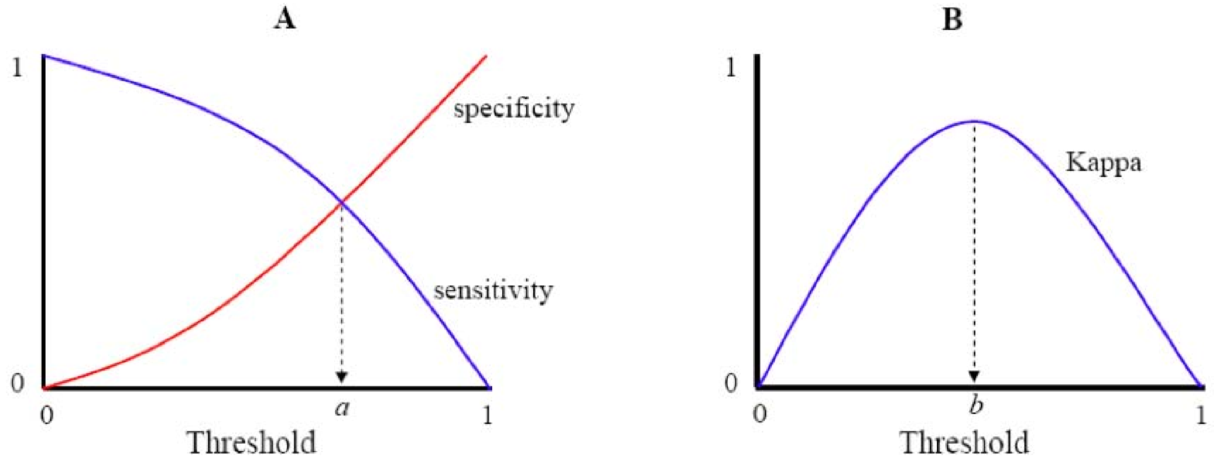


Figure 5. Plots showing changes in test statistics as the threshold of occurrence is adjusted. **A** Decrease in sensitivity and increase in specificity as the threshold is increased. The threshold labeled *a* corresponds to the specificity-sensitivity equality threshold. **B** Changes in the Kappa statistic as the threshold is adjusted. The threshold labeled *b* corresponds to that threshold at which Kappa is maximized.

mance (sensitivity, specificity, accuracy and Kappa), they concluded that the best methods for setting thresholds included maximizing the sum of sensitivity and specificity, using the average probability/suitability score, and setting equal prevalence between the calibration data and the prediction (Table 4). Maximizing Kappa did not perform well, and use of an arbitrary fixed value performed worst.

Methods that use only presence records for setting a threshold are required for cases in which absence data are unavailable. Presence-only methods can also be justified on the grounds that they avoid false absences (Section 3): it may be argued that we should be primarily concerned with maximizing the number of observed presences that are correctly predicted, rather than minimizing the number of absences that are incorrectly predicted as presences (since some absences may be recorded in apparently suitable environments; Pearson et al., 2006). For example, Pearson et al. (2007) used a dataset comprising very few presence-only records for geckos in Madagascar. Because confidence was high that the localities and species identification were correct, and because these species are not highly mobile and are therefore unlikely to be found in unsuitable habitat (i.e. sink habitat; see Section 2), omission of any occurrence record was considered a clear model error. Therefore, the minimum predicted value corresponding to an observed presence was selected as a threshold to ensure zero omission. Distribution models thus predicted that many regions of the study area were suitable although no presences had been detected there. This approach suited the aim of the study, which was to prioritize regions for future surveys by estimating the potential distribution (see Case Study 1).

As a final illustration of the importance of selecting an appropriate decision threshold, we can return to an example raised in Section 2. If the purpose of modeling is to identify areas within which disturbance may impact a species negatively (e.g. as part of an environmental impact assessment), then the threshold may be set low to identify a larger area of potentially suitable habitat. In contrast, if the model was intended to identify potential introduction or reintroduction sites for an endangered species or species of recreational value, then it would be appropriate to choose a relatively high threshold. Choosing a high threshold reduces the risk of choosing unsuitable sites by identifying those areas with highest suitability (Pearce and Ferrier, 2000).

Threshold-independent assessment

When model output is continuous, assessment of predictive

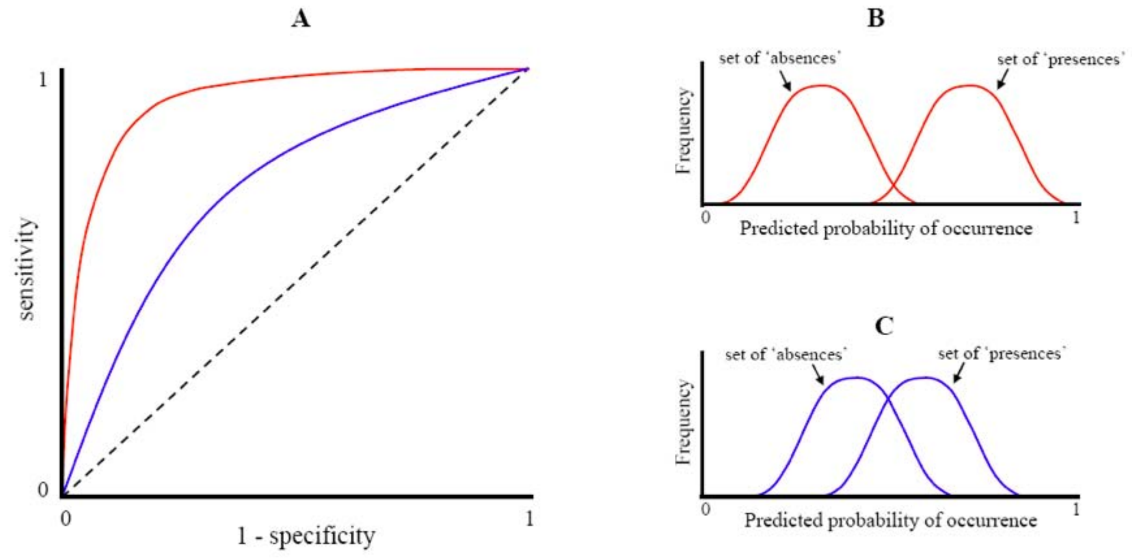


Figure 6. Example Receiver Operating Characteristic (ROC) Curves and illustrative frequency distributions. **A** ROC curves formed by plotting sensitivity against '1 – specificity'. Two ROC curves are shown, the upper curve (red) signifying superior predictive ability. The dashed 1:1 line signifies random predictive ability, whereby there is no ability to distinguish occupied and unoccupied sites. **B** and **C** show example frequency distributions of probabilities predicted by a model for observed 'presences' and 'absences'. The results shown in **B** reveal good ability to distinguish presence from absence, whilst results in **C** show more overlap between the frequency distributions thus revealing poorer classification ability. The case shown in **B** would produce an ROC curve similar to the upper (red) curve in **A**. The case shown in **C** would give an ROC curve more like the lower (blue) curve in **A**.

performance using statistics derived from the confusion matrix will be sensitive to the method used to select a threshold for creating a binary prediction. Furthermore, if predictions are binary, the assessment of performance does not take into account all of the information provided by the model (Fielding and Bell, 1997). Therefore, it is often useful to derive a test statistic that provides a single measure of predictive performance across the full range of possible thresholds. This can be achieved using a statistic known as AUC: the Area Under the Receiver Operating Characteristic Curve.

The AUC test is derived from the Receiver Operating Characteristic (ROC) Curve. The ROC curve is defined by plotting sensitivity against '1 – specificity' across the range of possible thresholds (Figure 6A). Sensitivity and specificity are used because these two measures take into account all four elements of the confusion matrix (true and false presences and absences). It is conventional to subtract specificity from 1 (i.e. 1 – specificity) so that both sensitivity and specificity vary in the same direction when the decision threshold is adjusted (Pearce and Ferrier, 2000). The ROC curve thus describes the relationship between the proportion of observed presences correctly predicted (sensitivity) and the proportion of observed absences incorrectly predicted (1 – specificity). Therefore, a model that predicts perfectly will generate an ROC curve that follows the left axis and top of the plot, whilst a model with predictions that are no better than random (i.e. is unable to classify accurately sites at which the species is present and absent) will generate a ROC curve that follows the 1:1 line (Figure 6A).

In order to summarize predictive performance across the full range of thresholds we can measure the area under the ROC curve (the AUC), expressed as a proportion of the total area of the square defined by the axes (Swets, 1988). The AUC thus ranges from 0.5 for models that are no better than random to 1.0 for models with perfect predictive ability. We can think of AUC in terms of the frequency distributions of probabilities predicted for locations at which we have empirical data on presence and absence (Figure 6, B and C). A high AUC score reflects that the model can discriminate accurately between locations at which the species is present or absent. In fact, AUC can be interpreted as the probability that a model will correctly distinguish between a presence record and an absence record if each record is selected randomly from the set of presences and absences. Thus, an AUC value of 0.8 means the probability is 0.8 that a record selected at random from the set of presences will have a predicted value greater than a record selected at random from the set of absences (Fielding and Bell, 1997; Pearce and Ferrier, 2000).

AUC is a test that uses both presence and absence records. However, Phillips et al. (2006) have demonstrated how the test can be applied using randomly selected 'pseudo-absence' records in lieu of observed absences. In this case, AUC tests whether the model classifies presence more accurately than a random prediction, rather than whether the model is able to accurately distinguish presence from absence.

A number of different methods can be used to compute the AUC (Pearce and Ferrier, 2000). Some distribution modeling programs automatically calculate the AUC (e.g. Maxent; Table 3). The statistic can also be calculated using numerous other free software packages (e.g. R, for free download see http://www.r-project.org/).

Choosing a suitable test statistic

The choice of test statistic depends largely on how the model will be applied. If the aim is to predict the actual distribution (Section 2) then use of a test that incorporates both presence and absence records may be preferable (e.g. accuracy or Kappa). A good model will successfully predict both presences and absences with equal frequency. However, when using this approach it is important to realize that predictions that an unoccupied area is environmentally suitable (type 3 predictions, Figure 3) are considered model errors. Because type 3 predictions are theoretically expected (Section 2), models may be judged poor although they make biologically sound predictions.

If the aim of the modeling is to estimate the potential distribution (Section 2) then presence-only assessment of model performance using sensitivity and statistical significance is likely to be preferable. In this case we cannot test if the model is *correct* (since we do not know the true potential distribution), but rather we test if the model is *useful*. Our criteria for usefulness are that the model successfully predicts presence in a high proportion of test localities (i.e., known occurrences) whilst not predicting that an excessively large proportion of the study area is suitable. Thus, a model that successfully predicts whether the species is present at all test localities whilst classifying most of the study area as suitable may be correct (the environment may truly be suitable throughout most of the study area); however, the model is not useful because it is not more informative than a random prediction.

Subjective guidelines can be used to decide what values of a test statistic correspond to 'good' model performance. For example, Landis and Koch (1977) suggested that Kappa scores >0.75 represent an 'excellent' model, whilst Swets (1988) classified any AUC score >0.9 as 'very good'. However, the only true test of the model is whether it is useful for a given application. There are numerous potential applications of these methods (Table 1) and the final part of this synthesis describes three representative case studies.

Case Studies

Case Study 1: Predicting Distributions of Known and Unknown Species in Madagascar (based on Raxworthy et al., 2003)

Our knowledge of the identity and distribution of species on Earth is remarkably poor, with many species yet to be described and catalogued. This problem has two key elements, which may be termed the 'Linnean' and 'Wallacean' shortfalls (Whittaker et al., 2005). The Linnean shortfall refers to our lack of knowledge of how many, and what kind, of species exist. The term is a reference to Carl Linnaeus, who laid the foundations of modern taxonomy and the 18th century. The

Linnean shortfall concerns our highly incomplete knowledge of the diversity of life that exists on Earth.

The Wallacean shortfall refers to our inadequate knowledge of the distributions of species. This term is a reference to Alfred Russel Wallace who, as well as contributing to the early development of evolutionary theory, was an expert on the geographical distribution of species (he is sometimes referred to as 'the father of biogeography'). The Wallacean shortfall thus refers to our poor knowledge of the biogeography of most species. Species distribution modeling offers a powerful tool to address both the Linnean and Wallacean shortfalls, as demonstrated in a study by Raxworthy et al. (2003).

Raxworthy et al. (2003) modeled the distributions of 11 species of chameleon that are endemic to the island of Madagascar. They used species occurrence records from recent surveys and from older specimens deposited in collections of natural history museums. No observed absence records were available for building the models. Environmental variables were derived from remote sensing data, from a digital elevation model, and from weather station data that had been interpolated to a grid (i.e. converted from point vector to raster data). In all, 25 GIS layers were used in the modeling, including environmental variables describing temperature, precipitation, land cover, and elevation. All analyses were undertaken at a resolution of 1 km^2. The modeling algorithm used was GARP (see Table 3), which generated an output ranging from 0 – 10 at increments of 1. Two alternative thresholds of occurrence were used: threshold = 1 (termed by the authors "any model predicts"), and threshold = 10 (termed "all models predict").

Predictive performance of the models was first evaluated by splitting the available data into two parts, 50% for calibrating the model and 50% for testing the model. The authors calculated the number of test localities at which the species was correctly predicted to be present and tested the statistical significance of the results using a chi-square test. Performance of the 11 models was generally good, with overall prediction success as high as 83%. Predictions usually were better than random. A second evaluation tested model performance using independent test data from herpetological surveys undertaken at 11 sites after the models had been built. In this case, model evaluation was based on both presence and absence records, since surveys were sufficiently thorough that detection probability was high. The success of these predictions was more than 70% and levels of statistical significance were uniformly high.

Raxworthy et al. (2003) thus demonstrated the potential for species' distribution models to be used to guide new field surveys toward areas in which the probability of species presence was high. This approach takes advantage of the type of model prediction illustrated by area *2* in Figure 3: the model identifies an area that is environmentally similar to where the species has already been found, but for which no occurrence data are available. The models can thus help to address the Wallacean shortfall, by improving our knowledge of the distributions of known species.

Raxworthy et al. (2003) also demonstrated that the models can help to address the Linnean shortfall by guiding field surveys toward areas where species new to science are most likely to be discovered. In this case, the approach makes use of the type of model prediction illustrated by area *3* in Figure 3: areas are identified that are unoccupied by the species being modeled, but where closely-related species that occupy similar environmental space are most likely to be found. By surveying sites identified by the distribution models for known species, Raxworthy et al. (2003) discovered seven new species, considerably greater than the number that would usually be expected on the basis of similar survey effort across a less-targeted area.

Case Study 2: Species' Distribution Modeling as a Tool for Predicting Invasions of Non-Native Plants (based on Thuiller et al., 2005)

Invasive species are increasingly a global concern, with invasions altering ecosystem functioning, threatening native biodiversity, and negatively impacting agriculture, forestry, and human health. Species' distribution modeling can be used to identify areas that are most likely to be colonized by a known invader. The general approach is to model the distribution of a species using occurrence records from its native range, then project the model into new regions to assess susceptibility to invasion. The approach makes use of the type of model prediction illustrated by area *3* in Figure 3: areas are identified that are part of the potential, but not actual, distribution. Thuiller et al. (2005) used this method to identify parts of the world that are potentially susceptible to invasion by plant species native to South Africa.

Thuiller and colleagues developed distribution models for 96 South African plant taxa that are invasive in other regions of the globe. Species distribution data were extracted from large databases of occurrence records that have been collated for South Africa. Since the surveys incorporated within these databases were fairly comprehensive, the authors argued that absences within the databases are reliable and the modeling was thus undertaken using presence and absence data. Four climate-related variables that are known to affect plant physiology and growth were developed and used as input to the models. These included two measures of temperature (growing degree days and temperature of the coldest month) along with indexes of humidity and plant productivity. All the analyses were undertaken at a resolution of 25x25 km, which was considered sufficiently fine to identify environmental differences between regions at a global scale. Generalized Additive Models (implemented using the Splus-based BIOMOD application; Table 3) were used to build the distribution models.

The first step in the study was to model each species' distribution based on its native range in South Africa. Models were calibrated using 70% of the available records for each species, with the remaining 30% retained for model testing. The AUC validation test was applied using presence and absence test data. Because AUC assesses predictive performance based on continuous model output, it was not necessary to set a threshold of occurrence. Validation statistics for the test data were generally very good, with a median AUC score across all 96 species of 0.94 (minimum = 0.68, maximum = 1.0).

The second step in the study was to project the calibrated models worldwide. The accuracy of the global predictions was assessed for three example species using presence-only records from their non-native distributions (for example, in Europe, Australia, and New Zealand). Absence records were not available for these tests, so model performance was assessed using chi-square tests (see section 5). For each of the species, predictions of potentially suitable areas outside South Africa showed considerable agreement with observed records of invasions (chi-square test, $P<0.05$).

In a further step, Thuiller and colleagues summed the probability surfaces for all 96 taxa to produce a global map for risk of invasion by species of South African origin. Parts of the world most susceptible to invasion included six biodiversity 'hotspots', including the Mediterranean Basin, California Floristic Province, and southwest Australia. This study demonstrates that species distribution modeling can be a valuable tool for identifying sites prone to invasion. Such sites may be prioritized for monitoring and quarantine measures can be put in place to help avoid the establishment of invasive species.

Case Study 3: Modeling the Potential Impacts of Climate Change on Species' Distributions in Britain and Ireland (based on Berry et al., 2002)

Climate change has the potential to significantly impact the distribution of species. Species' distribution models have been used in a number of studies that aim to predict the likely redistribution of species under projected climate change over the coming century. The general approach is to calibrate the models based on current distributions of species and then predict future distributions of those species across landscapes for which the environmental input variables have been perturbed to reflect expected changes.

Berry et al. (2002) modeled 54 species that were chosen to represent a range of habitats common in Britain and Ireland. Species distribution data were obtained for the whole of Europe (i.e. European extent) from range maps for European plants, amphibians, butterflies, and mammals. Since survey effort across Europe is generally very high, areas where a species had not been recorded were considered reliable measures of absence, and both presence and absence data were therefore used in the modeling. Five environmental variables describing temperature, precipitation, and soil type were used. These variables were generated using both current climate data and predictions of future climate from a General Circulation Model (GCM). GCMs are complex simulations that predict future climates using scenarios of greenhouse gas emissions. The environmental variables were calculated at a coarse resolution (~50x50 km) for Europe, and also at a finer resolution (10x10 km) for Britain and Ireland. The modeling algorithm was an Artificial Neural Network (Table 3), which gave predictions of relative suitability ranging from 0 to 1. These continuous predictions were converted into binary predictions of presence and absence by applying a threshold of occurrence that maximized the Kappa statistic.

An important part of Berry et al.'s (2002) method was that calibration of the distribution model was carried out at the European scale (large extent and coarse resolution) and then the model was used to predict distributions in Britain and Ireland (smaller extent and finer resolution). This approach ensured that when distributions were predicted under future climate scenarios in Britain and Ireland, the model was not required to extrapolate beyond the range of data for which it was calibrated (since future climates in Britain and Ireland are expected to be similar to conditions currently experienced elsewhere in Europe).

Evaluation of the models was undertaken at the European scale by comparing model predictions against a test dataset, which comprised one third of the available data that had been randomly selected and not used in model calibration. Since both presence and absence records were available, the Kappa statistic was applied. Results from these tests revealed generally good model performance, with 27 out of 54 species achieving Kappa scores >0.75, and only 7 of 54 species with Kappa scores <0.6.

The models calibrated for Europe were then used to predict distributions in Britain and Ireland under current and projected future climates. Berry et al. (2002) emphasized that they did not predict actual distributions, but rather 'bioclimate envelopes', or suitable climate space. It is important to remember that actual future distributions will be determined by many factors that are not taken into account in the modeling, including the ability of species to colonize areas that become suitable. Nevertheless, the distribution models enabled each species to be placed in one of three categories: those expected to lose suitable climate space, those expected to gain suitable climate space, and those showing little change. Species' distribution modeling thus enables preliminary assessments of the possible impacts of climate change to be made, providing information that may be valuable in developing conservation policies to address the threat.

Acknowledgements

Thorough comments from five anonymous reviewers greatly improved the module. The module also benefitted significantly from editorial input by Ian Harrison and the rest of the NCEP team. Extensive discussions with Robert Anderson, Catherine Graham, Steven Phillips, Town Peterson, Chris Raxworthy, Jorge Sóberon and members of the New York Species Distribution Modeling Discussion Group greatly contributed to the module. I am also indebted to comments and discussion by participants in courses I have taught through the American Museum of Natural History, and the Global Biodiversity Information Facility.

Terms of Use

Reproduction of this material is authorized by the recipient institution for non-profit/non-commercial educational use and distribution to students enrolled in course work at the institution. Distribution may be made by photocopying or via the institution's intranet restricted to enrolled students. Recipient agrees not to make commercial use, such as, without limitation, in publications distributed by a commercial publisher, without the prior express written consent of AMNH.

This material is based on work supported by the National Science Foundation under the Course, Curriculum and Laboratory Improvement program (NSF 0127506), and the United States Fish and Wildlife Service (Grant Agreement No. 98210-1-G017).

Literature Cited

Anderson, R.P., M. Gomez-Laverde, and A.T. Peterson. 2002a. Geographical distributions of spiny pocket mice in South America: Insights from predictive models. Global Ecology and Biogeography 11, 131-141.

Anderson, R. P., A.T. Peterson, and M. Gómez-Laverde. 2002b. Using niche-based GIS modeling to test geographic predictions of competitive exclusion and competitive release in South American pocket mice. Oikos 98: 3-16

Anderson, R.P., D. Lew, and A.T. Peterson. 2003. Evaluating predictive models of species' distributions: Criteria for selecting optimal models. Ecological Modelling 162: 211-232.

Anderson, R. P. and E. Martínez-Meyer. 2004. Modeling species' geographic distributions for conservation assessments: an implementation with the spiny pocket mice (Heteromys) of Ecuador. Biological Conservation 116: 167-179.

Araújo, M.B. and M. Luoto. 2007. The importance of biotic interactions for modelling species distributions under climate change. Global Ecology and Biogeography 16: 743-753.

Araújo, M.B., and R.G. Pearson. 2005. Equilibrium of species' distributions with climate. Ecography 28: 693-695.

Araújo, M.B., R.G. Pearson, W. Thuiller, and M. Erhard. 2005a. Validation of species-climate envelope models under climate change. Global Change Biology 11: 1504-1513.

Araújo, M.B., W. Thuiller, P.H. Williams, and I. Reginster. 2005b. Downscaling European species atlas distributions to a finer resolution: Implications for conservation planning. Global Ecology and Biogeography 14: 17-30.

Araújo, M. B., and P.H. Williams. 2000. Selecting areas for species persistence using occurrence data. Biological Conservation 96: 331-345.

Beerling, D.J., B. Huntley, and J.P. Bailey. 1995. Climate and the distribution of *fallopia japonica*: Use of an introduced species to test the predictive capacity of response surfaces. Journal of Vegetation Science 6: 269-282.

Berry, P.M., T.P. Dawson, P.A. Harrison, and R.G. Pearson. 2002. Modelling potential impacts of climate change on the bioclimatic envelope of species in Britain and Ireland. Global Ecology and Biogeography 11: 453-462.

Bourg, N.A., W.J. McShea, and D.E. Gill. 2005. Putting a cart before the search: Successful habitat prediction for a rare forest herb. Ecology 86: 2793-2804.

Boyce, M.S., P.R. Vernier, S.E. Nielsen, and F.K.A. Schmiegelow. 2002. Evaluating resource selection functions. Ecological Modelling 157: 281-300.

Brotons, L., W. Thuiller, M.B. Araújo, and A.H. Hirzel. 2004. Presence-absence versus presence-only modelling methods for predicting bird habitat suitability. Ecography 27: 437-448.

Buckland, S.T., and D.A. Elston. 1993. Empirical models for the spatial distribution of wildlife. Journal of Applied Ecology 30: 478-95.

Carpenter, G., A.N. Gillson, and J. Winter. 1993. DOMAIN: a flexible modeling procedure for mapping potential distributions of plants and animals. Biodiversity and Conservation 2: 667-680.

Chase, J.M., and M.A. Leibold. 2003. Ecological niches: Linking classical and contemporary approaches University of Chicago Press, Chicago, USA.

Chuine, I., and E.G. Beaubien. 2001. Phenology is a major determinant of temperate tree distributions. Ecology Letters 4, 500-510.

Cramer, J. S. 2003. Logit Models: From Economics and Other Fields. Cambridge University Press.

Elith, J., C. Graham, and the NCEAS species distribution modeling group. 2006. Novel methods improve prediction of species' distributions from occurrence data. Ecography 29, 129-151.

Elith, J. and Leathwick, J.R. 2007. Predicting species' distributions from museum and herbarium records using multiresponse models fitted with multivariate adaptive regression splines. Diversity and Distributions 13: 165-175.

Engler, R., A. Guisan, and L. Rechsteiner. 2004. An improved approach for predicting the distribution of rare and endangered species from occurrence and pseudo-absence data. Journal of Applied Ecology 41: 263-274.

Ferrier, S., G. Watson, J. Pearce, and Drielsma 2002. Extended statistical approaches to modelling spatial pattern in biodiversity in northeast new south wales. I. Species-level modelling. Biodiversity and Conservation 11: 2275-2307.

Fielding, A.H. and J.F. Bell. 1997. A review of methods for the assessment of prediction errors in conservation presence/absence models. Environmental Conservation 24: 38-49.

Fleishman, E., R. Mac Nally, and J.P. Fay. 2002. Validation tests of predictive models of butterfly occurrence based on environmental variables. Conservation Biology, 17: 806-817.

Fleishman, E., R. Mac Nally, J.P. Fray, and D.D. Murph. 2001. Modeling and predicting species occurrence using broad-scale environmental variables: An example with butterflies of the great basin. Conservation Biology 15: 1674-1685.

Gibbons, D.W., J.B. Reid, and R.A. Chapman. 1993. The new atlas of breeding birds in Britain and Ireland: 1988-1991 Poyser, London, UK.

Graham, C.H., S. Ferrier, F. Huettman, C. Moritz, and A.T. Peterson. 2004a. New developments in museum-based informatics and applications in biodiversity analysis. Trends in Ecology and Evolution 19: 497-503.

Graham, C.H., S. R. Ron, J.C. Santos, C.J. Schneider, and C. Moritz. 2004b. Integrating phylogenetics and environmental niche models to explore speciation mechanisms in Dendrobatid frogs. Evolution 58: 1781-1793.

Graham, C.H., C. Moritz, and S. E. Williams. 2006 Habitat history improves prediction of biodiversity in rainforest fauna. PNAS 103: 632-636.

Guisan, A., O. Broennimann, R. Engler, M. Vust, N.G. Yoccoz, A. Lehman, and N.E. Zimmermann. 2006. Using niche-based models to improve the sampling of rare species. Conservation Biology 20: 501-511.

Guisan, A., T.C. Edwards Jr., and T. Hastie. 2002. Generalized linear and generalized additive models in studies of species distributions: Setting the scene. Ecological Modelling 157: 89-100.

Guisan, A., and W. Thuiller. 2005. Predicting species distribution: Offering more than simple habitat models. Ecology Letters 8: 993-1009.

Hampe, A. 2004. Bioclimatic models: what they detect and what they hide. Global Ecology and Biogeography 11: 469-471.

Hannah, L., G.F. Midgley, G. Hughes, and B. Bomhard. 2005. The view from the cape: extinction risk, protected areas, and climate change. BioScience 55: 231-242.

Heikinnen, R.K., M. Luoto, R. Virkkala, R.G. Pearson, and J-H. Körber. 2007. Biotic interactions improve prediction of boreal bird distributions at macro-scales. Global Ecology and Biogeography 16: 754-763.

Higgins, S.I., D.M. Richardson, R.M. Cowling, and T.H. Trinder-Smith. 1999. Predicting the landscape-scale distribution of alien plants and their threat to plant diversity. Conservation Biology 13: 303-313.

Hijmans, R. J., S.E. Cameron, J.L. Parra, P.G. Jones, and A. Jarvis. 2005. Very high resolution interpolated climate surfaces for global land areas. International Journal of Climatology 25: 1965-1978.

Hirzel, A.H., J. Hausser, D. Chessel, and N. Perrin. 2002. Ecological-niche factor analysis: How to compute habitat-suitability map without absence data. Ecology 83: 2027-2036.

Huberty, C.J. 1994. Applied Discriminant Analysis. Wiley Interscience, New York, USA.

Hugall, A., C. Moritz, A. Moussalli and. J. Stanisic. 2002. Reconciling paleodistribution models and comparative phylogeography in the wey tropics rainforest land snail Gnarosophia bellendenkerensis. PNAS 99: 6112-6117.

Huntley, B., P.M. Berry, W. Cramer, and A.P. Mcdonald. 1995. Modelling present and potential future ranges of some European higher plants using climate response surfaces. Journal of Biogeography 22: 967-1001.

Hutchinson, G.E. 1957. Concluding remarks. Cold Spring Harbor Symposium on Quantitative Biology 22: 415-457.

Iverson, L.R., and A.M. Prasad. 1998. Predicting abundance of 80 tree species following climate change in the eastern United States. Ecological Monographs 68: 465-485.

Kozak, J.H. and J.J. Wiens. 2006. Does niche conservatism promote speciation? A case study in North American salamanders. Evolution 60: 2604-2621.

Landis, J.R., and G.C. Koch. 1977. The measurement of observer agreement for categorical data. Biometrics 33: 159-174.

Lawler, J. J., D. White, R.P. Neilson, and A.R. Blaustein. 2006. Predicting climate-induced range shifts: model differences and model reliability. Global Change Biology 12: 1568-1584.

Lawton, J. L. 2000. Concluding remarks: a review of some open questions. Pages 401-424 in M. J. Hutchings, E. John, and A. J. A. Stewart, editors. Ecological Consequences of Heterogeneity. Cambridge University Press.

Leathwick, J.R., D. Rowe, J. Richardson, J. Elith and T. Hastie. 2005. Using multivariate adaptive regression splines to predict the distribution of New Zealand's freshwater diadromous fish. Freshwater Biology 50, 2034-2052.

Leathwick, J.R., J. Elith, and T. Hastie. 2006. Comparative performance of generalized additive models and multivariate adaptive regression splines for statistical modelling of species distributions. Ecological Modelling 199: 188-196.

Lehman, A., J.M. Overton, and J.R. Leathwick. 2002. GRASP: generalized regression analysis and spatial prediction. Ecological Modelling 157: 189-207.

Lindenmayer, D. B., H.A. Nix, J.P. McMahon, M.F. Hutchinson, and M.T. Tanton. 1991. The conservation of Leadbeater's possum, *Gymnobelideus leadbeateri* (McCoy): a case study of the use of bioclimatic modelling. Journal of Biogeography 18: 371-383.

Liu, C., P.M. Berry, T.P. Dawson, and R.G. Pearson. 2005. Selecting thresholds of occurrence in the prediction of species distributions. Ecography 28: 385-393.

Loiselle, B.A., C.A. Howell, C.H. Graham, J.M. Goerck, T. Brooks, K.G. Smith, and P.H. Williams. 2003. Avoiding pitfalls of using species distribution models in conservation planning. Conservation Biology 17: 1591-1600.

Manel, S., J.M. Dias, and S.J. Ormerod. 1999. Comparing discriminant analysis, neural networks and logistic regression for prediction species distributions: a case study with a Himalayan river bird. Ecological Modelling 120: 337-347.

Manel, S., H.C. Williams, and S.J. Ormerod. 2001. Evaluating presences-absence models in ecology: the need to account for prevalence. Journal of Applied Ecology 38: 921-931.

Martinez-Meyer, E., A.T. Peterson, and W.W. Hargrove. 2004. Ecological niches as stable distributional constraints on mammal species, with implications for pleistocene extinctions and climate change projections for biodiversity. Global Ecology and Biogeography 13: 305-314.

Monserud, R.A., and R. Leemans. 1992. Comparing global vegetation maps with the kappa statistic. Ecological Modelling 62: 275-293.

Pearce, J., and D.B. Lindenmayer. 1998. Bioclimatic analysis to enhance reintroduction biology of the endangered helmeted honeyeater (*lichenostomus melanops cassidix*) in southeastern Australia. Restoration Ecology 6: 238-243.

Pearce, J., and S. Ferrier. 2000. Evaluating the predictive performance of habitat models developed using logistic regression. Ecological Modelling 133: 225-245.

Pearson, R.G., and T.P. Dawson. 2003. Predicting the impacts of climate change on the distribution of species: Are bioclimate envelope models useful? Global Ecology and Biogeography 12: 361-371.

Pearson, R.G., T.P. Dawson, P.M. Berry, and P.A. Harrison. 2002. Species: A spatial evaluation of climate impact on the envelope of species. Ecological Modelling 154: 289-300.

Pearson, R.G., T.P. Dawson, and C. Liu. 2004. Modelling species distributions in Britain: A hierarchical integration of climate and land-cover data. Ecography 27: 285-298.

Pearson, R.G., C.J. Raxworthy, M. Nakamura, and A.T. Peterson. 2007. Predicting species' distributions from small numbers of occurrence records: A test case using cryptic geckos in Madagascar. Journal of Biogeography 34: 102-117.

Pearson, R.G., W. Thuiller, M.B. Araújo, E. Martinez-Meyer, L. Brotons, C. McClean, L. Miles, P. Segurado, T.P. Dawson, and D. Lees. 2006. Model-based uncertainty in species' range prediction. Journal of Biogeography 33: 1704-1711.

Peterson, A.T. 2003. Predicting the geography of species' invasions via ecological niche modeling. Quarterly Review of Biology 78: 419-433.

Peterson, A.T., and J.J. Shaw. 2003. *Lutzomyia* vectors for cutaneous leishmaniasis in southern brazil: Ecological niche models, predicted geographic distributions, and climate change effects. International Journal of Parasitology 33: 919-931.

Peterson, A.T., B.W. Benz, and M. Pape. 2007. Highly pathogenic H5N1 avian influenza: Entry pathways into North America via bird migration. PLoSONE 2(2): e261. doi:10.1371/journal.pone.0000261.

Peterson, A. T., J. Soberon, and V. Sanchez-Cordero. 1999. Conservatism of Ecological Niches in Evolutionary Time. Science 285: 1265-1267.

Peterson, A. T., R. R. Lash, D. S. Carroll, and K. M. Johnson. 2006. Geographic potential for outbreaks of Marburg hemorrhagic fever. American Journal of Tropical Medicine & Hygiene 75: 9-15.

Peterson, A.T., M.A. Ortega-Huerta, J. Bartley, V. Sanchez-Cordero, J. Soberon, R.W. Buddemeier, and D.R.B. Stockwell. 2002. Future projections for mexican faunas under global climate change scenarios. Nature 416: 626-629.

Phillips, S.J., R.P. Anderson, and R.E. Schapire. 2006. Maximum entropy modeling of species geographic distributions. Ecological Modelling 190, 231-259.

Phillips, S.J., M. Dudik, and R.E. Schapire. 2004. A maximum entropy approach to species distribution modeling. In Proceedings of the 21st international conference on machine learning, pages 655-662. AMC Press, New York, USA.

Pulliam, H. R. 2000. On the relationship between niche and distribution. Ecology Letters, 3:349-361.

Raxworthy, C.J., C. Ingram, N. Rabibosa, and R.G. Pearson. 2007. Species delimitation applications for ecological niche modeling: a review and empirical evaluation using Phelsuma day gecko groups from Madagascar. Systematic Biology.

Raxworthy, C.J., E. Martinez-Meyer, N. Horning, R.A. Nussbaum, G.E. Schneider, M.A. Ortega-Huerta, and A.T. Peterson. 2003. Predicting distributions of known and unknown reptile species in Madagascar. Nature 426: 837-841.

Robertson, M. P., N. Caithness, and M.H. Villet. 2001. A PCA-based modeling technique for predicting environmental suitability for organisms from presence records. Diversity and Distributions 7: 15-27.

Segurado, P. and M.B. Araújo. 2004. An evaluation of methods for modelling species distributions. Journal of Biogeography.

Stockwell, D.R.B. and D.P. Peters. 1999. The GARP modelling system: Problems and solutions to automated spatial prediction. International Journal of Geographical Information Systems 13: 143-158.

Swets, J.A. 1988. Measuring the accuracy of diagnostic systems. Science 240: 1285-1293.

Thomas, C.D., A. Cameron, R.E. Green, M. Bakkenes, L.J. Beaumont, Y.C. Collingham, B.F. Erasmus, M.F. De Siqueira, A. Grainger, L. Hannah, L. Hughes, B. Huntley, A.S. Van Jaarsveld, G.F. Midgley, L. Miles, M.A. Ortega-Huerta, A.T. Peterson, O.L. Phillips, and S.E. Williams. 2004. Extinction risk from climate change. Nature 427: 145-148.

Thuiller, W. 2003. Biomod - optimizing predictions of species distributions and projecting potential future shifts under global change. Global Change Biology 9: 1353-1362.

Thuiller, W., M.B. Araújo, R.G. Pearson, R.J. Whittaker, L. Brotons, and S. Lavorel. 2004. Uncertainty in predictions of extinction risk. Nature 430: 33 (doi: 10.1038/Nature02716).

Thuiller, W., D.M. Richardson, P. Pysek, G.F. Midgley, G.O. Hughes, and M. Rouget. 2005. Niche-based modeling as a tool for predicting the global risk of alien plant invasions. Global Change Biology 11: 2234-2250.

Verbyla, D.L., and J.A. Litaitis. 1989. Resampling methods for evaluating classification accuracy of wildlife habitat models. Environmental Management 13: 783-787.

Whittaker, R.J., M.B. Araújo, P. Jepson, R.J. Ladle, J.E.M. Watson, and K.J. Willis. 2005. Conservation biogeography: Assessment and prospect. Diversity and Distributions 11: 3-23.

Woodward, F. I., and D. J. Beerling. 1997. The dynamics of vegetation change: health warnings for equilibrium 'dodo' models. Global Ecology and Biogeography 6: 413-418.

Zaniewski, A.E., A. Lehman, and J.M. Overton. 2002. Predicting species spatial distributions using presence-only data: A case study of native New Zealand ferns. Ecological Modelling 157: 261-280.

Zar, J.H. 1996. Biostatistical Analysis 3rd Edition. Prentice Hall, New Jersey, USA.

Biodiversity Conservation and Human Health

Andrés Gómez [*] and Elizabeth Nichols [†]

[*] The American Museum of Natural History, New York, NY, U.S.A., email agomez@amnh.org

[†] The American Museum of Natural History, New York, NY, U.S.A., email nichols@amnh.org

K. Frey

Biodiversity Conservation and Human Health

Andrés Gómez and Elizabeth Nichols

OVERVIEW

This Exercise is a practical application of the ideas discussed in the Synthesis and Presentation, focusing on the analysis of a hypothetical potential conflict between biodiversity conservation and human health.

There are at least five major diseases in Africa that have freshwater snails as intermediate hosts. These diseases are:

1. *Schistosomiasis:* An important parasitic disease of humans that resulted in over 44,000 deaths around the world in 2002; it is caused by a flatworm.
2. *Bovine schistosomiasis:* An important parasite of cattle, which is caused by species in the same genus as the human form of the disease; however, this parasite does not attack humans (i.e., there is no transmission from animals to humans).
3. *Paragonimiasis:* A fluke infection of the lungs that affects an estimated 22 million people worldwide.
4. *Paramphistomiasis:* Caused by a trematode infestation, this is an important disease of domestic cattle and wild ungulates.
5. *Fasciolasis:* A liver fluke that primarily infests cattle and sheep.

An understanding of the basic biology of a disease is critical for understanding how it affects humans, and how the transmission cycle is affected by other components of biodiversity. Here, we will use schistosomiasis as a model to describe the relevant life cycle and epidemiology for this group of related diseases.

Schistosomiasis (also known as bilharzia) is often a chronic illness caused by several species of parasitic flatworm in the genus *Schistosoma*, which are collectively known as schistosomes. It is estimated that over 250 million people worldwide carry this disease, especially in tropical countries in South America, Africa, and Asia. The symptoms of schistosomiasis include pain, fever, diarrhea, and fatigue, and can also lead to other symptoms in the digestive, urinary, and central nervous systems. It can be treated by a single oral dose of a common anti-parasite drug.

Schistosomes cycle in a snail-human life cycle (Figure 1). Snails belonging to three genera (*Biomphalaria*, *Bulinus*, and *Oncomelania*) are the obligate intermediate hosts. Humans become infected through the skin with the parasite's infective life stages (known as cercariae), while swimming or wading in bodies of water with infected adult snails (see Stage 6 in Figure 1). Uninfected snails become infected when untreated human excreta containing eggs reach water bodies (see Stage 10 in Figure 1). The life cycle therefore has three main components: 1) in water, where eggs are released and subsequently hatch; 2) in the intermediate hosts, the early stages of the parasite mature inside the snails; and 3) in the final host, when cercariae infect humans and cause the disease. The disease can be controlled by stopping the transmission cycle at any of the stages of the cycle.

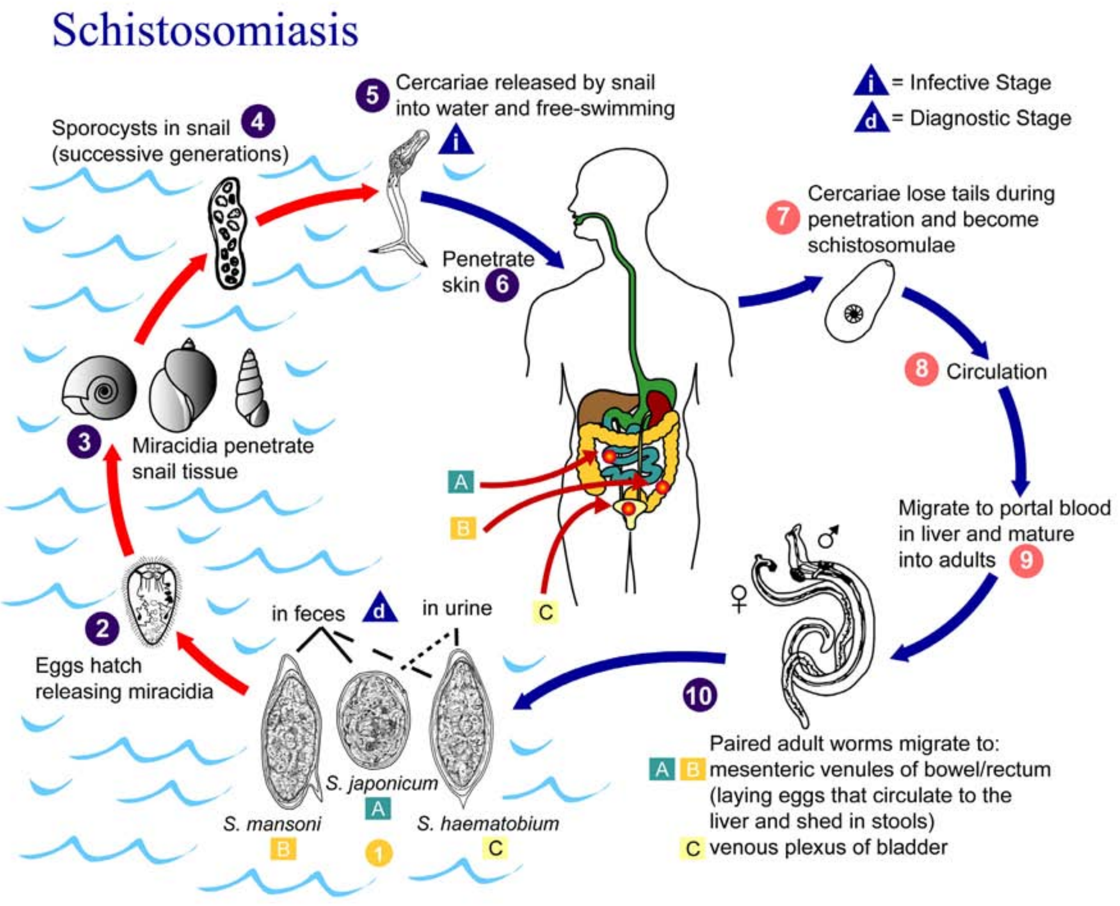


Figure 1. Schistosomiasis Life Cycle. This image was produced by the United States Department of Health and Human Services and is in the public domain.

OBJECTIVE

The objective is to analyze the human health and conservation implications of a common public health strategy.

PROCEDURE

You are the scientific advisor to the Minister of the Environment in an East African nation. A recent outbreak of schistosomiasis, both in humans and in cattle, has devastated several provinces across the country.

The Minister of Health has just offered the President a proposal to control the disease outbreak through widespread snail control across a series of natural and man-made waterways. Your employer (the Minister of the Environment) has asked you to provide a brief outlining the impacts of the proposed spraying. Specifically, she is interested in understanding the consequences of the plan on human health and biodiversity. The Minister suspects that, although the proposed plan will be effective in controlling these diseases, biodiversity, ecosystem functioning, and human health may be negatively affected by it, and thus an alternative plan would be ultimately preferable.

To better prepare your case, you have also been provided with the summary the Health Minister presented to the President:

> *In light of the burden on human and animal health, and the economic consequences of snail-borne diseases, it is the position of the Ministry of Health that a major control strategy should be undertaken immediately. We propose the widespread use of synthetic molluscicides along the four major watersheds of the country in order to eliminate the snails that act as intermediate hosts for these parasites. This will be a multi-year, multi-million dollar project that will continue until the prevalence of these five diseases has diminished significantly, and may be reinstated at any point where high incidence of any of these diseases is seen to increase. The economic losses caused by these diseases (measured in the costs of treatment, lost wages due to illness, and losses in productivity in livestock ranches) justify this measure and the control of such costs will offset the expenses for this program.*

You will write a two-page maximum policy brief for the Minister of the Environment analyzing the implications of the proposed spraying. The Minister will use your brief for her talking points in arguing against the Minister of Health's proposal. Your final product should contain:

1. A short analysis on the implications for biodiversity in the country
2. A concise explanation of the ways the plan will negatively impact overall health
3. An alternative control strategy

Use the information provided in the *Biodiversity Conservation and Human Health* lecture, the attached memo on "How to Write a Policy Brief," and the following additional resources to construct your arguments. You are encouraged to complete your picture of this imaginary country with any information you decide is relevant. For example, your group may include in the analysis that distinctive snail shell ornaments are very important for the cultural identity of the country's minorities, that snails are a significant protein source, or any other social, political, or economic factor that may be helpful in constructing a well-reasoned reply. If you make any such assumptions, be sure you define them, and logically connect them to your broader argument when you write your policy brief.

In addition to the relevant linkages outlined in the *Biodiversity Conservation and Human Health* lecture and synthesis, some points you may consider when doing your analysis include, but are not limited to:

- The effects of the plan on snail species (and other fauna) not involved in disease transmission.
- Ecosystem-wide consequences; e.g., what would happen to species that are above (such as fish) or below (such as plants) snails in the food web? What would be the consequences for ecosystem function? Which of these functions can impact human health?
- The effects of the molluscicides themselves.
- The estimated 2/3 of the 330 snail species in Africa that are under some level of threat (Kristensen and Brown, 1999); you may assume that 2/3 of the snail species in your country are threatened as well.

THE POLICY BRIEF[i]

Overview

The policy brief is a document, which outlines the rationale for choosing a particular policy alternative or course of action in a current policy debate. It is commonly produced in response to a request directly from a decision-maker who intends to advocate for the position detailed in the brief.

The brief may provide a targeted discussion of the current alternatives without arguing for a particular one or focus directly on providing an argument for the adoption of a specific policy alternative. In either case, as any policy debate is a marketplace of competing ideas, the purpose of the policy brief is to convince the target audience of the urgency of the current problem and the need to adopt the preferred alternative or course of action outlined and, therefore, serve as an impetus for action.

The most common audience for a policy brief is the decision-maker, but it is also not unusual to use the document to support broader advocacy initiatives targeting a wide but knowledgeable audience (e.g., decision-makers, journalists, diplomats, administrators, researchers). The policy brief is usually said to be the most common and effective written communication tool in a policy campaign. However, in balancing all of the criteria below, many analysts also find the brief the most difficult policy tool to write.

An effective policy brief is typically:

Focused – All aspects of the policy brief (from the message to the layout) need to be strategically focused on achieving the intended goal of convincing the target audience. The argument provided must build on what they do know about the problem, provide insight about what they don't know about the problem, and be presented in language that reflects their values, i.e., using ideas, evidence, and language that will convince them.

Professional, not academic – The common audience for a policy brief is not interested in the research/analysis procedures conducted to produce the evidence, but is very interested to know the writer's perspective on the problem and potential solutions based on the new evidence.

Evidence-based – The policy brief is a communication tool produced by policy analysts and, therefore, all potential audiences not only expect a rational argument, but will only be convinced by argumentation supported by evidence that the problem exists and the consequences of adopting particular alternatives.

Limited – To provide a comprehensive, yet targeted argument in limited space, the focus of the brief needs to be limited to a particular problem, or area of a problem, and succinct in achieving its goals.

Understandable – This not only refers to using clear and simple language (i.e., not the jargon and concepts of an academic discipline), but also to providing a well-explained and easy to follow argument targeting a wide, but knowledgeable audience.

Accessible – the writer of the policy brief should facilitate the ease of use of the document by the target audience and, therefore, should subdivide the text using clear descriptive titles to guide the reader.

COMMON STRUCTURAL ELEMENTS OF A POLICY BRIEF

As discussed above, policy briefs directly reflect the different roles that the policy analyst commonly plays, i.e., from researcher to advocate. The type of brief that we are focusing on is one from the more action-oriented, advocacy end of the continuum. Although there is much variation even at this end of the scale, the most common elements of the policy brief are as follows:

Title

The title aims to catch the attention of the reader and compel him/her to read on, and so needs to be descriptive, punchy, and relevant.

Context and importance of the problem

The purpose of this element of the brief is to convince the target audience that a current and urgent problem exists, which requires them to take action. The context and importance of the problem is both the introductory and first building block of the brief. As such, it usually includes the following:

- A clear **statement of the problem or issue** in focus
- A short **overview of the root causes** of the problem
- A clear **statement of the implications** of the problem

Critique of policy option(s)

The aim of this element is to detail shortcomings of the current approach or options being implemented and, therefore, illustrate both the need for change and the focus of where change needs to occur. This section may contain:

- A short **overview of the policy option(s)** in focus
- An **argument illustrating why and how** the current or proposed approach is failing; it is important, for the sake of credibility, to recognize all opinions expressed in the debate of the issue

Recommendations

The policy alternatives you are advocating.

Sources consulted or recommended

LITERATURE CITED

Kristensen, T. K. and D. S. Brown. 1999. Control of intermediate host snails for parasitic diseases – A threat to biodiversity in African freshwaters? Malacologia 41:379-391.

---

[i] This description of the policy brief was modified from Young and Quinn (2004) (http://www.policy.hu/ipf/fel-pubs/samples/PolicyBrief-described.pdf)

TERMS OF USE

"A. Gómez and E. Nichols. 2010. Biodiversity Conservation and Human Health. Exercise. American Museum of Natural History, *Lessons in Conservation*. Available at http://ncep.amnh.org/linc."

This material is based on work supported by the National Science Foundation under the Course, Curriculum and Laboratory Improvement program (NSF 0127506).

Ecological Decay on Isolated Forest Fragments

Madhu Rao

Wildlife Conservation Society, New York, NY, U.S.A., email mrao@wcs.org

K. Frey

Ecological Decay on Isolated Forest Fragments

Madhu Rao

ASSIGNMENT

For this exercise, imagine that you belong to a Non-Governmental Organization which advises the Government of Venezuela on all domestic environmental issues. Eleven years ago, the Government created a large hydroelectric reservoir by submerging a large area of pristine tropical forest leaving only islands of forest where the high elevation areas remain above water. The Department of Environment and Natural Resources (DENR) is now keen to evaluate the potential of developing the site for **ecotourism** and as an **ecological research station**. The DENR assumes that ecotourists would be particularly interested in viewing the highly diverse flora and fauna at the site. Large mammals (e.g. capuchin and howler monkeys, deer), reptiles (iguana), birds (macaws) and invertebrates (leaf-cutter ants) could be potentially attractive to ecotourists. Simultaneously, the DENR feels that the site has potential for development as an ecological research station for scientific investigation. The DENR seeks your professional technical advice on specific ecological issues that would influence its decision on whether it should pursue one or both options. Information from faunal inventories that were conducted on the islands last year (hence 10 years following inundation) is available to help your analysis.

Some of the background information in the exercise is real but everything else is hypothetical.

Lago Guri is a large hydroelectric reservoir that was created in 1986, following the construction of the Raul Leoni Dam along the Rio Caroni in east-central Venezuela (see Figure 1). The inundation of over 4300 km^2 of hilly terrain resulted in the conversion of a once continuous forested landscape into hundreds of isolated islands. The habitat of all islands and the mainland is semi-deciduous tropical forest.

Faunal inventories were conducted on six “small” islands (0.25 to 0.9 ha), four “medium” islands (4 to 12 ha), two “large” islands (> 150 ha), and two sites on the mainland ten years following inundation. Through these inventories, scientists measured the abundance at each site of selected vertebrates and invertebrates (Terborgh et al., 2001). The inventories revealed that small and medium islands already lacked more than 75% of the vertebrate species known to occur on the nearby mainland, whereas the two large islands retained nearly all species. **The scientists designated the two large islands along with the two stations on the mainland as control sites.** (Tables 1 and 2).

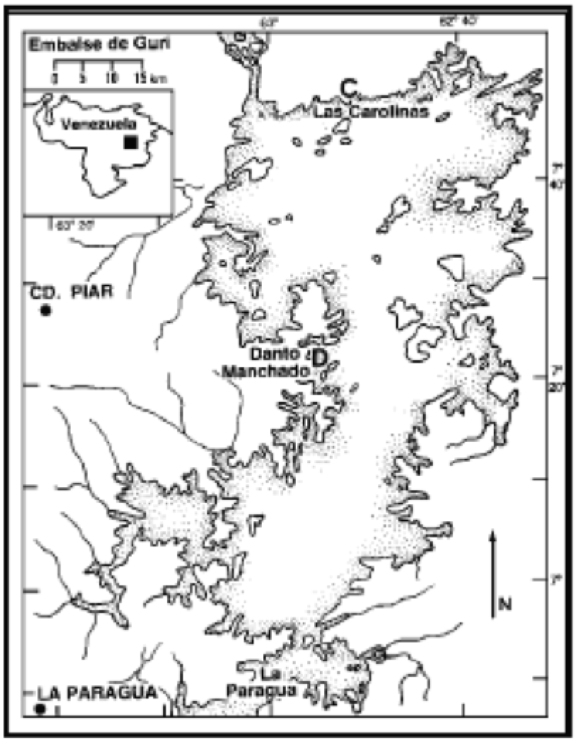


Figure 1. The Location of the dam and the hydroelectric plant are indicated by the narrow portion of the lake in the extreme northwest. The islands referred to in this exercise are located in the centre of the lake around Danto Machado (one of the large islands). Map by Gerardo Aymard.

Figure taken from: Norconk, M.A., and B.W Grafton. 2003. Changes in forest composition and potential feeding tree availability on a small land-bridge island in Lago Guri, Venezuela. Pages 211-228 in L.K. Marsh, editor. In Primates in Fragments: Ecology and Conservation. Kluwer/Plenum Press.

Table 1. Presence-absence Data for Various Trophic Levels on Small, Medium, and Large Guri Landmasses

Mainland	Large Islands	Medium Islands	Small Islands
FRUGIVORES (Seed dispersers)			
Capuchin monkeys, birds, bats, etc.	Capuchin monkeys, birds, bats, etc.		
PREDATORS OF VERTEBRATES			
Jaguar, puma, ocelot, other medium and small felids	Jaguar, puma, ocelot, other medium and small felids		
PREDATORS OF INVERTEBRATES			
Spiders, anurans, lizards, birds, armadillos	Spiders, anurans, lizards, birds, armadillos	Spiders, anurans, lizards, birds, armadillos	Spiders, anurans, lizards, birds
SEED PREDATORS			
Primates, rodents	Rodents	Rodents	Rodents
HERBIVORES			
Deer, tapirs, peccaries, howler monkeys, iguanas, leaf-cutter ants	Deer, tapirs, peccaries, howler monkeys, iguanas, leaf-cutter ants	Howler monkeys, iguanas, leaf-cutter ants	Howler monkeys, iguanas, leaf-cutter ants

Table 2. Faunal and Floral Inventories for Guri Landmasses of Different Sizes

Parameter	Small Islands					
	S1	S2	S3	S4	S5	S6
Area (ha)	0.3	0.7	0.3	0.6	0.5	0.9
No. of stems > 1m tall < 1cm DBH/500m^2	42	85	65	63	54	76
Howler monkeys per ha	4.0	8.6	7.1	6.9	–	–
No. leaf cutter colonies per ha	4.0	4.3	6.7	6.7	4.0	4.1
Number of Agoutis per ha	0	0	0	0	0	0
Number of adult *Hymenaea courbaril* trees/ha	4.3	4.0	3.9	4.4	5.0	5.3
Number of *Hymenaea courbaril* seedlings > 1m tall < 1cm DBH/500m^2	4	3	0	1	3	2

Parameter	Medium Islands				Controls			
	M1	M2	M3	M4	C1	C2	C3	C4
Area (ha)	8	5	11	12	350	150		
No. of stems > 1m tall < 1cm DBH/500m^2	214	311	375	236	304	321	379	340
Howler monkeys per ha	1	0.2	0.00	0.03	0.2	0	1	0.01
No. leaf cutter colonies per ha	0.25	0.20	0.18	0.17	0.1	0.01	0.2	0.25
Number of Agoutis per ha	1.4	1.9	1.2	1.0	8.2	7.9	8.3	7.6
Number of adult *Hymenaea courbaril* trees/ha	8.7	9.6	7.9	8.5	9.4	8.2	8.9	8.3
Number of *Hymenaea courbaril* seedlings > 1m tall < 1cm DBH/500m^2	4	2	1	2	25	18	22	19

Relict animal populations persisting on **small** islands fall into three trophic categories: predators of invertebrates (spiders, anurans, lizards, and birds), seed predators (small rodents), and herbivores (howler monkeys, common iguanas, and leafcutter ants). Leaf-cutter ants are dominant herbivores with the potential to harvest vast quantities of foliage from seedlings, saplings, and mature adult trees. These highly selective herbivores, whose populations are assumed to be limited primarily by predation by armadillos, avoid plant species with high levels of tannins (Rao, 2000; Rao et al., 2001). Red howler monkeys (*Alouatta seniculus*) are also herbivorous, feeding predominantly on canopy foliage (Milton, 1981). In a study at another site with similar habitat quality, scientists found that increased howler monkey density resulted in increased tree growth rates, potentially due to accelerated nutrient cycling and increased above-ground plant productivity (Feeley and Terborgh, 2005). Howler monkeys have poor nutrient assimilation rates, and hence their waste excretions tend to be nutrient rich (Nagy and Milton, 1979) and are quickly returned to the system (<24 hours) and made available for plant uptake. The scientists also discovered that increased above-ground plant productivity led to a positive, indirect effect on bird species richness; (Feeley and Terborgh, In Press). These results highlight the potential for disparate taxonomic groups to be related through indirect interactions and trophic cascades.

Medium islands harbor, in addition to the above, armadillos (*Dasypus* sp., predators of young leaf-cutter ant colonies) and agoutis (*Dasyprocta* sp., the only known effective seed dispersers of *Hymenaea courbaril* trees). Armadillos are known to occur at low densities on large and medium islands but do not occur on small islands

Density information for two of the three dominant herbivores found on the Guri landmasses are given in Table 2. Iguana populations on the islands are difficult to estimate accurately, since they are usually found high in the canopy and are well-camouflaged.

Patterns of Species Loss: What is Missing and Why?

LEVEL I

In order to develop the islands for ecotourism and as an ecological research station the Department of Environment needs information on the ecological integrity of the islands. Specifically, your mandate is to provide an analysis of the patterns of species loss on the islands. Which species are missing, why, and what are the potential consequences of their loss? In your report, explain 1) the absence of certain trophic levels on medium and small islands in Lago Guri, and 2) the factors that tend to be associated with high extinction risk (Use Table 1)?

LEVEL II

Altered Trophic Interactions

Altered diversity of species and in particular, the absence of certain trophic levels on medium and small islands has resulted in highly aberrant communities. Such highly altered communities are ideally suited for scientific investigations as they provide an excellent opportunity to systematically examine the consequences of species loss.

1) As an illustrative example of the value of the site as an ecological research station, can you predict and compare variation in levels of herbivory across small, medium, and large Guri landmasses using Tables 1 and 2? Using data in table 2, estimate average densities of the two herbivores (leaf-cutter ants and howler monkeys).

2) Trophic cascades in tropical systems are of broad scientific interest given the complexity of species-rich tropical ecosystems. Can you identify one example of a potential cascade involving three trophic levels from the system described above?

LEVEL III

Agoutis and Hymenaea Courbaril Recruitment

The Department of Environment and Natural Resources in Venezuela is particularly interested in conserving healthy populations of the neotropical canopy tree *Hymenaea courbaril* (Caesalpiniaceae) at the site. Many populations of *H. courbaril* are declining in other parts of Venezuela due to unknown causes; hence, DENR is keen to protect and conserve the species at the Guri site and seeks your advice on how they can do this.

Ecology of H. courbaril

It is hypothesized that the tree depends on a small rodent –the agouti (*Dasyprocta* spp., Rodentia) for the effective dispersal of its seeds. Fruiting *H. courbaril* trees produce 100 to 500 indehiscent, hard, 10±20-cm long pods, each of which contains up to seven seeds. Paca (*Agouti paca*), peccaries (*Peccari* spp.) and tapirs (*Tapirus* spp.) can open *H. coubaril* pods, but all are seed predators and are unlikely to be efficient dispersers (Hallwachs 1986). Any seeds dispersed by these animals usually remain on the soil surface, available to other seed predators such as small and medium-sized rodents. In contrast, rodents of the genus *Dasyprocta* bury scatter-hoarded pods and seeds below the soil surface where they are relatively safe from predation (Asquith et al., 1999). Moreover, *Dasyprocta* spp. disperse *H. courbaril* pods up to at least 225 m. For this exercise, we will assume that *H. courbaril* recruitment is largely influenced by Agoutis, which are so critical for the regeneration of this tree that in the absence of *Dasyprocta, H. courbaril* would go extinct in areas where its other seed predators are present.

Questions

Based on the information on the ecology of *Hymenaea courbaril* trees and hypothetical abundance information for adults and saplings of *H. courbaril* in Table 2, can you:

1. Examine the abundance patterns of adults and seedlings of *H. courbaril* trees on the islands?
2. Provide one potential explanation for observed variation in seedling density of *H. courbaril* trees on small, medium and large Guri landmasses?
3. Based on your response to (a) and (b) above, what would be your advice on conserving populations of *H. courbaril* trees on small and medium Guri islands?

LEVEL IV

Ecotourism or Ecological Research Station or Both?

Based on your findings and responses to questions 1-3 above, write up a brief report to the DENR with your recommendation on whether the site is more suitable for ecotourism or as an ecological research station for scientific investigation, or a combination of both? Clearly state the factors that support your recommendation.

Discussion question:

What might some other factors be that could influence this decision?

LITERATURE CITED

Asquith, N.M., J. Terborgh, A. E. Arnold, and C. M. Riveros. 1999. The fruits the agouti ate: *Hymenaea courbaril* seed fate when its disperser is absent. Journal of Tropical Ecology 15:229-235.

Didham, R. K., J. H. Lawton, P.M. Hammond, and P. Eggleton. 1998. Trophic structure stability and extinction dynamics of beetles (Coleoptera) in tropical forest fragments. Philosophical Transactions of the Royal Society, London Biological Sciences 353:437-451.

Duarte, C.M. 2000. Marine biodiversity and ecosystem services: an elusive link. Journal of Experimental Marine Biology and Ecology 250:117-131.

Duffy, E. 2003. Biodiversity loss, trophic skew and ecosystem functioning. Ecology Letters 6:680-687.

Estes, J. A., D.O. Duggins, and G.B. Rathbun.1989. The ecology of extinctions in kelp forest communities. Conservation Biology 3:251-264.

Feeley, K. J., and J. W. Terborgh. 2005. Elevated densities of herbivores (Alouatta seniculus) affect soil nutrient concentrations and tree growth on islands in Lago Guri, Venezuela. Ecology 86:116–124.

Feeley, K.J. and J.W. Terborgh. 2006. Habitat fragmentation and effects of herbivore (Howler Monkey) abundances on bird species richness. Ecology 87(1): 144-150.

Hallwachs, W. 1986. Agoutis (*Dasyprocta punctata*): the inheritors of guapinol (*Hymenaea courbaril*: Leguminosae). Pages 285-304 in A. Estrada and T.H. Fleming, editors. Frugivores and seed dispersal. W. Junk Publishers, Dordrecht, Netherlands.

Larsen, T. H., N. M. Williams and C. Kremen. 2005. Extinction order and altered community structure rapidly disrupt ecosystem functioning. Ecology Letters 8:538-547.

Milton, K. 1981. Food choice and digestive strategies of two sympatric primate species. American Naturalist 117:496–505.

Nagy, K. A. and K. Milton. 1979. Energy metabolism and food consumption by wild howler monkeys (*Alouatta palliata*). Ecology 60:475–480.

Pimm, S. L., H.L. Jones, and J. Diamond. 1988. On the risk of extinction. American Naturalist 132: 757-785.

Purvis, A., P.M. Agapow, J.L. Gittleman, and G.M. Mace. 2000. Nonrandom extinction and the loss of evolutionary history. Science 288:328-330.

Rao, M. 2000. Variation in leaf-cutter ant (*Atta* sp.) densities in forest isolates: the potential role of predation. Journal of Tropical Ecology 16:209–225.

Rao, M., J. Terborgh, and P. Nuñez. 2001. Increased herbivory in forest isolates: implications for plant community structure and composition. Conservation Biology 15:624–633.

Redord, K. H. and P. Feinsinger. 2001. Sustainable use and the ecology of interactions. Pages 370-399 in J. D. Reynolds, G.M. Mace, K.H. Redford, and J.G., Robinson, editors. Conservation of Exploited Species. Cambridge University Press, Conservation Biology Series 6.

Terborgh, J., L. Lopez, P. Nunez V., M. Rao, G. Shahabuddin, G. Orihuela, M. Riveros, R. Ascanio, G. H. Adler, T. D. Lambert, and L. Balbas. 2001. Ecological meltdown in predator-free forest fragments. Science 294:1923–1926.

Terborgh, J., L. Lopez, J. Tello, D. Yu, and A. R. Bruni. 1997. Transitory states in relaxing ecosystems of land bridge islands. Pages 256–274 in W. F. Laurence and R. O. Bierregaard, editors. Tropical forest remnants. University of Chicago Press, Chicago, Illinois, USA.

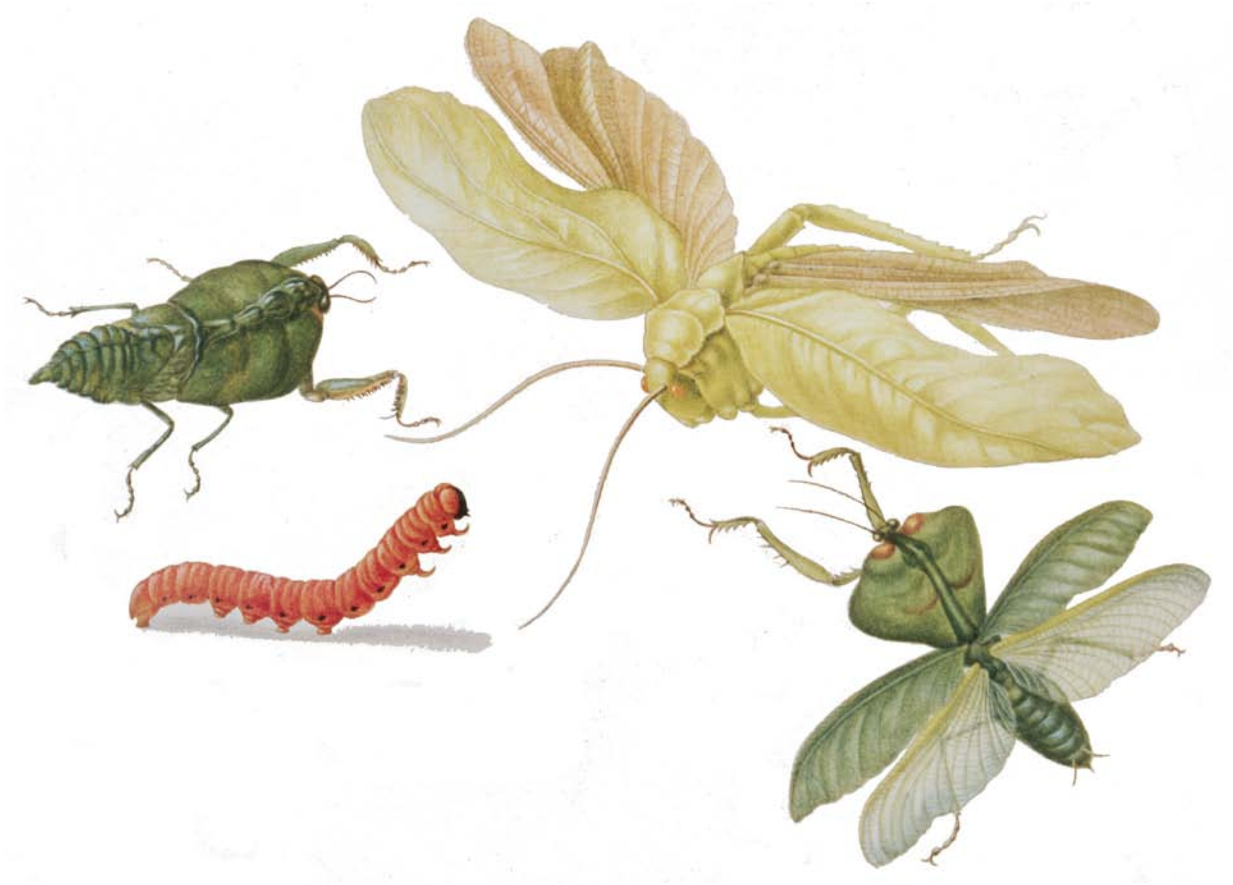
TERMS OF USE

Reproduction of this material is authorized by the recipient institution for non-profit/non-commercial educational use and distribution to students enrolled in course work at the institution. Distribution may be made by photocopying or via the institution's intranet restricted to enrolled students. Recipient agrees not to make commercial use, such as, without limitation, in publications distributed by a commercial publisher, without the prior express written consent of AMNH.

All reproduction or distribution must provide both full citation of the original work, and a copyright notice as follows:

"Madhu Rao. Ecological Consequences of Extinction. 2010. Exercise. American Museum of Natural History, Lessons in Conservation. Available at http://ncep.amnh.org/linc."

This material is based on work supported by the National Science Foundation under the Course, Curriculum and Laboratory Improvement program (NSF 0127506).



A Brief Tutorial on Maxent

Steven Phillips

AT&T Labs-Research, Florham Park, NJ, U.S.A., email phillips@research.att.com

S. Laube

A Brief Tutorial on Maxent

Steven Phillips

INTRODUCTION

This tutorial gives a basic introduction to use of the Maxent program for maximum entropy modelling of species' geographic distributions, written by Steven Phillips, Miro Dudik, and Rob Schapire, with support from AT&T Labs-Research, Princeton University, and the Center for Biodiversity and Conservation, American Museum of Natural History. For more details on the theory maximum entropy modeling as well as a description of the data used and the main types of statistical analysis used here, see:

Phillips, S.J., R.P. Anderson, and R.E. Schapire. 2006. Maximum entropy modeling of species geographic distributions. Ecological Modelling 190: 3-4 pp 231-259.

A second paper describing more recently-added features of the Maxent software is:

Phillips, S.J. and M. Dudik. 2008. Modeling of species distributions with Maxent: new extensions and a comprehensive evaluation. Ecography (31)2: 161-175.

The environmental data we will use consist of climatic and elevational data for South America, together with a potential vegetation layer. Our sample species will be *Bradypus variegatus*, the brown-throated three-toed sloth. This tutorial will assume that all the data files are located in the same directory as the Maxent program files; otherwise you will need to use the path (e.g., c:\data\maxent\tutorial) in front of the file names used here.

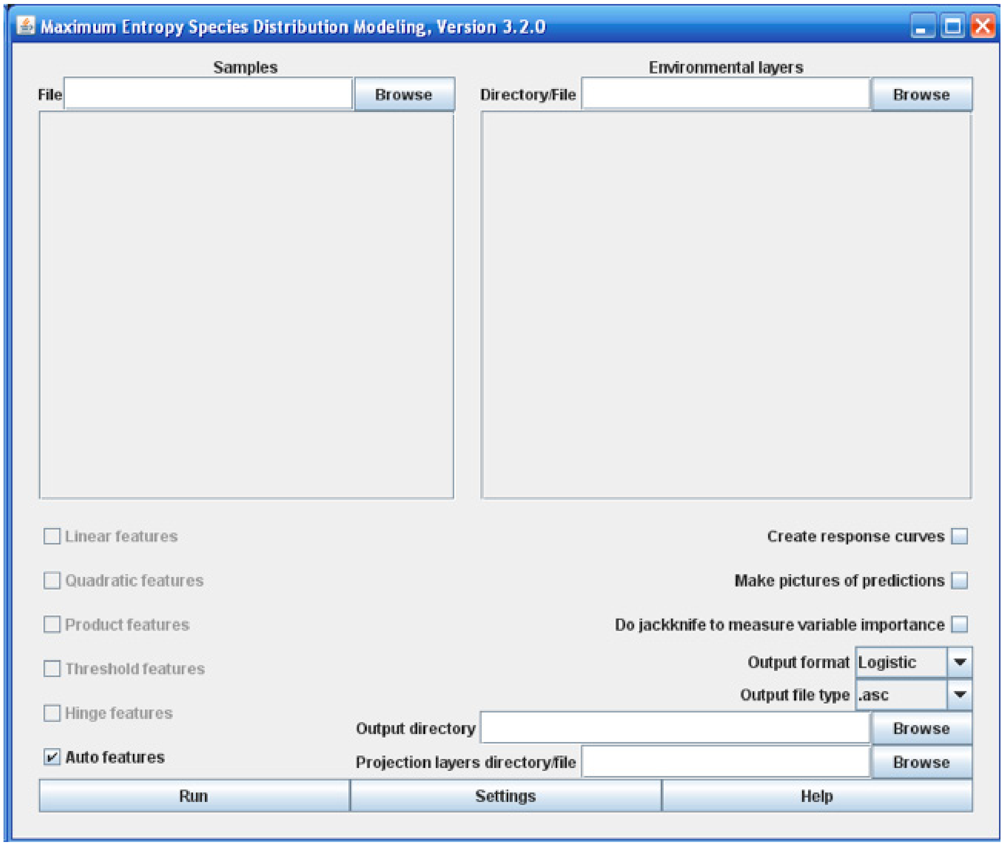
GETTING STARTED

Downloading

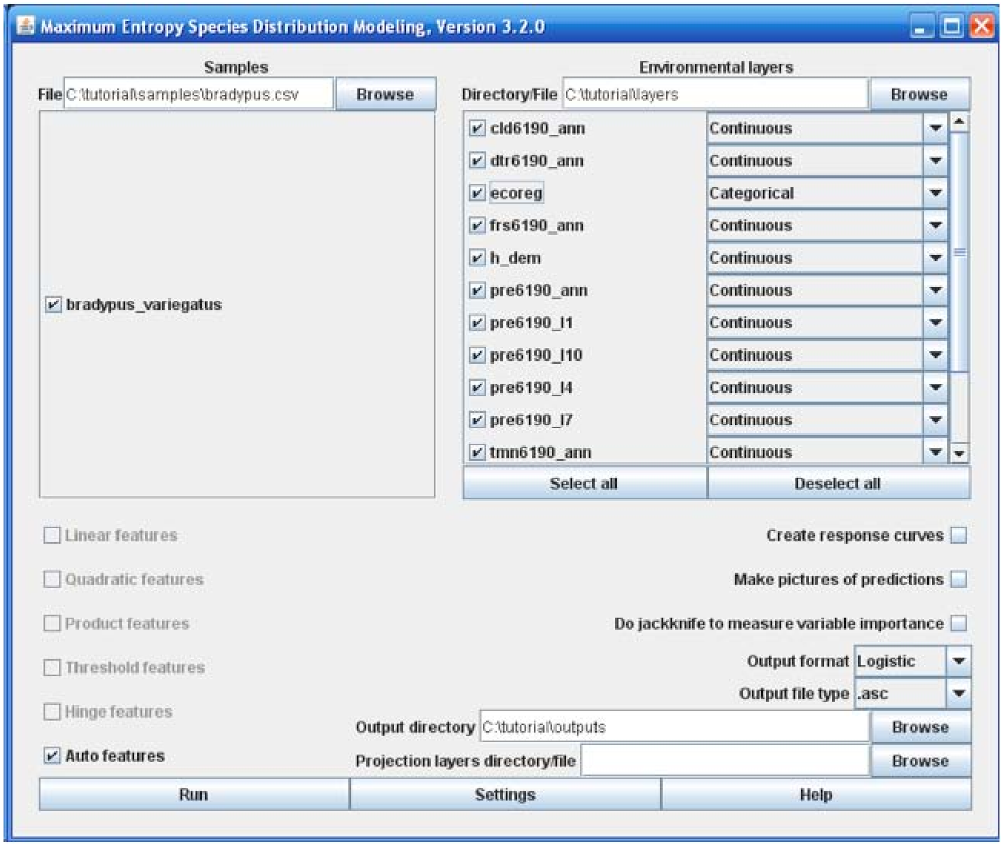
The software consists of a jar file, maxent.jar, which can be used on any computer running Java version 1.4 or later. Maxent can be downloaded, along with associated literature, from www.cs.princeton.edu/~schapire/maxent; the Java runtime environment can be obtained from java.sun.com/javase/downloads. If you are using Microsoft Windows (as we assume here), you should also download the file maxent.bat, and save it in the same directory as maxent.jar. The website has a file called "readme.txt", which contains instructions for installing the program on your computer.

Firing up

If you are using Microsoft Windows, simply click on the file maxent.bat. Otherwise, enter "java -mx512m -jar maxent.jar" in a command shell (where "512" can be replaced by the megabytes of memory you want made available to the program). The following screen will appear:



To perform a run, you need to supply a file containing presence localities ("samples"), a directory containing environmental variables, and an output directory. In our case, the presence localities are in the file "samples\bradypus.csv", the environmental layers are in the directory "layers", and the outputs are going to go in the directory "outputs". You can enter these locations by hand, or browse for them. While browsing for the environmental variables, remember that you are looking for the directory that contains them – you don't need to browse down to the files in the directory. After entering or browsing for the files for *Bradypus*, the program looks like this:



File C:\tutorial\samples\bradypus.csv

☑ bradypus_variegatus

Directory/File C:\tutorial\layers

☑ cld6190_ann	Continuous
☑ dtr6190_ann	Continuous
☑ ecoreg	Categorical
☑ frs6190_ann	Continuous
☑ h_dem	Continuous
☑ pre6190_ann	Continuous
☑ pre6190_l1	Continuous
☑ pre6190_l10	Continuous
☑ pre6190_l4	Continuous
☑ pre6190_l7	Continuous
☑ tmn6190_ann	Continuous

☐ Linear features
☐ Quadratic features
☐ Product features
☐ Threshold features
☐ Hinge features
☑ Auto features

Create response curves ☐
Make pictures of predictions ☐
Do jackknife to measure variable importance ☐
Output format Logistic
Output file type .asc
Output directory C:\tutorial\outputs
Projection layers directory/file

The file "samples\bradypus.csv" contains the presence localities in .csv format. The first few lines are as follows:

```
species,longitude,latitude
bradypus_variegatus,-65.4,-10.3833
bradypus_variegatus,-65.3833,-10.3833
bradypus_variegatus,-65.1333,-16.8
bradypus_variegatus,-63.6667,-17.45
bradypus_variegatus,-63.85,-17.4
```

There can be multiple species in the same samples file, in which case more species would appear in the panel, along with *Bradypus*. Coordinate systems other than latitude and longitude can be used provided that the samples file and environmental layers use the same coordinate system. The "x" coordinate (longitude, in our case) should come before the "y" coordinate (latitude) in the samples file. If the presence data has duplicate records (multiple records for the same species in the same grid cell), the duplicates can be removed by clicking on the "Settings" button and selecting "Delete duplicates".

The directory "layers" contains a number of ascii raster grids (in ESRI's .asc format), each of which describes an environmental variable. The grids must all have the same geographic bounds and cell size (i.e. all the ascii file headings must match each other perfectly). One of our variables, "ecoreg", is a categorical variable describing potential vegetation classes. The categories must be indicated by numbers, rather than letters or words. You must tell the program which variables are categorical, as has been done in the picture above.

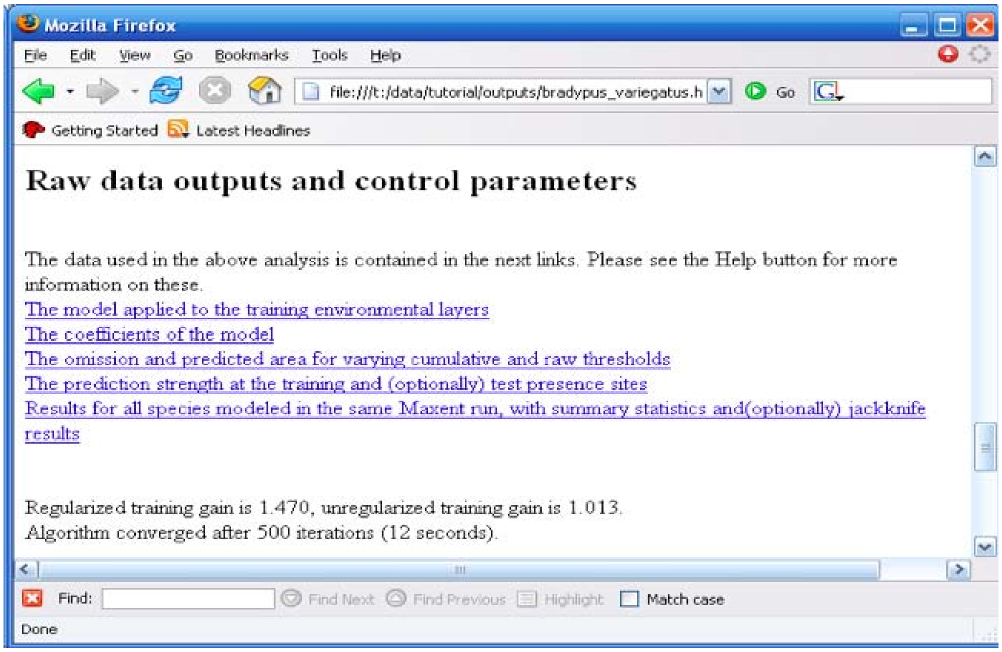
DOING A RUN

Simply press the "Run" button. A progress monitor describes the steps being taken. After the environmental layers are loaded and some initialization is done, progress towards training of the maxent model is shown like this:

bradypus_variegatus: Gain is 1.456168

The gain is closely related to deviance, a measure of goodness of fit used in generalized additive and generalized linear models. It starts at 0 and increases towards an asymptote during the run. During this process, Maxent is generating a probability distribution over pixels in the grid, starting from the uniform distribution and repeatedly improving the fit to the data. The gain is defined as the average log probability of the presence samples, minus a constant that makes the uniform distribution have zero gain. At the end of the run, the gain indicates how closely the model is concentrated around the presence samples; for example, if the gain is 2, it means that the average likelihood of the presence samples is exp(2) ≈ 7.4 times higher than that of a random background pixel. Note that Maxent isn't directly calculating "probability of occurrence". The probability it assigns to each pixel is typically very small, as the values must sum to 1 over all the pixels in the grid (though we return to this point when we compare output formats).

The run produces multiple output files, of which the most important for analyzing your model is an html file called "bradypus.html". Part of this file gives pointers to the other outputs, like this:



Raw data outputs and control parameters

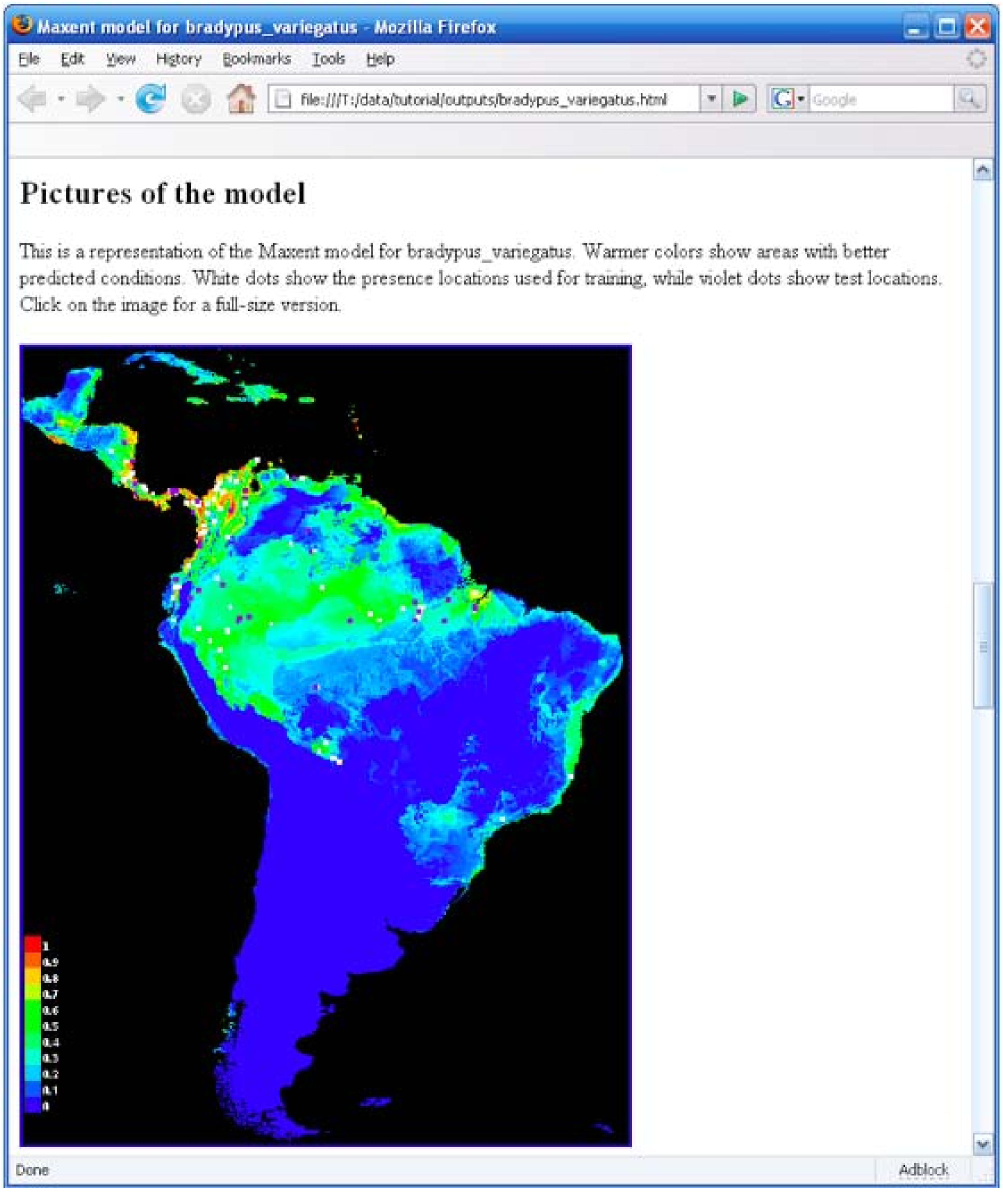
The data used in the above analysis is contained in the next links. Please see the Help button for more information on these.

The model applied to the training environmental layers
The coefficients of the model
The omission and predicted area for varying cumulative and raw thresholds
The prediction strength at the training and (optionally) test presence sites
Results for all species modeled in the same Maxent run, with summary statistics and(optionally) jackknife results

Regularized training gain is 1.470, unregularized training gain is 1.013.
Algorithm converged after 500 iterations (12 seconds).

LOOKING AT A PREDICTION

To see what other (more interesting) output there can be in bradpus.html, we will turn on a couple of options and rerun the model. Press the "Make pictures of predictions" button, then click on "Settings", and type "25" in the "Random test percentage" entry. Then, press the "Run" button again. After the run completes, the file bradypus.html contains a picture like this:



Pictures of the model

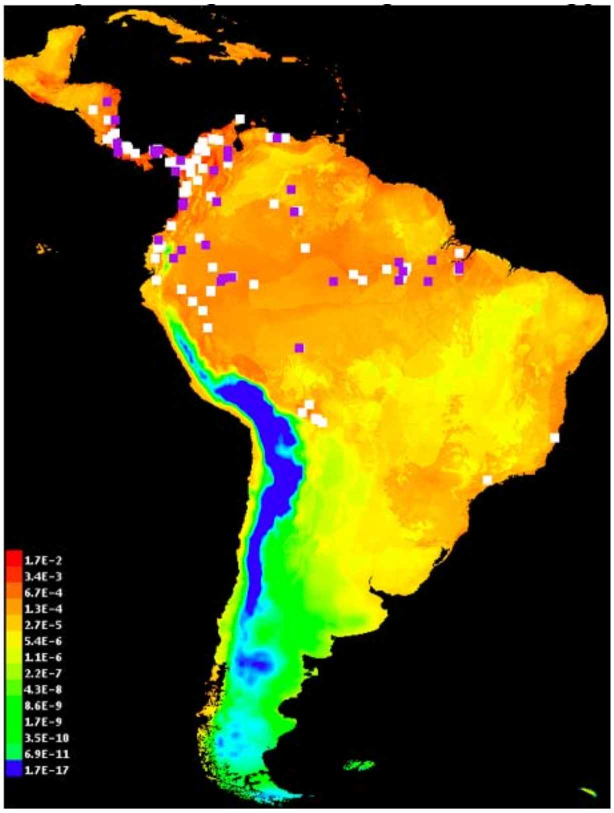
This is a representation of the Maxent model for bradypus_variegatus. Warmer colors show areas with better predicted conditions. White dots show the presence locations used for training, while violet dots show test locations. Click on the image for a full-size version.

The image uses colors to indicate predicted probability that conditions are suitable, with red indicating high probability of suitable conditions for the species, green indicating conditions typical of those where the species is found, and lighter shades of blue indicating low predicted probability of suitable conditions. For *Bradypus*, we see that suitable conditions are predicted to be highly probable through most of lowland Central America, wet lowland areas of northwestern South America, the Amazon basin, Caribean islands, and much of the Atlantic forests in south-eastern Brazil. The file pointed to is an image file (.png) that you can just click on (in Windows) or open in most image processing software. If you want to copy these images, or want to open them with other software, you will find the .png files in the directory called "plots" that has been created as an output during the run.

The test points are a random sample taken from the species presence localities. The same random sample is used each time you run Maxent on the same data set, unless you select the "random seed" option on the settings panel. Alternatively, test data for one or more species can be provided in a separate file, by giving the name of a "Test sample file" in the Settings panel.

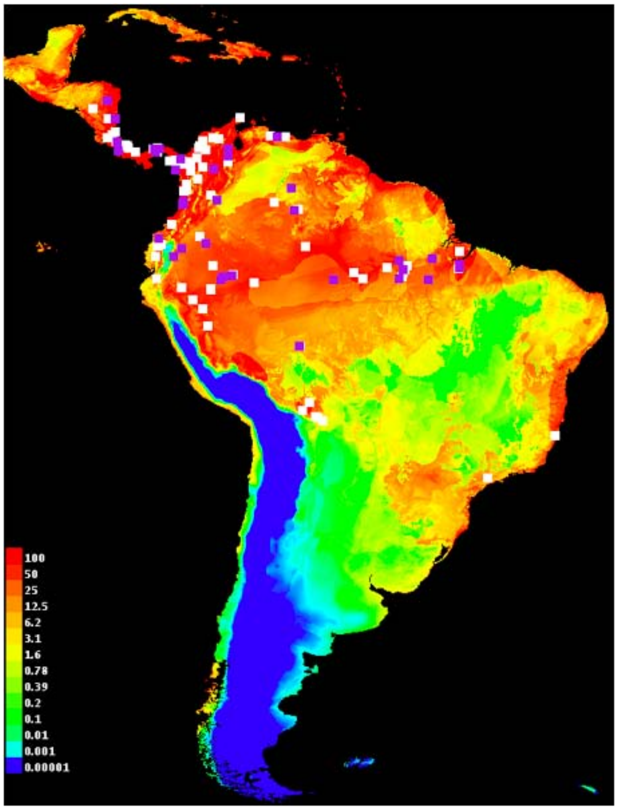
OUTPUT FORMATS

Maxent supports three output formats for model values: raw, cumulative, and logistic. First, the raw output is just the Maxent exponential model itself. Second, the cumulative value corresponding to a raw value of r is the percentage of the Maxent distribution with raw value at most r. Cumulative output is best interpreted in terms of predicted omission rate: if we set a cumulative threshold of c, the resulting binary prediction would have omission rate c% on samples drawn from the Maxent distribution itself, and we can predict a similar omission rate for samples drawn from the species distribution. Third, if c is the exponential of the entropy of the maxent distribution, then the logistic value corresponding to a raw value of r is $c \cdot r/(1+c \cdot r)$. This is a logistic function, because the raw value is an exponential function of the environmental variables. The three output formats are all monotonically related, but they are scaled differently, and have different interpretations. The default output is logistic, which is the easiest to conceptualize: it gives an estimate between 0 and 1 of probability of presence. Note that probability of presence depends on details of the sampling design, such as the plot size and (for vagile organisms) observation time; logistic output estimates probability of presence assuming that the sampling design is such that typical presence localities have probability of presence of about 0.5. The picture of the *Bradypus* model above uses the logistic format. In comparison, using the raw format gives the following picture:



Note that we have used a logarithmic scale for the colors. A linear scale would be mostly blue, with a few red pixels (you can verify this by deselecting "Logscale pictures" on the Settings panel) since the raw format typically gives a small number of sites relatively large values – this can be thought of as an artifact of the raw output being given by an exponential distribution.

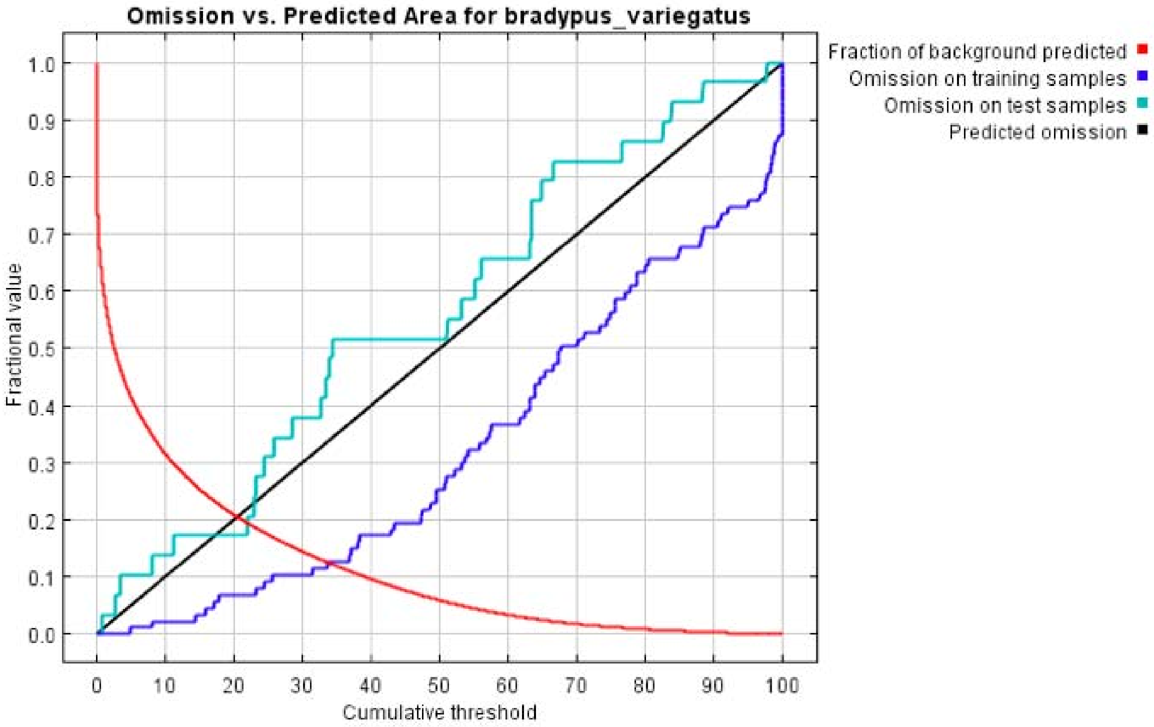
Using the cumulative output format gives the following picture:



As with the raw output, we have used a logarithmic scale for coloring the picture in order to emphasize differences between smaller values. Cumulative output can be interpreted as predicting suitable conditions for the species above a threshold in the approximate range of 1-20 (or yellow through orange, in this picture), depending on the level of predicted omission that is acceptable for the application.

STATISTICAL ANALYSIS

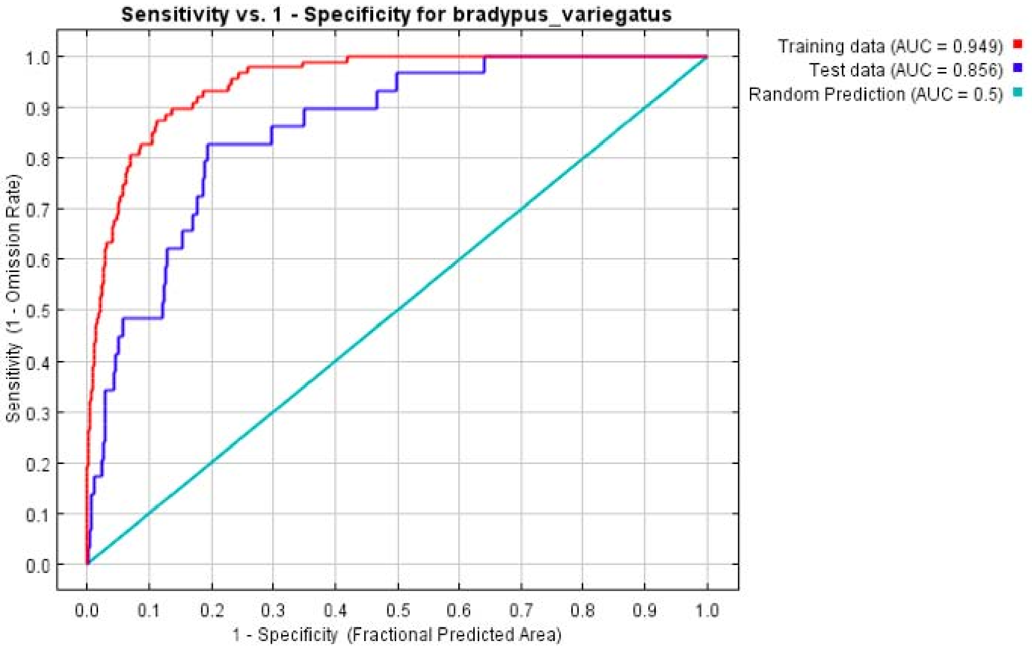
The “25” we entered for “random test percentage” told the program to randomly set aside 25% of the sample records for testing. This allows the program to do some simple statistical analysis. Much of the analysis made use of a threshold to make a binary prediction, with suitable conditions predicted above the threshold and unsuitable below. The first plot shows how testing and training omission and predicted area vary with the choice of cumulative threshold, as in the following graph:



Here, we see that the omission on test samples is a very good match to the predicted omission rate; the omission rate for test data drawn from the Maxent distribution itself. The predicted omission rate is a straight line, by definition of the cumulative output format. In some situations, the test omission line lies well below the predicted omission line: a common reason is that the test and training data are not independent, for example if they derive from the same spatially-autocorrelated presence data.

The next plot gives the receiver operating curve for both training and test data, shown below. The area under the ROC curve (AUC) is also given here; if test data are available, the standard error of the AUC on the test data is given later on in the web page.

If you use the same data for training and for testing then the red and blue lines will be identical. If you split your data into two partitions, one for training and one for testing

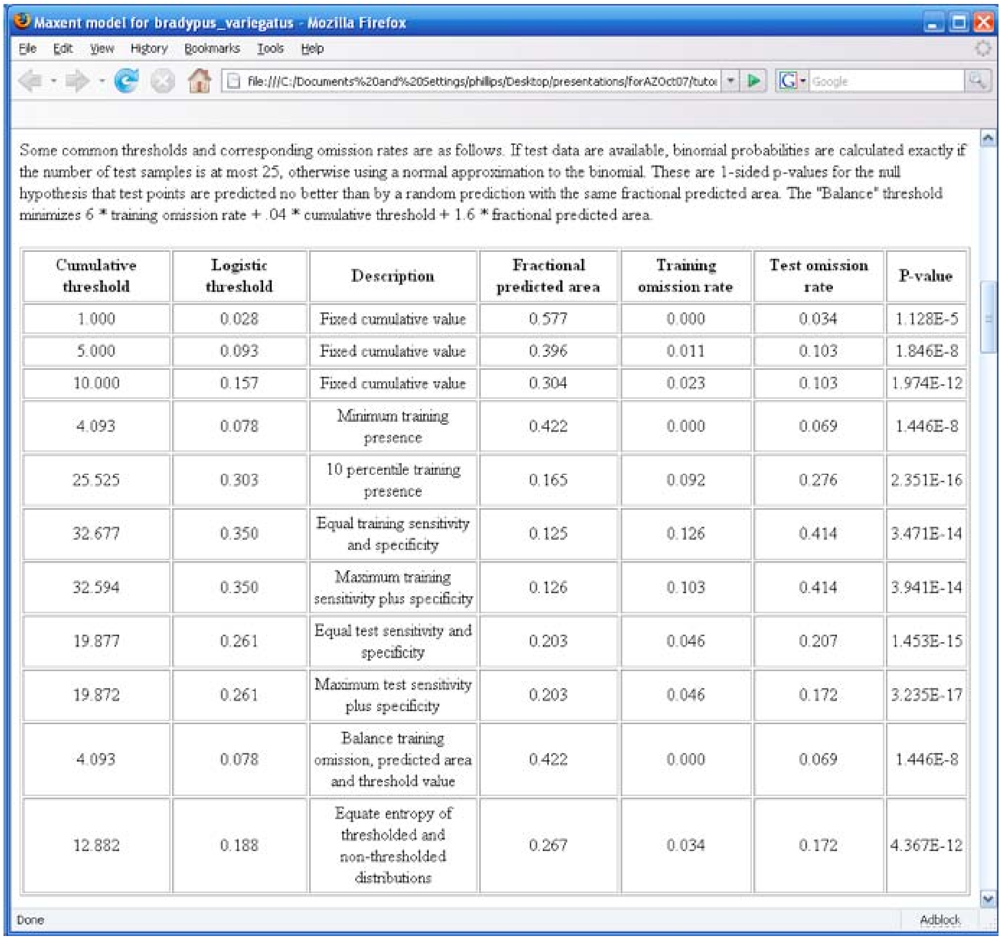


it is normal for the red (training) line to show a higher AUC than the blue (testing) line. The red (training) line shows the "fit" of the model to the training data. The blue (testing) line indicates the fit of the model to the testing data, and is the real test of the models predictive power. The turquoise line shows the line that you would expect if your model was no better than random. If the blue line (the test line) falls below the turquoise line then this indicates that your model performs worse than a random model would. The further towards the top left of the graph that the blue line is, the better the model is at predicting the presences contained in the test sample of the data. For more detailed information on the AUC statistic a good starting reference is:

Fielding, A.H. and J.F. Bell. 2007. A review of methods for the assessment of prediction errors in conservation presence/absence models. Environmental Conservation 24(1): 38-49.

Because we have only occurrence data and no absence data, "fractional predicted area" (the fraction of the total study area predicted present) is used instead of the more standard commission rate (fraction of absences predicted present). For more discussion of this choice, see the paper in *Ecological Modelling* mentioned on page 1 of this tutorial. It is important to note that AUC values tend to be higher for species with narrow ranges, relative to the study area described by the environmental data. This does not necessarily mean that the models are better; instead this behavior is an artifact of the AUC statistic.

If test data are available, the program automatically calculates the statistical significance of the prediction, using a binomial test of omission. For *Bradypus*, this gives:



Some common thresholds and corresponding omission rates are as follows. If test data are available, binomial probabilities are calculated exactly if the number of test samples is at most 25, otherwise using a normal approximation to the binomial. These are 1-sided p-values for the null hypothesis that test points are predicted no better than by a random prediction with the same fractional predicted area. The "Balance" threshold minimizes 6 * training omission rate + .04 * cumulative threshold + 1.6 * fractional predicted area.

Cumulative threshold	Logistic threshold	Description	Fractional predicted area	Training omission rate	Test omission rate	P-value
1.000	0.028	Fixed cumulative value	0.577	0.000	0.034	1.128E-5
5.000	0.093	Fixed cumulative value	0.396	0.011	0.103	1.846E-8
10.000	0.157	Fixed cumulative value	0.304	0.023	0.103	1.974E-12
4.093	0.078	Minimum training presence	0.422	0.000	0.069	1.446E-8
25.525	0.303	10 percentile training presence	0.165	0.092	0.276	2.351E-16
32.677	0.350	Equal training sensitivity and specificity	0.125	0.126	0.414	3.471E-14
32.594	0.350	Maximum training sensitivity plus specificity	0.126	0.103	0.414	3.941E-14
19.877	0.261	Equal test sensitivity and specificity	0.203	0.046	0.207	1.453E-15
19.872	0.261	Maximum test sensitivity plus specificity	0.203	0.046	0.172	3.235E-17
4.093	0.078	Balance training omission, predicted area and threshold value	0.422	0.000	0.069	1.446E-8
12.882	0.188	Equate entropy of thresholded and non-thresholded distributions	0.267	0.034	0.172	4.367E-12

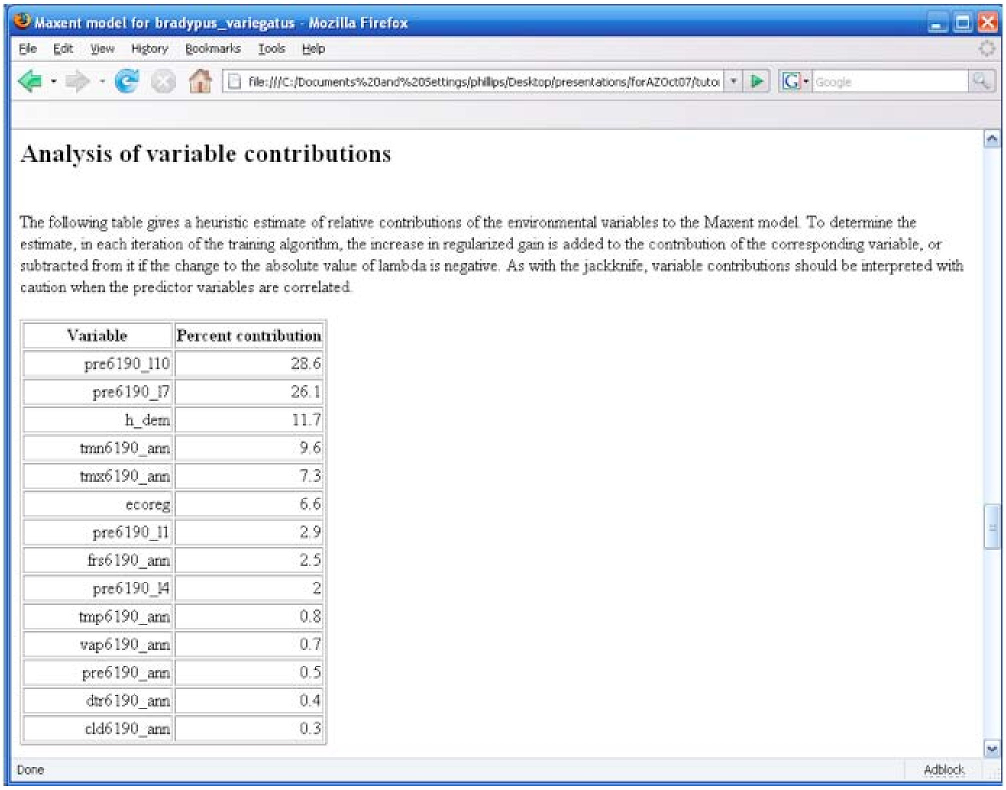
For more detailed information on the binomial statistic, see the *Ecological Modelling* paper mentioned above.

WHICH VARIABLES MATTER MOST?

A natural application of species distribution modeling is to answer the question, which variables matter most for the species being modeled? There is more than one way to answer this question; here we outline the possible ways in which Maxent can be used to address it.

While the Maxent model is being trained, we can keep track of which environmental variables are making the greatest contribution to the model. Each step of the Maxent algorithm increases the gain of the model by modifying the coefficient for a single fea–

ture; the program assigns the increase in the gain to the environmental variable(s) that the feature depends on. Converting to percentages at the end of the training process, we get the following table:



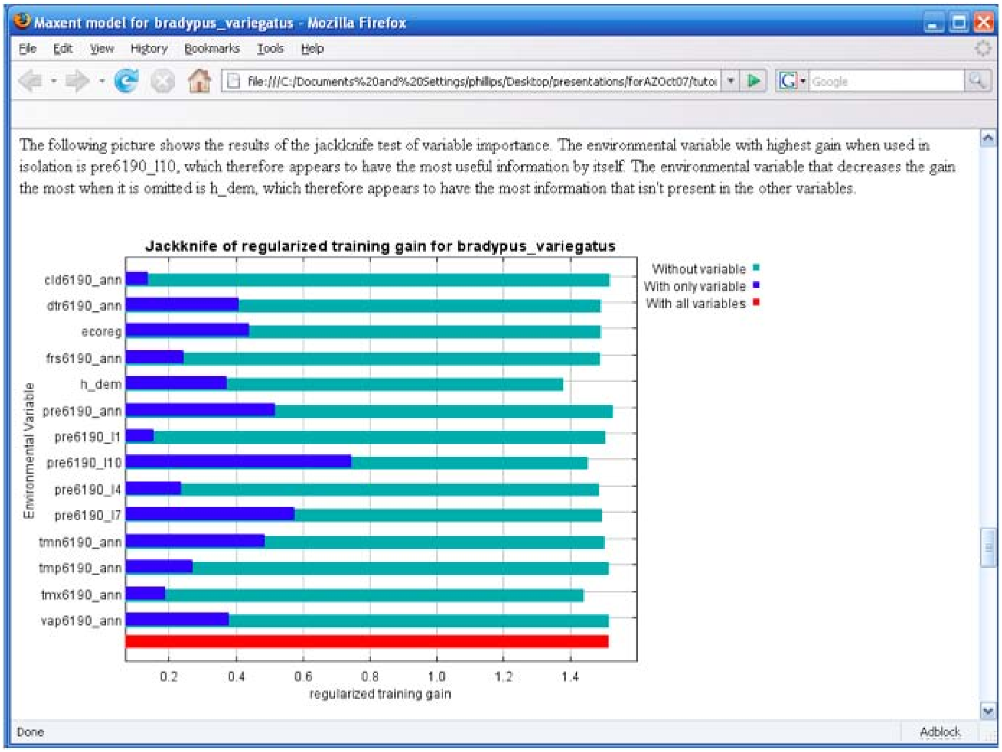
Analysis of variable contributions

The following table gives a heuristic estimate of relative contributions of the environmental variables to the Maxent model. To determine the estimate, in each iteration of the training algorithm, the increase in regularized gain is added to the contribution of the corresponding variable, or subtracted from it if the change to the absolute value of lambda is negative. As with the jackknife, variable contributions should be interpreted with caution when the predictor variables are correlated.

Variable	Percent contribution
pre6190_l10	28.6
pre6190_l7	26.1
h_dem	11.7
tmn6190_ann	9.6
tmx6190_ann	7.3
ecoreg	6.6
pre6190_l1	2.9
frs6190_ann	2.5
pre6190_l4	2
tmp6190_ann	0.8
vap6190_ann	0.7
pre6190_ann	0.5
dtr6190_ann	0.4
cld6190_ann	0.3

These percent contribution values are only heuristically defined: they depend on the particular path that the Maxent code uses to get to the optimal solution, and a different algorithm could get to the same solution via a different path, resulting in different percent contribution values. In addition, when there are highly correlated environmental variables, the percent contributions should be interpreted with caution. In our *Bradypus* example, annual precipitation is highly correlated with October and July precipitation. Although the above table shows that Maxent used the October precipitation variable more than any other, and hardly used annual precipitation at all, this does not necessarily imply that October precipitation is far more important to the species than annual precipitation.

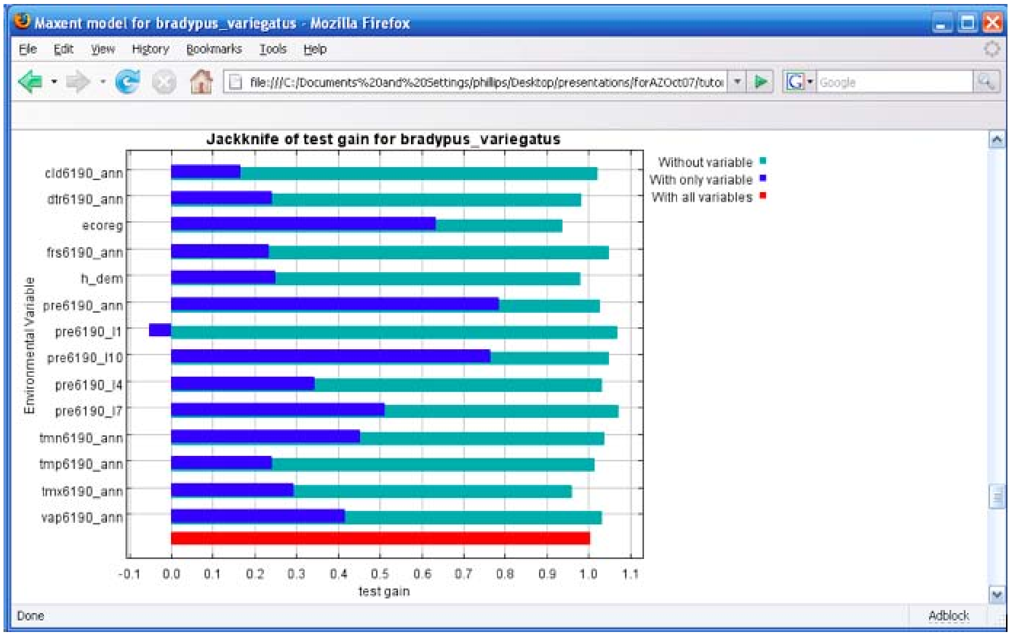
To get alternate estimates of which variables are most important in the model, we can also run a jackknife test by selecting the "Do jackknife to measure variable important" checkbox. When we press the "Run" button again, a number of models are created. Each variable is excluded in turn, and a model created with the remaining variables. Then a model is created using each variable in isolation. In addition, a model is created using all variables, as before. The results of the jackknife appear in the "bradypus.html" files in three bar charts, and the first of these is shown below.

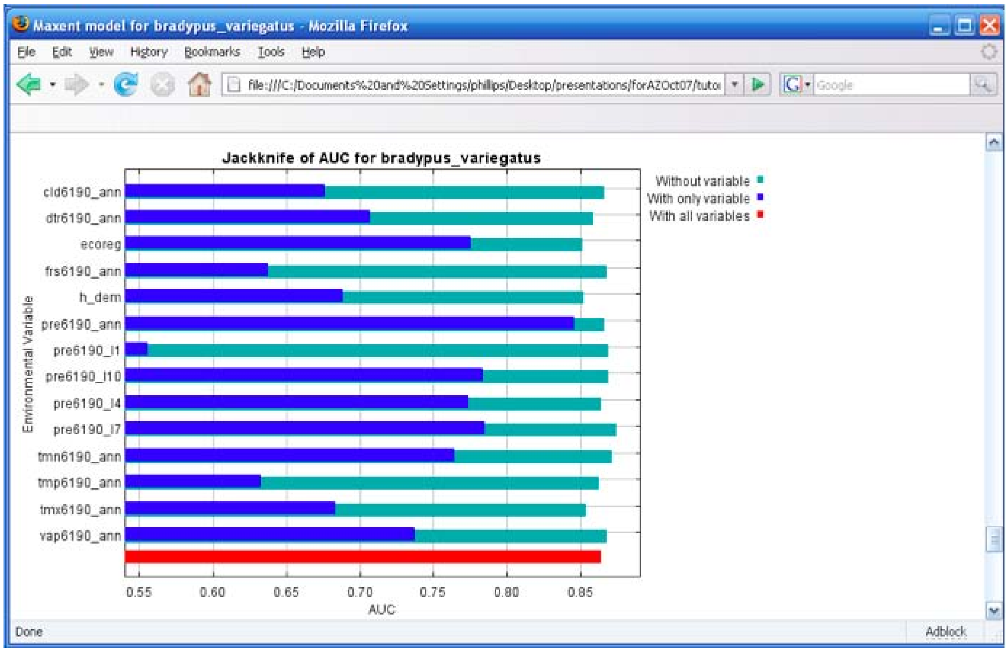


The following picture shows the results of the jackknife test of variable importance. The environmental variable with highest gain when used in isolation is pre6190_l10, which therefore appears to have the most useful information by itself. The environmental variable that decreases the gain the most when it is omitted is h_dem, which therefore appears to have the most information that isn't present in the other variables.

We see that if Maxent uses only pre6190_l1 (average January rainfall) it achieves almost no gain, so that variable is not (by itself) useful for estimating the distribution of *Bradypus*. On the other hand, October rainfall (pre6190_l10) allows a reasonably good fit to the training data. Turning to the lighter blue bars, it appears that no variable contains a substantial amount of useful information that is not already contained in the other variables, because omitting each variable in turn did not decrease the training gain considerably.

The bradypus.html file has two more jackknife plots, which use either test gain or AUC in place of training gain, shown below.





Comparing the three jackknife plots can be very informative. The AUC plot shows that annual precipitation (pre6190_ann) is the most effective single variable for predicting the distribution of the occurrence data that was set aside for testing, when predictive performance is measured using AUC, even though it was hardly used by the model built using all variables. The relative importance of annual precipitation also increases in the test gain plot, when compared against the training gain plot. In addition, in the test gain and AUC plots, some of the light blue bars (especially for the monthly precipitation variables) are longer than the red bar, showing that predictive performance improves when the corresponding variables are not used.

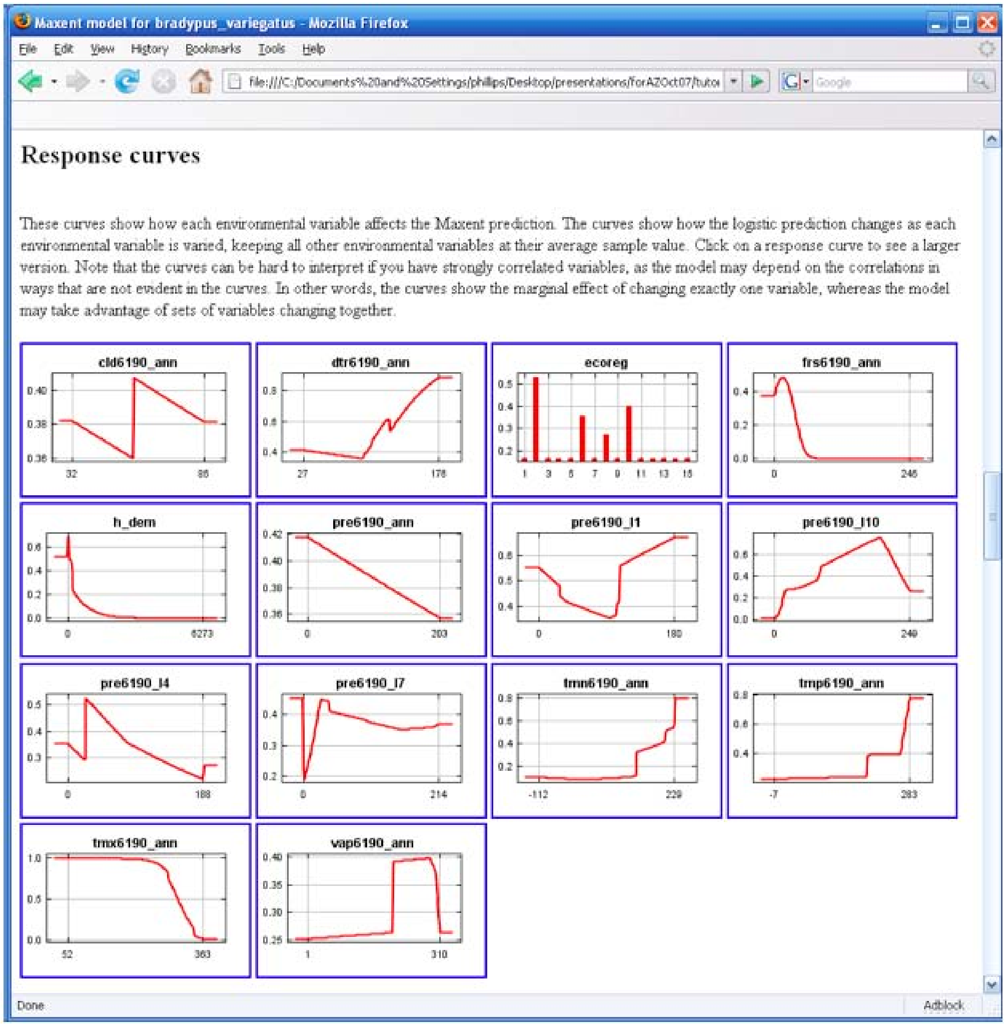
This tells us that monthly precipitation variables are helping Maxent to obtain a good fit to the training data, but the annual precipitation variable generalizes better, giving comparatively better results on the set-aside test data. Phrased differently, models made with the monthly precipitation variables appear to be less transferable. This is important if our goal is to transfer the model, for example by applying the model to future climate variables in order to estimate its future distribution under climate change. It makes sense that monthly precipitation values are less transferable: likely suitable conditions for *Bradypus* will depend not on precise rainfall values in selected months, but on the aggregate average rainfall, and perhaps on rainfall consistency or lack of extended dry periods. When we are modeling on a continental scale, there will probably be shifts in the precise timing of seasonal rainfall patterns, affecting the monthly precipitation but not suitable conditions for *Bradypus*.

In general, it would be better to use variables that are more likely to be directly relevant to the species being modeled. For example, the Worldclim website (www.worldclim.org) provides "BIOCLIM" variables, including derived variables such as "rainfall in the wettest quarter", rather than monthly values.

A last note on the jackknife outputs: the test gain plot shows that a model made only with January precipitation (pre6190_l1) results in a negative test gain. This means that the model is slightly worse than a null model (i.e., a uniform distribution) for predicting the distribution of occurrences set aside for testing. This can be regarded as more evidence that the monthly precipitation values are not the best choice for predictor variables.

HOW DOES THE PREDICTION DEPEND ON THE VARIABLES?

Now press the "Create response curves", deselect the jackknife option, and rerun the model. This results in the following section being added to the "bradypus.html" file:

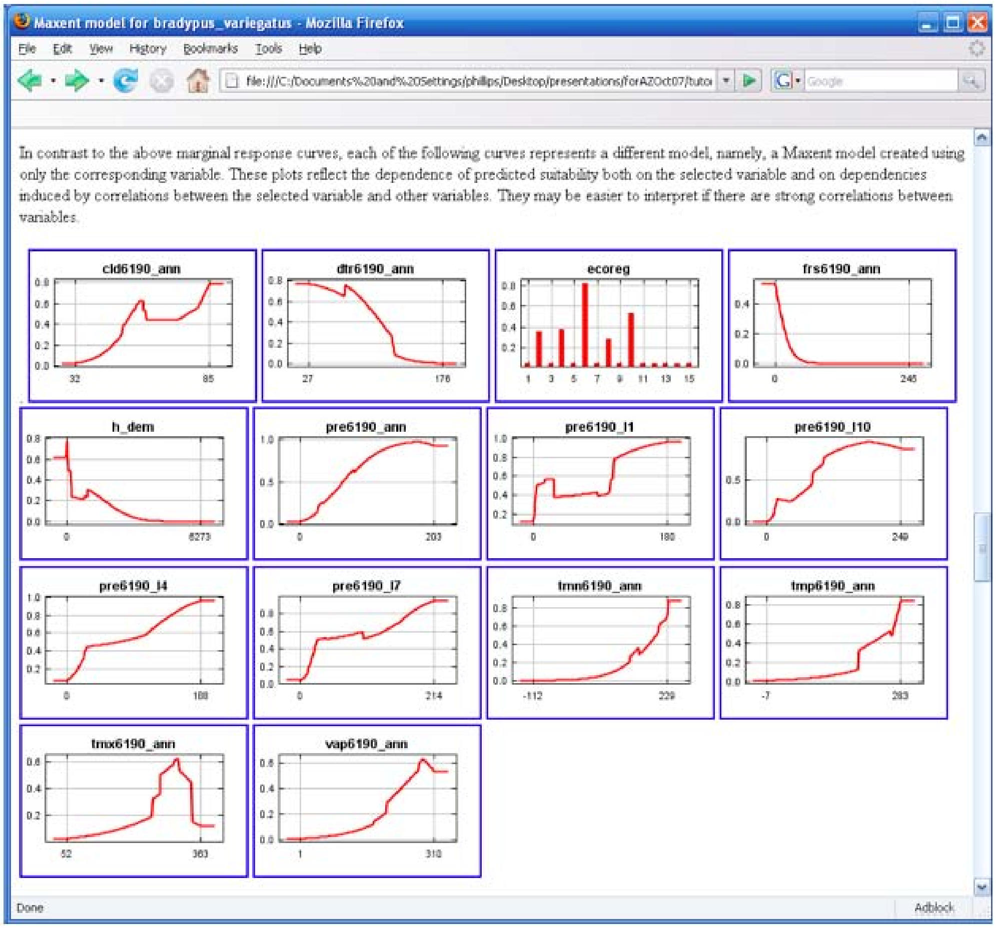


Response curves

These curves show how each environmental variable affects the Maxent prediction. The curves show how the logistic prediction changes as each environmental variable is varied, keeping all other environmental variables at their average sample value. Click on a response curve to see a larger version. Note that the curves can be hard to interpret if you have strongly correlated variables, as the model may depend on the correlations in ways that are not evident in the curves. In other words, the curves show the marginal effect of changing exactly one variable, whereas the model may take advantage of sets of variables changing together.

Each of the thumbnail images can be selected (by clicking on them) to obtain a more detailed plot, and if you would like to copy or open these plots with other software, the .png files can be found in the "plots" directory. Looking at vap6190_ann, we see that the response is low for values of vap6190_ann in the range 1-200, and is higher for values in the range 200-300. The value shown on the y-axis is predicted probability of suitable conditions, as given by the logistic output format, with all other variables set to their average value over the set of presence localities.

Note that if the environmental variables are correlated, as they are here, the marginal response curves can be misleading. For example, if two closely correlated variables have response curves that are near opposites of each other, then for most pixels, the combined effect of the two variables may be small. As another example, we see that predicted suitability is negatively correlated with annual precipitation (pre6190_ann), if all other variables are held fixed. In other words, once the effect of all the other variables has already been accounted for, the marginal effect of increasing annual precipitation is to decrease predicted suitability. However, annual precipitation is highly correlated with the monthly precipitation variables, so in reality we cannot easily hold the monthly values fixed while varying the annual value. The program therefore produces a second set of response curves, in which each curve is made by generating a model using only the corresponding variable, disregarding all other variables:

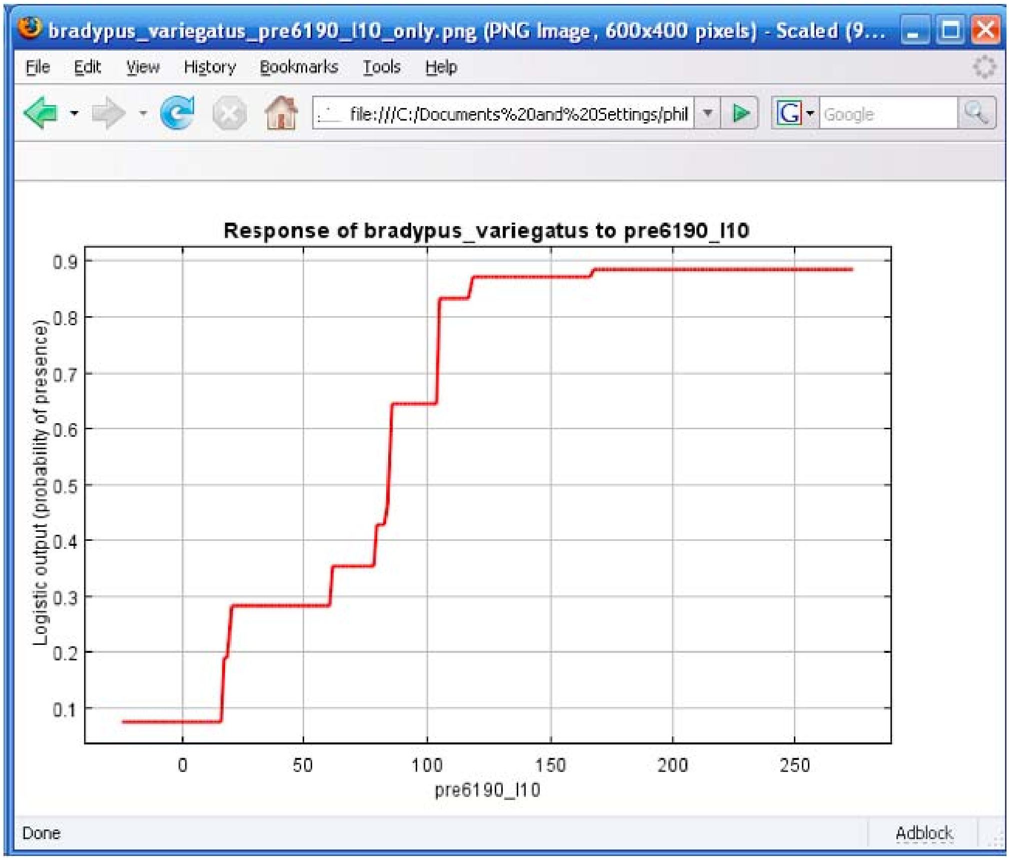


In contrast to the above marginal response curves, each of the following curves represents a different model, namely, a Maxent model created using only the corresponding variable. These plots reflect the dependence of predicted suitability both on the selected variable and on dependencies induced by correlations between the selected variable and other variables. They may be easier to interpret if there are strong correlations between variables.

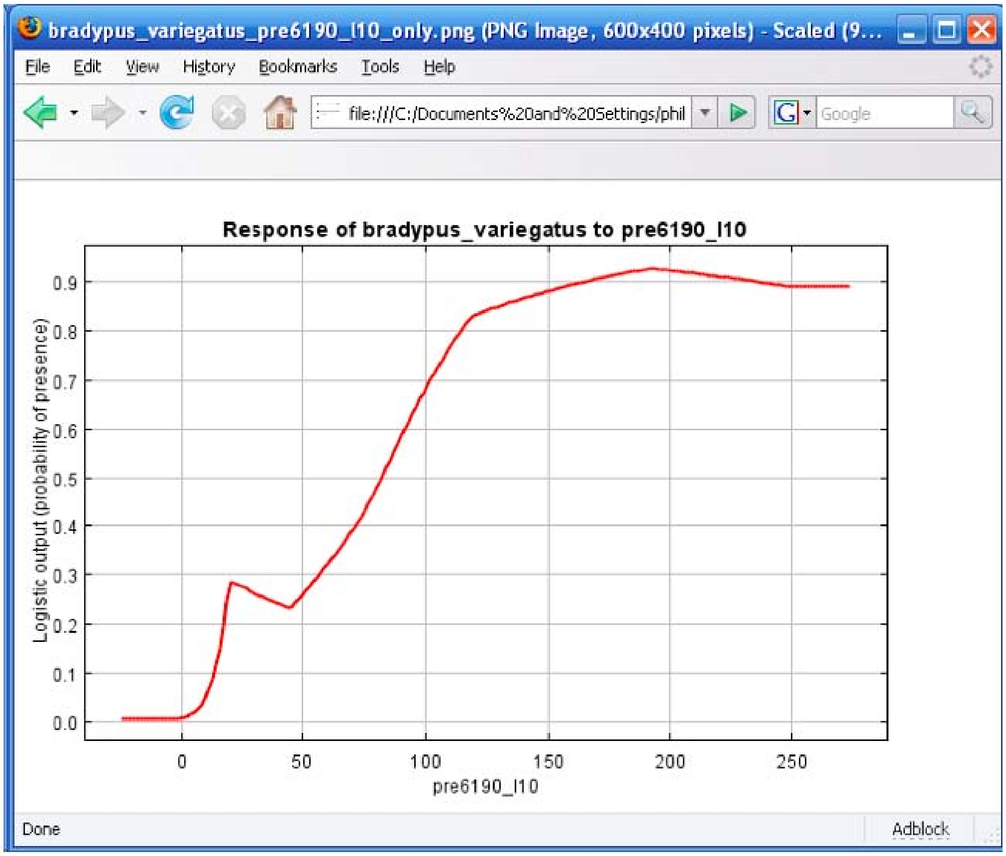
In contrast to the marginal response to annual precipitation in the first set of response curves, we now see that predicted suitability generally increases with increasing annual precipitation.

FEATURE TYPES AND RESPONSE CURVES

Response curves allow us to see the difference among different feature types. Deselect the "auto features", select "Threshold features", and press the "Run" button again. Take a look at the resulting feature profiles – you'll notice that they are all step functions, like this one for pre6190_l10:



If the same run is done using only hinge features, the resulting feature profile looks like this:



The outlines of the two profiles are similar, but they differ because different feature types allow different possible shapes of response curves. The exponent in a Maxent model is a sum of features, and a sum of threshold features is always a step function, so the logistic output is also a step function (as are the raw and cumulative outputs). In comparison, a sum of hinge features is always a piece-linear function, so if only hinge features are used, the Maxent exponent is piece-wise linear. This explains the sequence of connected line segments in the second response curve above. Using all classes together (the default, given enough samples) allows many complex responses to be accurately modeled. A deeper explanation of the various feature types can be found by clicking on the "help" button.

SWD FORMAT

Another input format can be very useful, especially when your environmental grids are very large. For lack of a better name, it's called "samples with data", or just SWD. The SWD version of our *Bradypus* file, called "bradypus_swd.csv", starts like this:

```
species,longitude,latitude,cld6190_ann,dtr6190_ann,ecoreg,frs6190_ann,h_dem,pre6190_ann,pre6190_l10,pre6190_l1,pre6190_l4,pre6190_l7,tmn6190_ann,tmp6190_ann,tmx6190_ann,vap6190_ann
bradypus_variegatus,-65.4,-10.3833,76.0,104.0,10.0,2.0,121.0,46.0,41.0,84.0,54.0,3.0,192.0,266.0,337.0,279.0
bradypus_variegatus,-65.3833,-10.3833,76.0,104.0,10.0,2.0,121.0,46.0,40.0,84.0,54.0,3.0,192.0,266.0,337.0,279.0
bradypus_variegatus,-65.1333,-16.8,57.0,114.0,10.0,1.0,211.0,65.0,56.0,129.0,58.0,34.0,140.0,244.0,321.0,221.0
bradypus_variegatus,-63.6667,-17.45,57.0,112.0,10.0,3.0,363.0,36.0,33.0,71.0,27.0,13.0,135.0,229.0,307.0,202.0
bradypus_variegatus,-63.85,-17.4,57.0,113.0,10.0,3.0,303.0,39.0,35.0,77.0,29.0,15.0,134.0,229.0,306.0,202.0
```

It can be used in place of an ordinary samples file. The difference is only that the program doesn't need to look in the environmental layers to obtain values for the variables at the sample points. The environmental layers are thus only used to get "background" pixels – pixels where the species hasn't necessarily been detected. In fact, the background pixels can also be specified in a SWD format file. The file "background.csv" contains 10,000 background data points. The first few look like this:

```
background,-61.775,6.175,60.0,100.0,10.0,0.0,747.0,55.0,24.0,57.0,45.0,81.0,182.0,239.0,300.0,232.0
background,-66.075,5.325,67.0,116.0,10.0,3.0,1038.0,75.0,16.0,68.0,64.0,145.0,181.0,246.0,331.0,234.0
background,-59.875,-26.325,47.0,129.0,9.0,1.0,73.0,31.0,43.0,32.0,43.0,10.0,97.0,218.0,339.0,189.0
background,-68.375,-15.375,58.0,112.0,10.0,44.0,2039.0,33.0,67.0,31.0,30.0,6.0,101.0,181.0,251.0,133.0
background,-68.525,4.775,72.0,95.0,10.0,0.0,65.0,72.0,16.0,65.0,69.0,133.0,218.0,271.0,346.0,289.0
```

We can run Maxent with "bradypus_swd.csv" as the samples file and "background.csv" (both located in the "swd" directory) as the environmental layers file. Try running it – you'll notice that it runs much faster, because it doesn't have to load the large environmental grids. The downside is that it can't make pictures or output grids, because it doesn't have all the environmental data. The way to get around this is to use a "projection", described below.

BATCH RUNING

Sometimes you need to generate multiple models, perhaps with slight variations in the modeling parameters or the inputs. Generation of models can be automated with command-line arguments, obviating the need to click and type repetitively at the program interface. The command line arguments can either be given from a command window (a.k.a. shell), or they can be defined in a batch file. Take a look at the file "batchExample.bat" (for example, right click on the .bat file inWindows Explorer and open it using Notepad). It contains the following line:

```
java -mx512m -jar maxent.jar environmentallayers=layers togglelayertype=ecoreg
```

```
samplesfile=samples\bradypus.csv outputdirectory=outputs redoifexists autorun
```

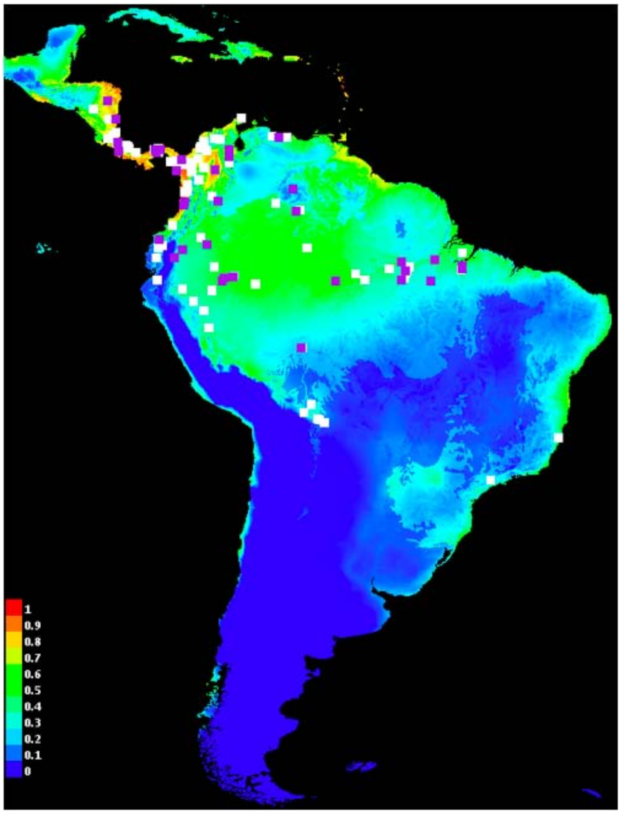
The effect is to tell the program where to find environmental layers and samples file and where to put outputs, to indicate that the ecoreg variable is categorical. The "autorun" flag tells the program to start running immediately, without waiting for the "Run" button to be pushed. Now, try double clicking on the file to see what it does.

Many aspects of the Maxent program can be controlled by command-line arguments – press the "Help" button to see all the possibilities. Multiple runs can appear in the same file, and they will simply be run one after the other. You can change the default values of most parameters by adding command-line arguments to the "maxent.bat" file. Many of the command-line arguments also have abbreviations, so the run described in batchExample.bat could also be initiated using this command:

```
java -mx512m -jar maxent.jar –e layers –t eco –s samples\bradypus.csv –o outputs –r -a
```

REGULARIZATION

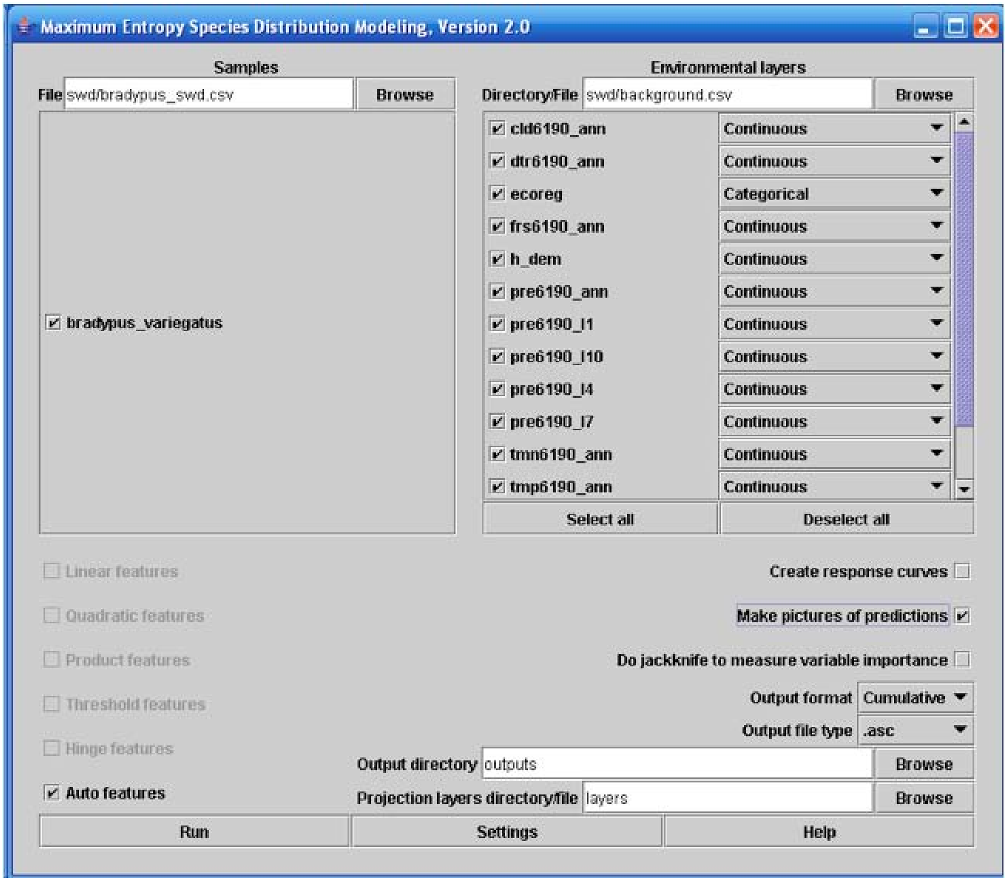
The "regularization multiplier" parameter on the settings panel affects how focused or closely-fitted the output distribution is – a smaller value than the default of 1.0 will result in a more localized output distribution that is a closer fit to the given presence records, but can result in overfitting (fitting so close to the training data that the model doesn't generalize well to independent test data). A larger regularization multiplier will give a more spread out, less localized prediction. Try changing the multiplier, and examine the pictures produced and changes in the AUC. As an example, setting the multiplier to 3 makes the following picture, showing a much more diffuse distribution than before:



The potential for overfitting increases as the model complexity increases. First try setting the multiplier very small (e.g. 0.01) with the default set of features to see a highly overfit model. Then try the same regularization multiplier with only linear and quadratic features.

PROJECTING

A model trained on one set of environmental layers (or SWD file) can be "projected" by applying it to another set of environmental layers (or SWD file). Situations where projections are needed include modeling species distributions under changing climate conditions, applying a model of the native distribution of an invasive species to assess invasive risk in a different geographic area, or simply evaluating the model at a set of test locations in order to do further statistical analysis. Here, we're going to use projection for a very simple task: to make an output ascii grid and associated picture when the samples and background are in SWD format. Type in, or browse for, the samples file "swd\bradypus_swd.csv" and the environmental layers in "swd\background.csv", then enter the "layers" directory in the "Projection Layers Directory", as pictured below. (More generally, the projection layers directory would contain grids with the same names as the variables used for training the model, but describing a different geographic region or different climatic conditions.)



Maximum Entropy Species Distribution Modeling, Version 2.0

Samples

File swd/bradypus_swd.csv

☑ bradypus_variegatus

Environmental layers

Directory/File swd/background.csv

☑ cld6190_ann	Continuous
☑ dtr6190_ann	Continuous
☑ ecoreg	Categorical
☑ frs6190_ann	Continuous
☑ h_dem	Continuous
☑ pre6190_ann	Continuous
☑ pre6190_l1	Continuous
☑ pre6190_l10	Continuous
☑ pre6190_l4	Continuous
☑ pre6190_l7	Continuous
☑ tmn6190_ann	Continuous
☑ tmp6190_ann	Continuous

☐ Linear features

☐ Quadratic features

☐ Product features

☐ Threshold features

☐ Hinge features

☑ Auto features

Create response curves ☐

Make pictures of predictions ☑

Do jackknife to measure variable importance ☐

Output format Cumulative

Output file type .asc

Output directory outputs

Projection layers directory/file layers

When you press "Run", a model is trained on the SWD data, and then projected onto the full ascii grids in the "layers" directory. The output ascii grid is called "bradypus_variegatus_layers.asc", and in general, the projection directory name is appended to the species name, in order to distinguish it from the standard (un-projected) output. If "make pictures of predictions" is selected, a picture of the projected model will appear in the "bradypus.html" file.

ANALYZING MAXENT OUTPUT IN R

Maxent produces a number of output files for each run. Some of these files can be imported into other programs if you want to do your own analysis of the predictions. Here, we demonstrate the use of the free statistical package R on Maxent outputs: this section is intended for users who have experience with R. We will use the following two files produced by Maxent:

bradypus_variegatus.csv
bradypus_variegatus_samplePredictions.csv

The first of these is produced when the background data are given in SWD format, and the second is always produced. Make sure you have test data (for example, by setting the random test percentage to 25); we will be evaluating the Maxent outputs using the same test data Maxent used. First, we start R, and install some packages (assuming this is the first time we're using them) and then load them by typing (or pasting):

```
install.packages("ROCR", dependencies=TRUE)
install.packages("vcd",  dependencies=TRUE)
library(ROCR)
library(vcd)
library(boot)
```

Throughout this section, we will use blue text to show R code and commands and green to show R outputs. Next we change directory to where the Maxent outputs are, for example:

```
setwd("c:/maxent/tutorial/outputs")
```

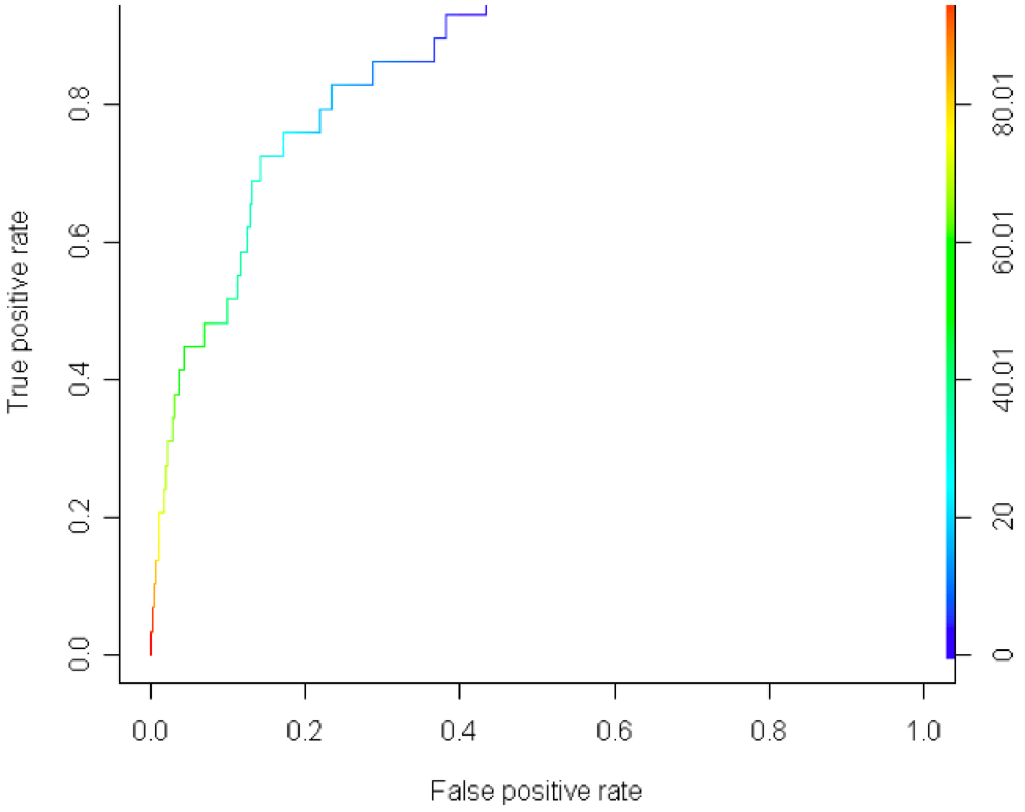
then read the Maxent predictions at the presence and background points, and extract the columns we need:

```
presence <- read.csv(“bradypus_variegatus_samplePredictions.csv”)
background <- read.csv(“bradypus_variegatus.csv”)
pp <- presence$Cumulative.prediction              # get the column of predictions
testpp <- pp[presence$Test.or.train==”test”]       # select only test points
trainpp <- pp[presence$Test.or.train==”train”]   # select only test points
bb <- background$Maxent.cumulative.values.at.background.points
```

Now we can put the prediction values into the format required by ROCR, the package we will use to do some ROC analysis, and generate the ROC curve:

```
combined <- c(testpp, bb)                                    # combine into a single vector
label <- c(rep(1,length(testpp)),rep(0,length(bb)))  # labels: 1=present, 0=random
pred <- prediction(combined, label)                     # labeled predictions
perf <- performance(pred, “tpr”, “fpr”)               # True / false positives, for ROC curve
plot(perf, colorize=TRUE)                                  # Show the ROC curve
performance(pred, “auc”)@y.values[[1]]             # Calculate the AUC
```

The plot command gives the following result:



while the "performance" command gives an AUC value of 0.8677759, consistent with the AUC reported by Maxent. Next, as an example of a test available in R but not in Maxent, we will make a bootstrap estimate of the standard deviation of the AUC.

```
AUC <- function(p,ind) {
pres <- p[ind]
combined <- c(pres, bb)
label <- c(rep(1,length(pres)),rep(0,length(bb)))
predic <- prediction(combined, label)
return(performance(predic, "auc")@y.values[[1]])
}

b1 <- boot(testpp, AUC, 100)  # do 100 bootstrap AUC calculations
b1                             # gives estimates of standard error and bias
```

This gives the following output:

```
ORDINARY NONPARAMETRIC BOOTSTRAP

 Call:
 boot(data = testpp, statistic = AUC, R = 100)

 Bootstrap Statistics :
      original      bias     std. error
t1* 0.8677759 -0.0003724138  0.02972513
```

and we see that the bootstrap estimate of standard error (0.02972513) is close to the standard error computed by Maxent (0.028). The bootstrap results can also be used to determine confidence intervals for the AUC:

```
boot.ci(b1)
```

gives the following four estimates – see the resources section at the end of this tutorial for references that define and compare these estimates.

```
Intervals :
 Level      Normal              Basic
 95%   ( 0.8099,  0.9264 )   ( 0.8104,  0.9291 )
```

```
Level     Percentile          BCa
95%   ( 0.8064,  0.9252 )  ( 0.7786,  0.9191 )
```

Those familiar with use of the bootstrap will notice that we are bootstrapping only the presence values here. We could also bootstrap the background values, but the results would not change much, given the very large number of background values (10000).

As a final example, we will investigate the calculation of binomial and Cohen's Kappa statistics for some example threshold rules. First, the following R code calculates Kappa for the threshold given by the minimum presence prediction:

```
confusion <- function(thresh) {
    return(cbind(c(length(testpp[testpp>=thresh]), length(testpp[testpp<thresh])),
               c(length(bb[bb>=thresh]), length(bb[bb<thresh]))))
  }
  mykappa <- function(thresh) {
    return(Kappa(confusion(thresh)))
  }
  mykappa(min(trainpp))
```

which gives a value of 0.0072. If we want to use the threshold that minimizes the sum of sensitivity and specificity on the test data, we can do the following, using the true positive rate and false positive rate values from the "performance" object used above to plot the ROC curve:

```
fpr = perf@x.values[[1]]
 tpr = perf@y.values[[1]]
 maxsum = 0
 for (i in 1:length(perf@alpha.values[[1]])) {
   sum = tpr[[i]] + (1-fpr[[i]])
   if (sum > maxsum) {
     maxsum = sum
     cutoff = perf@alpha.values[[1]][[i]]
     index = i
   }
 }
 mykappa(cutoff)
```

This gives a Kappa value of 0.0144. To determine binomial probabilities for these two threshold values, we can do:

```
mybinomial <- function(thresh) {
    conf <- confusion(thresh)
    trials <- length(testpp)
    return(binom.test(conf[[1]][[1]], trials, conf[[1,2]] / length(bb), "greater"))
  }
  mybinomial(min(trainpp))
  mybinomial(cutoff)
```

This gives p-values of 5.979e-09 and 2.397e-11 respectively, which are both slightly larger than the p-values given by Maxent. The reason for the difference is that the number of test samples is greater than 25, the threshold above which Maxent uses a normal approximation to calculate binomial p-values.

R RESOURSES

Some good introductory material on using R can be found at:

http://spider.stat.umn.edu/R/doc/manual/R-intro.html, and other pages at the same site.

http://www.math.ilstu.edu/dhkim/Rstuff/Rtutor.html

TERMS OF USE

Reproduction of this material is authorized by the recipient institution for non-profit/non-commercial educational use and distribution to students enrolled in course work at the institution. Distribution may be made by photocopying or via the institution's intranet restricted to enrolled students. Recipient agrees not to make commercial use, such as, without limitation, in publications distributed by a commercial publisher, without the prior express written consent of AMNH.

All reproduction or distribution must provide both full citation of the original work, and a copyright notice as follows:

"S. Phillips, Species' Distribution Modeling for Conservation Educators and Practitioners. 2009. Exercise. American Museum of Natural History, *Lessons in Conservation*. Available at http://ncep.amnh.org/linc."

This material is based on work supported by the National Science Foundation under the Course, Curriculum and Laboratory Improvement program (NSF 0127506).

NCEP gratefully acknowledges the support of the following organizations, institutions, and individuals:

The John D. and Catherine T. MacArthur Foundation
Joseph and Joan Cullman Conservation Foundation, Inc.
Dr. Kathryn Hearst
The New York Community Trust
Peggy Condron
Peter A. and Marion W. Schwartz Family Foundation
S. Malcolm and Elizabeth Gillis
Strachan and Vivian Donnelley
The United States National Science Foundation Course Curriculum and Laboratory Improvement Program DUE-0127506, DUE 0442490
The United States National Oceanic and Atmospheric Administration NA04OAR4700191
The United States Fish and Wildlife Service 98210-1-G017

We welcome your comments and feedback. To write to NCEP or for more information, contact the Network of Conservation Educators and Practitioners at:
American Museum of Natural History
Center for Biodiversity and Conservation
79th Street at Central Park West
New York, New York 10024
ncep@amnh.org

Lessons in Conservation is available electronically at http://ncep.amnh.org/linc

Network of Conservation Educators and Practitioners
Center for Biodiversity and Conservation
AMERICAN MUSEUM OF NATURAL HISTORY

LINC

http://ncep.amnh.org/linc